UNIVERSITÀ
DI PAVIA

**Faculty of Engineering**

**Department of Civil Engineering and Architecture**

**PhD in Design, modelling and simulation in engineering**

**Cycle XXXVII**

**PhD Thesis**

# Advanced flow routing and water quality modeling of urban catchments during wet weather periods

Candidate:

Mohammed N. Assaf

Supervisors:

Prof. Sara Todeschini

Prof. Enrico Creaco

Co-supervisor

Prof. Sauro Manenti

Academic Year 2024/2025

## Acknowledgements

I would like to express my sincere gratitude to **Prof. Sara Todeschini** and **Prof. Sauro Manenti** for their supervision, guidance, and constant support throughout my PhD journey. Their scientific rigor, insightful feedback, and availability were fundamental to the development of this work and to my growth as a researcher.

I am deeply grateful to **Prof. Enrico Creaco** for selecting me to join the PhD program and for supporting and supervising me from the very beginning. His trust, encouragement, and guidance provided a strong foundation for my doctoral path, and his continuous support has been essential throughout these years.

I would also like to acknowledge **Prof. Lorenzo Tamellini**, whose collaboration over these years has been invaluable. His expertise, openness to discussion, and constructive contributions significantly enriched this research.

## Abstract

Urbanization and climate change intensify stormwater challenges by accelerating runoff and increasing pollutant loads to receiving waters. This thesis advances event-based calibration strategies for stormwater quantity and quality modelling within EPA-SWMM. The work evaluates hydrological formulations, investigates how event selection and objective functions influence calibration outcomes, and develops integrated strategies for pollutant build-up and wash-off. A Global Sensitivity Analysis (GSA) is employed to identify influential parameters, assess how rainfall characteristics affect parameter sensitivity, and reduce dimensionality and equifinality. Results show that calibration design including model choice, event selection, sensitivity analysis, and objective functions, strongly determines predictive reliability. The thesis contributes methodological innovations and practical guidance to improve the robustness, efficiency, and applicability of urban stormwater models.

# Table of Contents

# Chapter 1: Introduction

## 1.1 Research Background

Urbanization causes changes to the hydrological cycle through substitution of permeable surfaces with impervious ones like asphalt and concrete. The result of this conversion is a decrease in infiltration, an expansion in effective impervious space, and an intensification in hydrologic connectivity, which causes faster and greater runoff responses to rainfall events. Urban catchment hydrographs usually have short lag times, steep peaks and lower baseflow contributions than natural systems. This leads to more flooding potential, sewer surcharge potential and stress on infrastructure. Climate change also increases the complexity of the issue by making rainfall extremes more severe and changing precipitation patterns and putting even more strain on stormwater systems. In this regard, proper stormwater modelling is to be used in order to guide urban drainage network design, operation and long-term planning.

Conceptual hydrological models have become one of the most widely used tools for representing these processes. Its semi-distributed conceptual framework allows representation of catchment heterogeneity and coupling of hydrology, hydraulics, and water quality. However, the predictive reliability of such models depends critically on calibration. Calibration aligns model parameters with observed data, but it is complicated by parameter uncertainty, equifinality, and data limitations. Furthermore, calibration outcomes are sensitive to methodological choices such as model formulation, storm event selection, and objective functions. For water quality modelling, challenges are amplified: pollutant build-up and wash-off are complex, highly variable, and tightly coupled to the underlying hydrological simulation.

## 1.2 Problem Statement

Despite its widespread use, urban stormwater modelling continues to face several unresolved challenges.

1. Model structure choices

Choices in model structure (e.g., non-linear reservoir vs. unit hydrograph) can produce different results (hydrographs) even under the same calibration dataset, highlighting the influence of model structural choices on predictive robustness.

2. Calibration event selection

Storm events are often selected for calibration without systematic criteria. Yet, event characteristics (e.g., rainfall depth, intensity, temporal distribution) strongly influence calibration outcomes and the accuracy of runoff metrics (volume, peak, timing). Yet guidance on how to select storms or criteria systematically remains scarce.

3. Objective function dependence

Calibration results are highly sensitive to the objective function used. Traditional squared-error criteria tend to overweight large storms or long-duration events, whereas alternative formulations may better balance performance across multiple metrics.

4. Stormwater quality calibration limitations

Water quality calibration is particularly problematic: pollutant wash-off depends directly on flow simulations, and traditional event-by-event calibration and instantaneous concentration objectives often yield inconsistent or non-physical parameter sets.

5. Equifinality and parameter identifiability

Equifinality and uncertainty remain persistent, with multiple parameter sets producing equally acceptable fits, limiting confidence in model transferability.

6. Lack of integrated calibration strategies

Current practice rarely integrates sensitivity analysis, multi-event calibration, and global-metric objectives into a coherent workflow. Without such integration, calibration remains fragmented and often inadequate for reliable, decision-relevant modelling.

## 1.3 Motivation and Research Gap

Urban stormwater management is a critical component of resilient cities, yet modelling practices often face a tension between theoretical sophistication and practical usability.

Models such as SWMM offer flexibility to represent hydrology, hydraulics, and water quality in a unified framework. Conversely, the calibration procedure is highly conditioned by their predictive success and is not only related to the availability of data but also to the methodological decisions. Practically, model calibration is often conducted with ad hoc approaches: a small number of storm events are picked, common error functions are chosen, and model parameters are varied until a reasonable fit is obtained. Although these methods might be adequate in the short-term design, they are not always robust, transferable, and have a clear rationale. This makes predictions of models less reliable, especially when they are relied upon to guide regulatory compliance, infrastructure design, or long-term planning.

Calibration outcomes are often ambiguous, and the influence of methodological choices, such as which storm events are used or which objective functions are optimized, can dominate over the model structure itself. Although parameter uncertainty has been studied individually, the interaction of event selection and objective functions to influence the calibration outcomes in urban rainfall-runoff modelling have not been systematically investigated.

An equally important practical gap exists in stormwater quality modelling. Pollutant build-up and wash-off processes are inherently dependent on simulated runoff, meaning that water quality predictions are only as reliable as the hydrological simulations that drive them. Few instructions are given on how to structure calibration strategies that explicitly couple modelling quantity and quality with multiple storms and performance balance between pollutant measures of cumulative mass and peak pollutant flux.

Collectively, these concerns indicate a research gap between the rigour of the methods used and the practical use of stormwater modelling. In particular, a gap in knowledge currently exists regarding the impact of hydrological formulation options, event selection criteria, and objective functions on calibration process. Filling this gap will be important in designing calibration workflows that are scientifically justifiable, and directly applicable to urban drainage design and management. This thesis addresses that requirement by assessing explicitly these dimensions by a combination of sensitivity analysis, optimization and integrated event calibration approach applied to a well-instrumented urban catchment.

## 1.4 Objectives and Research Questions

The overarching objective of the thesis is to advance calibration strategies for stormwater quantity and quality modelling using SWMM, with a focus on event-based approaches**.**

Specific objectives and research questions are:

1. Evaluate the impact of hydrological formulation on runoff simulation**.**
    - RQ1: How do the non-linear reservoir and unit hydrograph formulations compare in reproducing event-scale hydrographs and key metrics (peak, volume, time-to-peak)?
2. Assess the influence of calibration design on hydrological modelling.
    - RQ2a: How do different event-selection criteria affect calibration outcomes and model robustness across hydrological metrics?
    - RQ2b: How do alternative objective functions influence calibration performance, and to what extent can newly developed objective functions improve the balance between peak flow and volume reproduction?
3. Explore parameter sensitivity through Global Sensitivity Analysis (GSA)
    - RQ3: How can Global Sensitivity Analysis (GSA) be used to identify influential parameters and streamline the calibration process, and how do different rainfall characteristics (e.g., depth, intensity, temporal distribution) affect parameter sensitivity?
4. Develop improved calibration strategies for stormwater quality.
    - RQ4: How can pollutant build-up and wash-off parameters be calibrated more effectively by integrating hydrological foundations, event integration, and alternative objective functions?

## 1.5 Thesis Outline

The thesis is organized into eight chapters:

**Chapter 1** introduces the research background, problem statement, motivation, objectives, and significance.

**Chapter 2** reviews rainfall–runoff modelling in urban catchments, focusing on hydrological processes, SWMM, and calibration challenges.

**Chapter 3** reviews stormwater quality modelling, including pollutant sources, build-up and wash-off dynamics, and modelling approaches.

**Chapter 4** investigates the influence of different hydrological formulations in SWMM, comparing their ability to reproduce event-based runoff hydrographs and key performance metrics.

**Chapter 5** develops and applies optimization strategies for stormwater quality modelling, with emphasis on pollutant build-up and wash-off processes, calibration approaches, and the integration with validated hydrological results.

**Chapter 6** examines the effects of calibration data selection and objective functions on model performance, analyzing their implications for robustness and parameter identifiability in rainfall–runoff modelling.

**Chapter 7** concludes the thesis, outlining the main contributions, limitations, and directions for future research.

# Chapter 2 Rainfall-Runoff Water Quantity Modelling

## 2.1 Introduction

Urban stormwater management is a critical component of modern urban planning, as uncontrolled runoff can lead to flooding, infrastructure damage, and degradation of water quality in receiving waters. Rapid urbanization has dramatically altered the natural hydrologic cycle in cities. When landscapes are urbanized, natural vegetated or permeable surfaces are replaced with impervious surfaces (roads, rooftops, parking lots), which drastically reduce infiltration and increase surface runoff volumes (O'Driscoll et al., 2010).

Studies have long shown that urbanization intensifies the runoff response to rainfall generating higher peak discharges and more rapid streamflow rise than in undeveloped basins (Rose and Peters, 2001; Guan, Sillanpää and Koivusalo, 2016; Arnone et al., 2018). Roughly 90% of rainfall falling on impervious urban surfaces becomes direct runoff, as these surfaces offer little opportunity for water to infiltrate or be retained (Fletcher et al., 2024). Therefore, running the water away quickly from cities can relieve local flooding, but it causes more downstream floods and reduces recharge of the groundwater.

The problems caused by urbanization are made worse by the climate change consequences . The rising temperatures are causing more intense and frequent rainfall in several parts of the world (Nodine et al., 2024; Balistrocchi and Grossi, 2020). Climate models and the observations show that more intense storms will happen more often in many parts of the globe (Huang and Swain, 2022; Hoerling et al., 2016). In urban areas, these intense storms can overwhelm drainage infrastructure that was designed based on historical rainfall patterns. Expected hydrologic changes under climate change include greater stormwater runoff volumes, higher peak flow rates, and more frequent events exceeding the capacity of stormwater systems (Zahmatkesh et al., 2015; Zhou et al., 2019; Wu et al., 2017; Todeschini, 2012). In effect, climate change is magnifying the existing flood hazard in cities. Studies project that average annual runoff in urban watersheds could increase significantly (on the order of 20–30% in coming decades) due to the increased frequency of heavy storms, absent any adaptive measures (Nodine et al., 2024). These changes threaten to exacerbate urban flooding, put additional stress on aging drainage infrastructure, and increase pollutant loads delivered to rivers and streams. Indeed, urban

pluvial flooding (surface flooding from intense rainfall) is now recognized as a serious and growing risk in many cities, one that is expected to worsen with climate change.

Stormwater management aims to limit the effects of these problems by controlling runoff at the source, along its path, and at the end point. Managing the quantity of runoff (peak flows and volumes) is essential for flood prevention and protecting infrastructure, while managing quality is key to reducing pollutant impacts. Urban flooding causes big economic problems for many countries annually, and it is expected that these issues will increase as cities grow and climate risks rise (Hallegatte et al., 2013; Muis et al., 2015). By 2050, it is predicted that 68% of the world's population will be living in cities, which means more people and assets will face the risks of urban flooding. Also, a moderate increase in temperature (1.5 °C) could result in design rainfall intensities for a 10-year to 50-year storm going up by about ~12%, increasing the peak flow even more if no action is taken (Fletcher et al., 2024). These statistics underline the urgency of improving stormwater management to enhance urban resilience. Recognizing these challenges, cities are increasingly exploring adaptation strategies such as enlarged drainage capacities, decentralized green infrastructure, and updated design standards that incorporate climate projections (Spahr et al., 2021; Zellner et al., 2016; Venkataramanan et al., 2020). The main aim is to establish systems that can safely manage extra runoff from major storms and at the same time help remove pollutants and reuse water.

In this context, rainfall-runoff modeling plays a pivotal role. Accurate models allow engineers and scientists to simulate how stormwater moves through urban catchments and to evaluate the effectiveness of various management strategies. The objectives of urban stormwater quantity modeling include: (a) predicting runoff peaks and volumes for drainage design, (b) mapping areas that could be flooded under various storms, (c) assessing the impact of land use changes or climate change on urban hydrology, and (d) forecasting or adjusting the city's drainage systems to avoid heavy rainwater damage and to mitigate the negative impact of combined sewer overflows. In actual use, urban authorities and consultants from different countries turn to simulation models for infrastructure planning. One good example is the Storm Water Management Model (SWMM) from the U.S. EPA, which is now commonly used for urban drainage projects. A great number of peer-reviewed studies have focused on the use of SWMM for designing

storm sewers, studying the value of green infrastructure, analyzing combined sewer overflow controls, and a variety of other applications (Niazi et al., 2017). The fact that SWMM and similar models are widely used means that quantity modeling plays a key role in today's stormwater management since they help evaluate different scenarios and support decisions about urban water infrastructure.

In summary, effective urban stormwater management is crucial for reducing flood risk and protecting water resources in cities. Urbanization has increased runoff and flooding by creating impervious landscapes and climate change threatens to further intensify extreme rainfall and overwhelm existing drainage systems. Urban rainfall-runoff modeling addresses these challenges by offering a means to understand and predict hydrological responses in urban catchments, and to design interventions that achieve the goals of flood mitigation, adequate recharge of aquifers and water quality protection. The following sections provide a detailed review of the hydrologic processes in urban catchments (Section 1.2), the various modeling approaches developed to represent those processes (Section 1.3), and the applications and limitations of urban stormwater models (Section 1.4).

## 2.2 Hydrological Processes in Urban Catchments

Key processes in urban rainfall-runoff include infiltration (which in urban areas occurs primarily on pervious fragments like parks, lawns, or through engineered infiltration facilities), surface runoff (over impervious surfaces and saturated ground), surface storage/detention (in depression storage or constructed basins), and channel/pipe routing.

Urban catchments have unique hydrological mechanisms because there are plenty of impervious surfaces and human-made drainage systems. Rainfall-runoff generation in urban areas is dominated by fast surface runoff because impervious surfaces (such as asphalt and concrete) prevent water from infiltrating into the soil. Most rain in a natural landscape with vegetation is used by the soil and plants; just a little part of it flows into nearby streams. By contrast, in a fully urbanized catchment with high impervious cover, most rainfall becomes immediate runoff. Studies indicate that increasing impervious surface area reduces infiltration and surface storage while dramatically increasing stormwater runoff (Meyer, Paul and Taulbee, 2005; Walsh et al., 2005; Todeschini, 2016). The consequence is a much larger runoff volume as well as a quicker response time. Field measurements in various regions have confirmed that as watershed imperviousness

increases, baseflow (dry-weather flow fed by groundwater) tends to decline and storm hydrographs become flashier (sharper and more peaked) (Arnold and Gibbons, 1996).

The role of impervious areas is central to urban hydrology. Impervious surfaces not only generate more runoff by reducing infiltration, but they also accelerate the timing of runoff delivery. With smooth, connected surfaces and artificial drainage, water moves quickly into channels or pipes. This reduces the lag time between rainfall and runoff peak. It has been observed that urbanization can reduce the time of concentration (the time for runoff to travel to the catchment outlet), resulting in rapid rises in streamflow and shorter flood durations (Wong and Li, 1998).

Urban drainage systems play a crucial role in routing this runoff. Most cities depend on storm drains, sewer pipes, and sometimes open channels or tunnels to remove extra water from the streets and properties. They create a well-linked drainage route that carries water from the land to the water bodies. Even though their main purpose is to prevent local flooding by carrying runoff properly, this system has a side effect: it quickly moves stormwater and associated pollutants into nearby streams.

Research has shown that hydrologic response is strongly governed by the degree of hydrologic connectivity in urban areas – “effective impervious area” (the portion of impervious area directly connected to drains) is often a better predictor of runoff volume than total impervious area. The highest rises in peak flows occur in areas with a lot of impervious surfaces. On the other hand, if impervious surfaces are separated (draining into other areas or storage spots), the effects on peak runoff can be lessened. Urban hydrology is therefore shaped not just by the fraction of impervious cover, but by how that cover is connected to the drainage network.

The next section will review how these processes are represented in different modeling approaches, ranging from simple empirical methods to sophisticated distributed models.

## 2.3 Overview of Modeling Approaches

Over the past century, a wide range of hydrological models have been developed to simulate the rainfall-runoff process in catchments, including urban areas. These models vary in complexity and underlying philosophy. Broadly, rainfall-runoff models can be categorized as empirical, conceptual, or physically based, and they can be further

distinguished by whether they treat the catchment in a lumped or distributed manner, and whether they simulate single events or continuous time series. Each approach has its merits and limitations, and often the choice of model depends on the target and data availability. In this section, we provide an overview of these modeling approaches, with emphasis on how they have been applied to urban water quantity modeling.

Empirical models (also known as black-box or data-driven models) rely on direct relationships drawn from data, rather than explicitly simulating physical processes. They are often the simplest modeling approach. One of the earliest examples is the rational method, which is used for estimating the peak discharge (Q) in the sizing of small drainage systems (Mulvaney, 1850). The rational formula Q = C*I*A computes the peak runoff (Q) as a function of rainfall intensity (I), catchment area (A), and a runoff coefficient (C) (an empirically derived factor representing the fraction of rainfall that becomes runoff). This method represents an empirical understanding that peak flow is proportional to rainfall rate and area, with C adjusting for land surface characteristics (with higher C for impervious urban areas). Another example of an empirical model is the Curve Number (CN) method, developed by the U.S. Soil Conservation Service (SCS) (United States Soil Conservation Service, 1972). The CN method provides an empirical relationship to estimate total runoff volume from rainfall depth, based on land use and soil type categories. Both the rational method and SCS-CN do not resolve internal catchment processes or variation in space, but directly produce an aggregate runoff response. Other empirical approaches in urban hydrology include regression equations derived from observed flood records that correlate rainfall to runoff using training data.

Machine learning (ML) models have become increasingly popular for rainfall-runoff modeling due to their ability to handle complex and nonlinear relationships in hydrological processes (Adnan et al., 2021). Such models can effectively represent runoff without using detailed information about physical processes, so they are useful when little or incomplete physical data is available (Herath, Chadalawada and Babovic, 2021). Nevertheless, some major problems still prevent them from being more effective and widely used. One of the main issues is the lack of interpretability in many common ML methods, which decreases the transparency of models, reducing their acceptance among hydrologists (Herath, Chadalawada and Babovic, 2021; Mohammadi, 2021). Additionally, these models tend to

have problems with physical consistency as they do not use specific hydrological knowledge from particular domains, leading to results that are not fully in line with what happens in watersheds (Chadalawada, Herath and Babovic, 2020; Herath, Chadalawada and Babovic, 2021). This issue is exacerbated in large and spatially heterogeneous catchments, which can reduce the dependability and accuracy of ML-based runoff predictions (Herath, Chadalawada and Babovic, 2021). Also, building and adjusting ML frameworks is often very demanding, calling for significant time and need the involvement of highly skilled experts, which makes the modelling process both time-consuming and complex (Chadalawada, Herath and Babovic, 2020; Jehanzaib et al., 2022).

Conceptual models represent a catchment using simplified, conceptual storages and fluxes that mimic the behavior of real hydrologic processes. They attempt to strike a balance between empirical simplicity and physical realism by using equations that have a physical interpretation but are not tied to exact physical geometry. Typically, conceptual rainfall-runoff models consist of one or more interconnected reservoirs (tanks) representing depression storage, soil moisture, groundwater, etc., with empirical or semi-empirical formulas governing the flows between reservoirs (infiltration, evaporation, surface runoff, baseflow). These models require calibration of their parameters to match observed runoff data. A well-known example is the Nash cascade model (Nash, 1957), where a series of linear reservoirs produces a runoff hydrograph, or the unit hydrograph approach more generally (representing the catchment response as a transfer function applied to effective rainfall). Many conceptual models were developed in the late 20th century for watershed simulation, for example, the HSPF (Donigian, Bicknell and Imhoff, 1995), the HBV model (Bergström, 1976). Flexibility and parsimony are major advantages of conceptual models, they usually have modest data requirements and relatively fast run times (Ali, Rahman and Shaik, 2024). On the downside, because they simplify processes, their parameters may lack direct physical meaning and are often site-specific. Nonetheless, conceptual models are widely used for urban stormwater modeling, as they often achieve reasonable accuracy with fewer parameters to estimate. In practice, many software tools (including SWMM, InfoWorks, HEC-HMS) operate largely as conceptual models that require calibration.

Physically based models attempt to represent the actual physics of hydrological processes using governing equations and detailed catchment properties. These models are

sometimes called process-based or mechanistic models. In an ideal physically-based model, one would solve equations like Richards' equation for unsaturated flow in soil, Saint-Venant equations for surface and channel flow, and so on, using measurable parameters (e.g. soil hydraulic conductivity, surface roughness, etc.) without needing calibration. Several physical based models were developed and tested in urban modeling such as Systeme Hydrologique Europeen (SHE), WAterSHed model for 1-D, 2-D, and 3-D domains (WASH123D) and ParFlow.

In urban hydrology, fully physically-based models can be very complex: for instance, a model might couple a two-dimensional surface flow solver with a dynamic pipe flow model and explicit computation of infiltration in every grid cell. In principle, such models can capture the fine-scale variability of urban runoff processes and allow high-resolution flood simulations. They are particularly useful for simulating distributed phenomena like overland flooding in urban streets (where a 2D representation of flow is needed) or the interactions between surface water and groundwater. For example, they can use physically based model for development of dual drainage models that couple 1D sewer models with 2D overland flow models for better representation of urban flood extents (Burns et al., 2015). However, in practice, even physically-based models require calibration to some degree, because certain inputs (e.g. spatial distribution of soil properties, exact geometry of overland flow paths blocked by buildings, etc.) are uncertain. They also require a great deal of data such as a high-resolution topographic data, detailed sewer network schematics, soil maps, land use, etc., and significant computational resources. Because of this, fully distributed physically-based models have historically been less common for urban water modeling.

Another important classification in modeling is lumped vs. distributed representation of space. A lumped model treats the entire catchment (or large sub-areas) as a single unit with averaged parameters, without resolving internal spatial variability. Lumped models assume rainfall is uniform and that the catchment behaves as one homogenous block. In contrast, a distributed model divides the catchment into many smaller units (grid cells or subcatchments) and allows parameters or inputs to vary spatially. Distributed models can range from semi-distributed (e.g. dividing a basin into subcatchments, each treated as lumped) to fully distributed (e.g. gridded models with a cell-to-cell routing scheme). The

choice between lumped and distributed has practical implications. Lumped models have fewer parameters and are easier to calibrate but cannot explicitly account for spatial heterogeneity of rainfall or land surface. Distributed models can capture, for instance, that one part of a city might get more rain (from a convective storm) than another, or that some neighborhoods have more impervious cover than others, and route flows accordingly. They are thus more realistic in principle, but require more input data (e.g. rainfall at multiple points or radar rainfall fields, detailed land use maps, etc.). SWMM is typically applied in a semi-distributed manner: the user subdivides the urban area into a number of subcatchments, each with its own parameters (area, percentage impervious, slope, etc.), and runoff from each subcatchment is routed through a network of channels/pipes. This allows some spatial detail (e.g. different land use areas can be parameterized differently ) but it is not as fully resolved as a gridded model.

Finally, models can be classified to event based (EB) or continuous simulation (CS) based on temporal representation. EB models simulate an individual storm event, typically over a period of hours or days, focusing on the runoff generated from that event; this approach is often preferred when the focus is on key design metrics such as runoff volume and peak flow of an extreme hydrograph (Hossain, Hewa and Wella-Hewage, 2019). They often require initial conditions (like initial soil moisture) to be specified, and they do not inherently account for what happens between events. CS models, on the other hand, simulate a long-term record (months or years) including both storms and intervening dry periods. CS modeling requires representation of processes like evapotranspiration, soil moisture recovery, groundwater flow, etc., so that the model's state evolves realistically over time. This makes CS particularly relevant when checking the performance of nature-based solutions (NBS) / green infrastructure, because their retention and recovery between storms (antecedent moisture and storage refill) strongly affect performance (Muleta and Knolmar, 2025; Avellaneda et al., 2017). However, continuous models demand more input data (continuous meteorological time series) and often more calibration to get all aspects of the water balance correct. They also typically require longer simulation runs, meaning higher computational effort. Various studies have shown that the EB approach works better than the CS approach in their specific catchments, whereas CS is commonly adopted for more complete, long-term evaluation of stormwater controls and NBS performance under

variable climatic conditions (Jabeen et al., 2025). Chu and Steinman (2009) compared both methods in the Mona Lake Watershed in West Michigan and noticed that the EB model produced a better fit. De Silva, Weerakoon and Herath (2014) also looked into the Kelani River basin in Sri Lanka and found that the EB model performed better in reproducing extreme events. Sampath et al. (2014) investigated both models to forecast stream flows in the Tittawella catchment in Sri Lanka and found that the EB approach gave better results during the relatively short study time. According to Azmat et al. (2017), the EB model performed better than CS when researchers studied a high-altitude watershed. Also, it has been observed that EB modeling is very accurate in generating realistic hydrographs, mainly for storm-specific runoff responses. In the case of Scott Creek catchments, EB modeling was found to be more accurate and reliable in predicting the hydrographs and the peaks of runoff than CS modeling (Hossain, Hewa and Wella-Hewage, 2019; Jabeen et al., 2025).

## 2.4 Applications and Limitations

Applications of urban rainfall-runoff models in the literature are diverse, reflecting the multifaceted challenges of stormwater management. A primary application is in the design and sizing of urban drainage infrastructure. Engineers use models to estimate design flows (e.g. the 5-year or 100-year peak discharge) which inform the capacity requirements for storm sewers, culverts, detention ponds, and other facilities (Myronidis and Ivanova, 2020). In urban flood-prone areas, models help design flood control measures such as retention basins, infiltration trenches, or enlarged stormwater pipes by testing their performance under extreme event simulations (Jehanzaib et al., 2022). Urban runoff models are also extensively used for flood risk assessment and mapping. By coupling hydrologic models with hydraulic models (or using integrated 1D/2D models), researchers and practitioners simulate inundation extents and depths for extreme rainfall events in urban areas. This is crucial for identifying flood hotspots, emergency planning, and designing flood mitigation projects. Models have been applied at scales ranging from small urban catchments (single neighborhoods) to city-wide assessments. A global review noted that urban flood models at the meso-scale (city scale) have evolved with increasing resolution and data integration in recent years reflecting better availability of high-

resolution terrain data and computing power (Xu, Randall and Fryd, 2023). Additional prominent application is real-time flood forecasting for cities (García et al., 2015).

Another important application area is combined sewer overflow (CSO) and water quality modeling. Many cities have combined sewer systems that overflow during heavy rain, causing pollution of waterways. Models like SWMM, which can simulate both runoff quantity and quality (pollutant buildup/wash-off and sewer flow routing), are widely used to design CSO control strategies (Sakson, 2023; Todeschini et al. 2018; Liao et al., 2015; Cipolla, Maglionico and Stojkov, 2016). For example, cities under regulatory mandates often model their sewer systems to evaluate where to build storage tunnels or enlarge pipes to reduce overflow volumes. Stormwater quality models predict concentrations and loads of contaminants (such as sediments, nutrients, heavy metals) washed off from urban surfaces and conveyed via stormwater.

Urban rainfall-runoff models also support strategic planning and scenario analysis. With growing concerns about climate change and urbanization trends, models are used to project how future conditions might affect urban hydrology. Researchers have taken ensembles of climate model projections (which give possible changes in extreme rainfall statistics) and run stormwater models to quantify future flood risks and the adequacy of current infrastructure (Nodine et al., 2024). Similarly, scenario modeling is done for land use change: e.g., "what if this suburban area doubles in impervious cover over the next 20 years – how will peak flows change?" By comparing model simulations of current land use vs. future land use scenarios, planners can estimate the need for new drainage facilities or preservation of green spaces (Dada et al., 2021).

In summary, urban runoff models are applied for: (1) Designing drainage infrastructure (sizing pipes, culverts, ponds) to meet desired flood protection levels. (2) Flood risk mapping and mitigation planning (identifying flood-prone areas and testing solutions), (3) Real-time flood forecasting and operational control of drainage networks. (4) Evaluating the performance of stormwater management practices (comparing grey infrastructure and nature-based solutions in controlling runoff and pollution), (5) Regulatory compliance analysis (e.g., demonstrating that a new development's runoff will not exceed limits, or that a CSO control plan will meet pollution reduction targets), and (6) Long-term strategic

assessments for climate adaptation and urban planning (examining future scenarios of rainfall or land use).

However, urban stormwater models have several limitations and challenges. One fundamental limitation is the simplification of real-world processes. No model can capture the full complexity of an urban hydrological system – there will always be idealizations (e.g., assuming uniform rainfall over a subcatchment, or using a single infiltration parameter for an area with heterogeneous soils). These simplifications mean that model predictions are inherently uncertain and may misrepresent some local behaviors. Deletic et al. (2012) demonstrated how different sources of uncertainty (input uncertainty, parameter uncertainty, model structural uncertainty) each contribute to overall prediction uncertainty in an integrated urban drainage model.

A notorious issue in hydrological modeling is the problem of equifinality (literally, "equal endings"), as identified by Beven and Binley (1992). Equifinality means that many different combinations of model parameter values can produce similarly good fits to observed data, making it impossible to identify a single "correct" parameter set. This is especially true in complex models with many parameters, such as fully distributed models, where some parameters can compensate for others. In urban stormwater modeling, equifinality has been observed when different calibrations might match the hydrograph as numerous parameter sets yielded similar output behavior (Knighton et al., 2016).

Closely related is the issue of parameter uncertainty and identifiability. Urban models often require parameters that are not directly measurable or only crudely estimated (e.g., Manning's roughness for an "average" urban subcatchment, or percent impervious that is "effective"). Even physically-based models, which use ostensibly measured parameters, face uncertainty because urban environments are heterogeneous and measurements are scarce. The result is that modelers must calibrate parameters using observed rainfall-runoff data, and even then, some parameters may not be well constrained (identifiable).

## References


Adnan, R.M., Petroselli, A., Heddam, S., Santos, C.A.G. and Kisi, O., 2021. Comparison of different methodologies for rainfall–runoff modeling: machine learning vs conceptual approach. *Natural Hazards*, 105, pp.2987–3011.

Ali, S., Rahman, A. and Shaik, R., 2024. A review of event-based conceptual rainfall-runoff models: A case for Australia. *Encyclopedia*, 4(2), pp.966–983.

Arnold Jr, C.L. and Gibbons, C.J., 1996. Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American Planning Association*, 62(2), pp.243–258.

Arnone, E., Pumo, D., Francipane, A., La Loggia, G. and Noto, L.V., 2018. The role of urban growth, climate change, and their interplay in altering runoff extremes. Hydrological Processes, 32(12), pp.1755-1770.

Avellaneda, P.M., Jefferson, A.J., Grieser, J.M. and Bush, S.A., 2017. Simulation of the cumulative hydrological response to green infrastructure. *Water Resources Research*, 53(4), pp.3087–3101.

Azmat, M., Qamar, M.U., Ahmed, S., Hussain, E. and Umair, M., 2017. Application of HEC-HMS for the event and continuous simulation in high altitude scarcely-gauged catchment under changing climate. *European Water*, 57, pp.77–84.

Balistrocchi, M. and Grossi, G., 2020. Predicting the impact of climate change on urban drainage systems in northwestern Italy by a copula-based approach. *Journal of Hydrology: Regional Studies, 28*, p.100670.

Bergström, S., 1976. *Development and application of a conceptual runoff model for Scandinavian catchments*. PhD thesis. University of Lund, Sweden.

Beven, K. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6(3), pp.279–298.

Chadalawada, J., Herath, H.M.V.V. and Babovic, V., 2020. Hydrologically informed machine learning for rainfall-runoff modeling: A genetic programming-based toolkit for automatic model induction. *Water Resources Research*, 56(4), p.e2019WR026933.

Chu, X. and Steinman, A., 2009. Event and continuous hydrologic modeling with HEC-HMS. *Journal of Irrigation and Drainage Engineering*, 135(1), pp.119–124.

Dada, A., Urich, C., Berteni, F., Pezzagno, M., Piro, P. and Grossi, G., 2021. Water sensitive cities: An integrated approach to enhance urban flood resilience in Parma (Northern Italy). Climate, 9(10), p.152.

De Silva, M.M.G.T., Weerakoon, S.B. and Herath, S., 2014. Modeling of event and continuous flow hydrographs with HEC–HMS: case study in the Kelani River Basin, Sri Lanka. *Journal of Hydrologic Engineering*, 19(4), pp.800–806.

Deletic, A., Dotto, C.B.S., McCarthy, D.T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T.D., Rauch, W. and Bertrand-Krajewski, J.L., 2012. Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth, Parts A/B/C*, 42, pp.3–10.

Donigian, A.S. Jr., Bicknell, B.R. and Imhoff, J.C., 1995. Hydrological Simulation Program – Fortran (HSPF). In: V.P. Singh, ed. *Computer Models of Watershed Hydrology*. Colorado, USA: Water Resources Publications, pp.395–442.

Fletcher, T.D., Burns, M.J., Russell, K.L., Hamel, P., Duchesne, S., Cherqui, F. and Roy, A.H., 2024. Concepts and evolution of urban hydrology. Nature Reviews Earth & Environment, pp.1-13.

García, L., Barreiro-Gomez, J., Escobar, E., Téllez, D., Quijano, N. and Ocampo-Martínez, C., 2015. Modeling and real-time control of urban drainage systems: A review. *Advances in Water Resources*, 85, pp.120–132.

Guan, M., Sillanpää, N. and Koivusalo, H., 2016. Storm runoff response to rainfall pattern, magnitude and urbanization in a developing urban catchment. Hydrological Processes, 30(4), pp.543-557.

Hallegatte, S., Green, C., Nicholls, R.J. and Corfee-Morlot, J., 2013. Future flood losses in major coastal cities. *Nature Climate Change*, 3(9), pp.802–806.

Herath, H.M.V.V., Chadalawada, J. and Babovic, V., 2021. Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. *Hydrology and Earth System Sciences*, 25(8), pp.4373–4401.

Herath, H.M.V.V., Chadalawada, J. and Babovic, V., 2021. Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. *Hydrology and Earth System Sciences*, 25(8), pp.4373–4401.

Hoerling, M., Eischeid, J., Perlwitz, J., Quan, X.W., Wolter, K. and Cheng, L., 2016. Characterizing recent trends in US heavy precipitation. *Journal of Climate*, 29(7), pp.2313–2332.

Hossain, S., Hewa, G.A. and Wella-Hewage, S., 2019. A comparison of continuous and event-based rainfall–runoff (RR) modelling using EPA-SWMM. *Water*, 11(3), p.611.

Huang, X. and Swain, D.L., 2022. Climate change is increasing the risk of a California megaflood. Science Advances, 8(31), p.eabq0995.

Jabeen, K., Grossi, G., Turco, M., Dada, A., Palermo, S.A., Pirouz, B., Piro, P., Gnecco, I. and Palla, A., 2025. Continuous simulations for predicting green roof hydrologic performance for future climate scenarios. Hydrology, 12(2), p.41.

Jehanzaib, M., Ajmal, M., Achite, M. and Kim, T.W., 2022. Comprehensive review: Advancements in rainfall-runoff modelling for flood mitigation. *Climate*, 10(10), p.147.

Jehanzaib, M., Ajmal, M., Achite, M. and Kim, T.W., 2022. Comprehensive review: Advancements in rainfall-runoff modelling for flood mitigation. *Climate*, 10(10), p.147.

Knighton, J., Lennon, E., Bastidas, L. and White, E., 2016. Stormwater detention system parameter sensitivity and uncertainty analysis using SWMM. *Journal of Hydrologic Engineering*, 21(8), p.05016014.

Liao, Z.L., Zhang, G.Q., Wu, Z.H., He, Y. and Chen, H., 2015. Combined sewer overflow control with LID based on SWMM: an example in Shanghai, China. *Water Science and Technology*, 71(8), pp.1136–1142.

Meyer, J.L., Paul, M.J. and Taulbee, W.K., 2005. Stream ecosystem function in urbanizing landscapes. *Journal of the North American Benthological Society*, 24(3), pp.602-612.

Mohammadi, B., 2021. A review on the applications of machine learning for runoff modeling. *Sustainable Water Resources Management*, 7(6), p.98.

Muis, S., Güneralp, B., Jongman, B., Aerts, J.C. and Ward, P.J., 2015. Flood risk and adaptation strategies under climate change and urban expansion: A probabilistic analysis using global data. *Science of the Total Environment*, 538, pp.445–457.

Mulvaney, T.J., 1850. On the use of self-registering rain and flood gauges. *Proceedings of the Institution of Civil Engineers*, 4, pp.1–8.

Muleta, T.N. and Knolmar, M., 2025. Green Stormwater Infrastructure (GSI) performance assessment for climate change resilience in storm sewer network. *Water*, 17(17), p.2510.

Myronidis, D. and Ivanova, E., 2020. Generating regional models for estimating the peak flows and environmental flows magnitude for the Bulgarian-Greek Rhodope mountain range torrential watersheds. *Water*, 12(3), p.784.

Nash, J.E., 1957. The form of the instantaneous unit hydrograph. International Association of Scientific Hydrology Publication, 45, pp.114–121.

Niazi, M., Nietch, C., Maghrebi, M., Jackson, N., Bennett, B.R., Tryby, M. and Massoudieh, A., 2017. Storm Water Management Model: Performance review and gap analysis. Journal of Sustainable Water in the Built Environment, 3(2), p.04017002.

Nodine, T.G., Conley, G., Riihimaki, C.A., Holland, C. and Beck, N.G., 2024. Modeling the impact of future rainfall changes on the effectiveness of urban stormwater control measures. Scientific reports, 14(1), p.4082.

O'Driscoll, M., Clinton, S., Jefferson, A., Manda, A. and McMillan, S., 2010. Urbanization effects on watershed hydrology and in-stream processes in the southern United States. Water, 2(3), pp.605-648. https://doi.org/10.3390/w2030605

Rose, S. and Peters, N.E., 2001. Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. Hydrological processes, 15(8), pp.1441-1457.

Sakson, G., 2023. Analysis of suspended solids emissions from a combined sewage system using the Stormwater Management Model (SWMM). *Geology, Geophysics and Environment*, 49(4), pp.389–399.

Sampath, D.S., Weerakoon, S.B., Mishra, B.K. and Herath, S., 2014. HEC-HMS model for runoff simulation from Tittawella tank catchment. In: *Proceedings of the 1st Ruhuna International Science & Technology Conference*, University of Ruhuna, Matara, Sri Lanka, 22–23 January 2014.

Spahr, K.M., Bell, C.D., Gallo, E.M., McCray, J.E. and Hogue, T.S., 2021. Incorporating a multiple-benefit analysis into a stormwater decision-support tool at planning level. *Journal of Sustainable Water in the Built Environment*, 7(3), p.04021011.

Todeschini, S., 2012. Trends in long daily rainfall series of Lombardia (Northern Italy) affecting urban stormwater control. *International Journal of Climatology*, 32(6), p.900-919.

Todeschini, S., 2016. Hydrologic and Environmental Impacts of Imperviousness in an Industrial Catchment of Northern Italy. Journal of Hydrologic Engineering, 21(7), p.05016013.

Todeschini, S., Papiri, S., Ciaponi, C. 2018. Placement strategies and cumulative effects of wet-weather control practices for intermunicipal sewerage systems. *Water Resources Management*, 32(8), p.2885-2900

United States Soil Conservation Service, 1972. *SCS National Engineering Handbook, Section 4: Hydrology*. Washington, D.C.: U.S. Department of Agriculture.

Venkataramanan, V., Lopez, D., McCuskey, D.J., Kiefus, D., McDonald, R.I., Miller, W.M., Packman, A.I. and Young, S.L., 2020. Knowledge, attitudes, intentions, and behavior related to green infrastructure for flood management: A systematic literature review. *Science of the Total Environment*, 720, p.137606.

Walsh, C.J., Roy, A.H., Feminella, J.W., Cottingham, P.D., Groffman, P.M. and Morgan, R.P., 2005. The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society*, *24*(3), pp.706-723.

Wong, T.S. and Li, Y., 1998. Assessment of changes in overland time of concentration for two opposing urbanization sequences. *Hydrological sciences journal*, 43(1), pp.115-130.

Wu, Z., Chen, X., Lu, G., Xiao, H., He, H. and Zhang, J., 2017. Regional response of runoff in CMIP5 multi-model climate projections of Jiangsu Province, China. *Stochastic Environmental Research and Risk Assessment*, 31, pp.2627–2643.

Xu, H., Randall, M. and Fryd, O., 2023. Urban stormwater management at the meso-level: A review of trends, challenges and approaches. *Journal of Environmental Management*, 331, p.117255.

Zahmatkesh, Z., Karamouz, M., Goharian, E. and Burian, S.J., 2015. Analysis of the effects of climate change on urban storm water runoff using statistically downscaled

precipitation data and a change factor approach. *Journal of Hydrologic Engineering*, 20(7), p.05014022.

Zellner, M., Massey, D., Minor, E. and Gonzalez-Meler, M., 2016. Exploring the effects of green infrastructure placement on neighborhood-level flooding via spatially explicit simulations. *Computers, Environment and Urban Systems*, 59, pp.116–128.

Zhou, Q., Leng, G., Su, J. and Ren, Y., 2019. Comparison of urbanization and climate change impacts on urban flood volumes: Importance of urban planning and drainage adaptation. *Science of the Total Environment*, 658, pp.24–33.

# Chapter 3: Rainfall-Runoff Water Quality Modelling

## 3.1 Introduction

Urban stormwater runoff has been long considered as a source of a great amount of water pollution in the developed regions. Issues relating to poor quality of urban runoff were first raised in the 1970s when research started to be conducted on the increased levels of sediments, nutrients and heavy metals in stormwater, which led to stormwater quality becoming an issue of interest to regulators (Drapper et al., 2022). Unlike treated sanitary wastewater, urban runoff is usually released with little or no treatment and a highly mixed load of contaminants that may impair the quality of receiving water bodies. Runoff has the potential to carry with it pollutants that may accumulate on impervious surfaces (e.g. roads, rooftops, and parking lots) into the storm drains, streams, and rivers. It is currently known that this is one of the primary ways through which anthropogenic activities can affect aquatic environments (Anil et al., 2019). The deterioration of the receiving water quality as a result of urban runoff has ecological and human health impacts: nutrients in the runoff can cause eutrophication, algal blooms, and fish kills (Conley et al., 2009), and toxic heavy metals and hydrocarbons may accumulate in the soil and aquatic organisms there, and even trace amounts pose a threat to aquatic life (Pamuru et al., 2022). Due to these effects, stormwater runoff is regarded as one of the primary causes of water quality degradation in most areas (Yang and Lusk, 2018; Mohiuddin et al., 2011; Na Nagara, Sarkar and Datta, 2022) and enhancing its quality is a priority of urban water management. This challenge is exacerbated by the growth of waterproof (impervious) surfaces due to urbanization, which reduces infiltration while increasing runoff rates/volumes and degrading water quality (Altobelli, Cipolla and Maglionico, 2020). Urban stormwater generally has a wide variety of pollutants, including both traditional and emerging pollutants. Typical categories of pollutants are: total suspended solids (TSS) (particles of sediment eroded or deposited on surfaces), nutrients (especially forms of nitrogen and phosphorus in fertilizers, organic matter, or atmospheric deposition), heavy metals (such as zinc, copper, lead, cadmium due to vehicular wear, building materials, etc. ), hydrocarbons (e.g. oils, greases, polycyclic aromatic hydrocarbons due to vehicular exhaust and industrial processes), and pathogens (bacteria from animal waste) (Werbowski et al., 2021; Huber, Welker and Helmreich,

2016; Herath, Chadalawada and Babovic, 2021). In the urban areas, the pollutants are produced by a wide range of human and natural sources (see Section 2.2) and accumulate on surfaces during dry periods. When it rains, the first runoff usually has particularly high concentrations of such accumulated pollutants; a phenomenon known as the first flush. The first flush is the initial portion of runoff that can carry a disproportionately large fraction of the total pollutant load; in many cases, 60–80% of a pollutant's total event load is transported in the first 20–30% of runoff volume (Mamun, Shams and Nuruzzaman, 2020; Barco, Papiri and Stenstrom, 2008). Unfortunately, the notion of first flush has been adopted for a plurality of detection methodologies and most of the procedures imply a subjective and arbitrary estimation of the first flush volume. The lack of standardization implies that these methodologies cannot be used as an appropriate basis for the design and the sizing stormwater quality control measures. For this reason, sound quantitative methodologies have been proposed and tested (Bach et al., 2010; Todeschini et al., 2019; Niazkar et al., 2024; Todeschini 2024). The first flush effect is especially pronounced in smaller, highly impervious catchments and for certain contaminants (e.g. particulate pollutants), and it underscores the importance of managing the "initial burst" of urban runoff pollution (Lee et al., 2002).

Given the environmental significance of stormwater pollution, there is a strong impetus to model and manage rainfall-runoff water quality in urban areas. Accurate modeling allows researchers and practitioners to predict pollutant loads, identify pollution "hotspots," and evaluate control strategies before implementation. Urban stormwater quality models have become essential tools in planning green infrastructure (e.g. biofilters, ponds), designing regulatory programs (such as Total Maximum Daily Loads, TMDLs), and integrating stormwater management into urban development. These models range from simple empirical approaches to advanced process-based simulations, each with advantages and limitations. The following sections of this chapter provide a detailed review of (1) the sources and processes that govern pollutant accumulation and wash-off in urban runoff (Section 3.2), (2) the major approaches to modeling stormwater quality (Section 3.3), (3) a focused discussion on build-up and wash-off modeling formulations (Section 3.4), and (4) the practical applications of these models and their limitations (Section 3.5).

## 3.2 Sources and Dynamics of Urban Stormwater Pollutants

Sources of Pollutants: Urban runoff pollutants originate from a variety of sources, broadly including atmospheric deposition, traffic and transportation activities, building materials, and natural or landscaped areas. A recent comprehensive review by Müller et al. (2020) identified atmospheric deposition, vehicular traffic, and metallic building surfaces as three major contributor categories of contaminants in urban runoff. During dry weather, particles and chemicals from the atmosphere settle on surfaces (dry deposition), adding to the pool of surface pollutants. These can include dust and soil particles, combustion by-products (e.g. soot, tire wear particles, brake dust), and nutrients like ammonia or nitrate from air pollution. Rainfall can also bring pollutants from the air (wet deposition), washing out atmospheric contaminants directly into catchments. In densely urban environments, atmospheric deposition serves as a continuous background source of heavy metals and nutrients on impervious surfaces (Na Nagara, Sarkar and Datta, 2022).

Perhaps the most significant anthropogenic sources are related to vehicular traffic and road infrastructure. Roads accumulate a mix of pollutants from vehicles: tire wear releases zinc, cadmium, and micro-rubber particles; brake pad wear is a well-known source of copper, antimony, lead, and fine particulates; drips of motor oil and hydraulic fluids contribute petroleum hydrocarbons; vehicle exhaust deposits PAHs and trace metals; and road surface wear (asphalt or concrete) can release particulate matter (McKenzie et al., 2009; Müller et al. 2020). These pollutants concentrate on highways and busy streets, making transportation corridors hotspots for runoff contamination. Road maintenance activities (e.g. application of de-icing salts or sanding in winter, repainting of road markings) also introduce chemicals that can later appear in runoff. It has been noted that heavy metals commonly found in stormwater runoff ( Zn, Cu, Pb ) are often attributable in large part to traffic-related sources (tires, brakes, fuel combustion) as well as roof runoff and atmospheric fallout, which underscores the multiple pathways through which metals enter the urban environment (Huber, Welker and Helmreich, 2016; Pamuru et al., 2022).

Buildings and infrastructure themselves are another key group of pollutant sources**.** Roof materials (galvanized steel, copper, bitumen shingles) can leach metals (Zn from galvanized coatings, Cu from copper gutters or roofing) especially during the first flush of rain on a dry roof. Building siding, paint, and exposed metals (e.g. facades, flashing) may

similarly contribute pollutants through weathering. Urban landscapes and green spaces contribute nutrients and organic matter: fertilizers applied to lawns and gardens add nitrate and phosphate; leaf litter and grass clippings contribute organic nitrogen, phosphorus, and carbon compounds when they decay or are washed into drains (Wen, Chang and Wanielista, 2020). Soil erosion from construction sites or uncovered soil patches can be a major source of sediment (TSS) in runoff, and with the sediment often come bound nutrients and metals. Even pet wastes on streets or lawns can add bacterial contamination (pathogens) and nutrients to stormwater. In essence, urban land use and human activities dictate the type and amount of pollutants available for wash-off. Residential areas, for example, often show higher nutrients (from lawn care) and moderate TSS, while commercial or industrial areas can exhibit higher metals and hydrocarbons (Simpson, Winston and Brooker, 2022). Land use, climate, and imperviousness collectively influence runoff quality; global meta-analysis has shown that road-dominated and highly urbanized land uses yield some of the highest pollutant concentrations in runoff, while climate factors (e.g. arid climates with long dry periods) also lead to greater pollutant accumulation and subsequent runoff peaks (Bell, Tague and McMillan, 2019).

Dynamics of Pollutant Accumulation**:** Between rain events, pollutants gradually accumulate on surfaces through the processes mentioned above. This build-up phase during dry weather involves deposition of particles and droplets on impervious areas (and also pervious areas, although pollutants on pervious surfaces may infiltrate or be immobilized to some extent). The accumulation is not unlimited, many studies have observed that pollutant build-up tends to approach an upper limit over time, as processes like wind removal, street sweeping, or chemical degradation counteract further accumulation after a point. In conceptual modeling terms, a saturation mass (often denoted as $C_{max}$) can be defined, representing the maximum pollutant load that can accumulate on a unit area if dry weather persists indefinitely. Empirical observations on streets have shown, for example, that solids build-up may increase rapidly in the first few dry days and then plateau as the surface reaches an equilibrium layer of dust and debris (Wang, Huang and Li, 2019). The actual time to reach this asymptote can vary by context, in high-traffic areas, saturation might be reached in a matter of days, whereas in cleaner or periodically cleaned areas, the accumulation may remain far below any saturation. Nevertheless, the concept of build-up

reaching a saturation limit is widely used: pollutants accumulate at some rate (often approximated as linear or exponential growth initially) until a balance between deposition and removal is achieved (Wang, Huang and Li, 2019). Key parameters describing this process include the accumulation rate (the rate at which pollutant mass is added per unit area per unit time, often in mg/m² per day) and the saturation load $C_{max}$ in mg/m². These parameters are pollutant-specific and surface-specific. For instance, a high-traffic roadway might have a high daily accumulation rate of TSS and metals, but it may also get saturated quickly due to continual disturbances, whereas a low-traffic road might accumulate more slowly but reach a higher maximum load if undisturbed for long periods (Daly, Bach and Deletic, 2014).

Dynamics of Pollutant Wash-off: When rainfall commences, the accumulated pollutants become mobilized, a process referred to as wash-off. The initial stormwater runoff will pick up the dust, particles, and soluble deposits from surfaces, creating a pulse of contaminant transport. Wash-off is governed by complex interactions between rainfall characteristics (intensity, duration, drop size), surface conditions, and pollutant properties (particle size distribution, solubility, attachment to the surface). In general, higher rainfall intensities and runoff rates exert greater shear force and detachment energy on surface deposits, resulting in more vigorous wash-off of particulate pollutants (Miguntanna, 2009). Thus, a short but intense downpour can strip a surface of loose particles more effectively than a long, light event. However, even high-intensity storms have a finite capacity to wash-off pollutants. Coarser or more adherent particles may remain even after heavy rain, forming part of the residual that contributes to the next event's first flush. Egodawatta et al. (2007) demonstrated that the fraction of available pollutants washed off in an event varies primarily with rainfall kinetic energy and intensity, and that common wash-off models might need modification to account for an upper limit or "capacity" of wash-off in extremely intense storms. In their experiments on urban road surfaces, even very intense rainfall could not remove all the accumulated contaminants; a significant portion (often 20–50% of the initial load) remained, indicating strong adhesion or sheltering of some particles. This is known as wash-off capacity factor, which represents the maximum fraction of the accumulated load that a storm can remove under given conditions.

Transport and Transformation in Drainage Systems: After pollutants are washed off surfaces, they enter the stormwater drainage network (which could be a separate storm sewer system or a combined sewer that carries both runoff and sanitary wastewater). Within the conveyance system, transport processes and transformations further influence pollutant fate. A significant process is sediment deposition and resuspension in sewers or channels. As runoff travels through pipes or open drains, the flow velocity can vary; during the rising limb of a storm, velocities increase, potentially carrying a high load of suspended solids. When flow decreases (toward the end of the storm or during dry weather flow in combined sewers), heavier particles may settle out and form sediment deposits on pipe invert or channel beds. These sediments often contain attached pollutants (e.g. heavy metals, PAHs bound to particles, nutrients attached to organic matter). If another surge of high flow comes (either later in the same storm or in a subsequent event), those deposits can be re-entrained (resuspended), leading to a sudden spike in pollutant concentrations even after the surface wash-off has waned. In combined sewers, this phenomenon contributes to high pollutant loads in the first flush of combined sewer overflows (CSOs) and can cause operational challenges at treatment plants (a surge of settled solids being scoured out by increasing flow). Studies on combined sewer sediment dynamics note that during storms, sediments in sewers can be eroded or remolded into new bed forms, and each successive storm can remobilize previously deposited material. Thus, the pollutograph (pollutant concentration vs. time) at an outfall is not solely a function of surface wash-off; it also depends on in-sewer deposition/scour processes.

Chemical and biological transformations can also occur during transport. In detention basins or slow-flow sections, organic pollutants may degrade and nutrients may transform. In combined sewers during dry weather, microbial activity can cause deposited organic matter to decay, producing soluble pollutants that are later flushed out. Most stormwater models assume conservative transport, focusing on physical processes like advection, dilution, settling, and resuspension. This assumption is reasonable for fast travel times but can oversimplify cases with longer residence times or in-network treatment devices.

In summary, urban stormwater pollutant dynamics are characterized by intermittent accumulation of contaminants during dry periods followed by rapid mobilization in wet periods. Pollutant sources are diverse (from cars to construction dust to atmospheric

fallout) and thus the makeup of the pollutant mix can vary from one location or storm to another. The first flush phenomenon is a signature feature of urban runoff, reflecting the coupling of build-up and wash-off processes. And as runoff drains through engineered systems, processes like sedimentation and resuspension can further modulate pollutant delivery. These complexities pose challenges for modeling, since a model must attempt to represent these source-loading and transport processes with sufficient fidelity. The next sections delve into how different modeling approaches tackle (or simplify) these dynamics, and how build-up and wash-off are mathematically formulated in typical models.

## 3.3 Overview of Stormwater Quality Modeling Approaches

Modeling urban stormwater quality requires balancing physical realism with practical constraints (data availability, computational complexity). Over the decades, a variety of modeling approaches have been developed, generally falling into three broad categories: empirical models, conceptual (deterministic) models, and physically-based models (Obropta and Kardos, 2007). Empirical models rely on observed data and statistical relationships; conceptual models represent the system with simplified but mechanistic components (like reservoirs and fluxes); physically-based models attempt to simulate processes using fundamental physical and chemical equations with minimal simplification. Additionally, some literature distinguishes stochastic models (which explicitly incorporate randomness or probability distributions) and hybrid models (combining deterministic and stochastic elements) (Obropta and Kardos, 2007).

Empirical Models (e.g., Event Mean Concentration, pollutant rating curves, and machine learning models) use statistical descriptions of stormwater quality derived from monitoring data, without explicitly simulating the underlying processes of pollutant build-up and wash-off. The most common empirical model is the Event Mean Concentration (EMC) method. EMC is the flow-weighted average concentration of a pollutant over the course of a runoff event. Many studies and databases have reported typical EMC values for different land uses and pollutant types (Yu, Min and Kim, 2011; Sharma et al., 2012). To use the EMC approach, a modeler multiplies an EMC by the total runoff volume of an event (or period) to estimate pollutant load. This approach is widely used in practice for planning-level estimates and regulatory analyses because of its simplicity and minimal data requirements. The advantages of the empirical EMC model are that it requires no time-step

simulation and very few parameters; one essentially needs rainfall-runoff volume and an assumed concentration. It is particularly useful for long-term load estimation (e.g., annual pollutant load exported from a catchment) and for comparing loads under different land use scenarios. However, its limitations are significant: 1)it does not capture intra-event variations (it ignores first flush effects, peak concentrations, timing of pollutant delivery) (Costa, Carvalho and Koide, 2021; Maniquiz, Lee and Kim, 2010), 2) it oversimplifies complex stormwater processes, which can lead to inaccuracies in predicting pollutant concentrations (Sharifi, Massoudieh and Kayhanian, 2011; Norris, 2019), and 3) it lacks of consistent correlation between stormwater constituent concentrations and antecedent dry periods (ADP), which limits the predictive power of EMC models (Costa, Carvalho and Koide, 2021; Behrouz et al., 2022). Another empirical approach related to EMC is the use of regression models where pollutant load or concentration is related to storm characteristics (rainfall depth, peak flow, etc.) or watershed attributes (percent imperviousness, land use proportions). These regression models are calibrated on data and can provide quick predictions with uncertainty bounds. They are essentially a statistical form of modeling and similarly suffer if applied outside the range of data used to develop them.

Conceptual (Deterministic) Models: Conceptual models represent a middle ground, attempting to capture the essence of rainfall-runoff pollutant processes with simplified, lumped formulations. They are termed deterministic because they use cause-effect equations (often ordinary differential equations or algebraic equations) to describe processes like pollutant build-up and wash-off, usually without random components (though parameters may be calibrated to data). A prime example of conceptual models are the classic build-up/wash-off models, which are implemented in widely used software such as the U.S. EPA's Storm Water Management Model (SWMM) and many other urban drainage modeling tools (Egodawatta, Thomas and Goonetilleke, 2007). In such models, the catchment (or each land use within the catchment) is conceptualized as accumulating pollutants during dry periods and then losing them during rain according to some mathematical function (Al Ali, Bonhomme and Chebbo, 2016). These models typically operate at the scale of an individual catchment or even sub-catchments (with different land uses), often on an event or continuous simulation basis with many events. Deterministic

conceptual models have been extremely influential and are the basis of most urban stormwater quality simulation software. They allow users to input catchment characteristics (area, imperviousness, land use, etc.) and calibrate a few key parameters (like accumulation rates, wash-off coefficients). Because they explicitly simulate the processes in a temporal manner (e.g., computing runoff and pollutant discharge at each time step), they can capture phenomena like the first flush, the effect of varying rainfall intensity, and the depletion of pollutants through consecutive storms (Di Modugno et al., 2015). Conceptual models can also be extended to include terminals like treatment devices or receiving water bodies, making them versatile for planning and design. Deterministic models are the most commonly used and have the benefit of directly leveraging our understanding of processes, but they also come with prediction errors and uncertainties due to the simplified representation of highly variable processes (Obropta and Kardos (2007)). Indeed, while conceptual models are more process-faithful than pure empirical ones, they still make broad assumptions (e.g., treating the entire impervious area load as one homogeneous pollutant stock, or assuming immediate mixing of washed-off pollutants in runoff). As a result, calibration is usually necessary to adjust the model to local conditions (Wijesiri et al., 2016; Behrouz et al., 2024). Despite this, these models are central to stormwater management, for example, they enable scenario analysis such as "if we add street sweeping or pervious pavement, how much reduction in pollutant loads can we expect?" by altering the source or wash-off parameters in the model. These help urban planners integrate water quality considerations (often termed Water Sensitive Urban Design or Low Impact Development) by providing a relatively easy-to-use yet process-based framework.

Physically-Based Models: Physically-based models aim to represent processes with high fidelity to physical, chemical, and even biological mechanisms. In practice, such a comprehensive model is not feasible for routine application as the urban environment is exceedingly complex, and data to drive or validate such detailed simulations are lacking in most cases. However, certain components of physically-based modeling are sometimes implemented. For instance, the hydrologic and hydraulic part of stormwater models may use physically-based equations like the Saint-Venant equations for channel flow routing or even for overland flow routing in advanced cases.

When it comes to water quality, however, most models revert to conceptual or simplified approaches (like assuming pollutants are uniformly mixed in each pipe reach and using decay coefficients). Based on literatures, no broad urban catchment model currently simulates pollutant build-up and wash-off from first principles (e.g., deriving deposition rates from atmospheric pollutant concentrations and turbulence, or computing wash-off by solving Navier-Stokes equations for each particle) (Yuan, Sinshaw and Forshay, 2020; Hong et al., 2016). Instead, when we refer to "physically-based" in the context of catchment models, we usually mean models that have a higher degree of spatial and temporal resolution and use fundamental equations where possible. An example is the integration of distributed hydrology models (like fully distributed rainfall-runoff models) with water quality routing.

There is, however, emerging research on more physically based pollutant wash-off formulations, including raindrop/flow-driven detachment frameworks and computational fluid dynamics (CFD) at feature scale (Tong et al., 2025). These approaches show promise for advancing process understanding but are not yet integrated as full, city-scale pollutant modules into standard catchment software used by practitioners. Consequently, mainstream models remain hybrids: they combine physically based flow simulation with simplified pollutant representations.

The benefit of physically-based models is their potential accuracy and realism when provided with high-quality input data (Dotto et al., 2010). They are better at extrapolation (predicting outside of past observed conditions) because they rely on fundamental principles. For instance, if rainfall intensities increase under climate change, a physically-based erosion model could predict increased wash-off without needing new calibration, as long as the physics holds (Wijesiri et al., 2020). However, the drawbacks are significant: they require a large number of inputs (many of which may be unknown or highly uncertain in urban settings, like detailed land surface roughness, particle size distribution of street dust, etc.), and they can be computationally demanding (Devia, Ganasri and Dwarakish, 2015). Moreover, model complexity does not always translate to better predictive performance because errors in many input parameters can accumulate (the "over-parameterization" problem) (Del Giudice et al., 2015). In many cases, simpler models

calibrated appropriately can perform just as well as complex models for the purpose of load prediction, due to the stochastic nature of storm events and pollutant variability.

Stochastic and Hybrid Approaches: There are models that incorporate stochastic elements or combine approaches. Stochastic models might treat pollutant wash-off as a random process characterized by probability distributions, for example, generating random realizations of EMCs based on statistical properties, or using Monte Carlo simulation to account for uncertainty in parameters. These models acknowledge the inherent variability and uncertainty in stormwater quality (which can be quite large, even under similar conditions). Hybrid models attempt to leverage strengths of both deterministic and stochastic approaches. According to Obropta and Kardos (2007), hybrid approaches show promise in reducing prediction error and uncertainty in stormwater quality modeling.

In summary, the landscape of stormwater quality models ranges from simple empirical methods like EMCs (easy to apply but limited in representing dynamics) to conceptual deterministic models which form the backbone of most urban water quality simulation, to more physically-based or advanced models which attempt a finer-scale representation of processes at the cost of complexity. The choice of modeling approach depends on the objectives: for screening or broad-scale load estimates, empirical models might suffice; for design and planning of specific interventions and understanding first flush behavior, conceptual models are typically favored; for research investigations into detailed processes or when addressing scenarios outside the range of available calibration data, more physically-based or stochastic models might be explored. Crucially, the performance of any model must be evaluated against observed data, and the next section (2.4) will focus on the core component common to many modeling approaches (the build-up and wash-off formulations) which lie at the heart of how we mathematically represent pollutant mobilization in urban runoff.

## 3.4 Build-up and Wash-off Modeling

A central component of many urban stormwater quality models is the build-up/wash-off framework, which conceptually divides pollutant behavior on catchment surfaces into two phases: accumulation during dry weather (build-up) and removal during rainfall (wash-off). This framework, first introduced in the 1970s through empirical observations of street cleaning and storm events (e.g., Sartor and Boyd, 1972), remains widely used in

contemporary models like SWMM, InfoWorks ICM, and others (Egodawatta et al. (2007; Hossain et al., 2010).

Conceptual Framework:

In the build-up/wash-off model, the catchment surface (or each land use area) is treated as a storage compartment for pollutants. During periods with no rainfall, pollutants from various sources deposit on the surface. The build-up phase can be influenced by time (duration of dry period) and other factors like traffic or human activity during dry weather. When rain starts, a portion of the stored pollutants is dislodged and transported with runoff, the wash-off phase. Not all the stored pollutant is typically removed; some fraction may remain attached or in pores, especially if the rain event is not intense enough to scour everything.

The core assumptions usually are: (1) Build-up can be described as a monotonic increasing function of dry time, approaching an asymptotic maximum (Cmax) if the dry period is very long. (2) Wash-off rate at any time during rainfall is proportional to some measure of runoff energy (rainfall intensity, runoff flow, or shear stress) and to the amount of pollutant remaining on the surface at that time.

**Common Build-up Formulations:**

Three of the most widely used build-up formulations are the exponential build-up function, the power law (or linear until saturation) function and saturation (half-saturation) function (Hossain et al., 2010; Rossman, 2004; Charbeneau and Barret, 1998):

1- Exponential Build-up:

$$B(t) = C_{max}(1 - e^{-k_b t})$$

Where

- *B(t)* is the accumulated pollutant mass per unit area at time (*t*)
- $C_{max}$ is the saturation mass (the asymptotic maximum build-up) (mass/area),
- and $k_b$ is a build-up rate constant (1/time).

This equation starts with B(0) equal to zero (immediately after a wash-off, no pollutant on surface) and increases toward $C_{max}$ as time approaches infinity. The time constant one divided by $k_b$ represents how quickly the accumulation reaches near-saturation. An

exponential build-up is attractive because it captures the diminishing accumulation over long dry periods (fast initial build-up that slows down later).

2- Power Law Build-up:

$$B(t) = at^b, \qquad B(t) \leq C_{max}$$

where

- $a$ is scaling coefficient (mass/area)/time$^b$
- $b$ is an exponent (dimensionless)

The power law can be more flexible than the exponential form; for instance, one might fit a power curve to observed accumulation data if available.

3- Saturation Function Build-up

$$B(t) = \frac{C_1 t}{p + t}$$

where

- $p$ is the half-saturation time (time to reach 0.5 $Cmax$).

This form increases linearly at small $t$ and approaches $C_{max}$ asymptotically, with $p$ providing a physically interpretable time scale.

Table 3.1 present a summarization of the three build-up formulations, with a comparison of their mathematical form, and key properties.

Table 3.1. Comparison of common pollutant build-up formulations in urban stormwater quality modelling.

| **Formulation** | **Equation** | **Key Properties** | **Typical Uses / Advantages** |
|---|---|---|---|
| Exponential | $B(t) = C_{max}(1 - e^{-k_b t})$ | Rapid early growth; slows as saturation approached; time constant $1/k_b$ controls approach speed | Widely used; simple interpretation of $k_b$; captures diminishing returns over long dry periods |

| | | | |
|---|---|---|---|
| Power-law | $B(t) = at^b$, $B(t) \leq C_{max}$ | Shape highly flexible ($b$<1 fast early growth; $b$>1 slow start); requires explicit cap | Fits a wider range of empirical curves; used where accumulation data suggest non-exponential shape |
| Saturation (half-time) | $B(t) = \frac{C_1 t}{p + t}$ | Linear at small t; $B(p)$=0.5 $C_{max}$; intuitive time-scale parameter | Parameters easy to interpret; useful when half-time is known from observation |

Empirical studies have reported a wide range of parameter values. For example, dust and dirt accumulation on streets may increase by tens to hundreds of kg/ha/day of total suspended solids (TSS) during the first few dry days, with accumulation typically plateauing after about one week if undisturbed (Gülbaz and Kazezyilmaz-Alhan, 2015). In practice, $C_{max}$ and the build-up rate constant ($k_b$, or analogous parameters in other formulations) are key model inputs that usually require calibration, or estimation from literature, for the specific land use and pollutant type.

**Common Wash-off Formulations**:

Wash-off is often described either by an exponential decay model or a rating curve model. A classic formulations used in literature are:

1- Exponential Wash-off (flow-dependent)

$$\frac{dB}{dt} = -k_w q(t)^n B(t)$$

Where:

- $B(t)$ = remaining pollutant mass per unit area on the surface at time $t$; (mass/length$^2$)
- $q(t)$ = runoff flow per unit area (length/time)
- $k_w$ = wash-off coefficient (mass/length$^2$/time)
- $n$ = dimensionless exponent (often $n$ = 1 in many applications)

This differential equation means the rate of pollutant removal ($dB/dt$, negative) is proportional to:

1. The amount of pollutant present (B)

2. Some power of runoff ($q^n$), which captures the fact that higher flows cause faster wash-off.

If $n = 1$, the equation integrates to an exponential decay:

$$B(t) = B(0)e^{-k_w \int_0^t q(t')dt'}$$

For constant rainfall intensity, it simplifies to:

$$B(t) = B(0)e^{-k_w qt}$$

The instantaneous wash-off load $W$ (mass/time) is simply the rate of change in $B$:

$$W(t) = -\frac{dB}{dt} = k_w q(t)^n B(t)$$

2- Event Mean Concentration (EMC) Wash-off formulation

The EMC approach assumes that the pollutant concentration in runoff remains constant throughout a rainfall event. The pollutant load is calculated as the product of this concentration and the total runoff volume (or instantaneous flow rate).

$$W\ (t) = C_{EMC} Q(t)$$

Where:

- $W\ (t)$= pollutant wash-off load rate (mass/time)
- $C_{EMC}$ = Event Mean Concentration (mass/volume)
- $Q$ = runoff flow rate (L/s), which can be calculated as

$$Q(t) = q(t) f_{lu} A$$

- $f_{lu}$ = fraction of land use,
- $A$ = catchment area (m²)

3- Rating Curve (concentration-based) Wash-off Models

A more empirical approach relates runoff concentration to flow rate via a regression:

$$C(t) = \alpha Q(t)^{\beta}$$

Where:

- $C(t)$ for instantaneous concentration
- $\alpha, \beta$ = regression coefficients

The corresponding wash-off load rate is:

$$W(t) = C(t)Q(t) = \alpha Q(t)^{\beta+1}$$

Table 3.2 present a summarization of the three wash-off formulations, with a comparison of their mathematical form, and key properties.

Table 3.2. Comparison of common pollutant build-up formulations in urban stormwater quality modelling.

| Formulation | Equation | Key Properties | Typical Uses / Advantages |
|---|---|---|---|
| Exponential | $W = k_w q^n B$ | Mass-conservative with build-up; captures first-flush via time-varying $q$; closed-form solution | Used where continuous simulation with mass balance is needed; suitable for linking to build-up models |
| EMC (Event Mean Concentration) | $W = C_{EMC} Q$ | Simple; aligns with monitoring summaries; minimal parameters | Useful when only event-level concentration and flow data are available |
| Rating curve | $W = \alpha Q^{\beta+1}$ | Empirically captures first-flush patterns; easy to fit from outlet data | Applied when first-flush analysis from outlet monitoring is desired |

The key parameters for wash-off include the wash-off coefficient ($k_w$, or $n$ in the equation above). These parameters encapsulate complex realities: they are effectively lumped representations of detachment efficiency, surface roughness effect, particle size distribution, etc. For instance, a smooth asphalt surface might have a higher wash-off efficiency (larger $k_w$) for the same pollutant compared to a grassy surface where pollutants are more retained. Similarly, coarse heavy particles might behave as if $k_w$ is small, requiring more runoff to move. Wash-off parameters are therefore often calibrated per pollutant and per surface type. Experimental studies, like that of Egodawatta et al. (2007), have tried to tie these parameters to measurable physical quantities (rainfall kinetic energy, etc.). They found that introducing a wash-off capacity concept improved the fit of models, essentially acknowledging that $k_w$ is not constant but can depend on rainfall intensity. Some recent

models incorporate such refinements by varying wash-off coefficients with intensity or using multiple size classes of particles, each with their own wash-off behavior (Wijesiri et al., 2016).

## 3.5 Applications and Limitations

Rainfall-runoff water quality models. Key applications include:

• Planning and Design of Pollution Control Strategies: Urban water quality models inform the design of stormwater control measures and best management practices (BMPs). By simulating scenarios with and without these controls, models help size facilities and optimize their placement. In practice, regulatory guidelines often specify performance targets, models are used to demonstrate that a proposed design meets such targets. On a strategic level, models support stormwater management plans by highlighting pollution "hotspots" and evaluating alternative strategies. Because they can predict long-term averages and extreme event concentrations, models are crucial for designing systems robust to various conditions.

• Risk Assessment and Environmental Impact Assessment: Modeled outputs assess the risk posed by urban runoff to receiving waters. Such analyses are a staple of EIAs. Models also contribute to TMDLs by quantifying sources and needed reductions. Regulatory programs (EU WFD, US CWA 303(d)) require apportioning loads and testing scenarios; outputs inform risk assessments. Improved stormwater quality models benefit CSO management, TMDLs, BMP design, land-use impact assessment, water quality trading, and integrated urban water management (Rezaei et al., 2019).

• Urban Development and Master Planning: Models support master planning and low-impact development by evaluating growth effects and the benefits of green infrastructure. Many cities apply a "no net increase in stormwater pollution" principle enforced through modeling. Stormwater modeling tools planning calculators help ensure developments meet discharge standard. Climate-change assessments use continuous simulation to quantify effects for adaptation planning.

• Policy and Regulatory Compliance: Under MS4 permits and the EU WFD, cities model loads, track reductions, and evaluate scenarios. Some jurisdictions use predictions for real-time operations. Outputs guide management decisions and investment

prioritization.

Given these applications, stormwater quality models are linked to a “digital twin” of the urban water system for scenario testing.

Limitations of Current Models:

• (a) Simplified Representation: Lumped build-up/wash-off omits many processes; sewer mixing and transformations are approximated, so models capture trends rather than all nuances. Parameters may lose physical meaning at larger scales (Bonhomme and Petrucci, 2017). Many models focus on few pollutants and simple decay while real biogeochemistry can be complex.

• (b) Uncertainty and Equifinality: High event-to-event variability and multiple acceptable parameter sets create predictive uncertainty. It is therefore essential to perform sensitivity and uncertainty analyses and provide proper ranges. Using models outside calibration bounds reduces reliability; data scarcity is a common constraint (Wijesiri et al., 2016).

• (c) Scale and Aggregation Issues: Homogeneous sub-catchment assumptions can mask hotspots; finer resolution helps but needs data.

• (d) Evolving Pollutant Concerns: Emerging contaminants (e.g., microplastics, pesticides, pharmaceuticals) require new parameterization/processes.

• (e) Calibration Bias and Human Factors: Analyst choices affect results; automated methods help but still require judgment.

In summary, stormwater quality models remain powerful decision-support tools, but their application must account for structural simplifications, parameter uncertainty, monitoring gaps, and emerging challenges (microplastics, PFAS, climate non-stationarity). Ongoing research is moving towards particle-size-aware modelling, hybrid machine learning, continuous monitoring, and stochastic–deterministic formulations to reduce uncertainty.

## References

Al Ali, S., Bonhomme, C. and Chebbo, G., 2016. Evaluation of the performance and the predictive capacity of build-up and wash-off models on different temporal scales. *Water*, 8(8), p.312.

Altobelli, M., Cipolla, S.S. and Maglionico, M., 2020. Combined application of real-time control and green technologies to urban drainage systems. Water, 12(12), p.3432.

Anil, I., Alagha, O., Blaisi, N.I., Mohamed, I.A., Barghouthi, M.H. and Manzar, M.S., 2019. Source identification of episodic rain pollutants by a new approach: combining satellite observations and backward air mass trajectories. *Aerosol and Air Quality Research*, 19(12), pp.2827–2843.

Bach, P.M., McCarthy, D.T., Deletic, A., 2010. Redefining the stormwater first flush phenomenon. *Water Res.* 44, pp.2487–2498.

Barco, J., Papiri, S. and Stenstrom, M.K., 2008. First flush in a combined sewer system. *Chemosphere*, 71(5), pp.827–833.

Behrouz, M.S., Nayeb Yazdi, M., Sample, D.J., Scott, D. and Owen, J.S. Jr, 2022. What are the relevant sources and factors affecting event mean concentrations (EMCs) of nutrients and sediment in stormwater? *Science of the Total Environment*, 828, p.154368.

Behrouz, M.S., Sample, D.J., Kisila, O.B., Harrison, M., Nayeb Yazdi, M. and Garna, R.K., 2024. Parameterization of nutrients and sediment build-up/wash-off processes for simulating stormwater quality from specific land uses. *Journal of Environmental Management*, 358, p.120768.

Bell, C.D., Tague, C.L. and McMillan, S.K., 2019. Modeling runoff and nitrogen loads from a watershed at different levels of impervious surface coverage and connectivity to storm water control measures. *Water Resources Research*, 55(4), pp.2690–2707.

Bonhomme, C. and Petrucci, G., 2017. Should we trust build-up/wash-off water quality models at the scale of urban catchments?. *Water Research*, 108, pp.422–431.

Charbeneau, R.J. and Barret, M.E., 1998. Evaluation of methods for estimating stormwater pollutant loads. *Water Environment Research*, 70(7), pp.1295–1302.

Cipolla, S.S., Maglionico, M. and Stojkov, I., 2016. A long-term hydrological modelling of an extensive green roof by means of SWMM. Ecological Engineering, 95, pp.876–887.

Conley, D.J., Paerl, H.W., Howarth, R.W., Boesch, D.F., Seitzinger, S.P., Havens, K.E., Lancelot, C. and Likens, G.E., 2009. Controlling eutrophication: nitrogen and phosphorus. *Science*, 323(5917), pp.1014–1015.

Costa, M.E.L., Carvalho, D.J. and Koide, S., 2021. Assessment of pollutants from diffuse pollution through the correlation between rainfall and runoff characteristics using EMC and first flush analysis. *Water*, 13(18), p.2552.

Daly, E., Bach, P.M. and Deletic, A., 2014. Stormwater pollutant runoff: A stochastic approach. *Advances in Water Resources*, 74, pp.148–155.

Del Giudice, D., Reichert, P., Bareš, V., Albert, C. and Rieckermann, J., 2015. Model bias and complexity – Understanding the effects of structural deficits and input errors on runoff predictions. *Environmental Modelling & Software*, 64, pp.205–214.

Devia, G.K., Ganasri, B.P. and Dwarakish, G.S., 2015. A review on hydrological models. *Aquatic Procedia*, 4, pp.1001–1007.

Di Modugno, M., Gioia, A., Gorgoglione, A., Iacobellis, V., La Forgia, G., Piccinni, A.F. and Ranieri, E., 2015. Build-up/wash-off monitoring and assessment for sustainable management of first flush in an urban area. *Sustainability*, 7(5), pp.5050–5070.

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T. and Rauch, W., 2010. Stormwater quality models: performance and sensitivity analysis. *Water Science and Technology*, 62(4), pp.837–843.

Drapper, D., Olive, K., McAlister, T., Coleman, R. and Lampard, J.L., 2022. A review of pollutant concentrations in urban stormwater across eastern Australia, after 20 years. Frontiers in Environmental Chemistry, 3, p.853764.

Drapper, D., Olive, K., McAlister, T., Coleman, R. and Lampard, J.-L., 2022. A review of pollutant concentrations in urban stormwater across eastern Australia, after 20 years. *Frontiers in Environmental Chemistry*, 3, p.853764.

Egodawatta, P., Thomas, E. and Goonetilleke, A., 2007. Mathematical interpretation of pollutant wash-off from urban road surfaces using simulated rainfall. *Water Research*, 41(13), pp.3025–3031.

Gülbaz, S. and Kazezyilmaz-Alhan, C.M., 2015. Investigating the effects of low impact development (LID) on surface runoff and TSS in a calibrated hydrodynamic model. *Journal of Urban and Environmental Engineering*, 9(2), pp.91–96.

Herath, H.M.V.V., Chadalawada, J. and Babovic, V., 2021. Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. *Hydrology and Earth System Sciences*, 25(8), pp.4373–4401.

Hong, Y., Bonhomme, C., Le, M.-H. and Chebbo, G., 2016. New insights into the urban washoff process with detailed physical modelling. *Science of the Total Environment*, 573, pp.924–936.

Hossain, I., Imteaz, M.A., Goonetilleke, A. and Egodawatta, P., 2010. Development of a catchment water quality model for continuous simulations of pollutant build-up and wash-off. *Environmental Modelling & Software*, 25(9), pp.999–1009.

Huber, M., Welker, A. and Helmreich, B., 2016. Critical review of heavy metal pollution of traffic area runoff: occurrence, influencing factors, and partitioning. *Science of the Total Environment*, 541, pp.895–919.

Huber, M., Welker, A. and Helmreich, B., 2016. Critical review of heavy metal pollution of traffic area runoff: Occurrence, influencing factors, and partitioning. *Science of the Total Environment*, 541, pp.895–919.

Lee, J.H., Bang, K.W., Ketchum, L.H. Jr, Choe, J.S. and Yu, M.J., 2002. First flush analysis of urban storm runoff. *Science of the Total Environment*, 293(1–3), pp.163–175.

Mamun, A.A., Shams, S. and Nuruzzaman, M., 2020. Review on uncertainty of the first-flush phenomenon in diffuse pollution control. *Applied Water Science*, 10(1), p.53.

Maniquiz, M.C., Lee, S. and Kim, L.-H., 2010. Multiple linear regression models of urban runoff pollutant load and event mean concentration considering rainfall variables. *Journal of Environmental Sciences*, 22(6), pp.946–952.

McKenzie, E.R., Money, J.E., Green, P.G. and Young, T.M., 2009. Metals associated with stormwater-relevant brake and tire samples. *Science of the Total Environment*, 407(22), pp.5855–5860.

Miguntanna, N.P., 2009. *Nutrients build-up and wash-off processes in urban land uses*. PhD thesis. Queensland University of Technology.

Mohiuddin, K.M., Ogawa, Y.Z.H.M., Zakir, H.M., Otomo, K. and Shikazono, N., 2011. Heavy metals contamination in water and sediments of an urban river in a developing country. *International Journal of Environmental Science & Technology*, 8(4), pp.723–736.

Na Nagara, V., Sarkar, D. and Datta, R., 2022. Phosphorus and heavy metals removal from stormwater runoff using granulated industrial waste for retrofitting catch basins. *Molecules*, 27(21), p.7169.

Niazkar, M., Evangelisti, M., Peruzzi, C., Galli, A., Maglionico, M. and Masseroni, D., 2024. Investigating first flush occurrence in agro-urban environments in Northern Italy. Water, 16(6), p.891.

Norris, T., 2019. *Evaluating the use of event mean concentration models for the management of urban drainage systems*. PhD thesis. University of Sheffield.

Obropta, C.C. and Kardos, J.S., 2007. Review of urban stormwater quality models: deterministic, stochastic, and hybrid approaches. *JAWRA Journal of the American Water Resources Association*, 43(6), pp.1508–1523.

Pamuru, S.T., Forgione, E., Croft, K., Kjellerup, B.V. and Davis, A.P., 2022. Chemical characterization of urban stormwater: Traditional and emerging contaminants. *Science of the Total Environment*, 813, p.151887. https://doi.org/10.1016/j.scitotenv.2021.151887

Rezaei, A.R., Ismail, Z., Niksokhan, M.H., Dayarian, M.A., Ramli, A.H. and Shirazi, S.M., 2019. A quantity–quality model to assess the effects of source control stormwater management on hydrology and water quality at the catchment scale. *Water*, 11(7), p.1415.

Rossman, L.A., 2004. *SWMM (Storm Water Management Model) Version 5 User's Manual*. Washington, D.C.: U.S. Environmental Protection Agency.

Sartor, J.D. and Boyd, G.B., 1972. *Water pollution aspects of street surface contaminants*. Vol. 2. Washington, D.C.: U.S. Government Printing Office.

Sharifi, S., Massoudieh, A. and Kayhanian, M., 2011. A stochastic stormwater quality volume-sizing method with first flush emphasis. *Water Environment Research*, 83(11), pp.2025–2035.

Simpson, I.M., Winston, R.J. and Brooker, M.R., 2022. Effects of land use, climate, and imperviousness on urban stormwater quality: A meta-analysis. *Science of the Total Environment*, 809, p.152206.

Todeschini, S., Manenti, S., Creaco, E. 2019. Testing an innovative first flush identification methodology against field data from an Italian catchment. *Journal of Environmental Management*, 246, pp.418-425.

Todeschini, S., 2024. Innovative and Reliable Assessment of Polluted Stormwater Runoff for Effective Stormwater Management. *Water* 16(1), p.16.

Tong, X., Liang, Q., Sander, G., Wang, G. and Lai, X., 2025. A physically based model for non-point source pollutant wash-off process over impervious surfaces. *Water Resources Research*, 61(6), p.e2024WR038791.

Wang, J., Huang, J.J. and Li, J., 2019. A study of the road sediment build-up process over a long dry period in a megacity of China. *Science of the Total Environment*, 696, p.133788.

Wen, D., Chang, N.-B. and Wanielista, M.P., 2020. Assessing nutrient removal in stormwater runoff for urban farming with iron filings-based green environmental media. *Scientific Reports*, 10(1), p.9379.

Werbowski, L.M., Gilbreath, A.N., Munno, K., Zhu, X., Grbic, J., Wu, T., Sutton, R., Sedlak, M.D., Deshpande, A.D. and Rochman, C.M., 2021. Urban stormwater runoff: a major pathway for anthropogenic particles, black rubbery fragments, and other types of microplastics to urban receiving waters. *ACS ES&T Water*, 1(6), pp.1420–1428.

Wijesiri, B., Bandala, E., Liu, A. and Goonetilleke, A., 2020. A framework for stormwater quality modelling under the effects of climate change to enhance reuse. *Sustainability*, 12(24), p.10463.

Wijesiri, B., Egodawatta, P., McGree, J. and Goonetilleke, A., 2016. Understanding the uncertainty associated with particle-bound pollutant build-up and wash-off: A critical review. *Water Research*, 101, pp.582–596.

Wijesiri, B., Egodawatta, P., McGree, J. and Goonetilleke, A., 2016. Assessing uncertainty in pollutant build-up and wash-off processes. *Environmental Pollution*, 212, pp.48–56.

Yang, Y.Y. and Lusk, M.G., 2018. Nutrients in urban stormwater runoff: Current state of the science and potential mitigation options. *Current Pollution Reports*, 4(2), pp.112–127.

Yuan, L., Sinshaw, T. and Forshay, K.J., 2020. Review of watershed-scale water quality and nonpoint source pollution models. *Geosciences*, 10(1), p.25.

# Chapter 4 – Evaluating Rainfall Runoff Simulation Using Different Hydrological Approaches

## 1. Introduction

As outlined in Chapter 2, urbanization increases impervious surfaces and hydrologic connectivity, resulting in more rapid and amplified runoff responses. Event-based modelling approaches are often preferred in such contexts for their ability to capture storm-scale dynamics and peak discharges. A major challenge, however, lies in selecting appropriate hydrologic formulations within semi-distributed conceptual models such as EPA-SWMM. Two widely used formulations, the non-linear reservoir (N-LR) and the unit hydrograph (UH, RTK method). The N-LR method conceptualizes subcatchments as interconnected storages defined by physically meaningful parameters, while the UH method describes rainfall–runoff response through triangular hydrographs defined by empirical coefficients. Each approach represents a different balance between physical detail and calibration complexity. This chapter investigates the implications of these modelling choices for urban runoff simulation, with a particular focus on calibration complexity, performance, and reliability.

## 2. Comparison of Non-Linear Reservoir and Unit Hydrograph Formulations in EPA-SWMM: Implications for Urban Runoff Simulation

The primary objective of this study was to evaluate and compare the performance of the N-LR and UH formulations in simulating event-based runoff in an urban catchment. Specifically, the study examined:

- how well each formulation reproduced outlet hydrographs across multiple storm events,
- their relative accuracy in predicting peak discharge, total volume, and time-to-peak, and
- the trade-offs between parameterization burden, calibration complexity, and data requirements.

The results of this analysis (adapted based on the paper) are detailed in the paper titled *"*Comparison of Non-Linear Reservoir and Unit Hydrograph Formulations in EPA-SWMM: Implications for Urban Runoff Simulation*"* published in the *Hydrology*.

*hydrology*

MDPI

*Article*

# Comparison of Nonlinear Reservoir and UH Algorithms for the Hydrological Modeling of a Real Urban Catchment with EPASWMM

**Carlo Giudicianni, Mohammed N. Assaf, Sara Todeschini and Enrico Creaco ***

Dipartimento di Ingegneria Civile e Architettura, Università degli Studi di Pavia, Via Ferrata 3, 27100 Pavia, Italy
* Correspondence: creaco@unipv.it

**Abstract:** This paper presents a comparative analysis of two hydrological models in the Storm Water Management Model (SWMM) software, namely, the non-linear reservoir (N-LR) and the unit hydrograph (UH), on the urban catchment of Cascina Scala, Pavia in Italy. The two models were applied for the simulation of the rainfall-runoff transformation in the 42 sub-catchments in Cascina Scala, while flow routing in the underground channels was simulated by means of the De Saint-Venant equations. A dataset of rainfall and runoff for 14 events from 2000 to 2003 was adopted for the calibration and validation of the models. The calibration was performed on 7 out of the 14 events by maximizing the fit of modeled-to-measured hydrographs in the final channel of the system. Prediction performance was assessed through different indices. Results from both models fit measured data well in terms of the total hydrograph. Whereas the time to peak was reliably predicted by both models, the N-LR was found to slightly outperform the UH in terms of total volume and peak flow prediction, though it requires a more detailed knowledge of the system for its implementation. Accordingly, the UH must be preferred in the case of a scarcity of data.

**Keywords:** hydrological modeling; rainfall-runoff; storm water management; unit hydrograph; calibration and validation; genetic algorithm



**Citation:** Giudicianni, C.; Assaf, M.N.; Todeschini, S.; Creaco, E. Comparison of Nonlinear Reservoir and UH Algorithms for the Hydrological Modeling of a Real Urban Catchment with EPASWMM. *Hydrology* **2023**, *10*, 24. https://doi.org/10.3390/hydrology10010024

Academic Editor: Andrea Petroselli

Received: 22 December 2022
Revised: 11 January 2023
Accepted: 12 January 2023
Published: 16 January 2023



## 1. Introduction

The high rate of urbanization and the transformation of vegetated areas into impervious ones have significantly impacted water hydrological processes [1]. An increase in urbanization potentially results in increased runoff volumes, flash floods, and intensive water quality degradation [2,3]. Moreover, climate change consequences are putting more pressure on water network systems through extreme precipitation events [4,5]. Therefore, the analysis and design of urban catchments has become a crucial subject to achieve sustainable water resource management. In this regard, an accurate estimation of rainfall-runoff processes is of crucial importance for proper drainage system analysis and planning, water storage design, and flood mitigation measures. However, rainfall-runoff transformation is a complex hydrological task involving several variables related to rainfall patterns, catchment characteristics, soil type, and physiography [6]. Indeed, several physical processes are present in urban hydrology, such as precipitation and interception, infiltration, water movement through saturated and unsaturated soil and groundwater, evaporation, transpiration, and runoff [7]. The degree of complexity varies with the catchment's spatial and temporal scales. Rainfall-runoff models can be categorized depending on their spatial representation (lumped or distributed) and/or temporal representation (event-based or continuous time) [8,9]. Lumped models consider space-average values of model parameters to represent the entire catchment [10]. Distributed models consider, instead, space-variable parameters in the rainfall-runoff transformation [11]. Although distributed models are helpful in comprehending the physics of hydrological processes, they need accurate calibration of several parameters, which represents a very challenging task [12].

The rainfall-runoff models can be calibrated using event-based or continuous storm event simulation. Event-based models consider a single rainfall event with a total duration of some hours. Continuous simulation models instead use an extended period made up of rain events and inter-event dry periods, with a total duration ranging from months to years. In general, for the event-based approach, losses include only the infiltration process during the simulation, whereas in continuous modeling, the evapotranspiration process is also taken into account [13]. Several research works have used continuous calibration models [14–18]. Continuous calibration delivers an adequate estimation in terms of total runoff volume [19,20]. Event-based calibration instead provides a better estimation of time to peak, peak flow rate, and overall hydrograph shape, which are critical prerequisites for developing an effective pollutant dynamic modeling.

Distributed models typically use conceptual hydrological modeling to simulate rainfall-runoff transformation outside the channel network. The unit hydrograph (UH) [21] and the non-linear reservoir (N-LR) [22] are two methods often adopted in this hydrological modeling. In the N-LR, the surface runoff is generated by modeling the study catchment area as a nonlinear reservoir. In the UH, the surface runoff is reconstructed with the convolution integral, starting from the response of the catchment to an instantaneous input of rainfall with a unitary depth.

The open-source software Storm Water Management Model, created by the United States Environmental Protection Agency (US EPA-SWMM), [23] is one of the most frequently used software programs to simulate the rainfall-runoff transformation and the related hydrological processes in urban catchments [24]. Either the N-LR or the UH can be used in the US EPA-SWMM to model the formation of runoff over external catchment surfaces. Then, the De Saint-Venant equations are used for modeling the flow routing in underground channels, which are fed by external catchment surfaces. EPA-SWMM can be used to perform event-based or continuous simulations and can be applied to combined sewer systems, sanitary sewer systems, and catchments containing storm drains [23,25]. The accuracy and reliability of rainfall-runoff models are highly dependent on how the parameter values are defined [26,27]. Some of these can be physically measured from the catchment characteristics (e.g., area and slope). Others require parameterization based on high-quality rainfall-runoff measured data. This parameterization consists of finding the set of parameters that maximize the fit of model results to experimental observation. Traditionally, this process is carried out manually by using the trial-and-error method. However, manual parameterization is time consuming and subjective, and it depends on expert judgement. Furthermore, it may fail to reach the global optimal solution [28]. Therefore, automatic calibration methods using computer-based tools have been widely developed and used to obtain a more robust and efficient estimation of the best-fit parameter values [29–31]. In general, automatic calibration algorithms can be classified into deterministic and stochastic methods. Deterministic optimization algorithms are developed to locate the optimal parameter set close to a starting solution. Therefore, they are generally fast local optimization methods. To avoid getting trapped in local optima, stochastic optimization methods can be used, though they generally feature a higher computational overhead. The presence of many parameters [25] makes stochastic optimization methods preferable in hydrological models [32–34]. The genetic algorithm (GA) [35] is one of the efficient optimization approaches known for its capability of detecting global optimal solutions. It is a population-based heuristic algorithm inspired by the concept of natural selection that drives biological evolution. The GA has been applied successfully to several environmental research problems [36–38]. It is also widely used to optimize the parameters of rainfall-runoff transformation [39,40].

This research work attempts to develop an automatic calibration method for the event-based rainfall-runoff transformation in a small urban catchment. The main novelty of the present work consists of the application and comparison of the two rainfall-runoff models N-LR and UH in the framework of the EPA-SWMM software. In fact, to the best of our

knowledge, no previous work has ever attempted to compare the performance of these models on a real case study.

The following sections describe the case study, methodology and applications, followed by the discussion and conclusions.

## 2. Case Study

The selected study area is Cascina Scala, which is located in northern Pavia, Italy [41] (Figure 1). It is an urban catchment hosting 1500 residents. The contribution area is 12.7 ha, where 7.9 ha (62%) is impervious, and 4.8 ha (38%) is pervious.

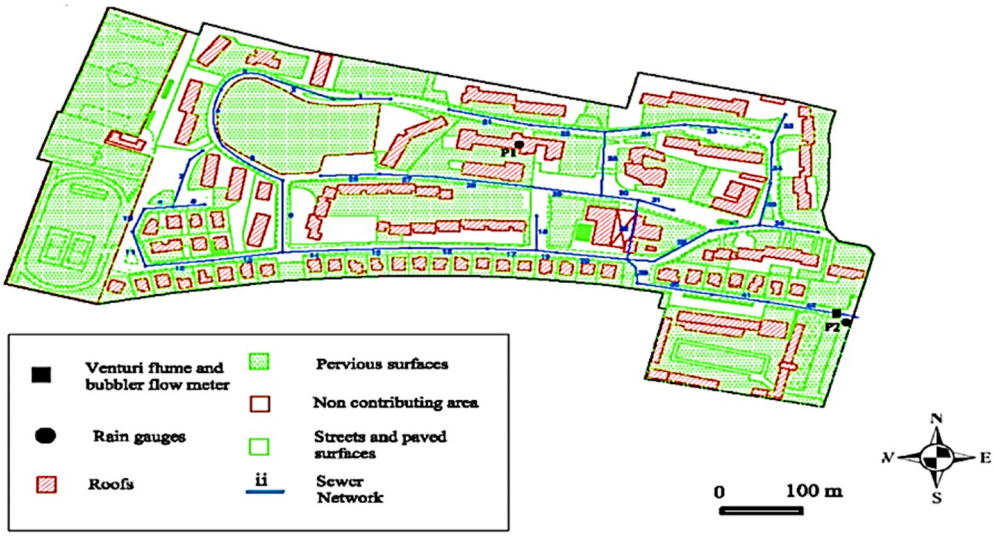



**Figure 1.** Cascina Scala experimental catchment, Pavia, Italy.

The impervious area is composed of 2.88 ha (22%) of roofs and 5.02 ha (39.6%) of streets and paved surfaces. The area is quite flat, with an average slope of 0.15% inclined from northwest to southeast and connected to the urban drainage system directly. The catchment can be separated into 42 sub-catchments, with a single pipe drain for each sub-catchment. Table 1 shows the characteristics (area, slope, and impervious area percentage divided into roofs and streets) of each sub-catchment.

After May 2001, a new area (1.33 ha) was connected to the urban drainage system, specifically to sub-catchment number 42. Therefore, the area of sub-catchment 42 was considered to grow from 0.22 to 1.55 ha starting from May 2001 (in Table 1, sub-catchment 42 * is the upgraded version of sub-catchment 42). Two tipping-bucket rain gauges were installed in the catchment area with 0.2 mm accuracy (P1 and P2 in Figure 1) to collect the rainfall data. The two rainfall gauges have a funnel with an area of 1000 $cm^2$. The presence of two gauges with a mutual distance of 310 m ensured that the spatial uniformity of the precipitation was controlled. Furthermore, it provided an accurate assessment of rainfall volume. To measure the runoff at the outlet of the channel network, an ISCO (ISCO, Genova, Italy) 4230 bubbler depth meter was placed upstream from a Venturi flume at the final reach of the channel network in the southeastern corner of the catchment (Figure 1). The system of underground channels in the Cascina Scala catchment is made of 42 concrete pipes. The upstream channels are circular conduits with a diameter ranging between 0.4 m and 0.6 m. The downstream channels are egg-shaped conduits with horizontal and vertical sizes ranging between 0.6–0.9 and 0.7–1.05 m, respectively. The conduit slope ranges between 0.15 and 1.01%, with an average value of 0.42 %. The pipe length ranges between 16 m and 75 m, with a total length of 2045 m. Twenty-three rainfall events were monitored for both rainfall-runoff transformation and water quality in the period between June 2000 and October 2003 (more details about the events can be found in [42]). As for the rainfall-runoff transformation, simultaneous measurements of rainfall intensity at the two rain-gauges and of the flow rate

in the final catchment channel were performed. A representative subset was selected for rainfall-runoff simulation, meeting the following requirements:

**Table 1.** Cascina Scala sub-catchment characteristics. See footnote for description of the symbols.

| ID | A (ha) | S (%) | IA | | ID | A (ha) | S (%) | IA | |
|---|---|---|---|---|---|---|---|---|---|
| | | | R (%) | S&S (%) | | | | R (%) | S&S (%) |
| 1 | 0.27 | 0.5 | 0 | 29.2 | 22 | 0.50 | 0.5 | 20.4 | 56.3 |
| 2 | 0.20 | 0.2 | 21.2 | 54.8 | 23 | 0.13 | 0.1 | 0 | 85.9 |
| 3 | 0.39 | 0.1 | 23.1 | 40.7 | 24 | 0.81 | 0.1 | 23.5 | 47.8 |
| 4 | 0.28 | 0.1 | 32.8 | 54.4 | 25 | 0.41 | 0.2 | 24.2 | 31.9 |
| 5 | 0.07 | 0.1 | 0 | 100 | 26 | 0.13 | 0.1 | 0 | 100 |
| 6 | 0.49 | 0.1 | 32.5 | 40 | 27 | 0.39 | 0.1 | 22.1 | 26 |
| 7 | 0.09 | 0.1 | 0 | 100 | 28 | 0.21 | 0.1 | 37.9 | 9 |
| 8 | 0.16 | 0.1 | 32.4 | 33.2 | 29 | 0.20 | 0.1 | 0 | 77.9 |
| 9 | 0.07 | 0.1 | 24.5 | 43.2 | 30 | 0.06 | 0.1 | 0 | 75.4 |
| 10 | 0.57 | 0.1 | 5.1 | 35.6 | 31 | 0.10 | 0.1 | 0 | 80.8 |
| 11 | 0.12 | 0.1 | 12.4 | 62.7 | 32 | 0.50 | 0.1 | 37.6 | 21.9 |
| 12 | 0.36 | 0.3 | 30.8 | 30.5 | 33 | 0.05 | 0.2 | 19.5 | 80.5 |
| 13 | 0.20 | 0.3 | 31.3 | 46.8 | 34 | 0.33 | 0.1 | 23.1 | 40.2 |
| 14 | 0.16 | 0.1 | 20.3 | 53 | 35 | 0.42 | 0.1 | 32 | 41.5 |
| 15 | 1.35 | 0.1 | 26.3 | 25.6 | 36 | 0.27 | 0.1 | 16 | 40.7 |
| 16 | 0.21 | 0.1 | 29.3 | 37.1 | 37 | 0.24 | 0.1 | 16.3 | 65.7 |
| 17 | 0.11 | 0.1 | 23.9 | 43.4 | 38 | 0.16 | 0.2 | 14.4 | 51.9 |
| 18 | 0.06 | 0.3 | 0 | 100 | 39 | 0.05 | 0.1 | 0 | 48 |
| 19 | 0.05 | 0.1 | 30.6 | 54.3 | 40 | 0.19 | 0.6 | 29.6 | 43.9 |
| 20 | 0.17 | 0.1 | 26.2 | 42 | 41 | 0.12 | 0.3 | 25.2 | 49.6 |
| 21 | 0.47 | 0.1 | 26.5 | 29.3 | 42 | 0.22 | 0.3 | 26.8 | 57.3 |
| | | | | | 42 * | 1.55 | 0.3 | 25.1 | 21.1 |

Footnote: For the 42 sub-catchments, ID: identifier of sub-catchments, A: area, S: slope, IA: impervious area, R: roof, S&S: streets and squares; 42 * indicates the modified catchment 42 after May 2001.

- Regular operation of rain gages and flow meter without any instrumentation failures or pressurized flow conditions;
- Total rainfall depth of at least 5 mm;
- Maximum rainfall intensity equal to or greater than 0.1 mm/min;
- Maximum rainfall depth of at least 2 mm over 15 min.

Fourteen events ($N_{r,tot}$) were found to fit the requirements (Table 2), having the following characteristics: total rainfall depth ($V_{total}$) ranging between 7 mm and 39.8 mm, rainfall duration ($T_{total}$) ranging between 50 and 1133 min, and peak flow rate ($Q_{Max}$) ranging between 0.06 $m^3/s$ and 0.55 $m^3/s$.

**Table 2.** Main characteristic of the selected rainfall events. The first seven events are for calibration. The following seven events are for validation.

| Event | Total Rainfall Depth $V_{total}$ (mm) | Rainfall Duration $T_{total}$ (min) | Peak Flow Rate $Q_{Max}$ ($m^3/s$) |
|---|---|---|---|
| 3 | 11.8 | 197 | 0.255 |
| 5 | 16.4 | 108 | 0.551 |
| 7 | 7 | 50 | 0.326 |
| 9 | 10.6 | 215 | 0.19 |
| 14 | 15.8 | 380 | 0.257 |
| 17 | 23.4 | 964 | 0.139 |
| 23 | 39.8 | 1133 | 0.149 |
| 8 | 11 | 64 | 0.376 |
| 11 | 26.2 | 478 | 0.281 |
| 12 | 18.6 | 443 | 0.263 |
| 13 | 8.4 | 111 | 0.157 |
| 19 | 12.6 | 248 | 0.234 |
| 20 | 16.2 | 231 | 0.245 |
| 21 | 7 | 286 | 0.06 |

The events were split into two subsets, i.e., the calibration and validation subsets, which were used for the optimization-based parameterization of the hydrological model and for the testing of the optimal set of parameters, respectively. The splitting was carried out while guaranteeing wide and representative ranges for $V_{total}$, $T_{total}$, and $Q_{Max}$ in the two subsets.

## 3. Materials and Methods

In this study, the distributed modeling approach was used for the rainfall-runoff transformation. For each sub-catchment of the Cascina Scala catchment (see previous Section 2), the N-LR and the UH models were constructed to simulate the runoff formation and routing, upstream from the channel network. The following sections describe the N-LR and UH models and their implementation in EPASWMM, followed by the genetic algorithm adopted for their parameterization.

### *3.1. Non-Linear Reservoir (N-LR)*

This method models the generic sub-catchment as a nonlinear reservoir, for which water storage is a function of inflow (precipitation) and outflow (infiltration, evaporation, and runoff), as shown in Figure 2. During the generic rain event, the reservoir storage starts to fill. Runoff starts only after water storage depth overcomes the depression storage depth associated with surface wetting, interception, and ponding.

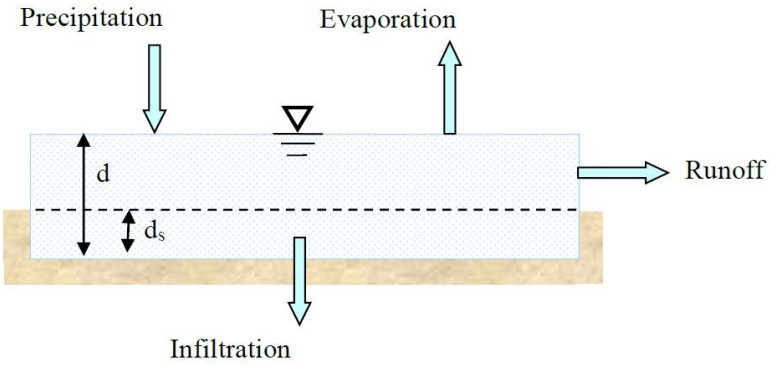



**Figure 2.** Nonlinear reservoir model of a sub-catchment [23].

In this study, the Horton model was used to model infiltration. The required parameters for Horton's approach include the maximum infiltration rate, the minimum infiltration rate, the decay coefficient, the maximum infiltration volume, and the regeneration coefficient. Evaporation is ignored in this work, since the temporal scale of event-based modeling is small.

As a result of mass conservation, the net change in water depth $d$ (m) in the sub-catchment per unit of time $t$ (s) is the difference between inflow (rainfall) and outflow (infiltration + runoff) according to the following equation:

$$\frac{\partial d}{\partial t} = i - f - r \tag{1}$$

where $i$ = rainfall intensity (m/s), $f$ = infiltration rate (m/s), and $r$ = runoff rate per unit of surface area (m/s).

Assuming a uniform flow over the sub-catchment surface, which features a width $W$ (m), a slope $S$ (-), and an area $A$ ($m^2$), $r$ can be computed by Manning's equation:

$$r = \frac{W(d - d_s)^{5/3} S^{1/2}}{n\ A} \tag{2}$$

where $d_s$ = depth of depression storage (m), and $n$ = Manning roughness coefficient (s/m$^{1/3}$), which depends on surface characteristics, being obtained as a function of the roughness of the various pervious and impervious areas present in the sub-catchment.

In the US EPASWMM, the use of the N-LR requires the following set of parameters to be defined for the external sub-catchments:

1. Slope (-);
2. Area (m$^2$);
3. Percentage of impervious and pervious area (-);
4. Manning roughness coefficient for impervious area (roof) (s/m$^{1/3}$);
5. Manning roughness coefficient for impervious area (street) (s/m$^{1/3}$);
6. Manning roughness coefficient for the pervious area (s/m$^{1/3}$);
7. Depth of depression storage on impervious area (mm);
8. Depth of depression storage on pervious area (mm);
9. Maximum infiltration rate (mm/h);
10. Minimum infiltration rate (mm/h);
11. Decay coefficient (1/h);
12. Maximum infiltration volume (mm);
13. Drying time (days);
14. Width coefficient (-).

In the applications of the present work, the optimal set of parameters for the N-LR method was derived as follows. Three parameters (1,2,3) were directly assigned based on the geometrical and land use characteristics of the sub-catchments (e.g., see Table 1 and [41]). After preliminary analyses, not reported in this paper, proving their poor impact on the results, the two parameters of the Horton model (12,13) were set to constant values of 0 mm and 7 days, respectively, according to the values suggested in [23]. The remaining nine parameters (4,5,6,7,8,9,10,11,14) were searched for by the genetic algorithm (see Section 3.3), assumed to be equal for all the sub-catchments. For parameter 14 (width coefficient), it was then multiplied by the length of the pipe in which the single sub-catchment drains the runoff, for calculating the equivalent width of each sub-catchment. Indeed, as indicated in [23], the width is a model parameter assumed to be proportional to the length of the drainage channel. For the channel network, the following set of parameters must be defined:

15. Manning roughness for conduit (s/m$^{1/3}$);
16. Conduit length (m);
17. Junction elevations (m);

In this case, two parameters (16,17) were directly assigned based on the available information about the sewerage system (see [41]), and the remaining parameter (15) was searched for by the genetic algorithm (see Section 3.3) and assumed to be equal for all the pipes, since they are all made of the same material (concrete).

### *3.2. Unit Hydrograph (UH)*

UH is traditionally used in surface water hydrology to estimate the direct runoff hydrograph resulting from a unit input of rainfall. UH techniques can be adapted to estimate rainfall-dependent infiltration/inflow (RDII) flows into a channel network. One of the widely used and most flexible UH approaches for estimating RDII is the UH-RTK method (the RTK abbreviation comes from the key three parameters that feature the unit hydrographs used by the technique, as shown in Figure 3). Notably:

- R is the fraction of rainfall entering the channel network, also known as runoff volumetric coefficient (-);
- T is the time to peak (h);
- K is the ratio of time to recession to time to peak (-).

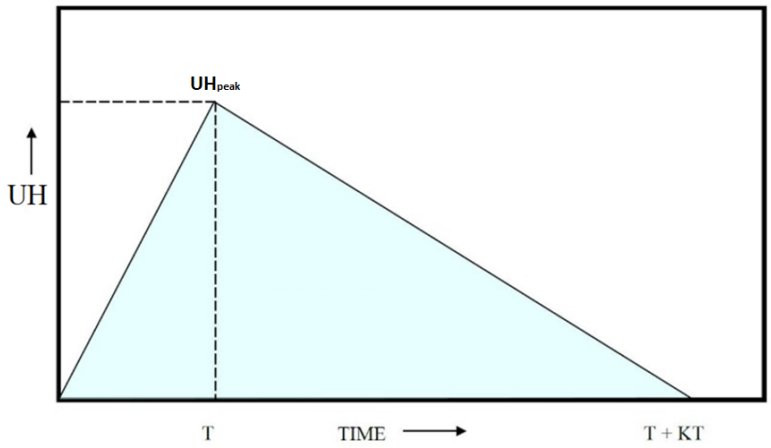



**Figure 3.** Example of a triangular unit hydrograph.

The RTK method was developed by [43] and became one of the most commonly used unit hydrograph methods in the storm water management field [44]. This method uses the single triangular unit hydrograph to represent the sub-catchment response to a rainfall event. These unit hydrographs can be employed for any event to generate the corresponding RDII flow rates. Figure 3 shows a single triangular UH assumed to represent the RDII flow induced by one unit of rainfall over a unit of time.

As the integral of the UH is equal to 1, its peak $UH_{peak}$ at time T is given by the following formula:

$$UH_{peak} = 2/(T + KT) \tag{3}$$

By applying the convolution integral, the runoff $r$ per unit of sub-catchment area at the generic time $t$ can be calculated as:

$$r(t) = \int_0^t R\, i(\tau) UH(t - \tau) d\tau \tag{4}$$

In the US EPASWMM, the use of the UH requires the following set of parameters to be defined for each external sub-catchment:

1. Area ($m^2$);
2. Parameter R (-);
3. Parameter T (h;
4. Parameter K (-);

In the applications of the present work, the optimal set of parameters for the UH method was derived as follows. One parameter (1) was directly assigned based on the geometrical characteristics of the sub-catchments (e.g., see Table 1), and the remaining three parameters (2, 3, 4) were searched by the genetic algorithm (see Section 3.3) for all the sub-catchments, for a total of (42 × 3 + 3 (for the sub-catchment 42 *)) = 129 variables. For the channel network, the following set of parameters must be defined:

5. Manning roughness for the conduit ($s/m^{1/3}$);
6. Conduit length (m);
7. Junction elevations (m).

Also in this case, two parameters (6,7) were directly assigned based on the available information about the sewerage system (see [41]), and the remaining parameter (5) was searched by the genetic algorithm (see Section 3.3) and assumed to be equal for all the pipes.

### *3.3. Genetic Optimization*

Inside the GA [35], a population of individuals exists, in which the generic individual is encoded in genes representing the decision variables of the problem. Following an initial random generation, the population evolves thanks to the processes of crossover and

mutation, till it converges towards a global optimum in terms of fitness, represented by the objective function (*OF*) to be optimized.

In the present work, the GA was used to optimize the parameters of the N-LR and UH algorithms for the 42 catchments of the Cascina Scala catchment (see Section 2) to maximize the fit of modeled-to-measured hydrographs in the final channel of the system. The parameters were optimized in a subset of seven rainfall events (calibration set) considering the following *OF* to be minimized:

$$OF = \sum_{i=1}^{N_r} \sum_{j=1}^{N_{t,i}} \left(Q_{p,i,j} - Q_{m,i,j}\right)^2 \quad (5)$$

in which $N_r$ and $N_{t,i}$ are the number of rain events (seven) considered for the optimization (the first seven of Table 2) and the number of time instants in the generic *i*-th rain event (discretization of measured water-discharge data 1 min), respectively. $Q_{m,i,j}$ ($m^3/s$) and $Q_{p,i,j}$ ($m^3/s$) are the measured and predicted water discharges (in the final channel of the system), respectively, at the *j*-th time and in the *i*-th rain event.

Notably, the application of the N-LR requires the calibration of 10 parameters (i.e., nine for the sub-catchments and one for the channels), whereas for the UH, the number of parameters to be calibrated is 130 (i.e., 129 for the sub-catchments and one for the channels).

In this work, the GA toolbox of MATLAB® was linked to the EPA-SWMM dll in order to perform the optimization for both the N-LR and the UH. For each individual of the GA, the sequence of instructions shown in Figure 4 was executed.

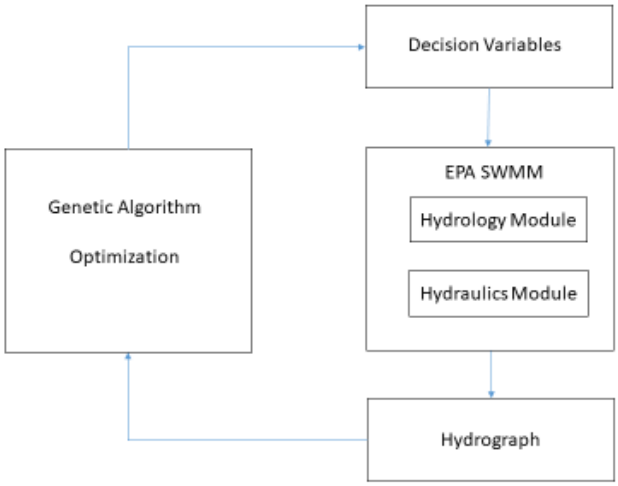



**Figure 4.** Flowchart of GA.

First, the parameters of either the N-LR or the UH were defined based on the decision variables encoded in individual genes. Then, the EPA-SWMM dll was run, i.e., the hydrological module first and then the hydraulic module, to estimate the runoff generated in each sub-catchment and to simulate the flow routing in the underground channels. Finally, the hydrograph at the exit of the Cascina Scala catchment was evaluated and compared with the measured one for each event of the calibration set, leading to the evaluation of the objective function *OF*.

### *3.4. Goodness-of-Fit Indices*

The performance of the calibration (carried out by minimizing Equation (5)) and validation steps was evaluated using goodness-of-fit indices, as suggested in the literature [45–47]. Three efficiency criteria were selected, the coefficient of efficiency (CE) (also

known as Nash–Sutcliffe efficiency), the coefficient of determination ($R^2$), and the root mean square error (RMSE), the equations of which are:

$$CE = 1 - \frac{\sum_{k=1}^{N}\left(V_{m,k} - V_{p,k}\right)^2}{\sum_{k=1}^{N}\left(V_{m,k} - \overline{V_m}\right)^2} \quad (6)$$

$$R^2 = \frac{\sum_{k=1}^{N}\left(V_{p,k} - \overline{V_m}\right)^2}{\sum_{k=1}^{N}\left(V_{m,k} - \overline{V_m}\right)^2} \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(V_{m,k} - V_{p,k}\right)^2} \quad (8)$$

where $\overline{V_m}$ is the mean of the measured variable $V_{m,k}$, $\overline{V_p}$ is the mean of the predicted variable $V_{p,k}$, $N$ is equal to $N_{t,i}$ or $N_{r,tot}$, depending on whether the generic goodness-of-fit index is calculated on the hydrograph or on one of the global variables *Total Volume*, peak flow $Q_{Max}$, and time to peak $T_{Peak}$.

## 4. Results

The results of the calibration step for the N-LR method are listed in Table 3, in which the range of variation (taken from [23]) and the final calibrated values for the 10 parameters are shown. The results of the calibration step for the UH method are listed in Table 4, in which the range of variation and the final calibrated values for the 129 parameters related to the sub-catchments are shown.

**Table 3.** The calibrated parameters of the N-LR method.

| Parameter | Range of Variation | Calibrated Value |
|---|---|---|
| Manning coefficient for roof ($s/m^{1/3}$) | 0.014–0.030 | 0.014 |
| Manning coefficient for street ($s/m^{1/3}$) | 0.010–0.055 | 0.011 |
| Manning coefficient for pervious area ($s/m^{1/3}$) | 0.10–0.35 | 0.321 |
| Depression storage on impervious area (mm) | 1.27–2.54 | 1.27 |
| Depression storage on pervious area (mm) | 2.54–5.08 | 3.08 |
| Maximum infiltration rate (mm/h) | 58–170 | 118 |
| Minimum infiltration rate (mm/h) | 1–57 | 50 |
| Decay coefficient (1/h) | 2–7 | 4.3 |
| Width coefficient (-) | 1–2 | 1.99 |
| Manning coefficient for conduit ($s/m^{1/3}$) | 0.012–0.02 | 0.012 |

Notably, for the parameter R, since it represents the runoff volumetric coefficient, its range of variation was set equal to ±20% of the ratio between the impervious area and the total area of each sub-catchment. For the parameter T, which represents the time to peak, it was allowed to vary, for all the sub-catchments, between 5 and 20 min, according to the topography and land use of the area. Finally, for the parameter K, according to the characteristics of the analyzed area, the range of variation was allowed to vary from 0.5 to 2 times the time to peak of the hydrograph. For the Manning coefficient of the conduit, the range of variations and the final calibrated value were equal to that of the N-LR model, and specifically $n$ = 0.012 $s/m^{1/3}$.

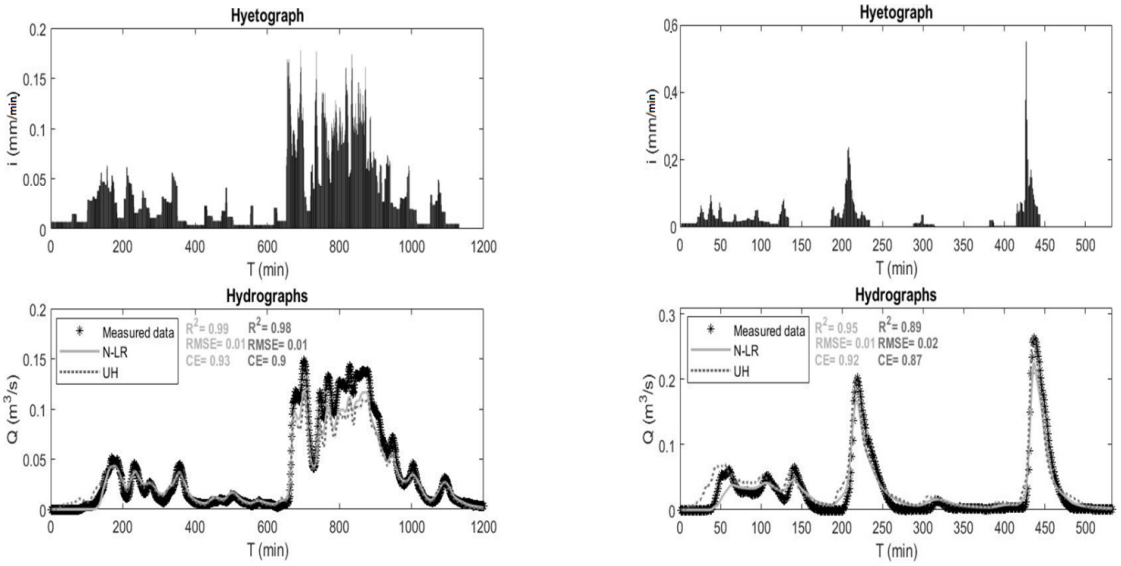
Without loss of generality, the hydrographs (with the corresponding hyetographs) for event 23 (from the calibration set) and event 12 (from the validation set), obtained with the N-LR (grey solid line) and the UH (dark-gray dotted line) models, are plotted in Figure 5a,b, respectively. Measured data are shown with star symbols. Results from both models were found to be in good agreement with the measured data, either in the calibration or the validation step, with a slight underestimation of the total runoff and peak flow, and with a very good prediction of the time to peak. Notably, as can be seen from the values of the $R^2$,

RMSE, and CE in Figure 5, the N-LR provides, either in calibration and validation, slightly better prediction performance:

- In *calibration*, the N-LR method yields $R^2$ = 0.99, RMSE = 0.01 $m^3/s$, and CE = 0.93, whereas the UH method yields $R^2$ = 0.98, RMSE = 0.01 $m^3/s$, and CE = 0.90;
- In *validation*, the N-LR method yields $R^2$ = 0.95, RMSE = 0.01 $m^3/s$, and CE = 0.92, whereas the UH method yields $R^2$ = 0.89, RMSE = 0.02 $m^3/s$, and CE = 0.87.

**Table 4.** The calibrated parameters of the UH method. (42 *) indicates the modified catchment 42 after May 2001.

| N | R (-) | | T (h) | | K (-) | |
|---|---|---|---|---|---|---|
| | Range of Variation | Calibrated Value | Range of Variation | Calibrated Value | Range of Variation | Calibrated Value |
| 1 | 0.23–0.35 | 0.24 | 0.08–0.33 | 0.33 | 0.5–2 | 1.99 |
| 2 | 0.61–0.91 | 0.63 | 0.08–0.33 | 0.32 | 0.5–2 | 1.46 |
| 3 | 0.51–0.77 | 0.51 | 0.08–0.33 | 0.08 | 0.5–2 | 0.57 |
| 4 | 0.7–1 | 0.73 | 0.08–0.33 | 0.09 | 0.5–2 | 0.53 |
| 5 | 0.8–1 | 0.8 | 0.08–0.33 | 0.33 | 0.5–2 | 2 |
| 6 | 0.58–0.87 | 0.58 | 0.08–0.33 | 0.33 | 0.5–2 | 1.88 |
| 7 | 0.8–1 | 0.8 | 0.08–0.33 | 0.17 | 0.5–2 | 0.62 |
| 8 | 0.52–0.79 | 0.53 | 0.08–0.33 | 0.1 | 0.5–2 | 0.69 |
| 9 | 0.54–0.81 | 0.61 | 0.08–0.33 | 0.09 | 0.5–2 | 0.52 |
| 10 | 0.33–0.49 | 0.37 | 0.08–0.33 | 0.33 | 0.5–2 | 2 |
| 11 | 0.6–0.9 | 0.67 | 0.08–0.33 | 0.08 | 0.5–2 | 0.57 |
| 12 | 0.49–0.74 | 0.55 | 0.08–0.33 | 0.08 | 0.5–2 | 0.56 |
| 13 | 0.62–0.94 | 0.64 | 0.08–0.33 | 0.33 | 0.5–2 | 1.97 |
| 14 | 0.59–0.88 | 0.59 | 0.08–0.33 | 0.08 | 0.5–2 | 0.56 |
| 15 | 0.42–0.62 | 0.53 | 0.08–0.33 | 0.09 | 0.5–2 | 0.55 |
| 16 | 0.53–0.8 | 0.69 | 0.08–0.33 | 0.08 | 0.5–2 | 0.52 |
| 17 | 0.54–0.81 | 0.57 | 0.08–0.33 | 0.09 | 0.5–2 | 0.56 |
| 18 | 0.8–1 | 0.91 | 0.08–0.33 | 0.08 | 0.5–2 | 1.5 |
| 19 | 0.68–1 | 1 | 0.08–0.33 | 0.08 | 0.5–2 | 0.5 |
| 20 | 0.55–0.82 | 0.82 | 0.08–0.33 | 0.09 | 0.5–2 | 0.5 |
| 21 | 0.45–0.67 | 0.48 | 0.08–0.33 | 0.27 | 0.5–2 | 1.91 |
| 22 | 0.61–0.92 | 0.62 | 0.08–0.33 | 0.33 | 0.5–2 | 1.68 |
| 23 | 0.69–1 | 0.72 | 0.08–0.33 | 0.09 | 0.5–2 | 1.37 |
| 24 | 0.57–0.86 | 0.62 | 0.08–0.33 | 0.08 | 0.5–2 | 0.5 |
| 25 | 0.45–0.67 | 0.58 | 0.08–0.33 | 0.08 | 0.5–2 | 0.5 |
| 26 | 0.8–1 | 0.94 | 0.08–0.33 | 0.08 | 0.5–2 | 0.5 |
| 27 | 0.38–0.58 | 0.58 | 0.08–0.33 | 0.3 | 0.5–2 | 1.2 |
| 28 | 0.38–0.56 | 0.44 | 0.08–0.33 | 0.08 | 0.5–2 | 0.89 |
| 29 | 0.62–0.93 | 0.63 | 0.08–0.33 | 0.33 | 0.5–2 | 1.86 |
| 30 | 0.6–0.9 | 0.72 | 0.08–0.33 | 0.1 | 0.5–2 | 0.52 |
| 31 | 0.65–0.97 | 0.7 | 0.08–0.33 | 0.33 | 0.5–2 | 1.7 |
| 32 | 0.48–0.71 | 0.48 | 0.08–0.33 | 0.08 | 0.5–2 | 1.16 |
| 33 | 0.8–1 | 0.97 | 0.08–0.33 | 0.08 | 0.5–2 | 0.52 |
| 34 | 0.51–0.76 | 0.7 | 0.08–0.33 | 0.08 | 0.5–2 | 0.51 |
| 35 | 0.59–0.88 | 0.59 | 0.08–0.33 | 0.1 | 0.5–2 | 0.79 |
| 36 | 0.45–0.68 | 0.67 | 0.08–0.33 | 0.08 | 0.5–2 | 0.58 |
| 37 | 0.66–0.98 | 0.98 | 0.08–0.33 | 0.08 | 0.5–2 | 0.59 |
| 38 | 0.53–0.8 | 0.59 | 0.08–0.33 | 0.08 | 0.5–2 | 0.59 |
| 39 | 0.38–0.58 | 0.4 | 0.08–0.33 | 0.16 | 0.5–2 | 0.5 |
| 40 | 0.59–0.88 | 0.88 | 0.08–0.33 | 0.09 | 0.5–2 | 0.6 |
| 41 | 0.6–0.9 | 0.9 | 0.08–0.33 | 0.08 | 0.5–2 | 0.83 |
| 42 | 0.37–0.55 | 0.55 | 0.08–0.33 | 0.21 | 0.5–2 | 1.5 |
| 42 * | 0.67–1 | 0.99 | 0.08–0.33 | 0.09 | 0.5–2 | 0.6 |




(**a**) (**b**)

**Figure 5.** Hyetographs (top) and comparison of the hydrographs (bottom) for N-LR (gray solid line) and UH (dark-gray dotted line) models with measured data (star symbol) for: (**a**) event 23 (from calibration set); and (**b**) event 12 (from validation set).

Table 5 reports the values of the three goodness-of-fit indices defined in Section 3.4, for all the events (from both the calibration and validation sets) and for both models, calculated considering measured and predicted water-discharge patterns at the catchment outlet. It is evident that the values of the RMSE are generally small, with the highest values for event 7 (in the calibration set) and event 8 (in the validation set), which are also the shortest rainfall events considered (durations of 50 min and 64 min, respectively). Therefore, both models seem to perform satisfactorily. In more detail:

- In *calibration*, looking at all the indices, the N-LR has a better performance in three events (7, 9, and 23); the UH prevails in two events (3 and 5); finally, the performance is similar in the remaining two events (14 and 17).
- In *validation*, looking at all the indices, the N-LR is better in four events (8, 11, 12, and 13); the UH prevails in two events (20 and 21); finally, the performance is similar in the remaining event (19).
- Globally, the N-RL seems to perform slightly better, also exhibiting higher performance in validation.

In Figures 6–8, the comparison between the prediction performance of the N-LR (light grey) and UH (dark grey) model, in the calibration (values indicated with solid circle) and validation (values indicated with solid triangle) sets, is shown in terms of total volume (Figure 6), peak flow (Figure 7), and time to peak (Figure 8). The three indices described in Section 3.4 were also applied to the total volume, peak flow, and time to peak considering both models on the whole dataset of fourteen rainfall events.

Notably, in Figure 6, the comparison between the measured and the predicted total volume is shown. The scattered dots along the bisector for both models confirm a good prediction capability either in calibration or validation. From the performance indices, it is clear that the N-LR model outperforms the UH in terms of $R^2$ (0.92 against 0.91), RMSE (183 $m^3$ against 222 $m^3$), and CE (0.91 against 0.86). Overall, both models overestimate the total volume, and only in one case (for the rainfall event with the largest total volume belonging to the validation set), they provide a remarkable underestimation.

**Table 5.** Simulation results. The yellow color is used to highlight the performance of the better model.

| Event | | $R^2$ (-) | | RSME ($m^3/s$) | | CE (-) | |
|---|---|---|---|---|---|---|---|
| | | **N-LR** | **UH** | **N-LR** | **UH** | **N-LR** | **UH** |
| Calibration | 3 | 0.78 | 0.87 | 0.03 | 0.02 | 0.77 | 0.85 |
| | 5 | 0.92 | 0.96 | 0.03 | 0.03 | 0.91 | 0.94 |
| | 7 | 0.51 | 0.42 | 0.07 | 0.08 | 0.50 | 0.30 |
| | 9 | 0.90 | 0.86 | 0.02 | 0.02 | 0.69 | 0.58 |
| | 14 | 0.94 | 0.94 | 0.02 | 0.02 | 0.88 | 0.86 |
| | 17 | 0.91 | 0.91 | 0.01 | 0.01 | 0.72 | 0.74 |
| | 23 | 0.99 | 0.98 | 0.01 | 0.01 | 0.93 | 0.90 |
| Validation | 8 | 0.81 | 0.68 | 0.04 | 0.06 | 0.81 | 0.57 |
| | 11 | 0.98 | 0.96 | 0.01 | 0.01 | 0.95 | 0.95 |
| | 12 | 0.95 | 0.89 | 0.01 | 0.02 | 0.92 | 0.87 |
| | 13 | 0.94 | 0.82 | 0.02 | 0.02 | 0.85 | 0.78 |
| | 19 | 0.97 | 0.97 | 0.02 | 0.02 | 0.91 | 0.92 |
| | 20 | 0.69 | 0.75 | 0.03 | 0.03 | 0.51 | 0.58 |
| | 21 | 0.85 | 0.88 | 0.01 | 0.01 | 0.49 | 0.44 |

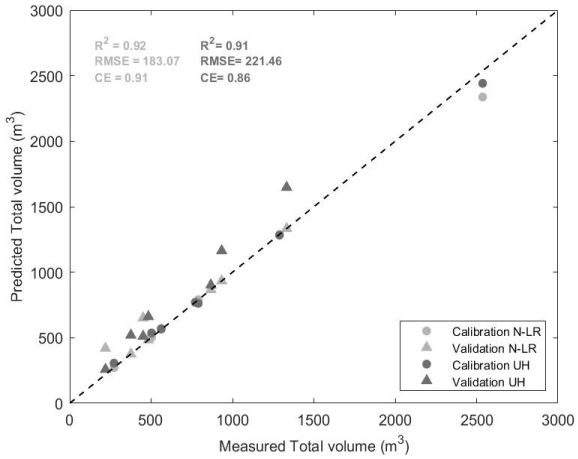



**Figure 6.** Comparison plot between measured and predicted total volume, for N-LR (light grey) and UH (dark grey) model in either calibration (solid circle) or validation (solid triangle) step.

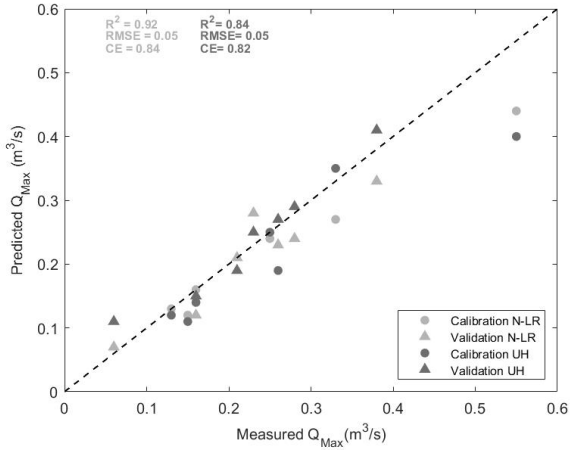



**Figure 7.** Comparison plot between measured and predicted peak flow $Q_{Max}$, for N-LR (light grey) and UH (dark grey) model in either calibration (solid circle) or validation (solid triangle) step.

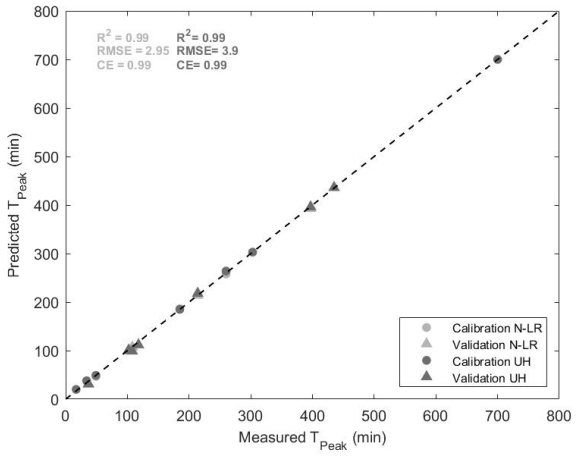



**Figure 8.** Comparison plot between measured and predicted time to peak $T_{peak}$, for N-LR (light grey) and UH (dark grey) model in either calibration (solid circle) or validation (solid triangle) step.

Figure 7 shows the comparison between the measured and the predicted peak flow $Q_{Max}$. Also in this case, the scattered dots along the bisector confirm the good agreement of the predicted values with the measured data, either in calibration or validation, with a general slight underestimation of the peak flow, especially for those events featuring higher values of the peak water discharge. From the performance indices, it is clear that the N-LR model generally outperforms the UH in terms of $R^2$ (0.92 against 0.84) and CE (0.84 against 0.82), while showing the same average error RMSE = 0.05 $m^3/s$.

Finally, in Figure 8, the comparison between the measured and the predicted time to peak is shown. From the performance indices and the alignment of the dots along the bisector, it is evident that both models, either in calibration or in validation, provide a very accurate estimation of the time in which the peak flow occurs, with an average error of just a few minutes (RMSE = 2.95 min for N-LR and RMSE = 3.9 min for the UH). The $R^2$ is excellent for both models ($R^2$ = 0.99).

## 5. Discussion

In the last decades, the Storm Water Management Model (SWMM) software of the EPA has been widely used for modeling the hydrology of urban and rural catchments due to its versatility and flexibility in reproducing the complex hydrological processes in various contexts. Overall, the good performance of this software was fully confirmed in this work on the Cascina Scala catchment. The results obtained considering both the adopted hydrological models, namely, the N-LR and the UH, were found to satisfactorily agree with the measured data in terms of the total hydrograph, total volume, peak flow, and time to peak at the catchment outlet. The N-LR showed a slightly better performance than the UH, at the price of a much more detailed amount of information required on the study area, mainly in terms of geometrical and morphological characteristics of pervious and impervious areas inside each sub-catchment. When this information is unavailable, the UH model can be conveniently adopted, as its parameterization does not require this detailed degree of knowledge. However, the easier applicability of the UH model is paid back by a larger parametrization burden (number of parameters equal to 10 and 130 for the N-LR and UH models, respectively). More specifically, both models provide a slight underestimation of the peak flow, especially for those events that feature the highest values of the peak water discharge and the shortest durations (see Figure 7 and Table 2). This could be because for modeling the catchment, a uniform spatial distribution of rainfall was assumed, and it is plausible that short and intense rainfall events are not uniformly distributed over the catchment. Furthermore, it must be highlighted that the results of the parameterization may be affected by the objective function adopted, as it was pointed out in [48], in which the authors investigated the influence of objective functions on the rainfall-runoff models' performance. In fact, the analysis of Table 5 points out that the best

performance, considering the three goodness-of-fit indices altogether, is obtained for the longest rainfall events, namely, 14, 17, and 23, for which $T_{total}$ = 380, 964, and 1133 min, respectively. This can be explained by the structure of the objective function used for the optimization (Equation (5)). Indeed, it minimizes the sum of the squared difference between the measured and predicted water discharges for each time instant of each event. Accordingly, this could drive the optimizer to favor parameter sets that fit better to longer rainfall events, which obviously feature a higher number of time instants and a larger number of terms to be summed as a result. By considering the events in the calibration step all together, the optimizer attempts to find a good compromise but implicitly gives a higher weight to longer events. Indeed, thanks to the multi-objective approach, the adoption of other objective functions considering various variables of interest, such as the total volume and peak flow, is expected to yield different sets of optimal parameters in comparison with those obtained in the present work. It is worth highlighting that this aspect is easily implementable in the proposed procedure, and therefore, this issue will be the topic of future investigations.

**Author Contributions:** Conceptualization, S.T. and E.C.; methodology, C.G., M.N.A. and E.C.; investigation, M.N.A.; writing—original draft preparation, C.G., M.N.A. and E.C.; writing—review and editing, S.T. and E.C.; data resources, S.T. funding acquisition, E.C. and S.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Italian Ministry for Research within the national project 'PRIN 2020—URCA!'.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** Support from the Italian MIUR and the University of Pavia is acknowledged within the program Dipartimenti di Eccellenza 2023–2027.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhu, Z.; Morales, V.; Garcia, M.H. Impact of combined sewer overflow on urban river hydrodynamic modelling: A case study of the Chicago waterway. *Urban Water J.* **2017**, *14*, 984–989. [CrossRef]
2. Quijano, J.C.; Zhu, Z.; Morales, V.; Landry, B.J.; Garcia, M.H. Three-dimensional model to capture the fate and transport of combined sewer overflow discharges: A case study in the Chicago Area Waterway System. *Sci. Total. Environ.* **2017**, *576*, 362–373. [CrossRef] [PubMed]
3. Todeschini, S. Hydrologic and environmental impacts of imperviousness in an industrial catchment of northern Italy. *J. Hydrol. Eng.* **2016**, *21*, 05016013. [CrossRef]
4. Kourtis, I.M.; Tsihrintzis, V.A. Adaptation of urban drainage networks to climate change: A review. *Sci. Total. Environ.* **2021**, *771*, 145431. [CrossRef]
5. Todeschini, S. Trends in long daily rainfall series of Lombardia (northern Italy) affecting urban storm water control. *Int. J. Climatol.* **2012**, *32*, 900–919. [CrossRef]
6. Caviedes-Voullième, D.; Ahmadinia, E.; Hinz, C. Interactions of Microtopography, Slope and Infiltration Cause Complex Rainfall-Runoff Behavior at the Hillslope Scale for Single Rainfall Events. *Water Resour. Res.* **2021**, *57*, e2020WR028127. [CrossRef]
7. Mohammadi, B. A review on the applications of machine learning for runoff modeling. *Sustain. Water Resour. Manag.* **2021**, *7*, 1–11. [CrossRef]
8. Brocca, L.; Melone, F.; Moramarco, T. Distributed rainfall-runoff modelling for flood frequency estimation and flood fore-casting. *Hydrol. Process.* **2011**, *25*, 2801–2813. [CrossRef]
9. Daniel, E.B.; Camp, J.V.; LeBoeuf, E.J.; Penrod, J.R.; Dobbins, J.P.; Abkowitz, M.D. Watershed modeling and its applications: A state-of-the-art review. *Open Hydrol. J.* **2011**, *5*, 27. [CrossRef]
10. Peel, M.C.; McMahon, T.A. Historical development of rainfall-runoff modeling. *WIREs Water* **2020**, *7*, 1471. [CrossRef]
11. Shin, M.J.; Guillaume, J.H.; Croke, B.F.; Jakeman, A.J. A review of foundational methods for checking the structural identi-fiability of models: Results for rainfall-runoff. *J. Hydrol.* **2015**, *520*, 1–16. [CrossRef]

12. Javadinejad, S.; Dara, R.; Jafary, F. Difference of rainfall-runoff models and effect on flood forecasting: A brief review. *Resour. Environ. Inf. Eng.* **2022**, *4*, 184–199. [CrossRef]
13. Kaffas, K.; Hrissanthou, V. Application of a continuous rainfall-runoff model to the basin of Kosynthos river using the hydrologic software HEC-HMS. *Glob. NEST J.* **2014**, *16*, 188–203.
14. Yang, J.; Castelli, F.; Chen, Y. Multiobjective sensitivity analysis and optimization of distributed hydrologic model MOBIDIC. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 4101–4112. [CrossRef]
15. Khazaei, M.R.; Zahabiyoun, B.; Saghafian, B.; Ahmadi, S. Development of an Automatic Calibration Tool Using Genetic Algorithm for the ARNO Conceptual Rainfall-Runoff Model. *Arab. J. Sci. Eng.* **2013**, *39*, 2535–2549. [CrossRef]
16. Paquet, E.; Garavaglia, F.; Garçon, R.; Gailhard, J. The SCHADEX method: A semi-continuous rainfall–runoff simulation for extreme flood estimation. *J. Hydrol.* **2013**, *495*, 23–37. [CrossRef]
17. Gamage, S.; Hewa, G.; Beecham, S. Modelling hydrological losses for varying rainfall and moisture conditions in South Australian catchments. *J. Hydrol. Reg. Stud.* **2015**, *4*, 1–21. [CrossRef]
18. Pinheiro, V.B.; Naghettini, M. Calibration of the Parameters of a Rainfall-Runoff Model in Ungauged Basins Using Synthetic Flow Duration Curves as Estimated by Regional Analysis. *J. Hydrol. Eng.* **2013**, *18*, 1617–1626. [CrossRef]
19. Tan, S.B.; Chua, L.H.; Shuy, E.B.; Lo, E.Y.-M.; Lim, L.W. Performances of Rainfall-Runoff Models Calibrated over Single and Continuous Storm Flow Events. *J. Hydrol. Eng.* **2008**, *13*, 597–607. [CrossRef]
20. Hossain, S.; Hewa, G.A.; Wella-Hewage, S. A Comparison of Continuous and Event-Based Rainfall–Runoff (RR) Modelling Using EPA-SWMM. *Water* **2019**, *11*, 611. [CrossRef]
21. Nash, J.E. The form of the instantaneous unit hydrograph. *Comptes Rendus Et Rapp. Assem. Gen. De Tor.* **1957**, *3*, 114–121.
22. Chen, C.W.; Shubinski, R.P. Computer Simulation of Urban Storm Water Runoff. *J. Hydraul. Div.* **1971**, *97*, 289–301. [CrossRef]
23. Rossman, L.A. *Storm Water Management Model User's Manual, Version 5.0*; National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency: Cincinnati, OH, USA, 2010.
24. Niazi, M.; Nietch, C.; Maghrebi, M.; Jackson, N.; Bennett, B.R.; Tryby, M.; Massoudieh, A. Storm Water Management Model: Performance Review and Gap Analysis. *J. Sustain. Water Built Environ.* **2017**, *3*, 04017002. [CrossRef]
25. Rossman, L.; Huber, W. *Storm Water Management Model Reference Manual*; EPA/600/R-15/162A; U.S. EPA Office of Research and Development: Washington, DC, USA, 2015; Volume 1.
26. Wang, W.-C.; Cheng, C.-T.; Chau, K.-W.; Xu, D.-M. Calibration of Xinanjiang model parameters using hybrid genetic algorithm based fuzzy optimal model. *J. Hydroinform.* **2011**, *14*, 784–799. [CrossRef]
27. Sarkar, A.; Kumar, R. Artificial Neural Networks for Event Based Rainfall-Runoff Modeling. *J. Water Resour. Prot.* **2012**, *04*, 891–897. [CrossRef]
28. Shamsi, U.M.S.; Koran, J. Continuous calibration. *J. Water Manag. Model.* **2017**, *25*, C414. [CrossRef]
29. Alamdari, N. Development of a robust automated tool for calibrating a SWMM watershed model. *World Environ. Water Resour. Congr.* **2016**, 221–228. [CrossRef]
30. Behrouz, M.S.; Zhu, Z.; Matott, L.S.; Rabideau, A.J. A new tool for automatic calibration of the Storm Water Management Model (SWMM). *J. Hydrol.* **2019**, *581*, 124436. [CrossRef]
31. Barco, J.; Wong, K.M.; Stenstrom, M.K. Automatic Calibration of the U.S. EPA SWMM Model for a Large Urban Catchment. *J. Hydraul. Eng.* **2008**, *134*, 466–474. [CrossRef]
32. Chow, M.F.; Yusop, Z.; Toriman, M.E. Modelling runoff quantity and quality in tropical urban catchments using Storm Water Management Model. *Int. J. Environ. Sci. Technol.* **2012**, *9*, 737–748. [CrossRef]
33. Xing, W.; Li, P.; Cao, S.-B.; Gan, L.-L.; Liu, F.-L.; Zuo, J.-E. Layout effects and optimization of runoff storage and filtration facilities based on SWMM simulation in a demonstration area. *Water Sci. Eng.* **2016**, *9*, 115–124. [CrossRef]
34. Masseroni, D.; Cislaghi, A. Green roof benefits for reducing flood risk at the catchment scale. *Environ. Earth Sci.* **2016**, *75*, 1–11. [CrossRef]
35. Holland, J.H. Genetic Algorithms and the Optimal Allocation of Trials. *SIAM J. Comput.* **1973**, *2*, 88–105. [CrossRef]
36. Huang, J.J.; Xiao, M.; Li, Y.; Yan, R.; Zhang, Q.; Sun, Y.; Zhao, T. The optimization of Low Impact Development placement considering life cycle cost using Genetic Algorithm. *J. Environ. Manag.* **2022**, *309*, 114700. [CrossRef] [PubMed]
37. Tetzlaff, D.; Soulsby, C.; Waldron, S.; Malcolm, I.A.; Bacon, P.J.; Dunn, S.M.; Lilly, A.; Youngson, A.F. Conceptualization of runoff processes using a geographical information system and tracers in a nested mesoscale catchment. *Hydrol. Process.* **2006**, *21*, 1289–1307. [CrossRef]
38. Gupta, A.K.; Shrivastava, R.K. Optimal design of water treatment plant under uncertainty using genetic algorithm. *Environ. Prog.* **2008**, *27*, 91–97. [CrossRef]
39. Chlumecký, M.; Buchtele, J.; Richta, K. Application of random number generators in genetic algorithms to improve rain-fall-runoff modelling. *J. Hydrol.* **2017**, *553*, 350–355. [CrossRef]
40. Lopes, M.D.; da Silva, G.B.L. An efficient simulation-optimization approach based on genetic algorithms and hydrologic modeling to assist in identifying optimal low impact development designs. *Landsc. Urban Plan.* **2021**, *216*, 104251. [CrossRef]
41. Papiri, S.; Ciaponi, C.; Todeschini, S. Il bacino urbano sperimentale di Cascina Scala (Pavia). *Aracne Via Raffaele Garofalo.* **2008**, *217*, 133.
42. Barco, J.; Papiri, S.; Stenstrom, M.K. First flush in a combined sewer system. *Chemosphere* **2008**, *71*, 827–833. [CrossRef] [PubMed]
43. Giguere, P.R.; Riek, G.C. Infiltration/Inflow Modeling for the East Bay (Oakland-Berkeley Area) I/I Study. In Proceedings of the 1983 International Symposium on Urban Hydrology, Hydraulics and Sediment Control, Lexington, KY, USA, 25–28 July 1983.

44. Vallabhaneni, S.; Chan, C.; Burgess, E.H. *Computer Tools for Sanitary Sewer System Capacity Analysis and Planning*; US Environmental Protection Agency, Office of Research Development: Washington, DC, USA, 2007.
45. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for sys-tematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
46. Krebs, G.; Kokkonen, T.; Valtanen, M.; Koivusalo, H.; Setälä, H. A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. *Urban Water J.* **2013**, *10*, 394–410. [CrossRef]
47. Zakizadeh, F.; Nia, A.M.; Salajegheh, A.; Sañudo-Fontaneda, L.A.; Alamdari, N. Efficient Urban Runoff Quantity and Quality Modelling Using SWMM Model and Field Data in an Urban Watershed of Tehran Metropolis. *Sustainability* **2022**, *14*, 1086. [CrossRef]
48. Fowler, K.; Peel, M.; Western, A.; Zhang, L. Improved rainfall-runoff calibration for drying climate: Choice of objective function. *Water Resour. Res.* **2018**, *54*, 3392–3408. [CrossRef]

# Chapter 5 – Calibration Strategies for Stormwater Quality Modelling

## 1. Introduction

Urban stormwater runoff transports pollutants from impervious surfaces to receiving waters, with consequences for ecological status and public health. Chapter 3 outlined how build-up and wash-off processes, coupled with event hydrology, give rise to characteristic palatographs and first-flush behavior, and reviewed the strengths and limits of the modelling spectrum—from empirical EMC approaches, through conceptual build-up/wash-off formulations (e.g., SWMM), to more physically based or hybrid models. It also highlighted recurring challenges: parameter uncertainty, equifinality, scale effects, and data scarcity, all of which complicate calibration and the generalizability of results.

Accurate stormwater quality predictions depend fundamentally on reliable rainfall–runoff modelling, since pollutant wash-off is driven by flow rates and runoff volumes. Hydrographs generated by hydrological models determine the timing and magnitude of pollutant transport, making robust flow simulation a prerequisite for meaningful water quality analysis. Accordingly, the stormwater quality modelling presented in this chapter builds directly on calibrated hydrological simulations from the Cascina Scala catchment, ensuring that peak flow, total volume, and time-to-peak are well reproduced. This strong linkage between quantity and quality modelling provides a solid foundation for evaluating pollutant wash-off behavior and predicting concentrations at the catchment outlet.

## 2. New optimization strategies for SWMM modeling of stormwater quality applications in urban area

The central objective of this study was to improve the robustness and applicability of stormwater quality calibration in urban catchments by developing new optimization strategies within the SWMM framework.

Specifically, the study examined:

- how direct optimization of build-up and wash-off parameters in SWMM could be achieved using observed TSS data,
- whether an integrated multi-event calibration strategy improves parameter consistency compared to event-by-event approaches,

- the performance of two objective functions: OF1 (instantaneous TSS concentrations) versus OF2 (global event descriptors: total mass, peak mass flow, time-to-peak),
- how different parameter range definitions (narrow, moderate, wide) affect calibration robustness, identifiability, and plausibility,
- the influence of calibration design on key quality metrics (cumulative load, peak pollutant mass flow, time-to-peak), and
- the computational efficiency of alternative optimization strategies in terms of convergence behavior and evaluation counts.

The results of this analysis are detailed in the paper titled *"*New optimization strategies for SWMM modeling of stormwater quality applications in urban area*"* published in the *Journal of Environmental Management*.

Journal of Environmental Management 361 (2024) 121244

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

# New optimization strategies for SWMM modeling of stormwater quality applications in urban area

Mohammed N. Assaf [a,*], Sauro Manenti [a,b], Enrico Creaco [a,b], Carlo Giudicianni [a], Lorenzo Tamellini [c], Sara Todeschini [a,b]

[a] Department of Civil Engineering and Architecture (DICAr), University of Pavia, Pavia, Italy
[b] Interdepartmental Centre for Water Research (CRA), University of Pavia, Pavia, Italy
[c] CNR-IMATI, National Research Council - Institute for Applied Mathematics and Information Technologies, Pavia, Italy

ARTICLE INFO

Handling editor: Dr. Lixiao Zhang

*Keywords:*
Stormwater quality modeling
Urban catchment
Total suspended solids
Build-up/wash-off model
Integrated event calibration approach
Mat-SWMM

ABSTRACT

Build-up/wash-off models were originally developed for small-scale laboratory facilities with uniform properties. The effective translation of these models to catchment scale necessitates the meticulous calibration of model parameters. The present study combines the Mat-SWMM tool with a genetic algorithm (GA) to improve the calibration of build-up and wash-off parameters. For this purpose, Mat-SWMM was modified to equip it with the capacity to provide comprehensive water quality analysis outcomes. Additionally, this research also conducts a comparative examination of two distinct types of objective functions in the optimization. Rather than depending on previous literature, this study undertook a numerical campaign to ascertain an appropriate range for the relevant parameters within the case study, thereby ensuring the optimization algorithm's efficient functionality. This research also implements an integrated event calibration approach, i.e., a novel method that calibrates all rainfall events collectively, thus improving systemic interaction representation and model robustness. The findings indicate that employing this methodology significantly enhances the reliability of the outcomes, thereby establishing a more robust procedure. The first objective function (TSS instantaneous less squared difference function, OF 1), which is widely employed in the literature, was designed to minimize the difference between observed and predicted instantaneous Total Suspended Solids (TSS) concentrations. In contrast, the second function (mass and mass peak consistency function, OF 2) considers integral model outputs, i.e., the overall mass balance, the time of the peak mass flow rate, and its intensity. The analysis of the outputs revealed that both objective functions demonstrated sufficient performance. OF 1 provided slightly better performance in predicting the TSS concentrations, whereas OF 2 demonstrated superior ability in capturing global event characteristics. Notably, the optimal parameter set identified through OF 2 aligned with the physically plausible ranges traditionally recommended in technical manuals for urban catchments. In contrast, OF 1's optimal set necessitated an expansion in the acceptable parameter ranges. Finally, from a computational burden viewpoint, OF 1 demanded a significantly higher number of function evaluations, thus implying an escalating computational cost as the range expands. Conversely, OF 2 necessitated fewer evaluations to converge toward the optimal solution.

## 1. Introduction

Anthropic activities related to urbanization and agricultural practices using fertilizers are among the activities responsible for altering the natural conditions of water bodies. Stormwater runoff from urban areas is one of the significant contributors to water pollution (Hong et al., 2019; Gao et al., 2023). The level of pollutant loads in catchments differs based on the catchment surface characteristics, the activities carried out on the catchment as well as the meteorological conditions (dry period, rainfall intensity, etc.) (Singh et al., 2022). Pollutants from the catchment surface are then transferred to waterways and water bodies through surface runoff. This runoff comes in contact with various natural and artificial compounds on catchment surfaces which enter the water flow as dissolved and suspended matter, resulting in water quality degradation. The severity of the impacts on water quality depends on the number and concentration of contaminants moved from the catchment.

* Corresponding author.
*E-mail addresses:* mohammed.assaf01@universitadipavia.it, MA.Mohammed.Assaf@Gmail.com (M.N. Assaf).

https://doi.org/10.1016/j.jenvman.2024.121244
Received 13 March 2024; Received in revised form 15 May 2024; Accepted 24 May 2024
Available online 29 May 2024

Therefore, estimates of stormwater runoff quantity and pollutant loads are necessary to evaluate the pollution effects on receiving water bodies and to design mitigation strategies (Xu et al., 2020; Todeschini et al., 2012). Computer models are widely used to assess and manage stormwater runoff quality in complex urban areas. However, appropriately identifying the land use of catchments is a major task in modeling water quality, as different land use and land cover types are connected with different pollutant accumulation processes which then are transported through stormwater runoff. These processes are not only influenced by the type of land use but also by the variability of human activities within those areas, such as vehicle movement and traffic, which significantly contribute to the characteristics of stormwater runoff. Additionally, rainfall spatial variability within the same catchment adds uncertainty aspects to the modeling complexity. Traditional rainfall-runoff models serve the investigation of water quantity issues, and their outputs are combined with further processing in order to investigate broader environmental concerns, such as water quality parameters (Yang and Chui, 2018; Simpson et al., 2022). Despite the existence of urban water quality models for many years prior to the implementation of environmental regulations, the last twenty years have seen a dramatic increase in their development. New stormwater quality standards (e.g., European Union (EU) Water Framework Directive (2000/60/CE); United States (US) National Pollutant Discharge Elimination System (NPDES), Phase I (US EPA, 1990) and Phase II (US EPA, 1999) have focused attention upon the need to predict the dynamics of pollutant load in urban water quality modeling. These standards have largely driven the development of many such models, providing theoretical frameworks and challenging applications for the hydrological and environmental engineering research community. Water quality models can serve the purpose of evaluating potential urban planning strategies for managing water and pollutant flows in scenarios requiring timely decision-making and assessments.

(Obropta and Kardos, 2007) classify urban water quality models at the catchment level into three categories: deterministic, stochastic, and hybrid.

1. Deterministic Models: These models establish a direct cause-and-effect relationship between variables, requiring parameter calibration to enhance accuracy in predicting observed data. Consisting of equations that represent various physical processes, they generate specific outputs for given inputs. These deterministic physical models aim to replicate the physical dynamics of water and pollutant movement in the catchment area, incorporating equations for accumulation and wash-off. Their parameters, assumed measurable, have physical interpretations. This category also includes "physical models," which are based on fluid mechanics' governing equations or employ physically meaningful parameters.
2. Stochastic Models: Based on empirical probabilistic relationships, these models generate outputs as distributions tied to stochastic inputs. As these models do not incorporate any physical parameter describing the catchment, they are also called "black box models," where the equations and parameters are entirely detached from the physical processes and catchment characteristics.
3. Hybrid Models: Combining deterministic and stochastic features, these models yield outputs as statistical distributions, integrating aspects of both model types.

Urban water quality models were predominantly statistical in the beginning, relying on single or multiple linear regressions between various parameters and pollutant concentrations or loads. These models were devised to replicate the variability of pollutant dynamic release at urban watershed outlets. More advanced regression methods, such as neural networks, were also tested, but they did not yield greater success than simple linear regression (MAY and SIVAKUMAR, 2008). However, the inability of statistical methods to predict pollutant changes due to factors such as evolving urbanization or climate conditions has prompted researchers to favor models that link pollutant concentrations in stormwater to conceptual processes and physical description of the urban environment (Sartor and Boyd, 1972). Physical models are based on a simplified representation of the processes occurring on urban surfaces, and the scientific literature is fraught with debates regarding the primary variables governing these processes (Rodak et al., 2020; Vaze and Chiew, 2003). In deterministic models, usually, the most delineated processes are 'build-up' and 'wash-off'. While these models are often termed 'conceptual' due to their reliance on deductive descriptions of urban runoff pollution processes, they are also considered 'physical' for their ability to represent the dynamics of pollutant transport and interactions. Specifically, these models incorporate key physical characteristics of the catchment, such as land slope, soil type, land use, impervious surface area, and drainage network configuration. Each characteristic significantly influences how pollutants accumulate and are transported by stormwater. Therefore, they were originally developed in plot-sized laboratory settings that simulate homogeneous urban environments, such as roads, and rely on two main equations (build-up and wash-off). Subsequently, they were incorporated into well-known urban hydrologic models (such as SWMM, CANOE, and MIKE URBAN), assuming that pollutants are deposited during dry periods and subsequently washed off during rainfall events.

When employed to study full-scale catchment, build-up/wash-off models simulate the average behavior of the catchment surfaces. The underlying assumption that a full-scale catchment behaves like a small-scale one deserves further scrutiny because of heterogeneity of spatial characteristics. It is important to note that, in most cases, these models are utilized in a lumped perspective, whereby a single set of parameters is used to characterize the build-up/wash-off dynamics of the entire catchment. For example, Vaze and Chiew (2003) adapted build-up and wash-off formulas at the scale of 100–200 ha of urban watersheds, which were considered uniform in terms of land use.

The availability of many geographic data bases recently has permitted detailed description of a given urban area's land uses. An effective means of arbitrating the allocation of parameter values, and honoring model spatial characteristics, could be to use these data to account for variability in land use across the model domain. The availability of these data implies that it could be possible to account for land use variability in water quality models, which would provide a more appropriate way to predict what is emitted by a given urban area. Nevertheless, the researchers demonstrated that the performance of these models at the catchment scale is inadequate in the absence of calibration. Furthermore, it is not easy to estimate the build-up/wash-off parameters a priori since their ranges vary considerably depending on the experimental case study. Consequently, reliable calibration of the model parameters is required for the lumped- and land-use-based water quality models to yield satisfactory performance at the catchment scale. In this context, automatic calibration procedures have been found quite helpful. Optimization procedures can adjust the model parameters by automatic procedures in a comparably short period (Barnhart et al., 2017). The calibration procedure conventionally was done by the iterative approximation method which is labor-intensive and subjective. This method, moreover, may not always reach the globally optimal solution (Alamdari, 2016). As a consequence, to surpass these challenges automated calibration techniques with the aid of computer-aided tools have been conceived and exercised significantly, and thus the process of determination of best-fit parameter value became more efficient and reliable (Alamdari, 2016; Behrouz et al., 2020; Barco et al., 2008b; Baker et al., 2022, 2023). A popular selection is the randomized optimization algorithms here (Chow et al., 2012; Masseroni and Cislaghi, 2016; Xing et al., 2016), because of the abundance of parameters (Tetzlaff et al., 2007) and the recognized capacity to avoid local optima. In addition to the commonly used optimization algorithms, Bayesian calibration methods also provide a robust framework for model calibration. These methods are particularly useful for their probabilistic approach to parameter estimation and their ability to incorporate prior knowledge through prior distributions (Baker et al., 2023). However, their

implementation is often computationally intensive and requires a solid foundation in statistical modeling, which may not be feasible for all applications (Rusanen et al., 2024). Of the large number of randomized optimization approaches, genetic algorithms (GAs) (Holland, 1973), is well-known for its ability to find global optima. This population-based heuristic algorithm draws inspiration from the principles of natural selection governing biological evolution. The GAs have proven successful in addressing numerous environmental research issues (Huang et al., 2022; Tetzlaff et al., 2007; Gupta and Shrivastava, 2008) and have been extensively employed to optimize parameters involved in hydrological applications (Chlumecký et al., 2017; Lopes and Gustavo Barbosa Lima da Silva, 2021; Giudicianni et al., 2023).

Two primary formatting approaches are prevalent in analyzing time-continuous datasets: using an extended data duration and dividing data into time-continuous segments. These approaches are commonly identified as continuous and event-based calibrations (Hossain et al., 2019). The strengths and weaknesses of each method have been thoroughly explored in various studies, Niazi et al. (2017), Hossain et al. (2019), and Shamsi and Koran, 2017. A significant advantage of the event-based calibration approach is its effectiveness in selecting a limited number of informative events, which, according to research by (Singh and Bárdossy, 2012) and Hossain et al. (2019), can yield better results than continuous or randomly chosen events. Additionally, event-based calibration methods offer an advantageous approach to compiling calibration and validation datasets. These datasets are statistically comparable and encompass a wide range of hydrological diversity. This level of diversity and similarity is challenging to attain when dealing with two continuous data segments.

In event-based stormwater modeling, the event-by-event approach is widely used, producing a unique set of calibrated parameters for each rainfall event. For instance, it would generate eight unique parameter sets in a scenario with eight distinct events. Subsequently, these sets are synthesized into a single calibrated parameter set through mathematical approaches such as averaging or weighted averaging. This approach presents several limitations that can impinge on its efficacy and applicability. Primarily, by generating multiple sets of calibrated parameters (each specific to a particular event) it inherently introduces variability and potential inconsistencies in the modeling process. Moreover, the synthesis process, based on potentially flawed assumptions like equal event representativeness, can lead to generalized parameters that may not capture specific event nuances.

Recent advances in modeling platforms and the development of versatile wrappers, such as the R package for SWMM, PySWMM, and MatSWMM (Leutnant et al., 2019; McDonnell et al., 2020; Riaño-Briceño et al., 2016), have facilitated the automation of event-based calibration. This progress has led to the development of a new integrated event calibration approach for stormwater qualitative modeling (Broekhuizen et al., 2020; Giudicianni et al., 2023). The new integrated event calibration approach introduced a significant advancement in stormwater modeling by calibrating all events collectively rather than individually. This unified approach directly addresses the fragmentation and inconsistency issues associated with the event-by-event approach. By considering all events simultaneously, it offers a more holistic representation of stormwater dynamics and eliminates the need for post-calibration synthesis, thereby enhancing model generalizability and applicability.

From a computational perspective, the integrated event calibration approach is more efficient, reducing the computational load and time investment compared to event-by-event calibration. This is especially beneficial for large datasets or complex modeling scenarios. However, to the best of our knowledge, no studies have yet assessed this calibration approach in the context of stormwater quality models. This work aims to implement and evaluate the integrated event calibration approach in stormwater quality modeling.

Therefore, in summary, the historical progression that provides the context for our study starts with the foundational work by Holland in 1973 on genetic algorithms. Subsequently, significant regulatory milestones such as the US EPA's NPDES standards and the EU Water Framework Directive have shaped research directions. Over the decades, advancements have been made from Vaze and Chiew's build-up and wash-off formula adaptations in 2003 to recent innovations in model calibration techniques by Giudicianni et al., in 2023. The timeline underscores the evolution from basic modeling approaches to sophisticated, integrated event calibration methods that our study builds upon.

Calibration entails the adjustment of parameters within a model to align model simulations with the data measured. From a mathematical standpoint, this process is typically translated as an optimization problem and executed automatically through optimization algorithms. Within this framework, objective functions play a crucial role in quantifying the degree of correspondence between the simulated outcomes and the actual observed variables. Objective functions in hydrological modeling have evolved considerably, now offering a more nuanced capture of model performance nuances, particularly in addressing discrepancies in timing and amplitude between simulated and observed hydrographs. Ewen (2011) introduces the Hydrograph Matching Method, which uses a dynamic programming algorithm to create visual performance measures that reflect how hydrologists visually interpret hydrographs. This method is particularly suitable for identifying differences in timing and amplitude, which are crucial for comprehensive model evaluation. Similarly, Seibert et al. (2016) enhance the Series Distance (SD) approach to distinguish timing and magnitude errors effectively, providing a more detailed visualization of these uncertainties in streamflow simulations. This method allows for a nuanced analysis of low flow and hydrological event dynamics, offering a sophisticated tool for model diagnostics. Additionally, Broekhuizen et al. (2021) explore alternative objective functions that account for timing errors in urban drainage models, showing that these can reduce uncertainties and improve parameter identifiability without sacrificing the reliability of model predictions. These studies collectively underscore the evolving nature of objective functions in hydrological modeling, pushing the boundary beyond traditional metrics to include more visually and temporally sensitive analysis techniques. The prevalent methodology involves calibrating models in alignment with "least squares" criteria, such as Nash Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970) or methodologies closely related, including Kling Gupta Efficiency (KGE; Gupta et al., 2009). More concerns are evolving that using "least squares" as an objective function is insufficient for encapsulating the full spectrum of critical attributes in the observational data (Fowler et al., 2018; Moussa and Chahinian, 2009). As a consequence, hydrologists have been using an increasing array of objective functions that accommodate multiple dimensions of hydrological responses (Sahraei et al., 2020; Tuo et al., 2018; Madsen et al., 2002; Fenicia et al., 2007). There is a move toward the development of a more holistic and multi-faceted approach to model calibration in hydrological studies. In the present study, we introduce a new objective function, which is designed to capture the global characteristics of the storm event. The method section includes a detailed discussion of the structure of this function and a critical comparison with the traditional 'least squares'.

MatSWMM, an open-source tool, is a multipurpose software package that enables users to easily manipulate simulation results for data analysis and system editing functionalities (Riaño-Briceño et al., 2016). It is compatible with two high-level programming languages (i.e., Matlab and Python) ensuring seamless implementation of interfaces and physical applications. MatSWMM benefits from matrix-oriented programming, plotting capabilities, optimization, and control toolboxes. MatSWMM serves as a co-simulation engine based on SWMM and is designed as a collection of functions to facilitate framework expansion. To the best of the authors' knowledge, MatSWMM has never been used before to perform optimization of water quality parameters. This is due to an intrinsic limitation of the MatSWMM tool that lacks a procedure for extracting water quality results from SWMM, which are however needed for parameters calibration through a proper and suitable approach, such

as a GA optimizer.

In the context of the vital challenge of managing and improving environmental quality, this study aims to contribute to the modelling, optimization and control of stormwater quality in drainage and environmental systems through enhanced calibration techniques for addressing the critical challenge of sustainable stormwater management in urban settings.

The main aim of this work is to perform accurate and reliable water quality analyses using build-up/wash-off model in an experimental urban catchment located in the city of Pavia (northern Italy). In order to reach this goal, the MatSWMM script has been modified, enabling it for the first time in water quality domain, to performing build-up and wash-off parameters calibration using GA optimizer. This adjustment represents a pivotal advancement in the application of MatSWMM for water quality analysis, making it an efficient tool in this field. This research adopts an integrated event calibration approach, distinct from traditional event-based methods, that calibrates parameters across multiple events for a comprehensive system-wide analysis. The optimization procedure has been carried out to investigate how the model response is affected by the size of the parameters range. Furthermore, a comparative analysis has been carried out exploring different types of objective functions.

## 2. Study case

The Cascina Scala catchment, located in the northern suburbs of Pavia, Lombardia, Italy, is a residential urban area that houses approximately 1500 residents. It spans over 12.68 ha, with 7.90 ha consisting of impervious surfaces like roofs and paved roads. The catchment is serviced by a combined sewerage system comprising 42 egg-shaped section concrete pipes, with an average slope of 0.042 m/m and a total length of approximately 2 km. The hydrological and hydraulic behavior of the catchment, along with the physical attributes of the catchment and drainage system, have been previously investigated (Barco et al., 2008a; Todeschini et al., 2019; Giudicianni et al., 2023).

The catchment is equipped with a comprehensive monitoring system, including two rainfall gauges, a flow gauge at the outlet, an ISCO 6700 SR refrigerated automatic grab sampler with a 24-bottle capacity (www.teledyneisco.com), and a YSI 600 multi-parameter water quality sonde (www.ysi.com). The pair of rainfall gauges (P1 and P2 shown in Fig. 1) were SIAP tipping bucket rain gauges (www.siapmicros.com), each ensuring an accuracy of 0.2 mm. These gauges, at a distance of 310 m, allow for spatial uniformity checks for precipitation and enable accurate evaluation of rainfall volume. At the sewer system's outlet located at the southeast corner of the catchment (Fig. 1), stormwater runoff measurements were taken using a Venturi tube and an ISCO 4230 bubbler flow meter (www.teledyneisco.com), while on-site temperature and conductivity were monitored using a YSI 600 multi-parameter sonde. The catchment, primarily composed of buildings and condominiums, had most of its impervious areas (approximately 70%) to the north and west, while the south and southeast comprised predominantly single-family dwellings.

Extensive investigation of the combined system has been conducted to estimate the dry-weather deposition contribution to the pollutant load. The sewerage was found to contain negligible sediment thickness. Indeed, it was designed to maintain higher velocity at low flow rates, thereby reducing sediment deposition (Bertrand-Krajewski, 2003). Moreover, continuous system simulations via US-EPA SWMM revealed that sediments deposited during the diurnal low flow periods are expelled during periods of peak diurnal flow rate. These findings, along with the conclusions in (Barco et al., 2008a) that first flush magnitude is independent of preceding dry periods, suggest that dry-weather sediments contribute minimally to the pollutant load of wet-weather flow rates.

Data on rainfall-runoff were collected from June 2000 to October 2003, with some gaps due to instrument malfunctioning. Water quality data were obtained for 23 rain events, resulting in a total of 281 samples. These samples were analyzed for total suspended solids (TSS), settleable solids (SetS), total nitrogen (TN), Ammonia ($NH_3$), Nitrogen (N), chemical oxygen demand (COD), Lead (Pb), Zinc (Zn), specific conductivity (SC), biochemical oxygen demand (BOD), phosphorus (P), and hydrocarbons (HC) in compliance with UNI EN ISO 9002 procedures.

A set of criteria was established to select the most representative storms to be investigated. These criteria, briefly recalled below, are crucial to ensure the accuracy and reliability of the analysis.

1 Data integrity: absence of instrumentation malfunctioning, missing samples, or laboratory problems. Instrumentation failures or missing samples can lead to gaps in the data, thus significantly affecting the results of the study. Similarly, laboratory problems can lead to inaccuracies in the measured parameters.
2 Absence of pressure flow conditions: it can alter the normal flow dynamics in the sewerage and may affect the representativeness of the storm event in relation to typical conditions.
3 Rainfall depth equal or greater than 5 mm: this criterion ensures that only events of significant rainfall are considered.


**Fig. 1.** Cascina Scala experimental catchment, Pavia, Italy.

4 Maximum rainfall intensity of at least 0.1 mm/min: this requirement further ensures that only significant rainfall events are included in the analysis.
5 Maximum rainfall depth of at least 2 mm over 15 min: this represents another threshold for selecting significant rainfall events.
6 Completeness of the hydrograph description: this requires the availability of samples distributed over the entire hydrograph, i.e., the plot of runoff rate vs. time for the event. This is necessary to accurately characterize the runoff and water quality response to the rainfall event.

Twelve events were identified to meet the above criteria. These events were subsequently subdivided into two groups: calibration and validation sets. To ensure balanced characteristics between the calibration and validation datasets, we systematically arranged the events by rainfall depth and alternated assignments, with one event designated for calibration and the subsequent event for validation, thereby maintaining consistent rainfall depth characteristics across both datasets. The main characteristics of the 12 event events are summarized in Table 1.

The datasets demonstrate comparable characteristics, with the average rainfall depth being 16.5 mm in the calibration dataset and 16.3 mm in the validation dataset. Additionally, the average TSS concentration is 432.7 mg/L for calibration and 493.1 mg/L for validation, and the average number of water samples collected is 17 both in the calibration and validation datasets. These similarities underscore the balanced nature of both subsets, ensuring that each represents the typical conditions encountered during the study period.

## 3. Methodology

The methodology of our study, as illustrated in Fig. 2, begins with the establishment of the Hydrology and Hydraulics Modules within the EPA SWMM for our case study. These modules are fundamental, as they simulate the hydrological responses and the water conveyance dynamics through urban drainage systems, respectively. The simulation from the Water Quality Module of the EPA SWMM produce a pollutograph, which details the pollutant concentration over the time evolution of the flow rate, providing vital information about peak pollutant loads and total pollutant mass.

This pollutograph is crucial as it is utilized for the fine-tuning of key parameters through a Genetic Algorithm implemented in MATLAB. The integration of SWMM with MATLAB is facilitated by Mat-SWMM, a tool that enables effective communication between the two platforms, allowing for seamless data exchange and parameter optimization. This setup ensures that the pollutant load outputs from SWMM are directly used in MATLAB for optimization, enhancing the precision and effectiveness of the stormwater quality management model. By seamlessly combining the simulation capabilities of SWMM with the robust optimization functions of MATLAB, our methodology offers a sophisticated approach to modeling urban stormwater systems, ensuring high accuracy and practical applicability in managing urban water quality.

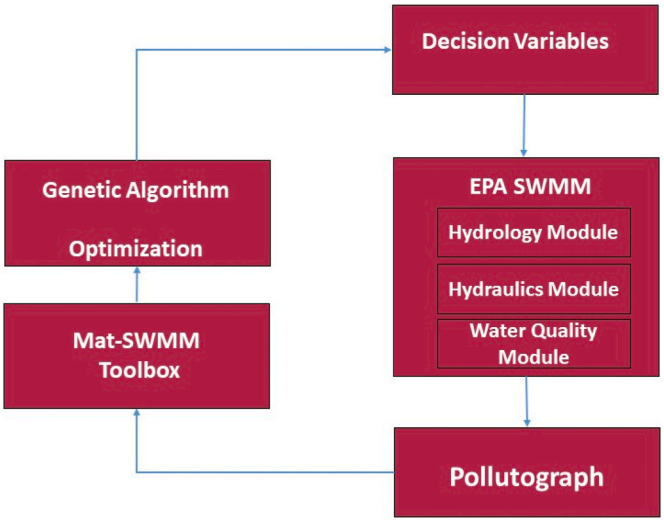



**Fig. 2.** Flowchart of the research methodology.

An overview of the tools, methodologies, and data types/sources used in the study is provided in Table S1.

### 3.1. Build-up and wash-off process

In stormwater quality studies, the pollutant dynamics in the urban environment is commonly represented through two main processes: pollutant build-up and wash-off. "Build-up" describes the generation and accumulation of pollutants on urban surfaces, influenced by factors like the length of the preceding dry period and various catchment characteristics, including impervious surfaces, traffic, and land use. "Wash-off" pertains to the movement and transport of these pollutants by stormwater runoff. While rainfall characteristics like intensity and duration are significant in determining the extent of pollutant wash-off, the physical characteristics of urban surfaces, such as slope and roughness, are also critical in influencing this process. For this study, events were specifically selected either following long/heavy rainfall, which likely washed away prior pollutant accumulations, or after extended dry periods, minimizing the impact of residual pollutants. Consequently, the initial pollutant mass for each event can be considered negligible, focusing analysis on the pollutant build-up and wash-off/transport processes of the specific events. Various mathematical techniques have been developed to model build-up and wash-off processes, as detailed in (Aryal et al., 2011). Among these techniques, the exponential function

**Table 1**
Main characteristics of the twelve selected rainfall events. The first six events are used for calibration. The remaining six events are used for validation.

| Event | Date [d/m/y] | Total Rainfall (mm) | Rainfall Duration (min) | Dry period [d] | Collected samples | TSS [mg/L] |
|---|---|---|---|---|---|---|
| 5 | June 23, 2000 | 16.4 | 108 | 10.8 | 18 | 80–890 |
| 7 | July 08, 2000 | 7 | 50 | 10 | 8 | 800–2080 |
| 8 | July 10, 2000 | 11 | 64 | 1.2 | 12 | 40–1000 |
| 14 | April 20, 2001 | 15.8 | 380 | 1[a] | 22 | 50–1190 |
| 17 | April 11, 2003 | 23.4 | 964 | 1.2 | 19 | 20–1400 |
| 23 | October 31, 2003 | 39.8 | 1133 | 1.6 | 22 | 45–560 |
| 11 | March 17, 2001 | 26.2 | 478 | 3.9 | 24 | 20–1280 |
| 12 | March 28, 2001 | 18.6 | 443 | 11 | 24 | 80–2360 |
| 13 | April 10, 2001 | 8.4 | 111 | 3.3 | 16 | 120–1420 |
| 19 | July 24, 2003 | 12.6 | 248 | 25.9 | 12 | 116–1770 |
| 20 | July 31, 2003 | 16.2 | 231 | 6.8 | 20 | 36–1470 |
| 21 | September 24, 2003 | 7 | 286 | 14.2 | 9 | 198–3880 |

[a] The revised analysis of the previous published data revealed that the preceding significant rainfall event was observed to end one day prior to event 14.

for build-up and power function for wash-off, developed by Rossman (2004) have gained wide acceptance showing high performance (Hossain et al., 2010; Chow et al., 2012; Tu and Smith, 2018; Gaut et al., 2019). Therefore, this study employs the exponential build-up (Eq. (1)) and the power wash-off (Eq. (2)) functions:

$$B = C_1\left(1 - e^{-C_2 t}\right) \quad (1)$$

where.

- $B$ = build-up [mass/length$^2$]
- $C_1$ = maximum build-up possible [mass/length$^2$];
- $C_2$ = rate constant [1/time];
- $t$ = antecedent dry period [time];

$$W = C_3 q^{C_4} B \quad (2)$$

where.

- $W$ = wash-off [mass/(time*length $^2$)]
- q = runoff rate over the subcatchment (length/time)
- $C_3$ = coefficient of wash-off [1/length]
- $C_4$ = exponent of wash-off [-]

Investigators have employed computational simulations to evaluate the values of parameters $C_1$ through $C_4$. A summary of several endeavors in this regard is presented in Table 2 (for studies on TSS), which shows that the build-up and wash-off parameters are highly site-specific and dependent on a combination of local and climatic factors (Tsihrintzis and Hamid, 1997). Local factors encompass land use, drainage system configurations, maintenance practices, geographic and geologic characteristics, and traffic volume, while climatic factors include antecedent dry periods, rainfall volumes, patterns, and intensities.

On the other hand, experimental studies and technical manuals (Rossman, 2010; Wijesiri et al., 2016; Murphy et al., 2015) focused on urban areas have shown that the physical values of build-up and wash-off parameters fall within a narrower range compared to those mentioned by simulation studies (Table 2). The studies that used wider parameter ranges have, specifically Range 3, interpreted that to account for differences in the modeling certain factors — such as street sweeping, or sediment pollution from sewer systems.

Three distinct intervals for parameters C1–C4 were examined in the present study (Table 3) to ascertain the influence on the process of optimal parameter identification and the subsequent efficacy of model fitting performance. The three proposed ranges, as illustrated in Table 3, extend from a relatively narrow domain (Range 1) to a relatively wide one (Range 3). Range 1 reflects the physical implications of build-up and wash-off parameters in a residential land context. However, due to inherent uncertainties in the modeling process, such as the omission of certain factors like street sweeping and sediment pollution from sewer systems, a wider parameter range was also considered to account for these uncertainties.

**Table 3**
Ranges of calibration parameters $C_1$–$C_4$.

| Parameters | Range 1 | Range 2 | Range 3 |
|---|---|---|---|
| Max. build-up $C_1$ [kg/ha] | 5–35 | 5–50 | 5–70 |
| Rate constant $C_2$ [1/day] | 0.1–1 | 0.1–2 | 0.1–3 |
| Coefficient of wash-off $C_3$ [1/mm] | 0.1–0.5 | 0.01–0.5 | 0.001–1 |
| Exponent of wash-off $C_4$ [-] | 1.2–1.5 | 1–2 | 1–3 |

### 3.2. Mat-SWMM

EPA-SWMM is a free open-source software for modeling hydrologic-hydraulic processes in urban watersheds and for simulating the water quantity and quality. The EPA-SWMM was released in 1971 and progressively updated till the last version (i.e., 5.2–2022), which is used in this study. The EPA-SWMM simulates all aspects of the catchment hydrologic and quality processes, including overland flow, conveyance through the drainage system, storage, and receiving water effects (Huber et al., 1988; Rossman, 2004). Its peculiar features make it appropriate for application to multiple hydrologic urban watersheds for a single event or continuous time series events. Therefore, it is widely used for modeling stormwater runoff, stormwater treatment facilities, overflow of combined and sanitary sewers, and other drainage systems (Tuomela et al., 2019; Yang et al., 2021; Zakizadeh et al., 2022; Taghizadeh et al., 2021).

In our application of the SWMM, the catchment area is meticulously divided into 42 distinct sub-catchments, with each one directly drained by a single pipe. This detailed division is crucial as it allows for a nuanced understanding and simulation of both hydrological and water quality processes at a finely granulated spatial scale. The characteristics of each sub-catchment—including area, slope, and the percentage of impervious area divided into categories such as roofs and streets—are comprehensively detailed in our reference to Giudicianni et al. (2023).

MatSWMM is a tool that facilitates advanced management of the SWMM model. The MatSWMM structure contains a distinct organization of folders and work files that are optimized to work with the toolbox. MatSWMM structure consists of the three main parts.

1. Handling of the SWMM Files: MatSWMM stores the input, report and Output files associated with the SWMM simulation in one folder called swmm files. Before running a simulation, the input file created with SWMM should be placed in this folder and the path to this input file should be specified in the code.
2. Management of SWMM Simulations: The simulation management part of MatSWMM allows the user to run simulations on a set of input files. After the simulation is complete, the results are stored in four different sub-folders. These folders are the simulations time and the three main elements of the Urban Drainage System (UDS) the links, nodes and subcatchments.
3. Presentation of Results: The results of the simulations are saved as ". csv" files in their respective folders. The content of these files will vary depending on the type of object simulated (links, nodes, or

**Table 2**
Parameters variations for build-up and wash-off equations for TSS in urban catchment in literature review.

| Location | $C_1$ | $C_2$ [1/day] | $C_3$ [mm$^{-1}$] | $C_4$ [-] | Reference |
|---|---|---|---|---|---|
| Malaysia | 0.003 (kg/m curb) | 0.8 | 0.2 | 1.4 | Chow et al. (2012) |
| New Zealand | 13.4–27.6 (kg/ha) | 0.2–0.23 | 0.24–0.27 | 1 | Wicke et al. (2012) |
| Australia | 8.5–53 (kg/ha) | 0.122–0.382 | 0.0015–0.213 | 0.363–1.27 | Hossain et al. (2010) |
| Estonia | 25 (kg/ha) | 1 | 4.9 | 1.57 | Hood et al. (2007) |
| Spain | 17.5 (kg/ha) | 0.3 | 1.811 | 1 | Temprano et al. (2006) |
| US | 32.35–83.98 (kg/ha) | 0.3–2.58 | 0.02–0.3 | 1.3–2 | Tu and Smith (2018) |
| Iran | 29.8–74.5 (kg/ha) | 0.983–3.069 | 0.4–0.7 | 2–2.2 | Zakizadeh et al. (2022) |
| Malaysia | 1.5 (kg/m curb) | 0.3 | 0.4 | 0.9 | Rezaei et al. (2019) |
| Canada | 73 (kg/ha) | 0.52 | 0.014 | 1.65 | Gaume et al. (1998) |
| France | 2-300 (kg/ha) | 0.03–20 | 0.01–10 | 0.2–3 | Bonhomme and Petrucci (2017) |

subcatchments) and contain different attributes relevant to that object.

Detailed information regarding this process can be referenced in (Riaño-Briceño et al., 2016). The original version of MatSWMM can be used for parameter optimization based on water quantity simulated results (such as flow, depth, volume, capacity, inflow, flooding, etc.) in specific objects (links, nodes, or subcatchments), but lacks any features to display results related to water quality. In this work, the script file that manages the simulation was properly modified, thereby enabling extraction of the water quality results (i.e., the TSS concentration at the closure node) required for parameter optimization.

### 3.3. Genetic algorithms (GA)

In the GA framework (Holland, 1973), a population of individuals (i.e., a set of admissible solutions of the optimization problem) exists, where each individual is encoded with so-called "genes", representing the decision variables of the problem. Initially generated randomly, the population evolves through crossover and mutation processes (i.e., a sequence of possible solutions is iteratively generated, based on the solutions already considered), converging towards a global optimum in terms of fitness, as represented by the objective function (OF) to be optimized. GAs are widely used in hydrological applications such as rainfall-runoff modeling, streamflow prediction, and groundwater modeling. Danandeh Mehr et al., 2018 presented a comprehensive review on GA applications in hydrology and water engineering, while applications of GA in hydrological modelling are discussed in Fallah--Mehdipour and Haddad, 2015. Despite the extensive deployment of GA in a variety of hydrological modeling activities, their application in urban water quality parameterization, to the best of our knowledge, remains unprecedented.

In this study, the GA was applied to optimize the four parameters of the build-up and wash-off for the study area to maximize the fit between measured and simulated pollutograph. In this study, two objective functions were applied. The first one aiming to minimize the sum of the squared difference between measured and predicted instantaneous TSS concentrations as follows:

$$OF\,1 = \sum_{i=1}^{N_r} \sum_{j=1}^{N_{t,i}} \left(C_{p,i,j} - C_{m,i,j}\right)^2 \tag{3}$$

where $N_r$ and $N_{t,i}$ are the number of rain events considered for the optimization and the number of time instants in the generic $i$-th rain event, respectively, $C_{m,i,j}$ (mg/L) and $C_{p,i,j}$ (mg/L) are measured and predicted concentrations, respectively, at the $j$-th time and in the $i$-th rain event.

This objective function, which has already been used in various works (Tu and Smith, 2018; Dotto et al., 2014; Bonhomme and Petrucci, 2017), is focused on the temporal variability of instantaneous measurements (point-by-point comparison), and is sensitive to capturing the dynamics of TSS behavior during the storm event. However, it may be significantly influenced by the presence of outliers or measurement errors in the TSS data, or can overfit specific patterns in the data that are not representative of overall system behavior.

This approach is also compatible with studies based on the Nash-Sutcliffe Efficiency (NSE), e.g. (Petrucci and Bonhomme, 2014), where the objective is to maximize NSE between the measured and predicted TSS concentrations. Because NSE and the least squared difference are linearly related, their respective optimizations yield identical outcomes (Kavetski et al., 2006).

The second objective functions minimizes the sum of the squared difference between the measured and predicted total mass and peak mass rate.

$$OF\,2 = \sum_{i=1}^{N_r} \left(\frac{M_{p,i} - M_{m,i}}{M_{m,i}}\right)^2 + \left(\frac{Qm_{p,i} - Qm_{m,i}}{Qm_{m,i}}\right)^2 \tag{4}$$

$M_{m,i}$ (kg) and $M_{p,i}$ (kg) are measured and predicted total mass in the $i$-th rain event, respectively, while $Qm_{m,i}$ (kg/s) and $Qm_{p,i}$ (kg/s) are measured and predicted peak mass rate in the $i$-th rain event, respectively.

Eq. (4) is an objective function based on the global characteristics of the storm event. The total mass reflects the overall pollutant loading from the catchment during the storm event, while the peak mass rate provides an estimate for the intensity of pollutant flux.

This function is less sensitive to point-by-point differences, as it focuses on global trends in the data - specifically, the total mass of TSS transported and the time and magnitude of peak mass rate. This function might be less sensitive to outliers compared to the first objective function, as it is more focused on aggregate measures rather than instantaneous point estimates.

In summary, while Eq. (3) has been widely adopted in analogous studies of pollutant dynamics in urban catchments, Eq. (4) represents a new contribution of this study with the aim of combining the information on instantaneous pollutant concentrations (as in Eq. (3)) with the pollutograph. shape and global characteristics affecting the amount and rate of the routed mass.

The genetic algorithm (GA) execution ceases upon fulfilling any predefined termination criteria. These criteria encompass the maximum number of generations, function tolerance, and the maximum number of stall generations. The specific values employed for these criteria were 400 for the maximum number of generations, 1E-6 for function tolerance, and 50 for the maximum stall generations.

Water quality modeling necessitates prior water quantity modeling. In this study, the water quantity parameters were employed corresponding to the findings of Giudicianni et al. (2023), which utilized a non-linear reservoir for flow-routing modeling in the Cascina Scala catchment. The accurate predictions of the total hydrograph, total volume, peak flow, and time to peak flow rate achieved by Giudicianni et al. (2023) are vital for minimizing uncertainty in water quality modeling. The total volume, peak flow rate, and time to peak flow rate were evaluated, yielding coefficients of efficiency of 0.91, 0.8, and 0.99 respectively. The corresponding root mean square errors were 183 $m^3$ for total volume, 0.05 $m^3/s$ for peak flow rate, and 2.95 min for time to peak flow rate. This is particularly relevant as the flow-routing process responsible for transporting contaminants within the catchment area.

In this study, we employ the integrated event-based calibration approach, which represents a significant departure from the traditional event-by-event calibration methods commonly used in stormwater modeling. Traditionally, separate calibrations for each rainfall event would result in multiple sets of parameters, which are then synthesized into a single calibrated parameter set through averaging techniques. This can introduce inconsistencies since it does not adequately reflect the cumulative dynamics of multiple events. The integrated event-based calibration method overcomes these limitations by calibrating all events simultaneously. This holistic approach not only ensures consistency and coherence across the calibration process but also enhances computational efficiency and model applicability. It negates the need for post-calibration synthesis, thereby providing a more robust representation of stormwater behaviors across varying conditions. By implementing this approach, we aim to provide a more accurate, reliable, and generalizable model of stormwater quality dynamics, reflecting the true variability and interconnectedness of rainfall events.

The combination of MatSWMM with the GA allows performing the calibration step considering the entire calibration set at once (first six events in Table 1). This approach is new compared to the methods commonly used in the literature, which typically calibrate each event individually and then take the average of the calibrated parameters. This holistic approach, which was previously applied by Giudicianni et al.

(2023) for the flow-routing in the Cascina Scala catchment, facilitates the optimization process in identifying a more robust solution, as a large calibration dataset helps the optimizer in finding a parameter set more representative of the physics of the process being analyzed, rather than on measurement noises.

### 3.4. Goodness-of-fit indices

Statistical indices were adopted to assess the performance of calibration and validation events, employing recommended goodness-of-fit indices from relevant works (Moriasi et al., 2007; Krebs et al., 2013; Zakizadeh et al., 2022). Two efficiency criteria, widely recognized in hydrological modeling evaluation, were chosen for assessing model performance. Coefficient of Efficiency (CE), also known as Nash-Sutcliffe Efficiency, is a widely used statistical index for assessing the predictive power of hydrological models (Eq. (5)) (Althoff and Rodrigues, 2021; Rittler and Muñoz-Carpena, 2013; Gupta and Kling, 2011; McCuen et al., 2006). It ranges from -∞ to 1, where this upper limit corresponds to a perfect match of modeled discharge to the observed data. Root Mean Square Error (RMSE), which represents the square root of the mean of the squared differences between predicted and observed data (Eq. (6)), is also computed. Notably, the smaller the RMSE value, the better the prediction, as it indicates that the average difference between observed and predicted values is small.

$$CE = 1 - \frac{\sum_{i=1}^{n} \left[X_{m,i} - X_p\right]^2}{\sum_{i=1}^{n} \left[X_{m,i} - \underline{X_m}\right]^2} \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[X_{m,i} - X_{p,i}\right]^2} \tag{6}$$

where $\underline{X_m}$ is the mean of the measured variable $X_{m,i}$, $\underline{X_p}$ is the mean of the predicted variable $X_{p,i}$, $n$ is equal to total number of measurements.

## 4. Results & discussion

### 4.1. Parameterization results

The calibration process for build-up and wash-off parameters has been carried out by applying the two objective functions within the selected ranges (Table 3). Obtained results are shown in Table 4. The results obtained from the calibration exercises reveal distinct behaviors of1 and OF2 across different ranges. Specifically, the solutions for OF1 in Range 1 and Range 2 were found at the boundaries of these ranges, suggesting that the algorithm would have preferred even more extreme values if allowed. This indicates that these ranges are not sufficient for OF1, with only Range 3 being large enough to contain the minimum within its bounds. Given that Range 1 is considered the most physically relevant, this suggests caution when employing OF1 as the optimal parameters might be significantly outside from initial estimates range. In contrast, OF2 demonstrated greater alignment with the physically meaningful Range 1, suggesting its consistency with what is considered the most representative range for the model.

Additionally, the substantial difference between the sets of parameters derived from each objective function merits attention. This variance has significant implications, as it highlights the potential for diverse outcomes based on the chosen calibration approach. A comprehensive comparison of the model's performance under the various optimized parameter sets will be further investigated in Sections 4.2 and 4.3.

In the optimization process, the GA required 5405, 8071, and 11,656 function evaluations (FE) to converge to the optimal solution for OF 1 with the range varying from Range 1 to Range 3, respectively. For OF 2, a total of 4230 (FE) were required, as shown in Table 4. As the number of function evaluations is a direct measure of the computational effort needed by the algorithm, a lower value indicates a suitable balance between computational cost and accuracy. For OF 1, as moving from Range 1 to Range 3, the number of function evaluations increases from 5405 to 11,656. This indicates that, as the parameters' domain widens, the GA requires more evaluations to find the optimal solution, reflecting the increased search space. The OF 2 needed 4230 function evaluations on Range 1, which is about 22% less than the 5405 function evaluations needed by OF 1 on the same Range 1. Therefore, OF 2 has found the optimal solution with less computational effort than OF 1 for the same range. Also, this aspect can be explained based on the structure of the adopted objective functions. Indeed, aiming at fitting on global parameters (total mass and peak mass flow rate) for the OF 2, makes the convergence faster than fitting to the whole pollutograph by the OF 1.

**Table 4**
The calibrated parameters of build-up ($C_1$ and $C_2$) and wash-off ($C_3$ and $C_4$) for the two objective functions utilizing the proposed ranges, along with the number of function evaluations (FE).

| Objective function | Objective function 1 | | | Objective function 2 |
|---|---|---|---|---|
| | Range 1 | Range 2 | Range 3 | Range 1 to 3 |
| $C_1$ | 34.94 | 49.978 | 58.361 | 24.238 |
| $C_2$ | 0.991 | 1.999 | 2.753 | 0.616 |
| $C_3$ | 0.107 | 0.030 | 0.033 | 0.215 |
| $C_4$ | 1.477 | 1.801 | 1.676 | 1.485 |
| FE | 5405 | 8071 | 11,656 | 4230[a] |

[a] This number refers to the Function Evaluation (FE) used range 1.

### 4.2. Comparing the three ranges of 1

The calibration results of 1 show that the final calibrated values of build-up and wash-off parameter changed in response to variations in the width of the parameter range used for the calibration. Therefore, a comparison of model performance using different ranges for evaluating calibrated values is carried out in order to detect how the model performance will be change across the ranges. Fig. 3 displays the comparative analysis of the measured and predicted TSS concentration obtained using OF 1 with Range 1 (light grey), Range 2 (dark grey), and Range 3 (black), during the calibration (Fig. 3a) and validation (Fig. 3b) phases. We utilized fit-accuracy measures to assess the performance of the three ranges for OF 1 across the twelve rainfall events in the dataset (six events for calibration and six for validation).

OF 1 Range 1 exhibited poor performance, significantly underestimating the TSS concentrations in both calibration and validation stages. The scatterplot and goodness-of-fit indices suggest that OF 1 Range 3 showed superior performance compared to OF 1 Range 2 during the calibration phase. This is supported by a higher CE (0.74 against 0.70) and a lower RMSE (233.45 mg/L against 246.39 mg/L).

For the validation phase, OF 1 Range 2 performed much better than OF 1 Range 3 with higher higher CE (0.7 against 0.47) and a lower RMSE (232.42 mg/L against 310.46 mg/L). The results analysis indicates that OF 1 Range 3, having broader parameter range, showed sufficient performance during the calibration phase with a higher CE, along with a lower RMSE compared to OF 1 Range 2. However, such performance during the calibration data has not translated into satisfactory generalization capabilities, poor CE and higher RMSE when compared to OF 1 Range 2, on the validation data. This highlights the need to evaluate range sensitivity when using OF 1 and select the model's optimal parameters' domain that fairly balances data fitting during calibration and model performance during validation. The differential performance between the calibration and validation phases indicates a potential sensitivity of 1 to the specific events used for calibration. Such performance variability suggests that it may be necessary to investigate effect of selecting calibration events to ensure robustness and stabilize model performance. Thus, using cross-validation techniques, such as leave-

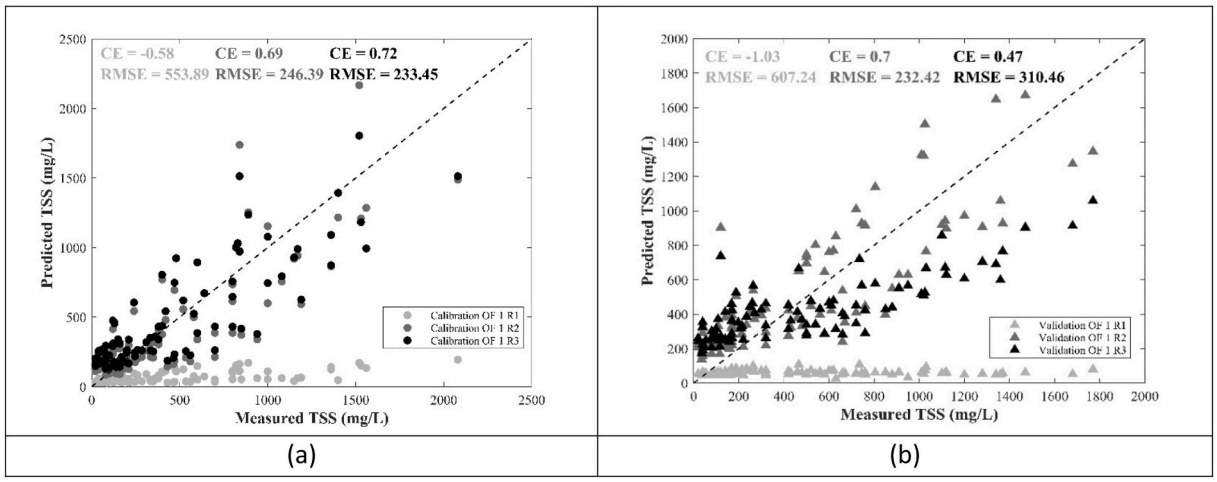



**Fig. 3.** Comparison of TSS concentration for OF 1 range 1 (light grey), OF 1 range 2 (dark grey) and OF 1 range 3 (black) in (a) calibration (solid circle); and (b) validation (solid triangle) step.

one-out or K-fold cross-validation, may offer more reliable assessment about the model's predictive capabilities.

Moreover, in some previous studies acceptable calibration results were obtained, while the validation results were poor (Dotto et al., 2010; Kleidorfer et al., 2009). The authors suggested that the spatial variability inherent in the build-up and wash-off processes accounts for the unsatisfactory results achieved. However, based on the findings of this study, an improper range of calibration parameters and a lack of model sensitivity testing to the parameter range may also lead to these types of issues.

To delve deeper into the comparison of the results from the three ranges of 1, the mass flow rate curve, as well as three global variables (i.e., total mass $M$, peak mass flow rate, $\dot{m}$ and time to peak mass flow $T_{peak}$), were also investigated.

The mass flow rate for event 5 (from the calibration set) and event 11 (from the validation set), obtained with the OF 1 and Range 1 (light grey solid line), Range 2 (dark-grey dotted line), Range 3 (black dotted line) are plotted in Fig. 4a and b, respectively. Measured data are shown with star symbols. Results from OF 1 Range 1 show poor performance in reproducing the curve shape and high underestimation of peak mass flow, either in calibration and validation events. The OF 1 Range 2 and OF 1 Range 3 results for event 5 show better performance regarding the overall measurements estimation and time to peak. For the event 11, both OF 1 Range 2 and OF 1 Range 3 outputs show similar performance, as they provide a good estimate of the time to peak mass flow but underestimate the peak mass flow rate.

In Fig. 5, the comparison between the prediction performance of the OF 1 Range 1 (light grey), OF 1 range 2 (dark grey), and OF 1 range 3 (black) in the calibration (solid circle markers) and validation (solid triangle markers) sets, is shown in terms of total mass ($M$) (Fig. 5a), peak mass flow rate ($\dot{m}$) (Fig. 5b), and time to peak mas flow ($T_{peak}$) (Fig. 5c). The goodness-of-fit indices for total mass, peak mass flow rate, and time to peak were computed, taking into account both calibration and validation datasets as delineated in Table 5.

Notably, Fig. 5a illustrates the comparison between the measured and the predicted total mass, highlighting the model's accuracy. The scattered dots far below the bisector show the poor performance of 1 Range 1 in both calibration (CE = − 1.03 and RMSE = 175.85 kg) and validation (CE = − 2.98 and RMSE = 157.02 kg) phases, as detailed in Table 5. Performance indices from Table 5 clearly demonstrate that the OF 1 Range 2 outperforms the OF 1 Range 3 in the validation set a higher CE (0.73 against 0.65) and RMSE (41.24 kg against 48.38 kg).

Fig. 5b shows the comparison between the measured and the predicted peak mass flow rate. Also in this case, the scattered dots far below the bisector show a poor performance of 1 Range 1. While for the OF 1 Range 2 and OF 1 Range 3, the scattered dots along the bisector confirm

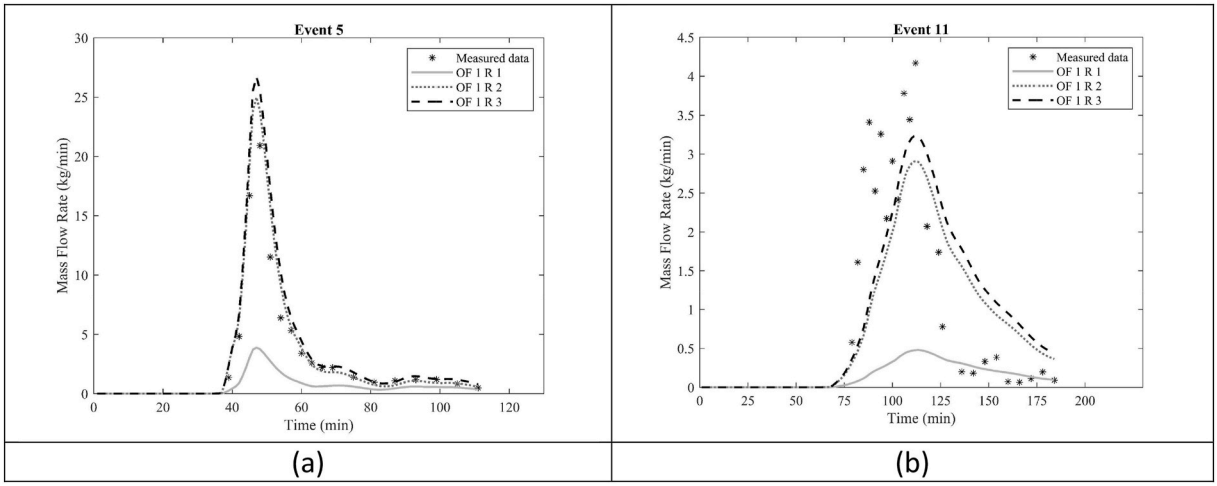



**Fig. 4.** Comparison of mass flow rate for OF 1Range 1 (light grey solid line), Range 2 (dark-grey dotted line), and Range 3 (black dotted line) with measured data (star symbol) for (a) event 5 (from calibration set); and (b) event 11 (from validation set).

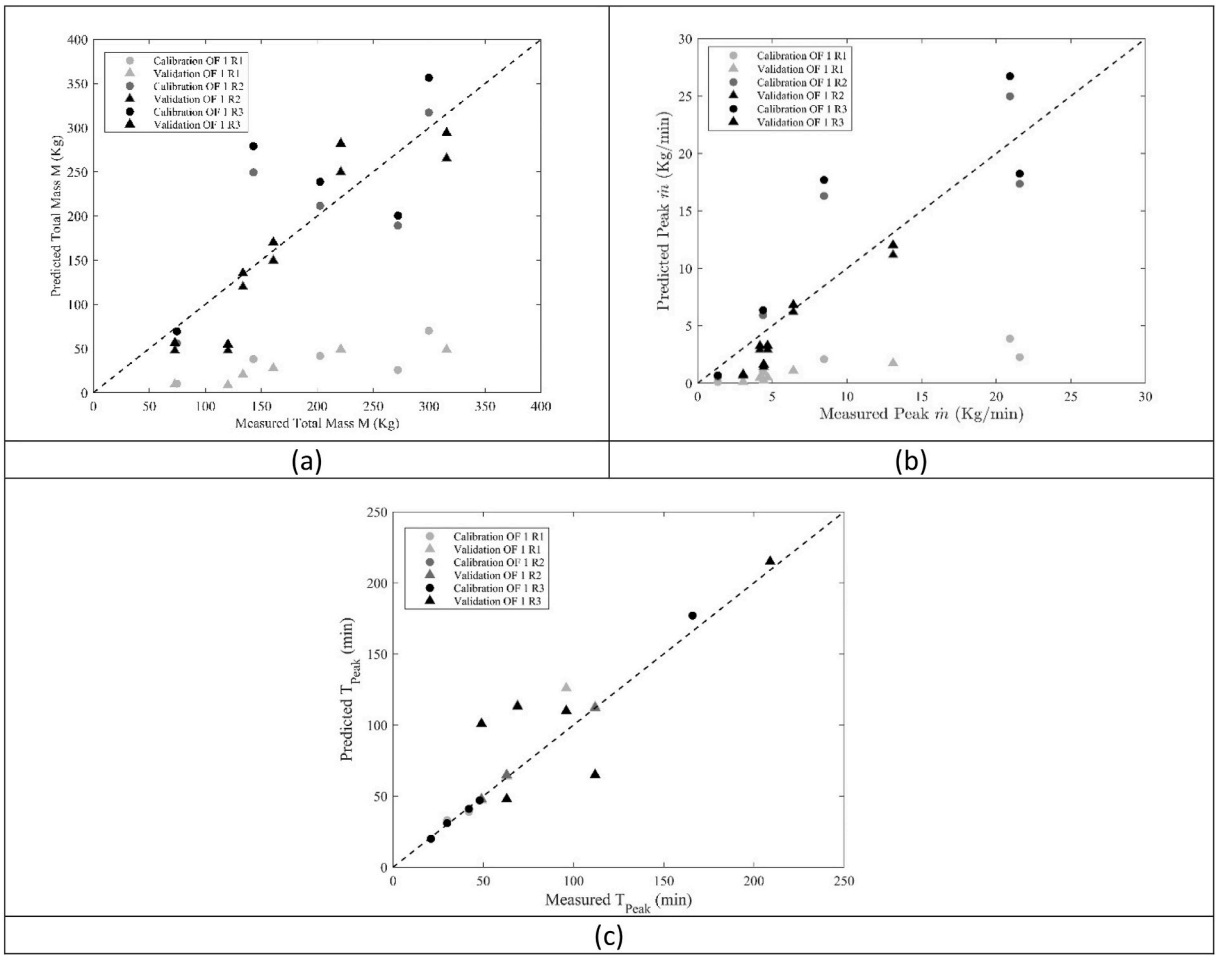



**Fig. 5.** Comparison plot between measured and predicted total mass (a), peak mass flow rate (b) and time to peak (c), for OF 1 range 1 (light grey), OF 1 range 2 (dark grey) and OF 1 range 3 (black) in either calibration (solid circle) or validation (solid triangle) step.

**Table 5**
The goodness-of-fit indices for total mass, peak mass flow rate, and time to peak across ranges of the two objective functions in calibration and validation set.

| | | OF 1 R1 | | OF 1 R2 | | OF 1 R3 | | OF 2 | |
|---|---|---|---|---|---|---|---|---|---|
| | | CE | RMSE | CE | RMSE | CE | RMSE | CE | RMSE |
| Measured Total Mass (kg) | Calibration | −1.03 | 175.85 | 0.66 | 75.14 | 0.72 | 61.54 | 0.85 | 66.84 |
| | Validation | −2.98 | 157.02 | 0.73 | 41.24 | 0.65 | 48.38 | 0.76 | 39.91 |
| Measured Peak (kg/min) | Calibration | −33.53 | 11.97 | 0.64 | 3.64 | 0.57 | 4.18 | 0.68 | 3.05 |
| | Validation | −2.21 | 5.97 | 0.74 | 1.69 | 0.63 | 1.96 | 0.89 | 1.12 |
| Measured Time to Peak (min) | Calibration | 0.98 | 5.31 | 0.98 | 5.04 | 0.98 | 5.42 | 0.99 | 1.48 |
| | Validation | 0.83 | 22.24 | 0.87 | 19.03 | 0.57 | 34.9 | 0.99 | 6.07 |

the good agreement of the predicted values with the measured data, either in calibration or validation. According to the performance indices in Table 5, OF 1 Range 2 generally excels over Range 3 in both calibration and validation sets, with higher CE (0.64 against 0.57 in calibration and 0.74 against 0.63 in validation) and lower RMSE (3.64 kg/min against 4.18 kg/min in calibration and 1.69 kg/min against 1.96 kg/min in validation).

Fig. 5c compares the measured and the predicted time to peak mass flow. The scatter plot and performance indices indicate that all three ranges of 1 provide reasonable estimates of the time to peak flow in both calibration and validation. Notably, OF 1 Range 2 demonstrates the best fit in the validation set with the highest CE (0.87) and the lowest RMSE (19.03 min) compared to the other ranges. Overall, the calibration parameters obtained with range 2 for OF 1 show better fitting performance comparing with those from the other two ranges in terms of total mass, peak mass flow rate, and time to peak mass flow. Therefore, the calibrated parameters values from range 2 for OF 1 are used in the subsequent comparison between OF 1 and OF 2 outputs.

### 4.3. Comparing of 1 range 2 and of 2

Fig. 6a and b, respectively, illustrate the comparative analysis of the measured and predicted TSS concentrations obtained with OF 1 Range 2 (light grey) and OF 2 (dark grey) during the calibration (solid circle) and validation (solid triangle) phases. The calibration results reveal that both objective functions deliver satisfactory performance. As expected, OF 1 slightly outperforms OF 2 in terms of CE for TSS concentration, scoring 0.7 compared to 0.68 for OF2. Furthermore, the OF 1 exhibits a notably lower RMSE for TSS concentration at 219.52 mg/, compared to 246.39 mg/L for OF 2).

The validation results also show a similar performance for CE (0.7 for both). In addition, OF 1 outperforms OF 2 in RMSE results (232.42 mg/L

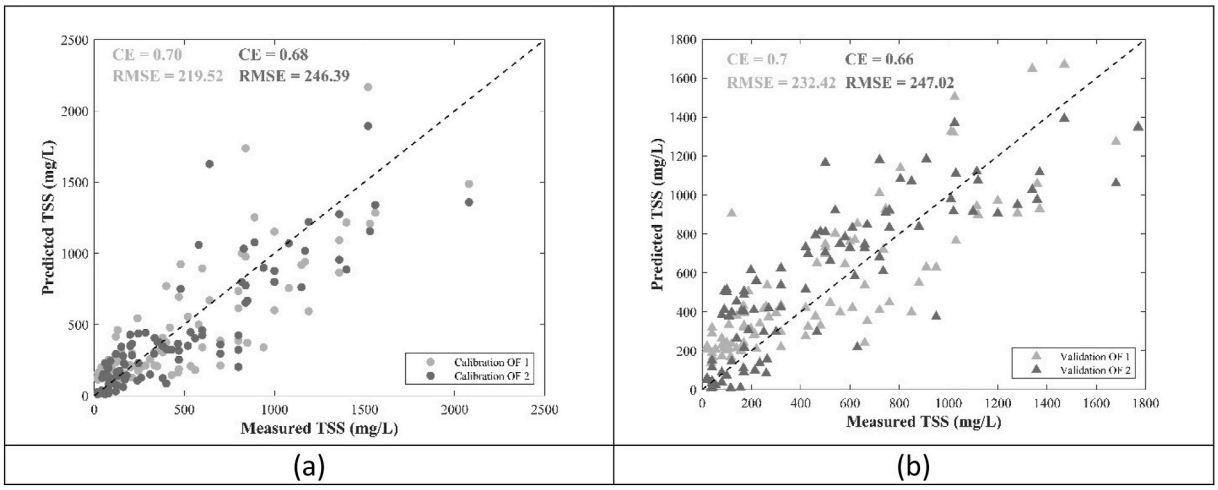



**Fig. 6.** Comparison of TSS concentration for OF 1 Range 2 (light grey) and OF 2 (dark grey) in (a) calibration (solid circle); and (b) validation (solid triangle) step.

against 247.02 mg/L). These outcomes suggest that both objective functions can adequately predict a model for estimating TSS concentrations, even if the optimized coefficients are quite different and OF2 was does not take TSS explicitly into account. The findings from this study show better results when compared to previous studies (Temprano et al., 2006; Nazahiyah et al., 2007), which primarily utilized OF 1 using the event-by-event calibration approach, without testing range robustness. This improvement can be attributed to the adoption of a novel integrated event calibration approach and the inclusion of a sensitivity analysis for the range of calibrated parameters. Recently, Bonhomme and Petrucci (2017) reported satisfactory calibration and validation results (up to 0.65 CE) by enhancing the spatial distribution of build-up/wash-off parameters using detailed geographical data. However, this type of analysis requires high-quality geographical data, which may not be readily available for all locations, and would increase model structural complexity.

The mass flow rate for event 5 (from the calibration set) and event 11 (from the validation set), obtained with the OF 1 Range 2 (light-grey solid line) and OF 2 Range 1 (dark-grey dotted line), are plotted in Fig. 6a and 5b, respectively. The data obtained from Event 5 suggests that OF 1 delivers a superior performance in terms of aligning with the measured data in the tail of the event. However, both models yield a satisfactory estimation of the time to peak mass flow. In contrast, for Event 11, OF 2 behaves better than OF 1 in terms of aligning with the measured data and capturing the overall curve shape. Additionally, OF 2 offers a more accurate estimate of the peak mass flow rate.Fig. 7

Fig. 8 provides a comparative analysis of the prediction capabilities of the OF 1 Range 2 (light grey) and OF 2 (dark grey). The figure includes results from both the calibration (solid circle markers) and validation (solid triangle markers) datasets. The fit-accuracy measures from Section 3.5 were employed to evaluate these parameters across the twelve rainfall events in the dataset for both objective functions, as detailed in Table 5.

The comparison between the measured and predicted total mass is shown in Fig. 8a. The scattering of dots along the bisector for both objective functions displays a satisfactory predictive ability in both calibration and validation phases, with a few exceptions of underestimation or overestimation. The performance indices show that the OF 1 exhibits superior performance over the OF 2 in terms of high CE (0.66 against 0.85 in calibration and 0.76 against 0.73 in validation) and lower RMSE (66.84 kg against 75.14 kg in calibration and 39.91 against 41.24 in validation).

Fig. 8b shows the comparison between the measured and predicted peak flow rate. The scattered dots along the bisector reveal the satisfactory fit of the model to the measured data. According to the performance indices, the OF 2 model exhibits overall better behavior than the OF 1 model in terms of CE (0.68 against 0.64 in calibration and 0.89 against 0.74 in validation) and RMSE (3.05 kg/min against 3.64 kg/min

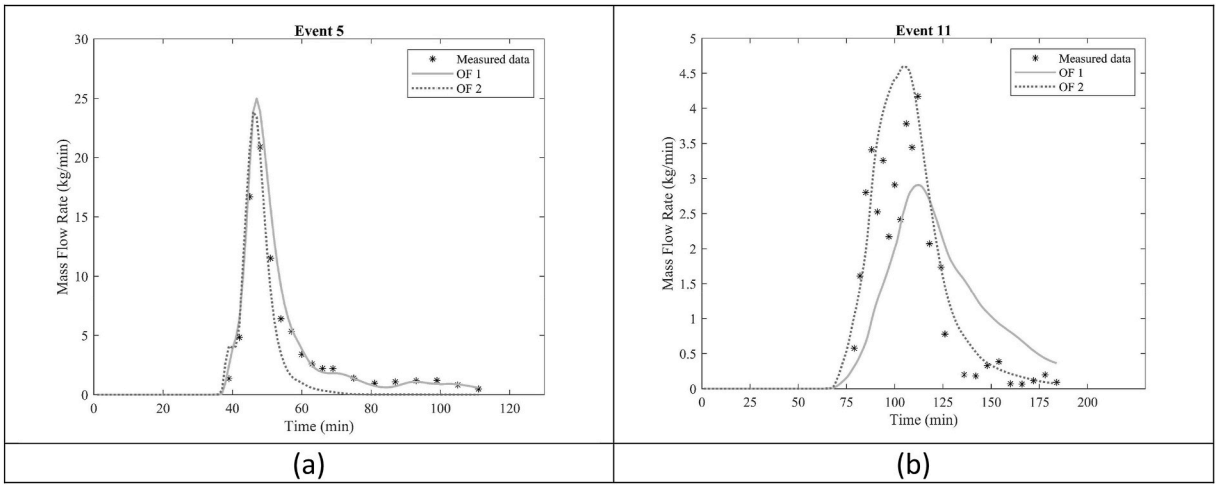



**Fig. 7.** Comparison of mass flow rate for OF 1 Range 2 (light-grey solid line) and OF 2 (dark-grey dotted line) with measured data (star symbol) for (a) event 5 (from calibration set); and (b) event 11 (from validation set).

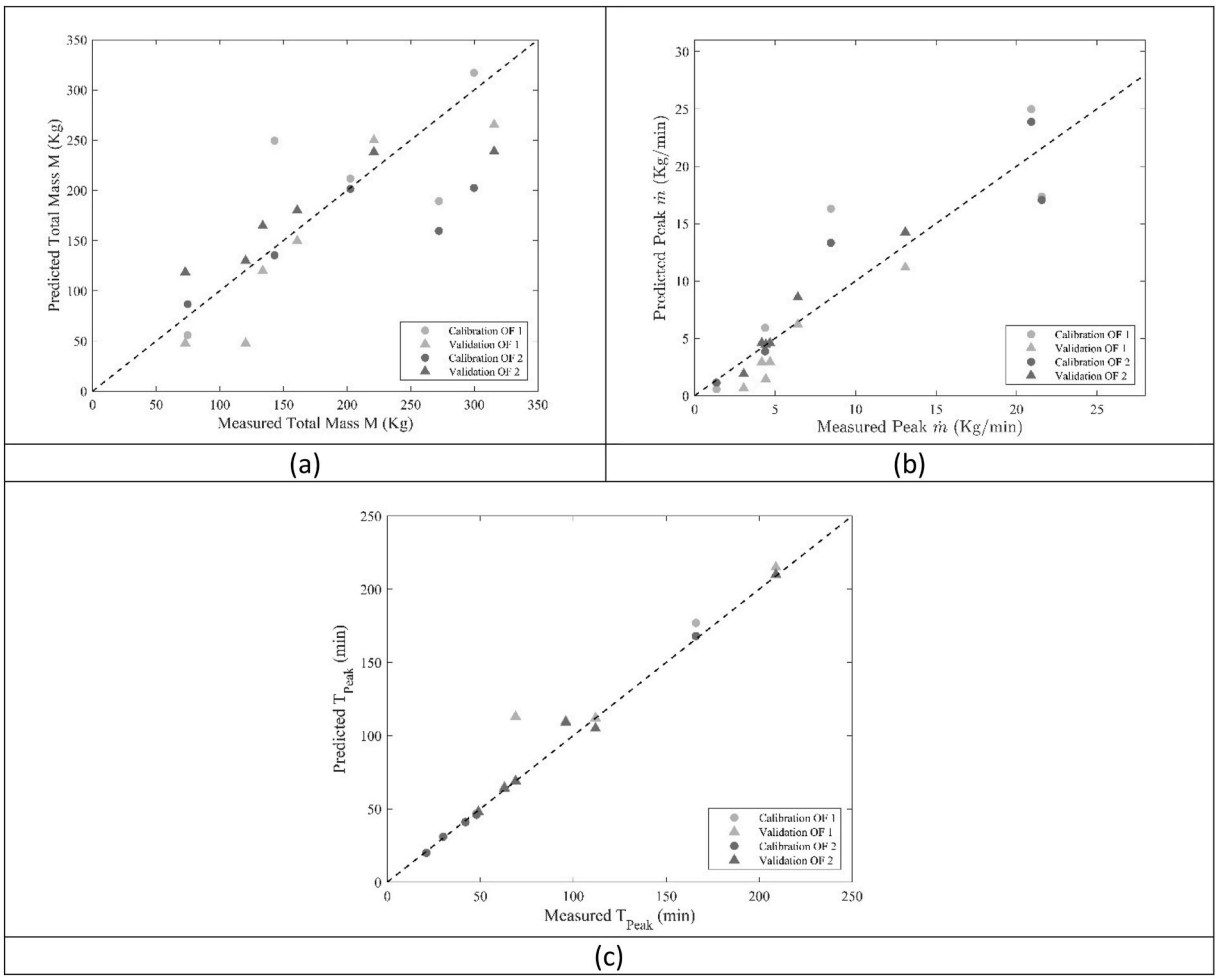



**Fig. 8.** Comparison plot between measured and predicted total mass (a), peak mass flow rate (b) and time to peak (c), for OF 1 Range 2 (light grey) and OF 2 (dark grey) in either calibration (solid circle) or validation (solid triangle) phases.

in the calibration and 1.12 kg/min against 1.69 kg/min in the validation).

Lastly, Fig. 8c compares the measured and predicted time to peak mass flow. The distribution of dots along the bisector and the performance indices highlight the accurate estimation provided by both models in predicting the time at which the peak flow occurs. OF 2 outperforms OF 1 in terms of CE (0.99 against 0.98 in calibration and 0.99 against 0.87 in the validation) and RMSE (5.04 min against 1.48 min in calibration and 6.07 against 19.03 in validation).

In summary, the performance of the two objective functions in estimating TSS concentration is comparable, with OF1 showing a slight improvement. Moreover, OF 1 provides a slightly better performance in predicting the total mass of TSS. However, OF 2 is insensitive to the width of parameter range and outperforms OF 1 in term of peak mass flow rate and time to peak mass flow estimate. On the other hand, comparative analysis of two objective functions, OF 1 Range 2 and OF 2, used in the optimization process demonstrates a distinct advantage of 2 over OF 1 in terms of computational efficiency. The OF 2 consistently required fewer function evaluations (4230) to converge to the optimal solution, regardless of the range, as opposed to the increasing function evaluations needed for OF 1 with increasing range width (5405 for Range 1, 8071 for Range 2, and 11,656 for Range 3). The reduction in function evaluations brings a decrease in computational effort, ostensibly making OF 2 a more efficient choice. To prove generalizability of the findings, it is recommended the application of the proposed methodology in future investigations extended to multiple case study areas and, importantly, to regions with more complex land use classifications. Such future developments would allow for a robust testing of the findings from this work, enhancing the understanding of its implications in wider and more complex circumstances. Our approach integrates traditional stormwater modeling techniques, such as the build-up and wash-off models, with cutting-edge advancements in optimization algorithms, showcasing the efficacy of blending established methodologies with innovative solutions in urban water management. Recent advancements exemplify this strategy; for instance, the application of the Non-Dominated Sorting Differential Evolution (NSDE) metaheuristic algorithm has significantly optimized the design of urban runoff systems and water distribution networks. These studies demonstrate substantial cost reductions and enhanced efficiencies, confirming the value of high-tech integrations into traditional frameworks (Cemiloglu et al., 2023a, 2023b). In forthcoming studies, we aim to systematically examine the selection of calibration events and the efficacy of different cross-validation methodologies in enhancing model performance. Moreover, in future studies, we aim to explore the integration of deep learning and machine learning techniques to further enhance the performance of the stormwater management system. These advanced computational methods hold potential for significantly improving predictive accuracy and operational efficiency by automatically identifying complex patterns and dependencies in environmental data that traditional models might overlook. By leveraging these technologies, we anticipate not only refining our current models but also driving substantial advancements in water quality management for scenarios requiring timely decision-making and assessment.

## 5. Conclusion

This study explores the potential of using Genetic Algorithms (GA) to calibrate build-up and wash-off parameters. Two objective functions were employed to optimize the build-up and wash-off parameters for the modeling of Total Suspended Solids (TSS) in an Italian catchment, Cascina Scala.

The first objective function (OF 1) uses instantaneous measurements at each time step, aiming to minimize the squared difference between the measured and predicted TSS concentrations. The second objective function (OF 2), a novel contribution of this study, aims to integrate information on instantaneous pollutant concentrations with pollutograph shape and global characteristics, which influence the amount and rate of the routed mass.

The analysis of both objective functions revealed distinct behaviors in response to changes in parameter ranges. Objective Function 1 (OF1) showed a tendency to produce varied final calibrated values depending on the range width. This variation suggests that the most suitable parameters for OF1 might extend beyond the initially estimated ranges, particularly in the case of the narrower Range 1 and Range 2. Only in Range 3 did the solutions appear adequately contained within the range, underscoring the need for broader parameter exploration when using OF1. In comparison, OF 2 aligned more closely with physically realistic parameters, implying a reduced computational effort for calibration due to less extensive parameter exploration (approximately 4200 Function Evaluations (FE) versus up to 11,600 FE for OF 1 depending on the range).

Model performance results indicate that both objective functions offer sufficient performance in estimating TSS concentrations. However, OF 1 appears slightly superior in estimating TSS concentrations. In contrast, OF 2 demonstrates improved capacity in capturing global parameters, especially peak mass flow rate and time-to-peak. Building on these findings, the integrated event calibration approach adopted in this study further enhances the model's robustness and efficiency. By collectively calibrating all rainfall events, it overcomes the limitations of event-by-event calibration, offering a more consistent and comprehensive model for stormwater quality analysis. This method contributes to the ongoing efforts to enhance stormwater modeling accuracy and offers valuable insights that could inform future research and practical applications in urban environmental quality management.

While our study demonstrated the effectiveness of the optimization strategies using Total Suspended Solids (TSS) as a model parameter, these methodologies are designed with the flexibility to be applied to a broad spectrum of water quality parameters, enhancing their utility across diverse hydrological and environmental contexts.

## CRediT authorship contribution statement

**Mohammed N. Assaf:** Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Sauro Manenti:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Enrico Creaco:** Writing – review & editing, Funding acquisition, Conceptualization. **Carlo Giudicianni:** Writing – review & editing. **Lorenzo Tamellini:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Sara Todeschini:** Writing – review & editing, Supervision, Methodology, Conceptualization, Funding acquisition.

## Declaration of competing interest

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

## Data availability

Data will be made available on request.

## Acknowledgements

This research was supported by the Italian Ministry of University and Research (MUR) through the project "PRIN 2020-URCA!, Prot. 20208TAK3H" and through the project "DORIAN Dipartimento di Eccellenza 23–27". Lorenzo Tamellini, Sara Todeschini, and Sauro Manenti have been supported by the PRIN 2022 PNRR project "Uncertainty Quantification of coupled models for water flow and contaminant transport" (No. P2022LXLYY), financed by the European Union - Next Generation EU.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2024.121244.

## References

Alamdari, Nasrin, 2016. Development of a robust automated tool for calibrating a SWMM watershed model. In: World Environmental and Water Resources Congress 2016, pp. 221–228.
Althoff, D., Rodrigues, L.N., 2021. Goodness-of-fit criteria for hydrological models: model calibration and performance assessment. J. Hydrol. 600, 126674.
Aryal, R., Kandasamy, J., Vigneswaran, S., Naidu, R., Lee, H.S., 2011. Review of stormwater quality, quantity and treatment methods part 2: stormwater: Quality modelling. Environmental Engineering Research 16, 143–149.
Baker, E.A., Cappato, A., Todeschini, S., Tamellini, L., Sangalli, G., Reali, A., Manenti, S., 2022. Combining the Morris method and multiple error metrics to assess aquifer characteristics and recharge in the lower Ticino Basin, in Italy. J. Hydrol. 614, 128536.
Baker, E.A., Manenti, S., Reali, A., Sangalli, G., Tamellini, L., Todeschini, S., 2023. Combining noisy well data and expert knowledge in a Bayesian calibration of a flow model under uncertainties: an application to solute transport in the Ticino basin. GEM - International Journal on Geomathematics 14, 8.
Barco, J., Papiri, S., Stenstrom, M.K., 2008a. First flush in a combined sewer system. Chemosphere 71, 827–833.
Barco, J., Wong, K.M., Stenstrom, M.K., 2008b. Automatic Calibration of the U.S. EPA SWMM Model for a Large Urban Catchment, vol. 134, pp. 466–474.
Barnhart, B.L., Sawicz, K.A., Ficklin, D.L., Whittaker, G.W., 2017. Moesha: a genetic algorithm for automatic calibration and estimation of parameter uncertainty and sensitivity of hydrologic models. Transactions of the ASABE 60, 1259–1269.
Behrouz, Mina Shahed, Zhu, Zhenduo, Shawn Matott, L., Rabideau, Alan J., 2020. A new tool for automatic calibration of the Storm Water Management Model (SWMM). J. Hydrol. 581, 124436.
Bertrand-Krajewski, J.L., 2003. Sewer sediment management: some historical aspects of egg-shaped sewers and flushing tanks. Water Sci. Technol. 109–122.
Bonhomme, C., Petrucci, G., 2017. Should we trust build-up/wash-off water quality models at the scale of urban catchments? Water Res. 108, 422–431.
Broekhuizen, Ico, Leonhardt, Günther, Marsalek, Jiri, Viklander, Maria, 2020. Event selection and two-stage approach for calibrating models of green urban drainage systems. Hydrol. Earth Syst. Sci. 24 (2), 869–885.
Broekhuizen, Ico, Leonhardt, Günther, Viklander, Maria, 2021. Reducing uncertainties in urban drainage models by explicitly accounting for timing errors in objective functions. Urban Water J. 18 (9), 740–749.
Cemiloglu, Ahmed, Zhu, Licai, Chen, Biyun, Lu, Li, Nanehkaran, Yaser A., 2023a. Enhancing urban surface runoff conveying system dimensions through optimization using the non-dominated sorting differential evolution (NSDE) metaheuristic algorithm. Water 15 (16), 2927.
Cemiloglu, Ahmed, Zhu, Licai, Abbas, Ugurenver, Nanehkaran, Yaser A., 2023b. Optimal exploitation of urban water supply networks based on pressure management with the nondominated sorting differential evolution (NSDE) algorithm. Water 15 (14), 2583.
Chlumecký, M., Buchtele, J., Richta, K., 2017. Application of random number generators in genetic algorithms to improve rainfall-runoff modelling. J. Hydrol. 553, 350–355.
Chow, M.F., Yusop, Z., Toriman, M.E., 2012. Modelling runoff quantity and quality in tropical urban catchments using Storm Water Management Model. Int. J. Environ. Sci. Technol. 9, 737–748.
Danandeh Mehr, A., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A.M.A., Yaseen, Z.M., 2018. Genetic programming in water resources engineering: a state-of-the-art review. J. Hydrol. 566, 643–667.
Dotto, C.B.S., Kleidorfer, M., Deletic, A., Fletcher, T.D., Mccarthy, D.T., Rauch, W., 2010. Stormwater quality models: performance and sensitivity analysis. Water Sci. Technol. 62, 837–843.
Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., Mccarthy, D.T., 2014. Impacts of measured data uncertainty on urban stormwater models. J. Hydrol. 508, 28–42.
Ewen, John, 2011. Hydrograph matching method for measuring model performance. J. Hydrol. 408 (1–2), 178–187.
Fallah-Mehdipour, E., Haddad, O.B., 2015. Application of genetic programming in hydrology. In: GANDOMI, A.H., ALAVI, A.H., RYAN, C. (Eds.), Handbook of Genetic Programming Applications. Springer International Publishing, Cham.

Fenicia, Fabrizio, Savenije, Hubert HG., Matgen, Patrick, Pfister, Laurent, 2007. A comparison of alternative multiobjective calibration strategies for hydrological modeling. Water Resour. Res. 43 (3).
Fowler, Keirnan, Murray, Peel, Western, Andrew, Zhang, Lu, 2018. Improved rainfall-runoff calibration for drying climate: choice of objective function. Water Resour. Res. 54 (5), 3392–3408.
Gao, Z., Zhang, Q., Li, J., Wang, Y., Dzakpasu, M., Wang, X.C., 2023. First flush stormwater pollution in urban catchments: a review of its characterization and quantification towards optimization of control measures. J. Environ. Manag. 340, 117976.
Gaume, E., Villeneuve, J.-P., Desbordes, M., 1998. Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. J. Hydrol. 210, 38–50.
Gaut, J., Chua, L.H.C., Irvine, K.N., LE, S.H., 2019. Modelling the washoff of pollutants in various forms from an urban catchment. J. Environ. Manag. 246, 374–383.
Giudicianni, C., Assaf, M.N., Todeschini, S., Creaco, E., 2023. Comparison of Nonlinear Reservoir and UH Algorithms for the Hydrological Modeling of a Real Urban Catchment with EPASWMM, vol. 10, p. 24.
Gupta, H.V., Kling, H., 2011. On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. Water Resour. Res. 47 (10).
Gupta, A.K., Shrivastava, R.K., 2008. Optimal Design of Water Treatment Plant under Uncertainty Using Genetic Algorithm, vol. 27, pp. 91–97.
Gupta, Hoshin V., Kling, Harald, Yilmaz, Koray K., Martinez, Guillermo F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. Journal of hydrology 377 (1–2), 80–91.
Holland, J.H., 1973. Genetic algorithms and the optimal allocation of trials. SIAM J. Comput. 2, 88–105.
Hong, Y., Liao, Q., Bonhomme, C., Chebbo, G., 2019. Physically-based urban stormwater quality modelling: an efficient approach for calibration and sensitivity analysis. J. Environ. Manag. 246, 462–471.
Hood, M., Reihan, A., Loigu, E., 2007. Modeling urban stormwater runoff pollution in Tallinn, Estonia. In: UWM, International Symposium on New Directions in Urban Water Management, Paris, pp. 12–14. September.
Hossain, Iqbal, Imteaz, Monzur, Gato-Trinidad, Shirley, Shanableh, Abdallah, 2010. Development of a catchment water quality model for continuous simulations of pollutants build-up and wash-off. International Journal of Environmental and Ecological Engineering 4 (1), 11–18.
Hossain, Sharif, Hewa, Guna Alankarage, Wella-Hewage, Subhashini, 2019. A comparison of continuous and event-based rainfall–runoff (RR) modelling using EPA-SWMM. Water 11 (3), 611.
Huang, J.J., Xiao, M., Li, Y., Yan, R., Zhang, Q., Sun, Y., Zhao, T., 2022. The optimization of Low Impact Development placement considering life cycle cost using Genetic Algorithm. J. Environ. Manag. 309, 114700.
Huber, Wayne C., Dickinson, Robert E., Barnwell Jr., Thomas O., Branch, Assessment, 1988. Storm Water Management Model; Version 4. Environmental Protection Agency, United States.
Kavetski, D., Kuczera, G., Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resour. Res. 42.
Kleidorfer, M., Deletic, A., Fletcher, T.D., Rauch, W., 2009. Impact of input data uncertainties on urban stormwater model parameters. Water Sci. Technol. : a journal of the International Association on Water Pollution Research 60, 1545–1554.
Krebs, G., Kokkonen, T., Valtanen, M., Koivusalo, H., Setälä, H., 2013. A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. Urban Water J. 10, 394–410.
Leutnant, Dominik, Döring, Anneke, Uhl, Mathias, 2019. swmmr-an R package to interface SWMM. Urban Water J. 16 (1), 68–76.
Lopes, Moana Duarte, Gustavo Barbosa Lima da Silva, 2021. An efficient simulation-optimization approach based on genetic algorithms and hydrologic modeling to assist in identifying optimal low impact development designs. Landsc. Urban Plann. 216, 104251.
Madsen, Henrik, Wilson, Geoffrey, Ammentorp, Hans Christian, 2002. Comparison of different automated strategies for calibration of rainfall-runoff models. Journal of hydrology 261 (1–4), 48–59.
Masseroni, D., Cislaghi, A., 2016. Green roof benefits for reducing flood risk at the catchment scale. Environ. Earth Sci. 75, 579.
May, D., Sivakumar, M., 2008. Comparison of artificial neural network and regression models in the prediction of urban stormwater quality. Water Environ. Res. 80, 4–9.
McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the nash–sutcliffe efficiency index. J. Hydrol. Eng. 11 (6), 597–602.
McDonnell, Bryant E., Ratliff, Katherine, Tryby, Michael E., Wu, Jennifer Jia Xin, Mullapudi, Abhiram, 2020. PySWMM: the python interface to stormwater management model (SWMM). J. Open Source Softw. 5 (52), 1.
Moriasi, D.N., Arnold, J.G., VAN Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50, 885–900.
Moussa, Roger, Chahinian, Nanée, 2009. Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events. Hydrol. Earth Syst. Sci. 13 (4), 519–535.
Murphy, L.U., Cochrane, T.A., O'Sullivan, A., 2015. Build-up and wash-off dynamics of atmospherically derived Cu, Pb, Zn and TSS in stormwater runoff as a function of meteorological characteristics. Sci. Total Environ. 508, 206–213.
Nash, J. Eamonn, Sutcliffe, Jonh V., 1970. River flow forecasting through conceptual models part I—a discussion of principles. Journal of hydrology 10 (3), 282–290.
Nazahiyah, R., Yusop, Z., Abustan, I., 2007. Stormwater quality and pollution loading from an urban residential catchment in Johor, Malaysia. Water Sci. Technol. 56, 1–9.
Niazi, Mehran, Nietch, Chris, Maghrebi, Mahdi, Jackson, Nicole, Bennett, Brittany R., Tryby, Michael, Massoudieh, Arash, 2017. Storm water management model: performance review and gap analysis. Journal of Sustainable Water in the Built Environment 3 (2), 04017002.
Obropta, C.C., Kardos, J.S., 2007. Review of urban stormwater quality models: deterministic, stochastic, and hybrid approaches. J. Am. Water Resour. Assoc. 43, 1508–1523.
Petrucci, G., Bonhomme, C., 2014. The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data. Journal of Hydrology 517, 997–1007. https://doi.org/10.1016/j.jhydrol.2014.06.019.
Rezaei, Abdul Razaq, Ismail, Zubaidah, Niksokhan, Mohammad Hossein, Amin Dayarian, Muhammad, Ramli, Abu Hanipah, Sharif, Moniruzzaman Shirazi, 2019. A quantity–quality model to assess the effects of source control stormwater management on hydrology and water quality at the catchment scale. Water 11 (7), 1415.
Riaño-Briceño, G., Barreiro-Gomez, J., Ramirez-Jaime, A., Quijano, N., Ocampo-Martinez, C., 2016. MatSWMM – an open-source toolbox for designing real-time control of urban drainage systems. Environ. Model. Software 83, 143–154.
Ritter, A., Muñoz-Carpena, R., 2013. Performance evaluation of hydrological models: statistical significance for reducing subjectivity in goodness-of-fit assessments. J. Hydrol. 480, 33–45.
Rodak, C.M., Jayakaran, A.D., Moore, T.L., David, R., Rhodes, E.R., Vogel, J.R., 2020. Urban stormwater characterization, control, and treatment. Water Environ. Res. 92, 1552–1586.
Rossman, L., 2004. Storm Water Management Model User's Manual Version 5.0. US Environmental Protection Agency, Washington, DC. EPA/600/R-05/040.
Rossman, L.A., 2010. Storm Water Management Model User's Manual, Version 5.0. National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency, Cincinnati, p. 276.
Rusanen, Anton, Björklund, Anton, Manousakas, Manousos I., Jiang, Jianhui, Kulmala, Markku T., Puolamäki, Kai, Daellenbach, Kaspar R., 2024. A novel probabilistic source apportionment approach: Bayesian auto-correlated matrix factorization. Atmos. Meas. Tech. 17 (4), 1251–1277.
Sahraei, Shahram, Asadzadeh, Masoud, Unduche, Fisaha, 2020. Signature-based multi-modelling and multi-objective calibration of hydrologic models: application in flood forecasting for Canadian Prairies. J. Hydrol. 588, 125095.
Sartor, James D., Boyd, Gail B., 1972. Water Pollution Aspects of Street Surface Contaminants, vol. 2. US Government Printing Office.
Seibert, Simon Paul, Ehret, Uwe, Zehe, Erwin, 2016. Disentangling timing and amplitude errors in streamflow simulations. Hydrol. Earth Syst. Sci. 20 (9), 3745–3763.
Shamsi, U.M.S., Koran, J., 2017. Continuous calibration. Journal of Water Management Modeling.
Simpson, I.M., Winston, R.J., Brooker, M.R., 2022. Effects of land use, climate, and imperviousness on urban stormwater quality: a meta-analysis. Sci. Total Environ. 809, 152206.
Singh, S.K., Bárdossy, A., 2012. Calibration of hydrological models on hydrologically unusual events. Advances in Water Resources 38, 81–91. https://doi.org/10.1016/j.advwatres.2011.12.006.
Singh, Nirankar, Poonia, Twinkle, Singh Siwal, Samarjeet, Srivastav, Arun Lal, Sharma, Harish Kumar, Mittal, Susheel K., 2022. Challenges of water contamination in urban areas. In Current directions in water scarcity research 6, 173–202. Elsevier.
Taghizadeh, S., Khani, S., Rajaee, T., 2021. Hybrid SWMM and particle swarm optimization model for urban runoff water quality control by using green infrastructures (LID-BMPs). Urban For. Urban Green. 60, 127032.
Temprano, J., Arango, Ó., Cagiao, J., Suárez, J., Tejero, I., 2006. Stormwater quality calibration by SWMM: a case study in Northern Spain. WaterSA 32, 55–63.
Tetzlaff, D., Soulsby, C., Waldron, S., Malcolm, I.A., Bacon, P.J., Dunn, S.M., Lilly, A., Youngson, A.F., 2007. Conceptualization of Runoff Processes Using a Geographical Information System and Tracers in a Nested Mesoscale Catchment, vol. 21, pp. 1289–1307.
Todeschini, S., Papiri, S., Ciaponi, C., 2012. Performance of stormwater detention tanks for urban drainage systems in northern Italy. J. Environ. Manag. 101, 33–45. https://doi.org/10.1016/j.jenvman.2012.02.003.
Todeschini, S., Manenti, S., Creaco, E., 2019. Testing an innovative first flush identification methodology against field data from an Italian catchment. J. Environ. Manag. 246, 418–425.
Tsihrintzis, V.A., Hamid, R., 1997. Modeling and management of urban stormwater runoff quality: a review. Water Resour. Manag. 11, 136–164.
Tu, M.-C., Smith, P., 2018. Modeling pollutant buildup and washoff parameters for SWMM based on land use in a semiarid urban watershed. Water, Air, Soil Pollut. 229, 121.
Tuo, Ye, Marcolini, Giorgia, Disse, Markus, Chiogna, Gabriele, 2018. A multi-objective approach to improve SWAT model calibration in alpine catchments. J. Hydrol. 559, 347–360.
Tuomela, C., Sillanpää, N., Koivusalo, H., 2019. Assessment of stormwater pollutant loads and source area contributions with storm water management model (SWMM). J. Environ. Manag. 233, 719–727.
Vaze, J., Chiew, F.H.S., 2003. Comparative evaluation of urban storm water quality models. Water Resour. Res. 39, SWC51–SWC510.
Wicke, D., Cochrane, T.A., O'Sullivan, A., 2012. Build-up dynamics of heavy metals deposited on impermeable urban surfaces. J. Environ. Manag. 113, 347–354.
Wijesiri, B., Egodawatta, P., McGree, J., Goonetilleke, A., 2016. Understanding the uncertainty associated with particle-bound pollutant build-up and wash-off: a critical review. Water Res. 101, 582–596.

Xing, W., Li, P., Cao, S.-B., Gan, L.-L., Liu, F.-L., Zuo, J.-E., 2016. Layout effects and optimization of runoff storage and filtration facilities based on SWMM simulation in a demonstration area. Water Sci. Eng. 9, 115–124.

Xu, D., Lee, L.Y., Lim, F.Y., Lyu, Z., Zhu, H., Ong, S.L., Hu, J., 2020. Water treatment residual: a critical review of its applications on pollutant removal from stormwater runoff and future perspectives. J. Environ. Manag. 259, 109649.

Yang, Y., Chui, T.F.M., 2018. Optimizing surface and contributing areas of bioretention cells for stormwater runoff quality and quantity management. J. Environ. Manag. 206, 1090–1103.

Yang, Y., Li, J., Huang, Q., Xia, J., Li, J., Liu, D., Tan, Q., 2021. Performance assessment of sponge city infrastructure on stormwater outflows using isochrone and SWMM models. J. Hydrol. 597, 126151.

Zakizadeh, F., Moghaddam Nia, A., Salajegheh, A., Sañudo-Fontaneda, L.A., Alamdari, N., 2022. Efficient urban runoff quantity and quality modelling using SWMM model and field data in an urban watershed of tehran metropolis. Sustainability 14, 1086.

US EPA , 1990, "National Polutant Discharge Elimination System Permit Application Regulations for Stormwater Discharges; Final Rule," 40 CFR Parts 122, 123, and 124, Federal Register 55(222) :47990-48091, November 16 (1990).

US EPA , 1999. "Evaluation of the toxicity and bioaccumulation of contaminants in sediment samples from Waukegan Harbor", Illinois. EPA series number pending, US EPA, Chicago, IL.

# Chapter 6 – Event Selection and Objective Functions in Stormwater Model Calibration

## 1. Introduction

Accurate rainfall–runoff modelling plays a central role in stormwater management, informing infrastructure design, flood risk assessment, and system optimization. The Storm Water Management Model (SWMM) is widely applied due to its semi-distributed structure, which balances urban heterogeneity with computational efficiency. However, the accuracy of SWMM results depends heavily on calibration, a process challenged by parameter uncertainty, data limitations, and equifinality. Since multiple parameter sets can reproduce observed hydrographs equally well, sensitivity analysis—particularly Global Sensitivity Analysis (GSA)—is essential for identifying the most influential parameters and improving calibration robustness.

Beyond parameter sensitivity, calibration outcomes are shaped by two critical methodological choices: event selection and objective function. The criteria by which events are chosen can therefore shape model parameterization and performance. Despite this, systematic investigations of event-selection strategies in urban catchments remain limited. Similarly, choice of the objective function used in calibration strongly influences parameter sets and model outputs.

## 2. Optimizing Hydrological Modelling on Real Urban Catchment: Impact of Calibration Data Selection and Objective Function

The objective of this study was to assess how calibration outcomes in urban rainfall–runoff modelling are influenced by two critical methodological choices: (i) the criteria used to select calibration events, and (ii) the choice of objective function employed in parameter optimization. By integrating global sensitivity analysis with event-based calibration, the study aimed to provide systematic evidence on how these factors affect model performance in terms of total runoff volume, peak discharge, and time-to-peak.

Specifically, the study examined:

- how different storm event selection criteria influence calibration outcomes in reproducing runoff volume, peak discharge, and time-to-peak,

- how various objective functions affect calibration robustness and performance across metrics
- the role of global sensitivity analysis (Morris method) in identifying the most influential SWMM parameters for optimization
- how these methodological choices impact parameter identifiability and overall model reliability**.**

The results of this analysis are detailed in the paper titled *"*Optimizing Hydrological Modelling on Real Urban Catchment: Impact of Calibration Data Selection and Objective Function*"* published in the *Journal of Hydrology*.

Journal of Hydrology 662 (2025) 133821

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

# Optimizing hydrological modelling on real urban catchment: impact of calibration data selection and objective function

Mohammed N. Assaf [a,*], Nicolò Salis [a], Sauro Manenti [a,b], Lorenzo Tamellini [c], Enrico Creaco [a,b], Sara Todeschini [a,b]

[a] Department of Civil Engineering and Architecture (DICAr), University of Pavia, Pavia, Italy
[b] Interdepartmental Centre for Water Research (CRA), University of Pavia, Pavia, Italy
[c] CNR-IMATI, National Research Council - Institute for Applied Mathematics and Information Technologies, Pavia, Italy

ARTICLE INFO

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Zhenxing Zhang, Associate Editor

*Keywords:*
Hydrological modelling
Model calibration
Objective function
Global sensitivity analysis
Rainfall–runoff model
Stormwater management
Urban hydrology

ABSTRACT

Hydrological models, like the Stormwater Management Model (SWMM), play a crucial role in water resource management. However, their effectiveness often depends on calibration based on historical rainfall-runoff events. Typically, calibration of urban drainage models is based on a select few observed rainfall-runoff events, which can be chosen from a broader dataset using various selection methods. This study explores the interaction between event selection criteria and objective functions in hydrological modelling. Four event selection approaches (Rainfall Depth (RD), Maximum Rainfall Intensity over a Five-Minute Interval (H5), Mean Intensity (MI), and Hyetograph Centre of Mass (HCM)) and six objective functions, were tested using high-resolution data from 42 rainfall-runoff events in an urban catchment area. Additionally, this research conducts a Global Sensitivity Analysis (GSA) of the drainage model parameters using the Morris method through an innovative integration of Mat-SWMM and the Sensitivity Analysis for Everybody (SAFE) Toolbox, creating a robust framework for assessing how various uncertain input parameters affect the hydrological model outputs. This integration facilitates the GSA process and allows researchers to conduct sensitivity studies on SWMM models using widely available tools. It provides a reproducible, scalable, and computationally efficient framework suitable for handling hydrological datasets of various sizes. Results reveal that sensitivity to model parameters varies significantly depending on the characteristics of the rainfall events. Only five influential parameters, out of the ten tested parameters, were identified and subsequently optimised using a genetic algorithm.

The study highlights the strengths of different event selection criteria for specific outputs: the RD criterion provided accurate estimates for total runoff volume, whereas the HCM and H5 criteria performed better in estimating peak flow rates. The results show that traditional objective functions, such as the squared difference, are particularly sensitive to the choice of the calibration event selection approach.

## 1. Introduction

Modelling hydrological processes serves as a critical tool for simulating watershed processes and supporting water resource management. However, their reliability is constrained by challenges in model identification, which stem from parameter uncertainty inherently associated with data limitations, model structural inadequacies, and observational errors (Manovam and Todini, 2006). Model identification refers to the process of defining a model structure and selecting appropriate parameter values that best represent the physical system under study. The inherent complexity of hydrological systems, compounded by varied land-use patterns, slopes, and coverage in the urbanized area, leads to complex interactions between different urban drainage structures and systems for wastewater and stormwater. These complexities are magnified by the varying temporal and spatial scales at which these interactions occur, which in turn escalate data requirements and adds to the model's complexity (Hamel and Tim, 2014). These factors, combined with the often-limited availability of detailed data, make it difficult to obtain a physically meaningful set of model parameters (Beven, 2006).

* Corresponding author.
*E-mail addresses:* mohammed.assaf@unipv.it, ma.mohammed.assaf@gmail.com (M.N. Assaf).

https://doi.org/10.1016/j.jhydrol.2025.133821
Received 31 March 2025; Received in revised form 5 June 2025; Accepted 2 July 2025
Available online 8 July 2025

Hydrological models are simplified representations of natural systems, so model parameters lose clear physical meaning and instead act as conceptual tools to simulate real-world behaviours. Because of this, there is no straightforward link between observed data and model parameters, which means parameter estimation process should be done indirectly. This model estimation process includes adjusting model parameters using location-specific historical data to improve the alignment between simulated and observed hydrological responses, which known as model calibration. The effectiveness of these calibration processes can be attributed to three main aspects: the approach for identifying optimal parameter sets, refinement of model optimisation, and the choice of data for calibration and validation (Beven, 2012).

One key challenge in model calibration is the difficulty of achieving parameter identifiability. This issue stems from three main sources of uncertainty: (i) data limitations, (ii) model uncertainty, and (iii) uncertainties in observations (Beven and Binley, 1992). Data limitations occur due to the limited availability of hydrological measurements across both space and time, often leading to overparameterization and equifinality, where different parameter combinations produce similarly accurate results (Beven, 2001). Model uncertainty arises from the simplifications made in representing real-world hydrological processes, as models use mathematical formulations that may not fully account for the intricate interactions between water, energy, and vegetation (Sorooshian and Gupta, 1983; Sivapalan et al., 2003). Lastly, uncertainties in observations result from measurement errors, equipment limitations, and the challenge of capturing hydrological variables at an appropriate scale (Blöschl and Sivapalan, 1995; Kuczera and Mroczkowski, 1998).

One key factor influencing parameter identifiability is the number of observations used for calibration. While increasing the number of calibration events can contribute to a broader representation of system behaviour, it only partially mitigates parameter equifinality. This is because the model outputs, such as total basin outflow, may be insensitive to certain parameters, meaning that variations in those parameters do not result in distinguishable changes in the output. When adjusting a large number of parameters with limited observational data, multiple parameter sets can produce similar model performance, leading to equifinality, non-uniqueness, and non-identifiability (Beven, 2001; Jakeman and Hornberger, 1993; Blöschl and Sivapalan, 1995; Moradkhani and Sorooshian, 2008; Beven and Freer, 2001). Studies have shown that increasing the number of observations can improve parameter identifiability and reduce equifinality (Kuczera and Mroczkowski, 1998, her and Chaubey, 2015). Recently, Wu et al. (2024) demonstrated that using 3–5 events in the calibration process can significantly decrease the model uncertainty and equifinality while providing proper model performance. Moreover, the resolution of land use cover data of the catchment area was found to have a high influence on model performance and uncertainty (Rahimi and Ebrahimian, 2024). Low-resolution land cover data (like data derived from coarse-resolution satellite imagery) may lead to overestimating or underestimating hydrological responses and increase model uncertainty (Hatt et al., 2004; Walsh et al., 2005; Balha et al., 2023). To mitigate the impact of data limitations and uncertainties in observations, this study draws on a rich 20 years dataset from the Cascina Scala experimental area. It includes long-term hydrological records and detailed land use data, capturing diverse hydrological responses. The large pool of high-resolution data strengthens model calibration and validation, reducing model uncertainty and improving parameter identifiability, making this site an ideal case study for various hydrological modelling research (Barco et al., 2008a; Wang and Altunkaynak, 2012; Granata et al., 2016; Azar et al., 2025).

Reduce the number of model parameters that need to be estimated (calibrated) is considered an effective strategy for addressing equifinality and enhancing the identifiability of parameter values (Hornberger et al., 1985; Jakeman & Hornberger, 1993; Wagener et al., 2001; Moreda et al., 2006). Parameter reduction also helps improve computational efficiency and faster the calibration process (Muleta and Nicklow, 2005; Knighton et al., 2016).

Sensitivity analysis is a fundamental procedure in hydrological modelling to determine the impact of various input parameters on the model outputs. This analysis is generally categorized into two main approaches: local and global. Local sensitivity analysis changes one input parameter at a time around a predefined baseline to see how it affects outputs (Song et al., 2015). However, this approach has significant limitations, particularly in complex models, where numerous nonlinear relationships exist between input and output variables (Shin et al., 2013; Zhan et al., 2013). The current work used global sensitivity analysis (GSA), which assesses the influence of all parameters simultaneously, thereby providing a more comprehensive understanding of their collective impact on output uncertainty (Baker et al., 2022; Baker et al., 2023).

Event-based calibration, which focuses on distinct rainfall-runoff events, is favoured over continuous method for its flexibility and ability to target informative events, leading to more robust and generalizable model calibration (Hossain et al., 2019; Singh and Bárdossy, 2012). A main challenge in event-based calibration is choosing which events to use for calibration and which for validation. The goal is to select events that reflect the full range of expected conditions, so the model remains accurate and reliable.

The impact of calibration event selection on urban drainage model performance has been subject of previous research (Tscheikner-Gratl et al. 2016; Kleidorfer et al. 2009; Schütze et al. 2002; and Mourad et al. 2005). Tscheikner-Gratl et al. (2016) calibrated water levels in a catchment's outflow pipe using 10 different rain events. In the results of this work, while some events were successfully replicated in calibration, the calibrated models varied in their ability to predict other events and produced different flooding volumes when applied to design storms. Kleidorfer et al. (2009) compared calibration results from the five longest duration events against the five highest peak flow events in combined sewer overflow (CSO) volumes. They found that using the longest duration events could reduce the number of required measurement sites for successful calibration. Schütze et al. (2002) noted that calibrating based on discrete events was time-efficient compared to full time series calibration, but it introduced additional uncertainties. Mourad et al. (2005) demonstrated the sensitivity of stormwater quality model calibration to the selection and number of events used.

Even though the studies cited above bring to the fore the effects of event selection on model calibration, it can also be noted that they have specific limitations in relation to the research methodologies employed. First, it can be noted that they consider only a limited range of generally available options when selecting calibration events. Secondly, they mainly employ the traditional event-based approach. Finally, it is common for these studies to depend on a limited number of events, which may not fully represent the range of scenarios that may be encountered by a model.

In hydrological modelling, manual calibration uses a trial-and-error to adjust parameters by comparing observed and simulated data, through metrics like NSE, and RMSE. While potentially accurate, it is time-consuming and subjective. To overcome these limitations, calibration is shifted to automated, framed as an optimisation problem aiming to determine parameter set that best aligns with specific optimisation objectives. These objective functions quantify how well the model replicates observed catchment behaviours.

Numerous research have explored hydrological models performance under numerous objective functions. Wu and Liu (2014) found that using sum of absolute errors is a viable option that delivers enhanced performance within an automated calibration framework. Fowler et al. (2018) examined eight objective functions across five conceptual model structures, focusing on multi-year droughts in the east and south of Australia. Houshmand Kouchi et al. (2017) conducted a study where they applied three different optimisation algorithms together with five different objective functions within a SWAT model to calibrate monthly

discharges in two watersheds in Iran. Moussa and Chahinian (2009) evaluated single and multi-objective functions for a lumped rainfall-runoff model in the Gardon catchment, Southern France, highlighting how parameter calibration heavily depends on the selected criteria. Garcia et al. (2017) assessed the sensitivity and robustness within rainfall-runoff model of an array of objective functions, each one focusing on varying discharge transformations or low-flow indices to amalgamations of such functions.

Even with the availability of the wide-ranging studies cited above, it can still be noted that there is a gap in research specifically paying attention to the estimated parameter values resulting from varying calibration event selection approaches and objective functions. Based on our best knowledge, no previous studies have examined how the interaction between event selection criteria and objective function choice influences both the calibration process and the resulting model behaviour. This gap extends to the influence of these parameters on the interpretation of hydrological processes in the watersheds under study. Understanding how different calibration event selection approaches and objective functions affect the model's calibrated parameters is crucial for ensuring model accuracy and applicability. Researchers have analysed event selection and chosen the objective function separately, but their effects on model calibration, sensitivity and performance together is not fully understood. This understanding not only impacts the model's performance but also informs our comprehension of the hydrological processes within the studied catchments. These parameters can yield insights into the watershed's behaviour, influencing decisions on water management and conservation strategies. This research gap indicates a potential area for further investigation. Delving into how these optimised parameters are derived and their implications for hydrological process interpretation could significantly enhance the effectiveness and applicability of hydrological models.

The primary aim of this research is to understand how different calibration event selection approaches and objective functions influence the calibrated parameters of hydrological models, which is essential for ensuring model accuracy and applicability. To the best of our knowledge, no prior studies in the literature have explored the interaction between event selection criteria and objective functions in hydrological modelling. This study examines various calibration event selection approaches and objective functions in an experimental urban catchment located in the city of Pavia (Northern Italy), using 42 high-resolution rainfall-runoff events with the Stormwater Management Model (SWMM). Additionally, a Global Sensitivity Analysis (GSA) of the drainage model parameters was conducted using the Morris method through a novel integration of Mat-SWMM and the Sensitivity Analysis for Everybody (SAFE) Toolbox. This framework provides a comprehensive approach to assess the impact of different uncertain input parameters on the model's outputs.

## 2. Case study and data

The Cascina Scala urban catchment, described more deeply in the studies by Barco et al. (2008b) and Todeschini et al. (2019), is located in Pavia, northern Italy (Fig. 1). This catchment, which is primarily used for residential purposes with approcimatly 1500 inhabitants, covers an area of 12.7 ha and is characterized by a slope from northwest to southeast, averaging 0.15 %. The land is connected to the sewer system, with around 62 % of the area being impervious, comprising 22.4 % roofs and 39.6 % streets and paved surfaces. The remaining 38 % of the area is pervious. The sewer system in this catchment is combined and constructed from standard concrete pipes, with a Manning's roughness coefficient n = 0.0147 s $m^{-1/3}$, which is within the typical range for concrete pipes reported in the literature (Davis, 1952). The catchment is divided into 42 subbasins, each drained by a pipe. The catchment design, featuring egg-shaped sewer sections and an average slope of 0.042 m/m, helps maintain high flow velocities and minimizes sediment deposition. The detailed characteristics of each sub-catchment, including slope, area, and ratio of pervious to impervious surfaces, can be found in Giudicianni et al. (2023).

Rainfall events at Cascina Scala area were monitored using two tipping-bucket rain gauges. These gauges possess a sensitivity of 0.2 mm and a funnel area of 1,000 $cm^2$. Additionally, runoff data at the terminal reach of the sewer network were captured using an ultrasonic depth meter strategically positioned upstream of a Venturi flume until June 2000 and after that, bubble flow meter was used instead of ultrasonic depth meter.

For the rainfall-runoff simulation in this study, a representative subset of rainfall events was meticulously chosen based on specific criteria to ensure the reliability and relevance of the data. These criteria include:

- Reliable Instrumentation: ensuring regular operation of rain gauges and the flow meter was crucial. This means that the selected events must report neither recorded failures in the instrumentation nor occurrences of pressurized flow conditions that could skew the data.
- Minimum Total Rainfall Depth: each event in the subset needed to have a total rainfall depth of at least 5 mm. This criterion ensures that the selected events have enough rainfall to impact the runoff, making them relevant for the study.
- Minimum Rainfall Intensity: the intensity of the rainfall during these events should be equal to or greater than 0.1 mm/min. This


**Fig. 1.** Cascina Scala experimental catchment, Pavia, Italy.

minimum intensity threshold is important for filtering out lighter rainfall events that might not affect the runoff dynamics in the urban watershed.

- Minimum Rainfall Depth over a Short Duration: The events should exhibit a maximum rainfall depth of at least 2 mm over a 15-minute period.

After reviewing a list of potential events, forty-two were selected as matching the criteria mentioned above. These events are described in detail in Table 1. The steps for selecting and splitting events for calibration and validation are explained in Section 3.3. In a related study, Todeschini et al. (2018) studied the placement strategies and cumulative effects of wet-weather control practices for intermunicipal sewerage systems, adopting for the analysis a continuous series of rainfall events acquired in Pavia from August 2006 to July 2007. The values of precipitation depth and duration for that year are consistent with the corresponding ranges for events reported in the current study. Moreover, in Todeschini et al. (2019), some of the events used in the current study were used for an analysis aimed at testing an innovative first flush identification methodology. The selected subset of storms proved effective in characterizing the dynamics of different types of pollutants in wet-weather runoff. These findings suggest that the rainfall events selected for this study are appropriate for urban catchments and can be used in wider hydrological and water quality studies.

## 3. Methodology

The methodological framework adopted in this study is depicted in Fig. 2. The first stage is the development of the Hydrology Module and the Hydraulics Module for the case study using the EPA Storm Water Management Model (SWMM). These modules are key parts of the SWMM and are designed to simulate hydrological processes and water movement through the system. The hydrograph is the output from these modules, displaying how wet weather flow or runoff changes over time in response to precipitation events. The methodology's next phase involves determining the influential parameters that have a significant effect on the hydrograph. To achieve this goal, the Morris method of sensitivity analysis methodically changes each parameter to determine its impact on the output of the model. This process is achieved using Mat-SWMM, together with the Sensitivity Analysis for Everybody (SAFE) Toolbox. Mat-SWMM plays the role of the interface linking MATLAB and SWMM to enable an exchange and analysis of data that is simple. Consequently, the SAFE Toolbox is employed to perform the Morris sensitivity analysis. This is an important phase for narrowing

**Table 1**
Main characteristic of the 42 rainfall events used in the study. Parameters include rainfall duration (minutes), rainfall depth (mm), maximum 5-minute rainfall intensity (mm), peak flow rate ($m^3/s$), antecedent dry period (days), and hyetograph centre of mass.

| Event No. | Rainfall duration (min) | Rainfall depth (mm) | Maximum intensity rainfall over a five-minute (mm) | Peak flow rate ($m^3/s$) | Antecedent dry period (d) | Centre of Mass |
|---|---|---|---|---|---|---|
| 1* | 11 | 12.8 | 8.8 | 0.96 | 4.2 | 0.68 |
| 2 | 197 | 11.8 | 2.7 | 0.25 | 0.5 | 0.27 |
| 3** | 539 | 22.6 | 1.16 | 0.18 | 36.2 | 0.55 |
| 4** | 108 | 16.4 | 4.5 | 0.55 | 10.8 | 0.53 |
| 5* | 23 | 15.6 | 7.5 | 0.88 | 4.8 | 0.46 |
| 6* | 50 | 7 | 1.8 | 0.33 | 10 | 0.28 |
| 7* | 64 | 11 | 3.4 | 0.38 | 1.8 | 0.46 |
| 8* | 215 | 10.6 | 1.3 | 0.16 | 0.9 | 0.53 |
| 9* | 247 | 3.8 | 0.4 | 0.03 | 0.3 | 0.65 |
| 10** | 443 | 18.6 | 1.7 | 0.26 | 11 | 0.55 |
| 11* | 111 | 8.4 | 1 | 0.16 | 3.3 | 0.63 |
| 12* | 380 | 15.8 | 2.3 | 0.26 | 0 | 0.55 |
| 13* | 100 | 4.8 | 0.8 | 0.11 | 0.6 | 0.58 |
| 14* | 862 | 18 | 0.4 | 0.03 | 5.7 | 0.52 |
| 15* | 964 | 23.4 | 0.8 | 0.13 | 1.2 | 0.40 |
| 16* | 248 | 12.6 | 2 | 0.23 | 25.9 | 0.75 |
| 17* | 231 | 16.2 | 1.5 | 0.21 | 6.8 | 0.50 |
| 18* | 286 | 7 | 5.1 | 0.06 | 14.2 | 0.40 |
| 19* | 666 | 10.4 | 0.3 | 0.03 | 1.6 | 0.57 |
| 20* | 487 | 19.2 | 1.64 | 0.22 | 0.7 | 0.41 |
| 21** | 14 | 12.8 | 8.8 | 0.97 | 14.1 | 0.97 |
| 22** | 308 | 12.8 | 1.6 | 0.13 | 3.8 | 0.46 |
| 23** | 237 | 16.2 | 1.5 | 0.21 | 6.3 | 0.55 |
| 24 | 161 | 11 | 3.9 | 0.39 | 11.7 | 0.28 |
| 25 | 93 | 7.8 | 4 | 0.27 | 3.5 | 0.19 |
| 26** | 17 | 5.8 | 4 | 0.30 | 10.0 | 0.43 |
| 27 | 28 | 7.4 | 2.7 | 0.24 | 1.0 | 0.61 |
| 28 | 110 | 8.4 | 1 | 0.16 | 3.3 | 0.67 |
| 29** | 445 | 8.6 | 1.8 | 0.06 | 14.3 | 0.40 |
| 30 | 51 | 10.6 | 2.3 | 0.26 | 2.1 | 0.64 |
| 31 | 155 | 16.2 | 2.8 | 0.29 | 2.1 | 0.43 |
| 32 | 18 | 25 | 3.9 | 0.25 | 3.8 | 0.50 |
| 33 | 22 | 7.4 | 4.4 | 0.30 | 14.8 | 0.44 |
| 34 | 102 | 14.4 | 6.7 | 0.43 | 5.9 | 0.33 |
| 35 | 62 | 5.8 | 3.2 | 0.30 | 0.2 | 0.43 |
| 36 | 38 | 15.2 | 7.9 | 0.34 | 2.5 | 0.40 |
| 37 | 70 | 15 | 6 | 0.45 | 7.0 | 0.23 |
| 38 | 115 | 7.2 | 1.4 | 0.16 | 30.7 | 0.34 |
| 39 | 32 | 8.4 | 3.1 | 0.28 | 18.7 | 0.56 |
| 40 | 36 | 21.2 | 7.4 | 0.53 | 0.8 | 0.58 |
| 41 | 22 | 14.2 | 8.4 | 0.44 | 10.0 | 0.72 |
| 42 | 47 | 14.4 | 5.8 | 0.21 | 4.1 | 0.49 |

* event used for calibration stage.
** event used for validation stage.

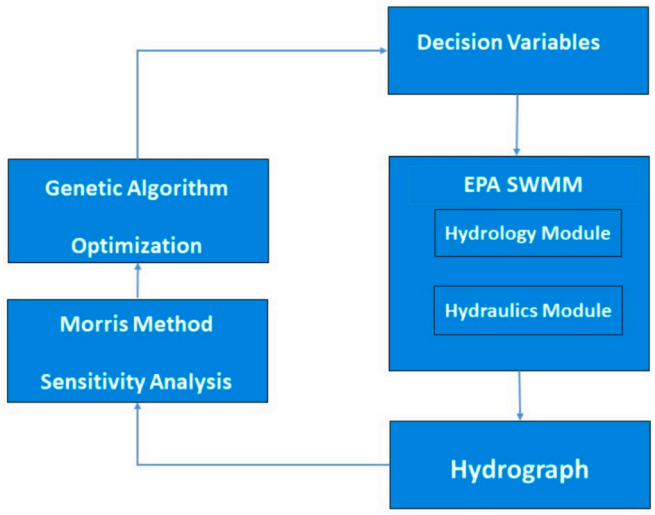



**Fig. 2.** Research methodology flowchart.

down the parameters that will be calibrated. After the sensitivity analysis, the influential parameters are fine-tuned through an optimisation process employing a Genetic Algorithm (GA), where the GA and SWMM integration is enabled by Mat-SWMM. These parameters are iteratively adjusted by the GA to optimise the performance of the model, with the objective of either maximising or minimising an objective function reflecting the hydrograph prediction's accuracy. Through the process of optimisation, critical parameters are calibrated precisely, which enhances the accuracy and reliability of the model. In this study, the model calibration process is framed as a hydrological model identification problem, in which the goal is to determine a set of parameter values that best reproduces observed hydrological responses under uncertainty. The problem is ill-posed due to equifinality, non-uniqueness, and uncertainties. To address these challenges, we adopt a structured methodology that integrates sensitivity analysis, parameter reduction, and optimisation using high-resolution, long-term data from an experimental catchment.

### 3.1. EPA-SWMM model description

The EPA-SWMM (Environmental Protection Agency – Storm Water Management Model) is a conceptual Runoff-Route model designed for both rural and urban settings. It is suitable for event-based modelling. It accommodates spatial variability by categorising catchment areas into homogeneous and smaller sub-catchments. The model simulates hydrological processes using three main algorithm groups: surface runoff algorithms, loss algorithms, and conveyance algorithms (Rossman, 2010; Rossman and Huber, 2016).

Surface runoff algorithms generate hydrographs from sub-catchments using nonlinear reservoir routing methods (Huber and Dickinson, 1988). Loss algorithms estimate stormwater losses due to depression storage, infiltration, and evaporation. Depression storage represents the initial losses that accumulate in surface depressions and do not directly contribute to runoff. In this study, evaporation is not taken into account, considering the limited temporal interval characteristic of the event-based modelling. The Horton technique was employed to calculate infiltration losses, where infiltration rate starts high and decreases over time until it reaches a minimal constant rate when the soil gets saturated. Overland flow calculations apply the law of conservation of mass, guaranteeing accurate accounting of water entering and leaving sub-catchments (Rossman and Huber, 2016). Additionally, the conveyance algorithms use dynamic wave routing, which consider flow's continuity and momentum, making a detailed analysis of the stormwater drainage system's flow characteristics available.

In SWMM, the role of the parameters is pivotal in defining the characteristics of the area under study and refining the rainfall-runoff simulations. They are generally divided into two types: inferred and measured parameters. The measured parameters denote the real measurements on the ground, like conduit dimensions and junction elevations. These parameters make quantifiable explicit data about the area being studied available. On the other hand, inferred parameters such as the Manning coefficient and depression storage are not measured directly in the field and introduce uncertainty in the model. Such parameters need careful adjustment to reduce inconsistencies in the rainfall-runoff simulations. Regarding the present study, ten inferred parameters are chosen for parameter estimation process, as shown in Table 2, along with their variability ranges, which are based on Rossman (2010). The sensitivity of these parameters is initially assessed using the Morris method to determine the parameters that are the most influential. Subsequently, a genetic algorithm is used to estimate the parameters.

As detailed by Riaño-Briceño et al. (2016), Mat-SWMM is an open-source, adaptable software package designed to enhance simulation results analysis and system editing. As an interface tool, Mat-SWMM integrated MATLAB functionalities into the SWMM code. It enables users to take advantage of the visualisation and computational prowess of MATLAB together with the simulation features of SWMM. In this study, Mat-SWMM was utilized to combine SWMM with MATLAB's genetic algorithm toolbox.

Mostly used in event-based calibration, the event-by-event approach usually generates unique calibrated parameters for each rainfall event. Nevertheless, this technique has numerous limitations because of its characteristic variability and possible inconsistencies. In most cases, this approach needs post-calibration synthesis, such as averaging, which may fail to accurately represent precise event nuances. Recent developments in modelling tools and automation, such as R package for SWMM, PySWMM, and Mat-SWMM, have led to the adoption of an integrated event calibration approach (Assaf et al. 2024; Giudicianni et al. 2023). This new method, which is the one used in this work, marks a significant advancement in stormwater modelling. It calibrates all events collectively, addressing the fragmentation and inconsistency of the traditional approach and offering a more holistic representation of stormwater dynamics. It also eliminates the need for post-calibration synthesis, enhancing model generalizability and applicability.

### 3.2. Morris method sensitivity analysis

In this study, sensitivity analysis was employed using the Morris method (Morris, 1991) to evaluate the influence of uncertainties in the input parameters on the rainfall-runoff model's performance. This method is effective for identifying key influential parameters, especially when the number of uncertain parameters is high and/or the model

**Table 2**
Model parameters selected for calibration process, including their notations and corresponding ranges of variation. The ranges are extracted from Rossman (2010).

| Parameter | Notation | Range of variation |
|---|---|---|
| Manning coefficient for roof ($s/m^{1/3}$) | $n_{roof}$ | 0.008–0.030 |
| Manning coefficient for street ($s/m^{1/3}$) | $n_{street}$ | 0.008–0.055 |
| Manning coefficient for the pervious area ($s/m^{1/3}$) | $n_{per}$ | 0.10–0.35 |
| Depression storage on impervious area (mm) | $S_{imp}$ | 1.00–3.54 |
| Depression storage on pervious area (mm) | $S_{per}$ | 2.54–5.08 |
| Maximum infiltration rate (mm/hr) | $I_{max}$ | 58–170 |
| Minimum infiltration rate (mm/hr) | $I_{min}$ | 1–57 |
| Decay coefficient (1/hr) | decay | 2–7 |
| Manning roughness for conduit ($s/m^{1/3}$) | $n_{cond}$ | 0.008–0.02 |
| Width coefficient (–) | $W_c$ | 1–2 |

computational time is expensive (Cappato et al. 2022). The Morris method evaluates how individual parameter variations affect model outputs by computing gradients at multiple randomly selected points in the parameter space. This is achieved by altering the value of each parameter ($p_i$) by a constant quantity ($\Delta_i$) at time along a specific path (trajectory j) in the model's parameter space. After each parameter is altered, a new initial point is selected for further adjustments along a different trajectory. when the model is run with these varying parameter combinations, the impact, known as the elementary effect (*EE*), of each parameter on the selected model output *f* (like total volume or peak flow rate) is computed using the following formula:

$$EE_{pi}(j) = \frac{f(p_1, \cdots, p_i + \Delta, \cdots, p_M) - f(\mathbf{p})}{\Delta} \quad (1)$$

The mean elementary effect ($\mu$) for each parameter is then calculated by averaging these effects, following the methodologies outlined by Morris (1991), Campolongo et al. (2007), Ruano et al. (2012), and Janetti et al. (2019). A larger $\mu$ value for a parameter indicates a higher sensitivity of the model output to that parameter. Furthermore, to account for the effects with opposing signs, which could reduce the magnitude of the elementary effect, the absolute mean elementary effect ($\mu^*$) is computed as follows:

$$\mu_{pi}^{*} = \frac{1}{r}\sum_{j=1}^{r} \left| EE_{pi}(j) \right| \quad (2)$$

where r denotes the aggregate number of trajectories. While the Morris method's standard approach uses multiple trajectory sizes, we examined various magnitudes to find out their impact. There is a need for the number of trajectories to strike a balance: such a number should be adequately large enough to produce significant results, lessening variance in μ*, while still being small enough to ensure efficient time for calculation, given that the model has evaluated $m_e = r(M + 1)$ times, with M representing the total number of input parameters considered in the sensitivity analysis. Our findings show that the employment of r = 1500 archives a desirable balance between efficiency and accuracy. This number of trajectories makes it possible to conduct a comparative analysis of the relative impact of each parameter on the output of the model (Reinecke et al., 2019; Campolongo et al., 2007). In addition, the elementary effects (σ) standard deviation is determined with the aim of assessing whether the impact of the parameter on the output of the model is linear or involves interactions with other parameters. When the σ is large compared to the μ*, it implies that there is a nonlinear interaction or relationship with other parameters (Reinecke et al., 2019; Feng et al., 2019; Morris (1991). Sensitive parameters are identified by computing the absolute elementary effects (μ) of each of the 10 parameters across all 42 rainfall events to ensure a comprehensive assessment of parameter behaviour across a wider range of conditions.

In this study, the sensitivity analysis is performed using an innovative integration of Mat-SWMM and the Sensitivity Analysis for Everybody (SAFE) Toolbox, creating a robust framework for assessing how various uncertain input parameters affect the hydrological model's outputs. Mat-SWMM acts as a bridge between SWMM and MATLAB, enabling smooth data exchange between the two. This integration allows the use of powerful analysis tools within MATLAB, including the SAFE Toolbox. The SAFE Toolbox, developed by Pianosi et al. (2015), is a flexible, standalone tool designed for global sensitivity analysis across different computational environments, including MATLAB. In this study, the SAFE Toolbox is used to systematically assess how changes in various input parameters affect the SWMM model's outputs, specifically runoff volume, peak flow rate, and time to peak flow rate. The innovative integration of Mat-SWMM with the SAFE Toolbox makes this process efficient and effective, enabling a thorough global sensitivity analysis within the SWMM framework. The integration of global sensitivity analysis into SWMM through this method is a novel contribution of the present work, providing an easy and efficient way to manage the complex interactions between model parameters. By combining the strengths of Mat-SWMM and the SAFE Toolbox, this study introduces a novel and powerful method for enhancing the performance of hydrological models. This integration marks a significant step forward in applying global sensitivity analysis to SWMM, offering a more streamlined and effective approach to improving model accuracy and reliability in hydrological predictions.

### 3.3. Calibration validation process

From the initial pool of 42 rainfall events (Table 1), sixteen events were selected for the model calibration phase based on hydrological and statistical criteria, ensuring representation of the overall rainfall events characteristics, including total rainfall depth, peak flow rate, varying seasons, dry periods and event durations. Four calibration selection strategies were employed, each selecting eight events from 16 events allocated for calibration. For the validation, eight events were chosen from the remaining 26 events and were selected to exhibit similar statistical characteristics to the sixteen events used for calibration. This careful selection ensured consistency in the characteristics of the dataset, allowing for valid and reliable comparison across different calibration strategies. This segregation of events into calibration and validation sets ensures that each calibration strategy was uniformly evaluated against an identical dataset, which ensures that any advantage that may have been gained from the overlapping events in both the calibration and validation sets was excluded.

In the present study, four distinct criteria are employed in calibration event selection: Rainfall Depth (RD), Maximum Rainfall Intensity over a Five-Minute Interval (H5), Mean Intensity (MI), and Hyetograph Centre of Mass (HCM). For each criterion, events were first ranked accordingly, for example, in the RD approach, events were ranked by rainfall depth, whereas in the H5 approach the rank was based on maximum rainfall intensity within a period of five minutes. Events were then alternately selected (1st, 3rd, 5th, and so on) to form the calibration set. This technique results in a dataset that preserves the characteristics defined by the selection criteria defined, which is selected to represent the calibration set for that specific criterion.

The first selection criterion is Rainfall Depth (RD), which ranks the calibration events based on the total accumulated rainfall. This approach ensures that the calibration dataset encompasses a wide range of precipitation events, from low to high rainfall depths, thereby offering a comprehensive representation of the different rainfall scenarios that could influence rainfall-runoff process.

The second selection criterion is Mean Intensity (MI), which ranks calibration events according to the average rainfall intensity over the duration of each event. The incorporation of a range of mean intensity scenarios enables the model to capture a wider spectrum of rainfall patterns, spanning from mild, extended events to more intense, short-duration occurrences. This variability in rainfall scenarios improves the model's capability to simulate and predict diverse hydrological responses, thereby enhancing its overall robustness and accuracy.

The third selection criterion is the Maximum Rainfall Intensity over a Five-Minute Interval (H5), which ranks calibration events according to the highest rainfall intensity observed within any five-minute segment of the event. Including various H5 scenarios ensures the model encounters to a spectrum of peak intensity events, ranging from moderate increases to extreme downpours. Incorporating these scenarios allows the model to effectively account for short-duration, high-intensity rainfall events, which, despite their short occurrence, can significantly influence hydrological responses.

The centre of mass of the hyetograph is the fourth criterion of event selection, which orders events based on coordinates identified as the mean precipitation and the time that equally divides the precipitation volume (ensuring equal volumes on the left and right). Values were standardised to introduce fairness to comparisons among events of various durations and levels of precipitation. This involves categorising

all heights based on maximum height and each time by the event's duration. Consequently, each event achieves a variation between 0 and 1 for both precipitation depth and duration. Such standardisation guarantees that both the duration of the event and precipitation depth are assigned equal importance. For instance, an event with a long duration but low precipitation will be comparable to an event with high intensity but short duration.

### 3.4. Objective functions

Based on an extensive review of the literature (Table 3), the objective functions applicable to studying deviations between simulated and observed values can be methodically categorised into four different groups, each of which is characterised by its distinct methodological approach to quantifying and minimising these deviations:

- Group I – In this group, the objective functions are designed to decrease the squared differences between the simulated and squared values. Squaring the deviations places greater weight on larger errors, making these criteria particularly useful when it is important to emphasize larger discrepancies more than smaller ones.
- Group II – Minimization of Absolute Deviations: In this category, the criteria do not square the deviations, reducing their sensitivity to larger discrepancies. Usually, this approach is employed in instances where the main objective is to reduce the general magnitude of errors without over-emphasising the larger errors.
- Group III: In this category, the objective functions use numerous transformations, including taking the square root or applying logarithmic transformations to both simulated and observed runoff values before the least squares computations are applied. Such a transformation can assist in stabilising the variance and rendering the data more normally distributed, which can give benefits to certain kinds of statistical analysis.
- Group IV – Aggregated Multi-Objective Approaches: multi-objective techniques in modelling offer a significant advantage by enabling the combination of various objective functions within a custom-built framework. These techniques can be employed through the derivation of aggregated or combined objective functions. This approach involves merging multiple objectives into a single composite metric, simplifying the optimisation process while still capturing the essence of multiple criteria.

**Table 3**
Description of each objective function tested for model calibration.

| Name | Group | Formula | Reference |
|---|---|---|---|
| Squared Difference (SD) | I | $SD = \sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}(Q_{p,i,j} - Q_{m,i,j})^2$ | Diskin and Simon (1977) |
| NSE | I | $NSE = 1 - \frac{\sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}(Q_{p,i,j} - Q_{m,i,j})^2}{\sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}(Q_{m,i,j} - \overline{Q_{m,i,j}})^2}$ | Nash and Sutcliffe (1970) |
| Absolute Error (Abs) | II | $Abs = \sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}\lvert Q_{p,i,j} - Q_{m,i,j}\rvert$ | Legates and McCabe (1999) |
| Index of Agreement (IoA) | II | $IoA = 1 - \frac{\sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}\lvert Q_{p,i,j} - Q_{m,i,j}\rvert}{2\sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}\lvert Q_{m,i,j} - \overline{Q_{m,i,j}}\rvert}$ | Willmott et al. (2012) |
| NSE- sqrt | III | $NSE - sqrt = 1 - \frac{\sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}(\sqrt{Q_{p,i,j}} - \sqrt{Q_{m,i,j}})^2}{\sum_{i=1}^{N_r}\sum_{j=1}^{N_{t,i}}(\sqrt{Q_{m,i,j}} - \overline{\sqrt{Q_{m,i,j}}})^2}$ | Chiew et al. (1995) |
| Volume and Peak (Q&P) | IV | $Q\&P = \sum_{i=1}^{N_r}\left(\frac{Peak_{p,i} - Peak_{m,i}}{Peak_{m,i}}\right)^2 + \left(\frac{Volume_{p,i} - Volume_{m,i}}{Volume_{m,i}}\right)^2$ | Barco et al. (2008a) |

$N_{t,i}$ and $N_r$ are the number of time instants in the generic *i*-th rain event and the number of rain events considered for optimisation, respectively. $Q_{m,i,j}$ and $Q_{p,i,j}$ are the measured and predicted flow rate, respectively, at the j-th time and in the i-th rain event. $\overline{Q_{m,i,j}}$ is the mean value of the measured flow rate. $Peak_m$ and $Peak_p$ are the measured and predicted peak flow rate, respectively. $Volume_{m,i}$ and $Volume_{p,i}$ are the measured and predicted volume, respectively.

In this study, at least one representative function of each group has been selected and tested. The selection was based on recommendations derived from the literature and preliminary analyses, which are not reported in this paper. Table 3 presents the objective functions described and the formulation group to which they belong.

### 3.5. Goodness-of-fitness

In the present study, the evaluation of the calibration and validation performance involves the employment of goodness-of-fit indices, as recommended by numerous scholars (Legates and McCabe, 1999). For this purpose, two main efficiency criteria were selected: the Coefficient of Efficiency (CE), also known as the Nash-Sutcliffe Efficiency (NSE), and the Root Mean Square Error (RMSE). For these indices, the formulas are as follows:

- Coefficient of Efficiency (CE) is given by the equation:

$$CE = 1 - \frac{\sum_{k=1}^{N_r}(V_{m,k} - V_{p,k})^2}{\sum_{k=1}^{N_r}(V_{m,k} - \overline{V_m})^2} \quad (3)$$

- Root Mean Square Error (RMSE) is calculated as:

$$RMSE = \sqrt{\frac{1}{N_r}\sum_{k=1}^{N_r}(V_{m,k} - V_{p,k})^2} \quad (4)$$

Where $V_{m,k}$ represents the measured variable, $V_{p,k}$ is the predicted variable, $\overline{V_m}$ is the mean of the measured variable and Nr denotes the total number of records.

Equations (3) and (4) (for CE and RMSE, respectively) quantify the model's predictive accuracy by conducting a comparison of the simulated (predicted) data against the observed (measured) data. The CE represents a normalised statistic that determines the comparative magnitude of the residual variance compared to the measured data variance. The RMSE is an available measure of the average magnitude of the error. In this study, CE and RMSE were applied as standard goodness-of-fit measures to evaluate the model's performance after calibration. Although these indices can also serve as optimisation criteria, our focus here is on their role in assessing how well the calibrated model reproduces observed behaviour. To ensure a consistent and comprehensive evaluation, these measures were applied to three relevant hydrological variables: total volume, peak flow rate, and time to peak. This consistent evaluation allows for a fair and comprehensive comparison of model performance across different objective functions and calibration strategies and ensures that all relevant aspects of hydrograph behaviour are assessed.

## 4. Results and discussion

### 4.1. Sensitivity analysis results

The Morris method is adopted for conducting sensitivity analysis of SWMM parameters. This method was employed to evaluate the sensitivity of 10 uncertain model parameters: $n_{roof}$, $n_{street}$, $n_{per}$, $S_{imp}$, $S_{per}$, $I_{max}$, $I_{min}$, decay, $n_{cond}$, and $W_c$ (as listed in Table 2). The analysis focuses on three hydrologic metrics: the total runoff volume, the peak flow rate and the time to peak flow rate. The focus on both global and instantaneous

flow-controlled variables ensures that the analysis captures both the cumulative water volume generated by rainfall events and the maximum instantaneous flow rate. The sensitive parameters are identified by computing the absolute elementary effects $\mu^*$ of each of the 10 parameters first and then taking their sum. A threshold is then established at 5 % of this total sum. Parameters with μ* values exceeding this threshold are classified as sensitive. The analysis is applied event by event across the entire dataset, which includes 42 events.

Without losing generality, we present the full results of sensitivity analysis for a selected event (Figs. 3 and 4) as representative examples. The complete set of sensitivity results across all events is available in the Supplementary Materials (S1–S3), allowing readers to assess the robustness and consistency of sensitivity analysis under various event conditions. Fig. 3 illustrates the convergence behaviour of the absolute value of the mean elementary effect ($\mu^*$) which has been computed for each model parameter across an increasing number of model evaluations (parameter trajectories) for a specific event (event 8). As detailed in Section 3.2, analyses were conducted using 1500 parameter trajectories to ensure the stability of parameter rankings. The plot shows that the $\mu^*$ values for all parameters tend to stabilise as the number of model evaluations increases for the three metrics, indicating robust parameter rankings. This convergence analysis underlines the stability of parameter rankings across the model evaluations, thereby validating the reliability of the sensitivity assessment conducted using the Morris method. This finding is also observed when analysing all events in the dataset.

Fig. 4 presents the absolute means and standard deviations of the EEs for the various model parameters concerning the total volume (Fig. 4.a), peak flow rate (Fig. 4.b), and time to peak flow rate (Fig. 4.c). For the total volume, $S_{imp}$ stands out with the highest μ*, approaching 1200, and a relatively small standard deviation, indicating its substantial influence and consistent impact on the total volume (Fig. 3.a; Fig. 4.a). $n_{street}$ and $W_c$ also show relatively high μ* values with moderate standard deviations, suggesting their notable effects on total volume. In contrast, the other seven parameters ($n_{roof}$, $n_{cond}$, $n_{per}$, $S_{per}$, $I_{max}$, $I_{min}$, and decay) show low μ* values, falling below the sensitivity threshold, signifying their minimal impact on this metric. The peak flow rate is most sensitive to $n_{street}$, followed by $W_c$, $n_{cond}$, and $n_{roof}$, respectively (Fig. 3.b; Fig. 4.b). The peak flow rate metric is insensitive to the $S_{imp}$, $n_{per}$, $S_{per}$, $I_{max}$, $I_{min}$, and decay in case of event 9. For the time to peak flow rate, $n_{street}$, $W_c$, $n_{cond}$, $n_{roof}$, and $S_{imp}$ exceed the sensitivity threshold, highlighting their noteworthy impact on peak flow timing.

To develop a comprehensive evaluation of model parameter sensitivity, the Morris sensitivity analysis method is applied to all events in the dataset. This approach allows the evaluation of the influence of each parameter on the three target metrics across various hydrological scenarios. The results of this sensitivity analysis are summarised in Table S1 for total volume, Table S2 for peak flow rate and Table S3 for time to peak flow rate. These tables present the μ* values for each parameter across the entire dataset event by event. Parameters that exceed the sensitivity threshold are highlighted in bold, indicating that they are crucial factors in the modelling process.

For the total volume (Table S1), $S_{imp}$ consistently exhibits the highest μ* across all events, highlighting that it is the most sensitive parameter, significantly influencing the total volume of discharge in the model. The parameters $n_{street}$ and $W_C$ also demonstrate significant sensitivity in various events, frequently exceeding the sensitivity threshold. Conversely, parameters such as $n_{per}$, $I_{max}$, $I_{min}$, and decay generally show low μ* values, often near zero or less than the sensitivity threshold, indicating their relatively minor impact on the total volume of discharge. The $n_{cond}$ parameter displays moderate sensitivity in some

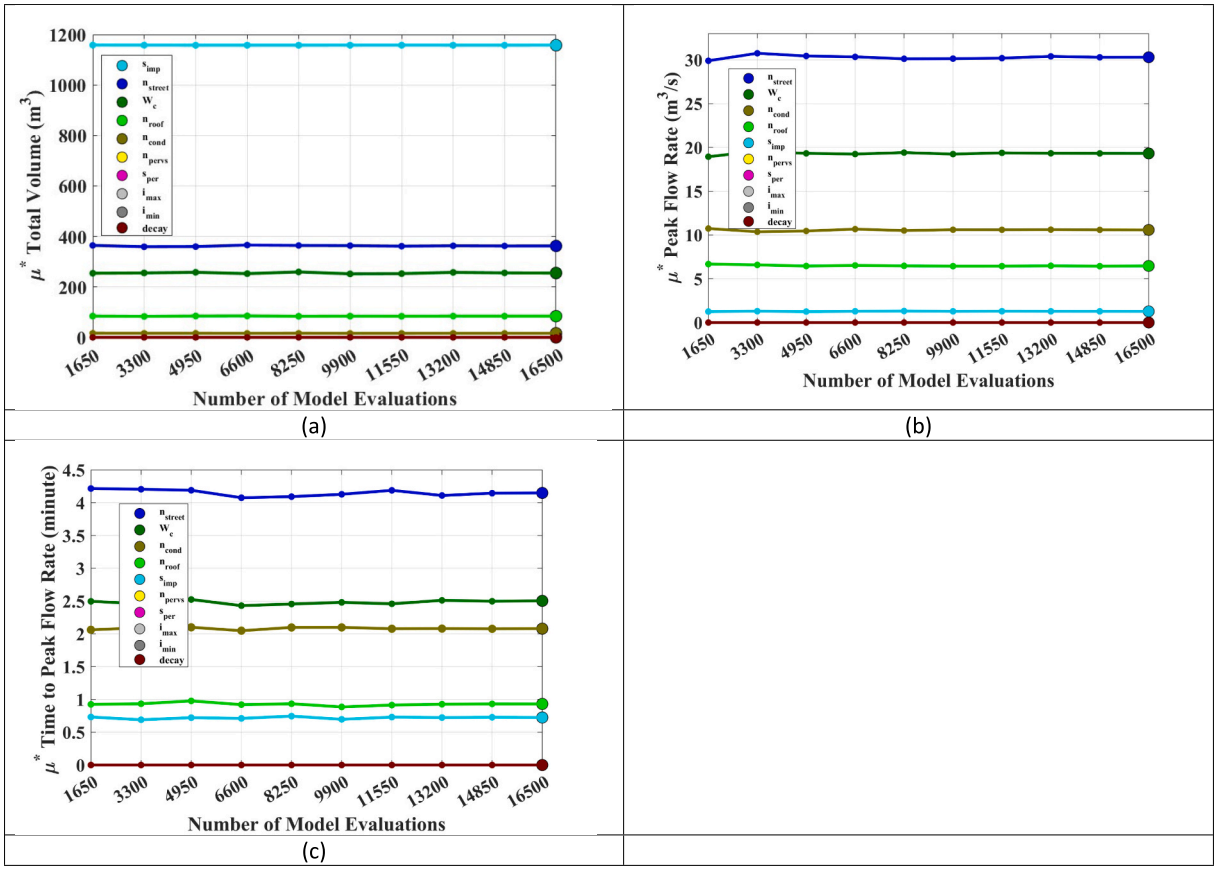



**Fig. 3.** Convergence of the mean elementary effect absolute value (μ*) of the 10 model parameters across increasing numbers of model evaluations for event 8. (a) μ* of the total volume, (b) μ* of the peak flow rate and (c) μ* of the time to peak flow rate. The parameters analysed include $S_{imp}$, $n_{street}$, $W_c$, $n_{roof}$, $n_{cond}$, $n_{per}$, $S_{per}$, $I_{max}$, $I_{min}$, and decay, as represented by different coloured markers.

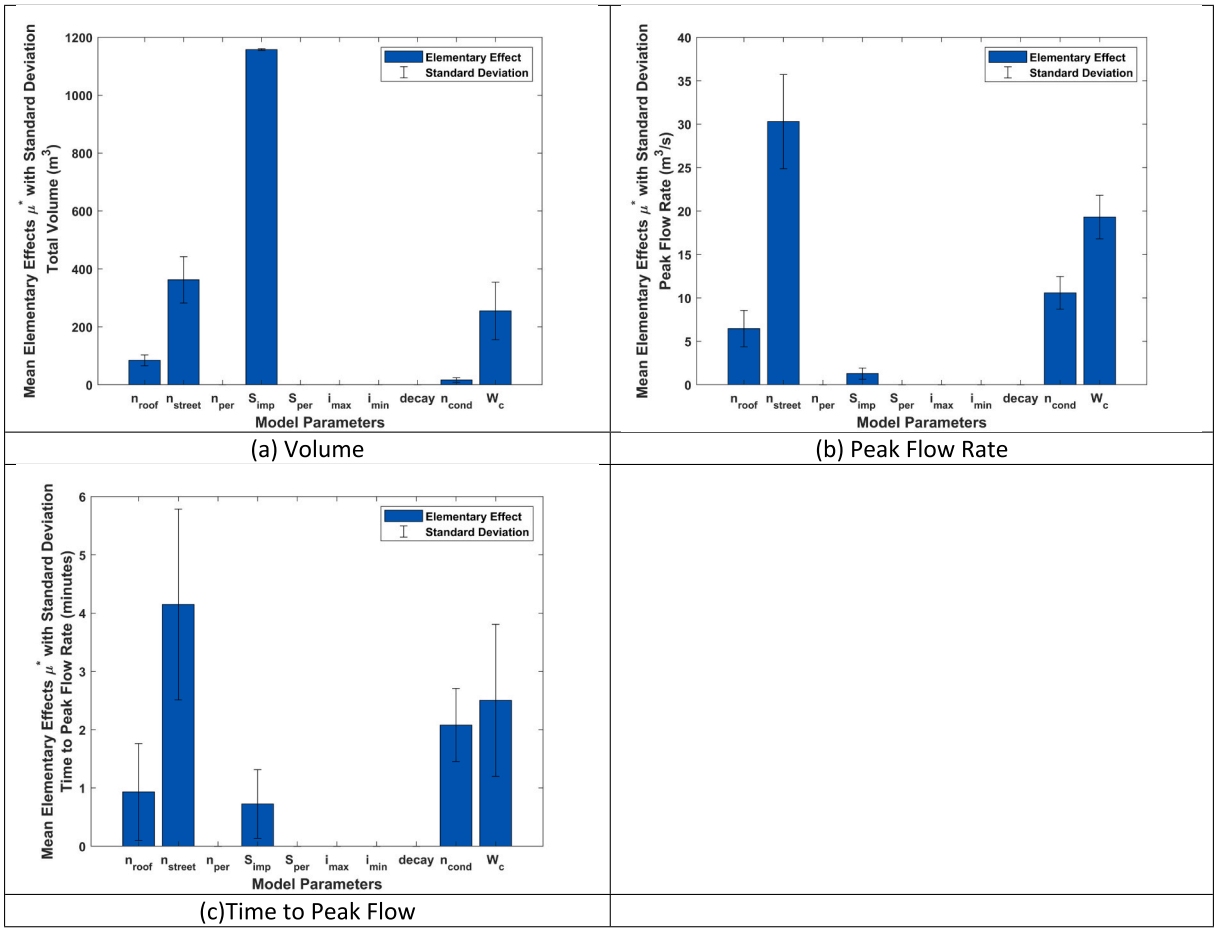



**Fig. 4.** Mean elementary effects (μ*) with standard deviations for model parameters evaluated for three key hydrological outputs: (a) total runoff volume ($m^3$), (b) peak flow rate ($m^3/s$), and (c) time to peak flow (minutes). The parameters include $S_{imp}$, $n_{street}$, $W_c$, $n_{roof}$, $n_{cond}$, $n_{per}$, $S_{per}$, $I_{max}$, $I_{min}$, and decay.

events. For instance, $n_{cond}$ has a μ* value of 714.7 at event 36 and 641.9 at event 27, indicating a notable impact on these events, though it is not as consistently influential as $S_{imp}$ or $n_{street}$. However, $n_{cond}$ show small μ*, under sensitivity threshold, in several events. These findings indicate that the sensitivity of certain parameters can vary across different events, highlighting the importance of conducting sensitivity analysis for all events in the dataset.

Table S2 presents the mean elementary effect absolute values (μ*) of the ten model parameters for peak flow rate across 42 events in the dataset. The analysis reveals that $n_{street}$ and $W_c$ consistently surpass the sensitivity threshold across all events. Additionally, $n_{cond}$ shows significant sensitivity in many events, although it falls below the sensitivity threshold in a few events. The parameters $n_{roof}$ and $S_{imp}$ also show significant sensitivity in certain events, while they remain insensitive in others. This variability in parameter sensitivity across different events underscores the complex nature of urban hydrological processes. Conversely, parameters like the $n_{per}$, $S_{per}$, $I_{max}$, $I_{min}$ and decay generally exhibit lower μ* values, suggesting they have a lesser impact on peak flow rate.

The sensitivity analysis for time to peak flow rate, as shown in Table S3, shows a similar pattern to the sensitivity results for the peak flow rate. $n_{street}$, $n_{cond}$ and $W_c$ consistently show significant sensitivity across all events. The parameters $S_{imp}$ and $n_{roof}$ demonstrate variable sensitivity from event to event; while they show moderate sensitivity in several events, they are also insensitive in some instances. Also, the $n_{per}$, $S_{per}$, $I_{max}$, $I_{min}$ and decay show low μ* values, falling below the sensitivity threshold for all events.

Table 4 summarizes the sensitivity of ten model parameters to three target metrics—total volume, peak flow rate, and time to peak flow—across all 42 events.

**Table 4**
Summary of parameter sensitivity for total volume, peak flow rate, and time to peak flow across 42 events. "Always" indicates consistent significant sensitivity across all events, "Sometime" denotes variable sensitivity from event to event, and "Never" reflects parameters that remain below the sensitivity threshold across all events.

| | Total Volume | Peak Flow | Time to Peak Flow |
|---|---|---|---|
| $n_{roof}$ | Never | Sometime | Sometime |
| $n_{street}$ | Always | Always | Always |
| $n_{per}$ | Never | Never | Never |
| $S_{imp}$ | Always | Sometime | Sometime |
| $S_{per}$ | Never | Never | Never |
| $I_{max}$ | Never | Never | Never |
| $I_{min}$ | Never | Never | Never |
| decay | Never | Never | Never |
| $n_{cond}$ | Sometime | Sometime | Always |
| $W_c$ | Always | Always | Always |

The sensitivity analysis results for the three metrics provide critical insights into the varying influences of model parameters on urban hydrological modelling. For total volume, $S_{imp}$ emerges as the dominant parameter, with $W_c$ and $n_{street}$ consistently showing significant sensitivity across all events. In contrast, for peak flow rate, $n_{street}$ and $W_c$ are more critical. Some parameters exhibit variability in sensitivity between events, showing μ* values above the sensitivity threshold in some cases and below it in others (e.g., $n_{cond}$ for the total volume metric, and $n_{roof}$, and $S_{imp}$ for peak flow rate). This variability in parameter sensitivity is

more pronounced in the peak flow rate analysis, suggesting that parameters affecting peak flow rate are more influenced by specific event characteristics such as rainfall intensity.

To better illustrate the influence of rainfall characteristics on parameter sensitivity, a summary table has been provided in the supplementary material (Table S4). This table outlines how the sensitivity of SWMM parameters varies with RD, H5, MI, and HCM, based on the results from 42 rainfall-runoff events.

The sensitivity analysis results reveal that model parameter sensitivity varies depending on the target metrics. This finding underscores the importance of conducting sensitivity analyses for different hydrological metrics to gain a comprehensive understanding of the influences of various parameters. Moreover, the analysis indicates that parameter sensitivity is not static but can change with varying external conditions, such as rainfall depth. For example, the sensitivity of infiltration parameters ($I_{max}$, $I_{min}$ and decay) increases with rainfall depth when total volume and peak flow rate are the metrics. These parameters exhibit zero $\mu^*$ values during low rainfall depth events, as the rainfall is easily absorbed by the soil in pervious areas, so that the infiltration capacity is not exceeded, and changes in $I_{max}$, $I_{min}$ and decay don't significantly alter the total volume or peak flow rate. However, when the rainfall depth increases, more water reaches the pervious surfaces, saturating the soil and activating infiltration limitations, causing their EEs to rise as the rainfall depth increases. Additionally, the sensitivity of $S_{imp}$ to peak flow rate also varies with changes in rainfall depth. During low rainfall depth events, a significant portion of rainfall is stored in surface depressions on impervious areas before contributing to runoff making small changes in $S_{imp}$ have a great effect on when runoff begins and the magnitude of peak discharge. Therefore, $S_{imp}$ shows the highest EE among all parameters, as it is the dominant parameter influencing peak flow rate in low depth events. However, as rainfall depth increases, the depression storage reaches capacity quickly, leading to immediate transfer of excess rainfall to runoff, which reduces the $S_{imp}$ relative importance. Instead, $W_c$ and $n_{street}$ take over as the primary controls of peak flow rate at high rainfall depths as these parameters determine how water flows across surfaces and through the drainage system.

Moreover, during events with high H5 (like events 1, 5 and 41), the parameters related to flow conveyance, such as $n_{street}$, $W_c$ and $n_{cond}$, increase in sensitivity due to critical role that roughness of street surfaces and conduits plays in determining how quickly and efficiently water is transported.

Overall, the Morris method was used to identify the most influential input parameters affecting various model outputs within their uncertainty ranges. This sensitivity analysis provides a robust framework for evaluating how input parameters contribute to model uncertainty, allowing parameters with minimal impact on model outputs to be assigned standard values, thereby reducing model uncertainty and computational complexity. Previous research has established the Morris method as an optimal sensitivity analysis technique due to its ability to easily remove or substitute failed simulations and its lower computational cost relative to other methods (Campolongo et al., 2007). A comparative study on the computational expense of sensitivity analyses for an energy balance model with similar simulation times to those used in the present paper (1–2 min per run) demonstrated that the Morris method is significantly more cost-effective, requiring only a few tens of runs compared to the hundreds or thousands needed by other methods (Nguyen and Sigrid, 2015).

The findings from the Morris analysis are instrumental in guiding model parameter estimation by identifying parameters to which the model is sensitive and insensitive (Herman et al., 2013). The analysis described in this work indicates that the five parameters ($n_{per}$, $S_{per}$, $I_{max}$, $I_{min}$ and decay) do not significantly affect model outcomes; thus, during calibration their values are fixed in accordance with default values from the SWMM manual (Rossman, 2015) which are respectively 0.15 $s/m^{1/3}$, 2.6 mm, 117 mm/hr, 17 mm/hr, and 5 $hr^{-1}$, allowing the optimisation efforts to focus on determining the values of the sensitive parameters. It worth noting that these default values are within the bounds covered with the Morris analysis, therefore being eligible as "reference values". Five parameters ($S_{imp}$, $W_c$, $n_{street}$, $n_{roof}$ and $n_{cond}$) are found to be sensitive based on the sensitivity analysis results for the three metrics across all events. Consequently, these five parameters are considered in the parameter estimation process. As discussed in previous studies, reducing the number of calibration parameters is a well-established strategy to mitigate equifinality (Beven, 2006; Beven and Binley, 1992). By limiting the calibration to the most sensitive parameters, the model becomes better constrained, reducing non-uniqueness and improving the likelihood of identifying parameter sets that are both statistically valid and hydrologically meaningful. This is particularly critical in the context of ill-posed inverse problems, where the number of observations is often insufficient to support the calibration of a large parameter set (Jakeman and Hornberger, 1993; Kuczera and Mroczkowski, 1998). Moreover, focusing on influential parameters helps enhance the efficiency and interpretability of the calibration process, while improving the model's predictive reliability across different hydrological conditions. In line with the broader concept of model identification, this strategy strengthens the correspondence between observed system behaviour and model representation, thus contributing to a more robust and defensible modelling framework under uncertainty (Wagener and Hoshin, 2005; Her and Chaubey, 2015).

A second sensitivity analysis was conducted for each of the six objective functions used in the study. The Morris method was applied to assess how changes in each parameter influence the objective function's value and to evaluate how parameter sensitivity varies across different functions. This analysis aims to ensure that the objective functions are sensitive to the parameters being optimised during calibration process, so that the differences observed in the calibrated parameter values are a direct result of the objective function characteristics, rather than being influenced by insensitive or irrelevant parameters. The analysis is applied on an event by event across the entire dataset, which includes 42 events, and a summary of results is presented in Table 5.

The results confirm that five parameters ($S_{imp}$, $n_{street}$, Wc, $n_{cond}$, and $n_{roof}$) were consistently identified as sensitive across all six objective functions. This result serves as a validation of the parameter selection process. Since the same set of influential parameters was used to calibrate the model under each objective function, any differences observed in the resulting parameter values or model outputs can be more likely attributed to the intrinsic characteristics of the objective functions.

### 4.2. Parameter optimisation results

The parameter estimation process focuses on the five identified sensitive parameters ($S_{imp}$, $W_c$, $n_{street}$, $n_{roof}$ and $n_{cond}$). This section presents the results of the optimisation process applied to the calibration sets, highlighting how different calibration event selection criteria and objective functions impact the estimation of these key parameters. Fig. 5 provides a comprehensive comparison of 24 different optimised values for each parameter, resulting from the combination of four calibration event selection criteria (RD, H5, MI, and HCM) and six objective functions (SD, Abs, NSE, NSE-sqrt, Q&P, and IoA).

The results for $n_{roof}$ (Fig. 5.a) highlight the variation in calibrated values across different combinations of event selection approaches and objective functions. The $n_{roof}$ values range from 0.0080 to 0.029 $s/m^{1/3}$, with the RD approach tending to produce higher values across most objective functions. Fig. 5.b shows the results for $n_{street}$, which indicate variation depending on the calibration event selection approach. The RD approach consistently produces higher $n_{street}$ values across most objective functions. In contrast, the H5, MI, and HCM approaches yield lower and more consistent $n_{street}$ values.

Fig. 5.c presents the optimisation results for $S_{imp}$, which indicate substantial variation in $S_{imp}$ values depending on the combination of calibration approach and objective function. The $S_{imp}$ values range from 1.1 to 3.5 mm, highlighting the critical and combined role of calibration

**Table 5**
Sensitivity analysis of the six objective functions to 10 SWMM parameters, across 42 events. The numbers represent the count of events in which each parameter exceeds the sensitivity threshold, with the corresponding percentage of total events shown in parentheses.

| | $n_{roof}$ | $n_{street}$ | $n_{per}$ | $S_{imp}$ | $S_{per}$ | $I_{max}$ | $I_{min}$ | decay | $n_{cond}$ | $W_c$ |
|---|---|---|---|---|---|---|---|---|---|---|
| SD | 38 (90.4 %) | 42 (100 %) | 0 | 42 (100 %) | 0 | 0 | 0 | 0 | 42 (100 %) | 39 (92.9 %) |
| Abs | 35 (83.3 %) | 42 (100 %) | 0 | 42 (100 %) | 0 | 0 | 0 | 0 | 40 (95.2 %) | 42 (100 %) |
| NSE | 37 (88.1 %) | 42 (100 %) | 0 | 42 (100 %) | 0 | 0 | 0 | 0 | 42 (100 %) | 41 (97.6 %) |
| NSE-sqrt | 36 (85.7 %) | 41 (97.6 %) | 0 | 42 (100 %) | 0 | 0 | 0 | 0 | 41 (97.6 %) | 40 (95.2 %) |
| Q&P | 29 (70.7 %) | 42 (100 %) | 0 | 42 (100 %) | 0 | 0 | 0 | 0 | 40 (95.2 %) | 42 (100 %) |
| IoA | 31 (73.8 %) | 42 (100 %) | 0 | 42 (100 %) | 0 | 0 | 0 | 0 | 40 (95.2 %) | 41 (97.6 %) |

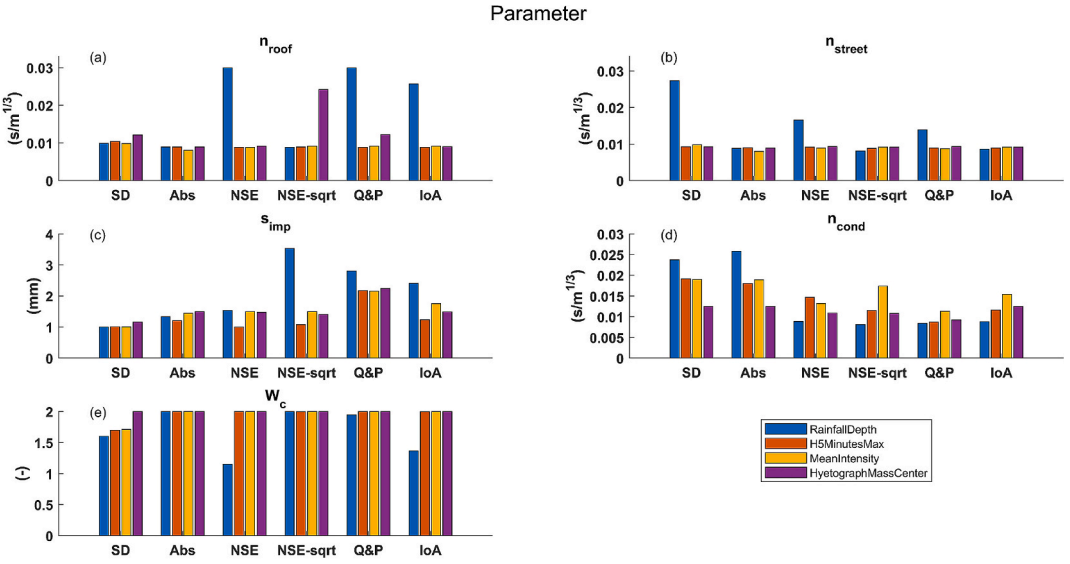



**Fig. 5.** Parameter estimation results for the five sensitive parameters identified in the study: (a) $n_{roof}$, (b) $n_{street}$, (c) $S_{imp}$, (d) $n_{cond}$ and (e) $W_c$. These results are based on the optimisation of the calibration set across different combinations of calibration event selection approaches: RD (blue), H5 (red), MI (yellow) and HCM (purple) and objective functions (SD, Abs, NSE, NSE-sqrt, Q&P, and IoA).

event selection approaches and objective functions in influencing the estimation of this key parameter, which has been identified as having the highest influence on total volume. Fig. 5.d underscores the significant influence of both calibration event selection approaches and objective functions on the estimation of $n_{cond}$. The SD and Abs objective functions tend to predict higher roughness values; however, the results from these objective functions do not show consistency over the different selection criteria.

The $W_c$ results (Fig. 5.e) show that the RD approach generally produces lower $W_c$ values, while the MI, H5, and HCM approaches often push the $W_c$ value to its upper limit of 2.0. Notably, it appears that the optimiser pushes the $W_c$ values toward this upper limit in many cases. However, this upper limit cannot be increased due to the physical meaning of $W_c$, which is constrained to a range between 1.0 and 2.0 (refer to Section 3.1). These results underscore the significant impact of both the calibration event selection approaches and the objective functions on the estimation of these key parameters, which directly influence the hydrological predictions of the SWMM model.

Fig. 6 presents the normalised parameter values across different calibration event selection approaches and objective functions. These normalised values are obtained by dividing each parameter's value by the upper limit of its corresponding range, resulting in values that range from 0 to 1. This normalisation facilitates the comparison of different parameters on a common scale. The 32 different parameter sets displayed in Fig. 6 demonstrate the sensitivity of SWMM model parameters to the calibration event selection approach and objective functions used.

The results show that the choice of the calibration event selection method greatly impacts the consistency of the calibrated parameter sets within the same objective function. Specifically, when using the H5, MI, and HCM approaches, the resulting parameter sets tend to exhibit less variability across the same objective functions (except for $n_{roof}$ and, to some extent, for $n_{street}$). In contrast, the RD approach leads to significantly different parameter sets for the same objective functions compared to the H5, MI, and HCM approaches. In other words, while the H5, MI, and HCM approaches generally produce similar parameter sets within a given objective function, the RD approach introduces more variability, resulting in greater differences in the calibrated parameter sets. This difference can likely be attributed to the nature of the RD criterion, which can include both long, low-intensity storms and short, intense storms, introducing greater hydrological heterogeneity. This variability may activate different runoff generation mechanisms, leading to more diverse sensitivity patterns in model parameters, especially in urban drainage systems where storage, and surface runoff respond non-linearly to rainfall intensity and temporal distribution.

However, it should be noted that the relationships between the parameters are complex and non-linear, meaning that even small changes in the parameter set can lead to significant changes in model performance (Fatone et al., 2021; Peng et al., 2023). These nonlinear interactions can cause similar parameter sets to produce different model performance metrics when applied under different calibration approaches. A comprehensive comparison of model performance is presented in Section 4.3.

A cross-performance analysis between objective functions was conducted to evaluate how a parameter set optimised for one objective function performs when evaluated under alternative objective functions. The aim of this analysis is to determine if the differences in the parameter calibrated values obtained in this study come from the intrinsic characteristics of each objective function, associated with the

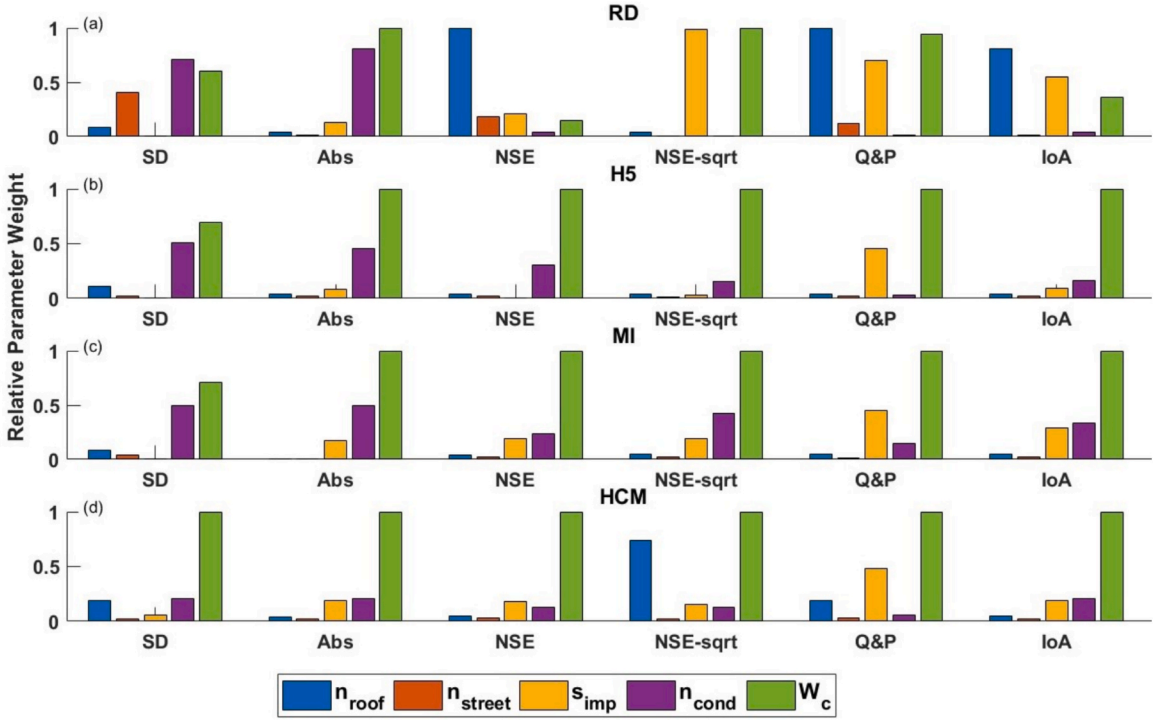



**Fig. 6.** Normalised parameter values across different calibration event selection approaches and objective functions. The figure is divided into four panels (a, b, c, d), each corresponding to one of the calibration event selection approaches: Rainfall Depth (panel a), H5MinutesMax (panel b), Mean Intensity (panel c), and Hyetograph Mass Center (panel d). Within each panel, the normalised parameter values of the five sensitive parameters—$n_{roof}$ (blue), $n_{street}$ (orange), $S_{imp}$ (yellow), $n_{cond}$ (purple), and $W_c$ (green)—are shown across eight objective functions: SD, Abs, NSE, NSE-sqrt, Q&P, and IoA.

random patterns in the optimisation, or weak identifiability in the optimisation. In this analysis, parameter sets optimised for one objective function (e.g., SD, Abs, etc.) are evaluated on all other objective functions. This analysis was applied to the four selection criteria (RD, H5, MI, and HCM). The performance of each optimised parameter set was normalised with respect to its own self-evaluation. For a parameter set optimised using objective function $f_i$, its performance when evaluated under another objective function $f_j$ was normalised using the following expression:

$$N_{ij} = \frac{f_j(p_i)}{f_i(p_i)}$$

where $f_j(p_i)$ is the performance value of the parameter set $p_i$ (optimised

**RD**

| | SD | Abs | IoA | NSE | NSE_sqrt | Q&P |
|---|---|---|---|---|---|---|
| SD | **1.00** | 1.21 | 2.16 | 1.76 | 3.28 | 2.16 |
| Abs | 1.18 | **1.00** | 1.19 | 1.16 | 1.34 | 1.18 |
| IoA | 1.31 | 1.20 | **1.00** | 1.20 | 1.23 | 1.19 |
| NSE | 1.32 | 1.22 | 1.21 | **1.00** | 1.51 | 1.27 |
| NSE_sqrt | 1.50 | 1.32 | 1.46 | 1.34 | **1.00** | 1.63 |
| Q&P | 1.78 | 1.68 | 1.17 | 1.38 | 1.29 | **1.00** |

**MI**

| | SD | Abs | IoA | NSE | NSE_sqrt | Q&P |
|---|---|---|---|---|---|---|
| SD | **1.00** | 1.17 | 1.16 | 1.15 | 1.35 | 1.43 |
| Abs | 1.16 | **1.00** | 1.16 | 1.15 | 1.19 | 1.20 |
| IoA | 1.18 | 1.26 | **1.00** | 1.17 | 1.16 | 1.16 |
| NSE | 1.25 | 1.35 | 1.20 | **1.00** | 1.21 | 1.24 |
| NSE_sqrt | 1.21 | 1.25 | 1.15 | 1.15 | **1.00** | 1.18 |
| Q&P | 1.40 | 2.10 | 1.13 | 1.22 | 1.19 | **1.00** |

**HMC**

| | SD | Abs | IoA | NSE | NSE_sqrt | Q&P |
|---|---|---|---|---|---|---|
| SD | **1.00** | 1.19 | 1.18 | 1.18 | 1.23 | 1.32 |
| Abs | 1.16 | **1.00** | 1.20 | 1.20 | 1.16 | 1.18 |
| IoA | 1.20 | 1.19 | **1.00** | 1.22 | 1.19 | 1.29 |
| NSE | 1.14 | 1.27 | 1.16 | **1.00** | 1.20 | 1.75 |
| NSE_sqrt | 1.18 | 1.19 | 1.16 | 1.21 | **1.00** | 1.54 |
| Q&P | 1.45 | 1.20 | 1.42 | 1.45 | 1.50 | **1.00** |

**H5**

| | SD | Abs | IoA | NSE | NSE_sqrt | Q&P |
|---|---|---|---|---|---|---|
| SD | **1.00** | 1.29 | 1.62 | 1.34 | 1.79 | 2.09 |
| Abs | 1.17 | **1.00** | 1.23 | 1.17 | 1.25 | 1.32 |
| IoA | 1.35 | 1.27 | **1.00** | 1.21 | 1.20 | 1.31 |
| NSE | 1.23 | 1.13 | 1.43 | **1.00** | 1.19 | 1.55 |
| NSE_sqrt | 1.30 | 1.19 | 1.38 | 1.16 | **1.00** | 1.46 |
| Q&P | 1.73 | 1.53 | 1.25 | 1.58 | 1.46 | **1.00** |

**Fig. 7.** Normalised cross-objective function performance matrices for parameter sets optimised under six different objective functions (SD, Abs, IoA, NSE, NSE_sqrt, and Q&P), evaluated using four event selection criteria: RD, MI, HCM, and H5. Each cell shows the normalised performance value, with diagonal values set to 1.00. Off-diagonal values indicate the relative performance of a parameter set optimised for one objective function when evaluated under another.

for $f_i$) evaluated under objective function $f_j$ and $f_i(p_i)$ is the original performance of the same parameter set under its own objective function.

Each matrix in Fig. 7 presents the normalised performance of optimised parameter sets for a specific event selection criterion when evaluated under different objective functions. Each diagonal element (values in bold) is set to 1.00, as it represents the self-evaluated performance for the corresponding function. The off-diagonal values indicate the relative performance of each optimised parameter set when tested against different objective functions.

The findings indicate that the performance of parameter sets optimised for one function degrades when evaluated under an alternative function. This degradation ranged from 13 % to approximately 200 % in some cases. This finding indicates that optimisation is function-specific, and one parameter set does not perform equivalently across different objectives. The observed variability in parameter values is thus more likely due the function-specific nature of the calibration process rather than equifinality or random effects introduced by the optimisation algorithm.

### 4.3. Model performance

The comparison of performance for the four calibration selection criteria (RD, H5, MI, and HCM) and six objective functions (SD, Abs, NSE, NSE-sqrt, Q&P, and IoA) is conducted using three key hydrological metrics: total volume, peak flow, and time to peak. Fig. 8 and Fig. 9 present the calibration and validation results for total volume, peak flow rate, and time to peak, using the CE and RMSE, respectively.

The calibration results for total volume (Fig. 8.a and 9.a) show that the calibration approach based on RD demonstrates the highest CE values (ranging from 0.77 to 0.88) and lowest RMSE (ranging from 113.7 $m^3$ to 93.8 $m^3$) across all objective functions. The RD approach achieves CE values of 0.86 and 0.88 for NSE-sqrt and IoA, respectively. The HCM approach also provides a high performance across all objective functions, especially for NSE-sqrt (CE = 0.82 and RMSE = 118.1 $m^3$) and IoA (CE = 0.82 and RMSE = 116.2 $m^3$). The H5 approach provides a moderate performance, with CE values ranging from 0.45 to 0.79 and RMSE values ranging from 185.5 to 109.2 $m^3$. In contrast, the MI approach typically produces the lowest CE values and the highest RMSE values, especially for the SD (CE = 0.43 and RMSE = 274.6 $m^3$), Abs (CE = 0.48 and RMSE = 246.8 $m^3$), and NSE (CE = 0.52 and RMSE = 178.7 $m^3$). This indicates that MI approach is less effective at accurately modelling the total volume compared to the other methods.

The validation results for total volume are presented in Fig. 8.b and 9.b using CE and RMSE, respectively. The results indicate that the RD approach, which performs well during the calibration stage, achieves the highest CE values and lowest RMSE values across several objective functions, particularly in NSE-sqrt (CE = 0.77 and RMSE = 138.4 $m^3$) and IoA (CE = 0.77 and RMSE = 136.3 $m^3$). These findings suggest that the RD approach is the most effective in capturing the total volume of runoff events. The H5 approach shows moderate performance during validation with CE values ranging from 0.43 to 0.72 and RMSE values ranging from 194.6 to 146.4 $m^3$. A similar trend is observed with the HCM approach, where CE values range from 0.44 to 0.6 and RMSE values from 192.2 to 162 $m^3$. The MI approach, again, demonstrates lower performance across most objective functions, with CE values as low as 0.31 and RMSE reaching up to 229 $m^3$ in the Abs objective function. Overall, the validation results are consistent with the calibration performance. The RD approach proves to be the most reliable approach access calibration and validation steps, especially with objective functions like NSE-sqrt and IoA, which show high CE values and low RMSE values.

Further analysis on the objective functions reveals distinct performance patterns across the different calibration event selection approaches. The objective functions of group II (SD and NSE) show variable effectiveness across different approaches, especially during the calibration phase, with CE values ranging from 0.43 for MI approach to 0.82 for RD approach. This finding indicates that SD and NSE are particularly sensitive to the choice of the calibration event selection approach. The Abs objective function follows a similar trend; it shows the highest CE value of 0.79 with RD and the lowest value at a CE value

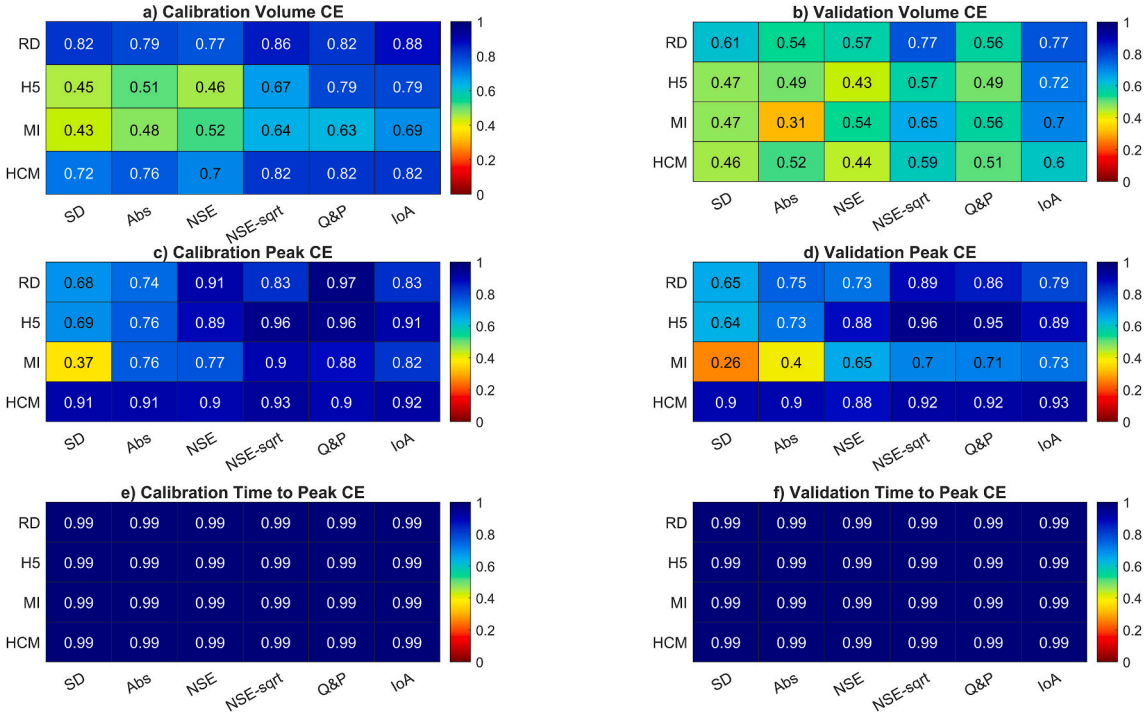



**Fig. 8.** The heatmap visualizes the performance metrics, where each cell's colour intensity corresponds to the CE value, following a gradient from 0 (red) to 1 (blue), as depicted by the colour bar to the right. The rows of the heatmap represent four distinct calibration event selection approaches: RD, H5, MI, and HCM. The columns signify the six objective functions: SD, Abs, NSE, NSE-sqrt, Q&P,and IoA.

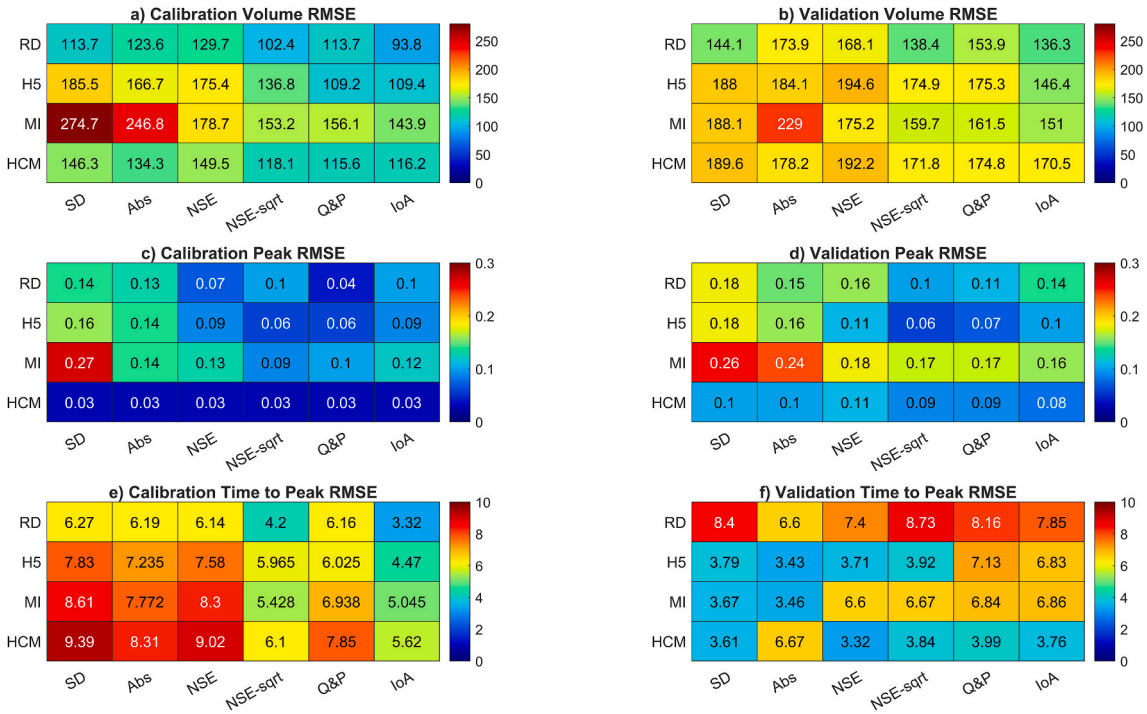



**Fig. 9.** The heatmap visualizes the performance metrics, where each cell's colour intensity corresponds to the RMSE value, following a gradient from 0 (red) to 1 (blue), as depicted by the colour bar to the right. The rows of the heatmap represent four distinct calibration event selection approaches: RD, H5, MI, and HCM. The columns signify the six objective functions: SD, Abs, NSE, NSE-sqrt, Q&P, and IoA.

of 0.48 with MI during the calibration phase. During validation, Abs shows CE values ranging from 0.31 for MI to 0.54 for RD. The Q&P objective function demonstrates good performance during the calibration and moderate performance during the validation, with stable performance across different calibration selection approaches. The NSE-sqrt and IoA objective functions provide high performance, especially with RD (CE of NSE-sqrt = 0.86 and CE of IoA = 0.89) and HCM (CE = 0.82 for both) during the calibration phase, and with RD (CE = 0.77 for both) during the validation phase. However, NSE-sqrt and IoA demonstrate a significant variable level of performance during calibration (CE ranging from 0.64 to 0.86 for NSE-sqrt, and from 0.69 to 0.88 for IoA) and validation (CE ranging from 0.57 to 0.77 for NSE-sqrt, and from 0.6 to 0.77 for IoA). This outcome signifies that NSE-sqrt and IoA are sensitive to the choice of calibration event selection approach. However, this sensitivity is less pronounced compared to that observed for SD, NSE, and Abs.

Fig. 8.c and 9.c present the calibration results for peak flow rate. The analysis shows that the HCM approach consistently achieves high performance across all objective functions, with CE values ranging from 0.9 to 0.92 and RMSE values of 0.03 $m^3/s$ for all objective functions. This indicates that the HCM approach allows effective prediction of the peak flow rates during calibration. The RD approach also performs well in NSE (with CE = 0.91 and RMSE = 0.07 $m^3/s$) and Q&P (with CE = 0.97 and RMSE = 0.04 $m^3/s$).

The H5 approach shows consistent and robust performance, with CE values reaching 0.96 for NSE-sqrt and Q&P, indicating high predictive accuracy. In contrast, MI approach has the lowest performance among the four approaches, particularly in SD (CE = 0.37 and RMSR = 0.27 $m^3/s$), Abs (CE = 0.76 and RMSE = 0.14 $m^3/s$) and NSE (CE = 0.77 and RMSE = 0.13 $m^3/s$). This indicates its relative ineffectiveness in accurately simulating peak flow rates during calibration. Although it shows moderate performance in some functions such as NSE-sqrt (CE = 0.9), overall, the MI approach lags behind the other approaches.

The validation results for peak flow rate (Fig. 8.d and 9.d) indicate that HCM approach, which also provides a well performance during the calibration, exhibits the highest performance across all objective functions with CE values ranging from 0.88 to and 0.93 and RMSE values ranging from 0.11 to 0.08 $m^3/s$. This suggests that HCM is highly effective in accurately predicting peak flow rates. The H5 approach also demonstrates robust performance, particularly in Q&P, NSE-sqrt, and IoA, where CE values reach 0.92, 0.92, and 0.93, respectively.

Conversely, the MI approach shows lower performance, particularly in SD (CE = 0.26 and RMSE = 0.26 $m^3/s$) and Abs (CE = 0.40 and RMSE = 0.24 $m^3/s$), suggesting challenges in accurately predicting peak flows during validation. However, MI performs moderately in functions like NSE-sqrt (CE = 0.7) and IoA (CE = 0.73), indicating some level of predictive capability. The RD approach shows moderate to high performance overall, with CE values ranging from 0.65 to 0.89 and RMSE values ranging from 0.18 to 0.10 $m^3/s$.

Examining the performance of different objective functions across the different calibration event selection approaches reveals that SD and Abs objective functions show a wide range of CE values from 0.26 for SD and 0.4 for Abs with MI to 0.90 for both functions with HCM. This high variability performance suggests that SD and Abs are highly sensitive to the choice of calibration approach. The similar pattern can be detected with NSE but with less variability in CE values that range from 0.65 to 0.88. The three objective functions (NSE-sqrt, Q&P, and IoA) consistently show superior performance (CE above 0.85) across different approaches, except for MI approach. The IoA demonstrates the highest performance across calibration approaches with CE ranging from 0.73 to 0.93. The high performance of IoA indicates that it is reliable and effective in accurately capturing the peak flow rate magnitude over different calibration selection approaches.

Fig. 8.e and 9.e present the calibration results for time to peak flow rate. The results show uniformly high performance across all approaches and objective functions, each achieving a CE value of 0.99. The results show that the RMSE values for time to peak flow rate are consistently low across all approaches and objective functions, indicating highly accurate prediction of peak flow timing. The highest RMSE value of 9.39 min, observed for the HCM approach with the SD objective function, is

still good given the long events duration. This demonstrates that even the highest RMSE values reflect strong predictive accuracy. This outstanding performance is also maintained in the validation results (Fig. 8.f and 9.f), where all combinations achieve CE value of 0.99 and RMSE values range from 8.73 to 3.32 min. These validation results further confirm the model's robustness and reliability in accurately simulating peak flow rate timing. This consistently good performance suggests that the SWMM model is highly effective at predicting the time to peak flow rate, regardless of the calibration event selection approach or objective function used. Also, this shows that the time to peak flow rate is less sensitive to the choice of calibration approach compared to other metrics like total volume or peak flow rate. The SWMM model robustness in estimating the time to peak can also be detected by the narrow range of change in μ* values for time to peak for all events (Table S3). This suggests that even with wide changes in parameter values, the model's prediction of the time to peak will remain largely unaffected. This narrow change in μ* values indicates that the time to peak is not sensitive to changes in the parameters, which indicates that the model is accurate in predicting this metric.

The comparison between the model optimisation results (Figs. 5 and 6) and the model performance (Figs. 8 and 9) reveals that different parameter sets can yield good model performance. This outcome is due to the inherently nonlinear nature of hydrological systems. In a nonlinear system, different parameter sets can interact with model inputs in complex ways, leading to similar outputs (Kouchi et al., (2017); Beven and Binley (1992)). This conclusion aligns with the concept of equifinality stated by Beven and Freer (2001), which suggests that various parameter sets can effectively simulate catchment characteristics. Similar findings have been observed in previous studies, such as Kouchi et al. (2017), who used three different optimisation algorithms and eight different objective functions to create calibrated parameter sets in two watersheds in Iran, Muleta (2012), who evaluated the performance of nine objective functions, and Moussa and Chahinian (2009), who compared the performance of various single and multi-objective functions on a catchment in Southern France.

The results show that RD approach is effective in providing an accurate estimation of total volume. The RD approach guarantees that the calibration dataset encompasses events with a wide range of precipitation magnitudes, which is important as the total volume is related to the amount of rainfall during the event. This diversity of light and heavy rainfall events allows the model to be calibrated under a wide variety of conditions, thereby enhancing its ability to accurately predict total runoff volume across different scenarios. For the peak flow rate estimation, the H5 approach demonstrates an accurate performance. This can be attributed to the fact that peak flow rate is highly sensitive to the intensity of rainfall over short durations, as the high intensity can cause rapid increases in runoff, which lead to peaks in flow rate. The H5 criterion includes a diverse array of peak intensity events in the calibration dataset, making the model better prepared to predict peak flow rates under various conditions. RD proved to be the most reliable selection criteria, offering the best performance for total volume estimation while also delivering reasonable accuracy in peak flow rate predictions.

The findings present a strong case for advancing beyond traditional objective functions like the SD or NSE. As mentioned in section 1, both SD and NSE are highly sensitive to large deviations because they square the differences between observed and predicted values, inherently giving disproportionate weights to larger errors. This makes them particularly vulnerable to outliers or significant errors. If a model prediction deviates significantly from the observed value, this deviation can disproportionately impact the overall score, potentially resulting in lower performance metrics. Also, this characteristic makes SD and NSE more sensitive to the specific characteristics of the calibration events, which are determined by the calibration event selection approach. In contrast, NSE-sqrt and IoA are designed to be less sensitive to extreme deviations between predicted and observed values. NSE-sqrt achieves this through the square root transformation, while IoA uses absolute differences. These approaches cause IoA and NSE-sqrt to reduce the disproportionate impact of large errors and outliers, which provides a balanced evaluation of model performance across the entire data range. As a result, NSE-sqrt and IoA demonstrate more consistent performance across different calibration event selection approaches.

Based on the discussed results, we can detect that those objective functions like NSE-sqrt and IoA, which consider instantaneous measurements at each time step, can give a better estimation for total volume and peak flow rate than objective functions that directly target the total volume and peak flow rate like Q&P objective function. This recommendation may initially seem confusing for some readers, as the objective function is typically selected to align closely with the model's primary goals. Consequently, many modellers might assume, albeit incorrectly, that optimising the same metric during the calibration period will maximize its performance during the validation. However, achieving robust simulation performance under varying conditions, relies on the accurate representation of underlying processes. Therefore, the calibration objective functions should be selected based on their ability to extract process-relevant information from the calibration data. This discussion underscores the importance of differentiating between calibration objective function(s) and modelling objectives.

### 4.4. Application and limitation

The calibration of urban hydrological models such as SWMM has traditionally relied on selecting a limited number of rainfall-runoff events based on subjective criteria or expert knowledge. These selections are usually made by event magnitude (e.g., largest events, highest peak events), without considering how well they represent key features of hydrological events (Kleidorfer et al. 2009; Mourad et al. 2005; Broekhuizen et al. 2020). The selection criteria for these approaches are often not clear and may not reflect the full range of hydrological variations under different rainfall conditions. In contrast, our work suggests a data-driven and replicable method for choosing events, evaluating four criteria, RD, H5, MI and HCM, to see how they affect parameter estimation. Most studies tend to use traditional SD or NSE metrics when selecting objective functions, without taking into account how sensitive these metrics are to various event characteristics. Our approach uses six objective functions and shows how their performance changes based on which events and calibration targets are chosen, showing that model performance can significantly change depending on the pairing between selection strategy and objective function.

In addition, most existing studies either skip sensitivity analysis, test with a small number of events or analyse only a single hydrological metric (Niazi et al., 2017). These approaches do not show how parameters change in different hydrological situations. Our integration of GSA using the Morris method, executed through an innovative coupling of Mat-SWMM and the SAFE toolbox, shows that parameter sensitivity varies substantially across rainfall types and target metrics.

The proposed methodology, combining event selection strategies, multi-objective calibration, and global sensitivity analysis, is designed to be applicable across a wide range of rainfall conditions, including high-intensity storm events. Its structure is inherently flexible and computationally efficient, making it suitable for broader applications in urban watershed modelling. We encourage future studies to apply and evaluate this framework under more extreme rainfall scenarios to further validate its robustness, particularly for flood-focused analyses and stormwater infrastructure design.

Although this study is based on rainfall-runoff events from a single urban catchment located in a Mediterranean climate, the methodology we propose is broadly applicable and not limited by these contextual factors. The modelling framework introduced in this study (statistical event selection criteria, objective function evaluation, and global sensitivity analysis), is designed to be modular and scalable. Each component is structured in a way that allows independent adaptation, making the overall approach flexible for application across various

urban catchment configurations. The event selection strategy is based on generalizable statistical criteria, and the multi-objective calibration approach can accommodate diverse performance goals. Additionally, the integration of open-source Mat-SWMM with SAFE toolbox facilitates efficient, automated sensitivity analysis, supporting large-scale simulation workflows.

The use of a well-characterized, relatively small residential catchment provides a controlled environment to develop and test the methodology under realistic urban hydrological conditions. This type of catchment is common in many cities globally and has been widely used in previous modelling studies (Broekhuizen et al., 2020; Liu et al., 2015; Zhu et al., 2019; Tamm et al..2023; Yang et al., 2023; Broekhuizen et al., 2021). Furthermore, although the selected events represent a diverse range of rainfall characteristics, including moderate and prolonged events, the proposed methodology is applicable and scalable to high-intensity storm events as well. We encourage future research to test the framework under more extreme rainfall conditions to further evaluate its robustness in flood-focused studies.

However, the successful application of this framework to larger and more complex urban catchments requires the availability of monitoring data. In particular, uniform and reliable runoff measurements and rainfall data across multiple locations are crucial to ensure accurate calibration and model interpretation. The acquired rainfall data should accurately describe the temporal and spatial variability of precipitation in the catchment. The available runoff data must have a fine discretisation to allow an exhaustive description of the experimental hydrographs and therefore a reliable evaluation of relevant hydrological variables, i.e. total volume, peak flow rate and time to peak. While the methodology is not inherently limited by catchment size, its effectiveness in more spatially heterogeneous settings will depend on the extent and resolution of the input data available. Despite these challenges, the core methodology remains applicable and offers a systematic, reproducible approach for parameter estimation and sensitivity analysis. Future research is encouraged to explore and refine the framework in larger-scale applications to assess its performance under more complex urban hydrological settings.

## 5. Conclusion

This study presents a comprehensive analysis of the interaction between calibration event selection criteria and objective functions for modelling rainfall-runoff in urban catchment area. Four event selection criteria (RD, H5, MI and HCM) and six objective functions (SD, Abs, NSE, NSE-sqrt, Q&P and IoA,) were tested.

The research conducts a GSA of SWMM model parameters using Morris method, introducing an innovative integration of Mat-SWMM and the Sensitivity Analysis for Everybody (SAFE) Toolbox. Using a comprehensive dataset of 42 high-resolution rainfall events, this work investigates how various rainfall characteristics impact the sensitivity of 10 uncertain model parameters. The analysis focuses on three hydrologic metrics: the total runoff volume, the peak flow rate and the time to peak flow rate. The results from the GSA showed that the sensitivity of model parameters is highly dependent on the characteristics of each rainfall event, highlighting considerable differences in parameter influence based on event type. Also, the results reveal that model parameter sensitivity varies depending on the target metrics. The analysis identified five influential parameters, which were then optimised using a genetic algorithm.

The results show that the RD approach is effective in providing accurate estimates of total volume. For the peak flow rate estimation, the H5 and HCM approaches demonstrate, instead, an accurate performance. These findings offer practical guidance for urban hydrologists and water managers: the RD method is recommended when accurate total volume estimation is prioritized, while H5 and HCM approaches are more appropriate for capturing peak flow dynamics.

The findings highlight the limitations of traditional objective functions like SD and NSE, which tend to emphasize large errors due to squaring differences, making them sensitive to outliers and specific calibration event characteristics. In contrast, NSE-sqrt and IoA reduce this sensitivity, NSE-sqrt by applying a square root transformation and IoA by using absolute differences, resulting in a more balanced evaluation of model performance across the dataset. Consequently, NSE-sqrt and IoA show more stable performance across various calibration event selection approaches.

## CRediT authorship contribution statement

**Mohammed N. Assaf:** Software, Formal analysis, Writing – original draft, Methodology, Conceptualization. **Nicolò Salis:** Software, Formal analysis, Writing – review & editing, Methodology. **Sauro Manenti:** Supervision, Conceptualization, Writing – review & editing, Methodology. **Lorenzo Tamellini:** Supervision, Funding acquisition, Writing – review & editing, Methodology. **Enrico Creaco:** Writing – review & editing, Conceptualization, Funding acquisition. **Sara Todeschini:** Supervision, Funding acquisition, Writing – review & editing, Methodology, Conceptualization.

## Declaration of competing interest

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

## Acknowledgements

This research was supported by the Italian Ministry of University and Research (MUR) through the project “PRIN 2020-URCA!, Prot. 20208TAK3H” and through the project “DORIAN of DICAr” - MUR Program “Dipartimento di Eccellenza 23-27”. Lorenzo Tamellini, Sara Todeschini, Sauro Manenti, and Mohammed N. Assaf have been supported by the PRIN 2022 PNRR project “Uncertainty Quantification of coupled models for water flow and contaminant transport” (No. P2022LXLYY), financed by the European Union - Next Generation EU.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2025.133821.

## Data availability

Data will be made available on request.

## References

Assaf, M.N., Manenti, S., Creaco, E., Giudicianni, C., Tamellini, L., Todeschini, S., 2024. New optimization strategies for SWMM modeling of stormwater quality applications in urban area. J. Environ. Manage. 361, 121244. https://doi.org/10.1016/j.jenvman.2024.121244.

Baker, E.A., Cappato, A., Todeschini, S., Tamellini, L., Sangalli, G., Reali, A., Manenti, S., 2022. Combining the Morris method and multiple error metrics to assess aquifer characteristics and recharge in the lower Ticino Basin, in Italy. J. Hydrol. 614, 128536. https://doi.org/10.1016/j.jhydrol.2022.128536.

Baker, E.A., Manenti, S., Reali, A., Sangalli, G., Tamellini, L., Todeschini, S., 2023. Combining noisy well data and expert knowledge in a Bayesian calibration of a flow model under uncertainties: an application to solute transport in the Ticino basin. GEM-International Journal on Geomathematics 14 (1), 8. https://doi.org/10.1007/s13137-023-00219-8.

Balha, A., Singh, A., Pandey, S., Kumar, R., Mallick, J., Singh, C.K., 2023. Assessing the impact of land-use dynamics to predict the changes in hydrological variables using effective impervious area (EIA). Water Resour. Manag. 37 (10), 3999–4014.

Barco, J., Wong, K.M., Stenstrom, M.K., 2008a. Automatic calibration of the US EPA SWMM model for a large urban catchment. J. Hydraul. Eng. 134 (4), 466–474.

Barco, J., Papiri, S., Stenstrom, M.K., 2008b. First flush in a combined sewer system. Chemosphere 71 (5), 827–833.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6 (3), 279–298. https://doi.org/10.1002/hyp.3360060305.
Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249 (1–4), 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8.
Beven, Keith J. *Rainfall-runoff modelling: the primer*. John Wiley & Sons, 2012.
Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320 (1–2), 18–36.
Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol. Earth Syst. Sci. 5 (1), 1–12.
Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrol. Process. 9 (3–4), 251–290.
Broekhuizen, I., Leonhardt, G., Viklander, M., 2021. Reducing uncertainties in urban drainage models by explicitly accounting for timing errors in objective functions. Urban Water J. 18 (9), 740–749.
Broekhuizen, I., Leonhardt, G., Marsalek, J., Viklander, M., 2020. Event selection and two-stage approach for calibrating models of green urban drainage systems. Hydrol. Earth Syst. Sci. 24 (2), 869–885.
Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Model. Software 22 (10), 1509–1518. https://doi.org/10.1016/j.envsoft.2006.10.004.
Cappato, A., Baker, E.A., Reali, A., Todeschini, S., Manenti, S., 2022. The role of modeling scheme and model input factors uncertainty in the analysis and mitigation of backwater induced urban flood-risk. J. Hydrol. 614, 128545. https://doi.org/10.1016/j.jhydrol.2022.128545.
Chiew, F.H.S., Whetton, P., McMahon, T., Pittock, A.B., 1995. Simulation of the impacts of climate change on runoff and soil moisture in Australian catchments. J. Hydrol. 167 (1–4), 121–147. https://doi.org/10.1016/0022-1694(94)02649-V.
Davis C.V. 1952. Handbook of Applied Hydraulics, Mcgraw-hill Book Company Inc., New York.
Diskin, M.H., Simon, E., 1977. A procedure for the selection of objective functions for hydrologic simulation models. J. Hydrol. 34 (1–2), 129–149. https://doi.org/10.1016/0022-1694(77)90066-X.
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 (10), 5493–5516. https://doi.org/10.5194/hess-25-5493-2021.
Feng, K., Lu, Z., Yang, C., 2019. Enhanced Morris method for global sensitivity analysis: good proxy of Sobol'index. Struct. Multidiscip. Optim. 59 (2), 373–387.
Fowler, K., Peel, M., Western, A., Zhang, L., 2018. Improved rainfall-runoff calibration for drying climate: Choice of objective function. Water Resour. Res. 54 (5), 3392–3408.
Garcia, F., Folton, N., Oudin, L., 2017. Which objective function to calibrate rainfall–runoff models for low-flow index simulations? Hydrol. Sci. J. 62 (7), 1149–1166.
Giudicianni, C., Assaf, M.N., Todeschini, S., Creaco, E., 2023. Comparison of nonlinear reservoir and UH algorithms for the hydrological modeling of a real urban catchment with EPASWMM. Hydrology 10 (1), 24. https://doi.org/10.3390/hydrology10010024.
Granata, F., Gargano, R., De Marinis, G., 2016. Support vector regression for rainfall-runoff modeling in urban drainage: a comparison with the EPA's storm water management model. Water 8 (3), 69.
Hamel, P., Tim, D.F., 2014. The impact of stormwater source-control strategies on the (low) flow regime of urban catchments. Water Sci. Technol. 69 (4), 739–745.
Hatt, B.E., Fletcher, T.D., Walsh, C.J., Taylor, S.L., 2004. The influence of urban density and drainage infrastructure on the concentrations and loads of pollutants in small streams. Environ. Manag. 34, 112–124.
Her, Y., Chaubey, I., 2015. Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty. Hydrol. Process. 29 (19), 4220–4237.
Herman, J.D., Kollat, J.B., Reed, P.M., Wagener, T., 2013. Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. Hydrol. Earth Syst. Sci. 17 (7), 2893–2903.
Hornberger, G.M., Beven, K.J., Cosby, B.J., Sappington, D.E., 1985. Shenandoah watershed study: calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. Water Resour. Res. 21 (12), 1841–1850.
Hossain, S., Hewa, G.A., Wella-Hewage, S., 2019. A comparison of continuous and event-based rainfall–runoff (RR) modelling using EPA-SWMM. Water 11 (3), 611.
Kouchi, H., Delaram, K.E., Faridhosseini, A., Sanaeinejad, S.H., Khalili, D., Abbaspour, K. C., 2017. Sensitivity of calibrated parameters and water resource estimates on different objective functions and optimization algorithms. Water 9 (6), 384.
Huber, W.C.; Dickinson, R.E. Storm Water Management Model User's Manual; Version 4.0; U.S. Environmental Protection Agency: Cincinnati, OH, USA, 1988.
Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? Water Resour. Res. 29 (8), 2637–2649.
Janetti, E.B., Guadagnini, L., Riva, M., Guadagnini, A., 2019. Global sensitivity analyses of multiple conceptual models with uncertain parameters driving groundwater flow in a regional-scale sedimentary aquifer. J. Hydrol. 544–556. https://doi.org/10.1016/j.jhydrol.2019.04.035.
Azar, K., Morteza, C.G., Creaco, E., 2025. Sensitivity analysis-aided calibration of urban drainage Modeling. Water 17 (5), 612.
Kleidorfer, M., Möderl, M., Fach, S., Rauch, W., 2009. Optimization of measurement campaigns for calibration of a conceptual sewer model. Water Sci. Technol. 59, 1523–1530. https://doi.org/10.2166/wst.2009.154.
Knighton, J., Lennon, E., Bastidas, L., White, E., 2016. Stormwater detention system parameter sensitivity and uncertainty analysis using SWMM. J. Hydrol. Eng. 21 (8), 05016014.
Kuczera, G., Mroczkowski, M., 1998. Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. Water Resour. Res. 34 (6), 1481–1489.
Legates, D.R., McCabe Jr, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241. https://doi.org/10.1029/1998WR900018.
Liu, L., Y. Liu, X. Wang, Dapeng Yu, K. Liu, H. Huang, and G. Hu. "Developing an effective 2-D urban flood inundation model for city emergency management based on cellular automata." Natural hazards and earth system sciences 15, no. 3 (2015): 381-391.
Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. J. Hydrol. 330 (1–2), 368–381.
Moradkhani, Hamid, and Soroosh Sorooshian. *General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis*. Springer Berlin Heidelberg, 2008.
Moreda, F., Koren, V., Zhang, Z., Reed, S., Smith, M., 2006. Parameterization of distributed hydrological models: learning from the experiences of lumped modeling. J. Hydrol. 320 (1–2), 218–237.
Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.
Mourad, M., Bertrand-Krajewski, J.-L., Chebbo, G., 2005. Stormwater quality models: sensitivity to calibration data. Water Sci. Technol. 52, 61–68. https://doi.org/10.2166/wst.2005.0110.
Moussa, R., Chahinian, N., 2009. Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events. Hydrol. Earth Syst. Sci. 13 (4), 519–535.
Muleta, M.K., 2012. Model performance sensitivity to objective function during automated calibrations. J. Hydrol. Eng. 17 (6), 756–767. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000497.
Muleta, M.K., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. J. Hydrol. 306 (1-4), 127–145.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
Nguyen, A.-T., Sigrid, R., 2015. A performance comparison of sensitivity analysis methods for building energy models. In: *Build. Simul.*, 8 Tsinghua University Press, pp. 651–664.
Niazi, M., Nietch, C., Maghrebi, M., Jackson, N., Bennett, B.R., Tryby, M., Massoudieh, A., 2017. Storm water management model: performance review and gap analysis. J. Sustain. Water Built Environ. 3 (2), 04017002. https://doi.org/10.1061/JSWBAY.0000817.
Peng, J., Zhao, H., Luo, Z., Ouyang, J., Yu, L., 2023. A new approach for sensitivity analysis of the stormwater management model applied in an airport. Water Science & Technology 88 (9), 2453–2464. https://doi.org/10.2166/wst.2023.335.
Pianosi, F., Sarrazin, F., Wagener, T., 2015. A matlab toolbox for global sensitivity analysis. Environ. Model. Software 70, 80–85.
Rahimi, A.Y., Ebrahimian, A., 2024. Urban impervious cover characterization for rainfall-runoff analysis: Impacts of data availability and resolution. J. Hydrol. 641, 131807.
Reinecke, R., Foglia, L., Mehl, S., Herman, J.D., Wachholz, A., Trautmann, T., Döll, P., 2019. Spatially distributed sensitivity of simulated global groundwater heads and flows to hydraulic conductivity, groundwater recharge, and surface water body parameterization. Hydrol. Earth Syst. Sci. 23 (11), 4561–4582.
Riaño-Briceño, G., Barreiro-Gomez, J., Ramirez-Jaime, A., Quijano, N., Ocampo-Martínez, C., 2016. MatSWMM–an open-source toolbox for designing real-time control of urban drainage systems. Environ. Modell. Software 83, 143–154.
Rossman, L.A. Storm Water Management Model User's Manual; Version 5.0; U.S. Environmental Protection Agency: Cincinnati, OH, USA, 2010.
Rossman, L.A. Storm Water Management Model User's Manual; Version 5.1; U.S. Environmental Protection Agency: Cincinnati, OH, USA, 2015.
Rossman, L.A., Huber, W., 2016. Storm water management model reference manual volume III–water quality. US EPA Office of Research and Development, Washington, DC. EPA/600/R-16/093.
Ruano, M.V., Ribes, J., Seco, A., Ferrer, J., 2012. An improved sampling strategy based on trajectory design for application of the Morris method to systems with many input factors. Environ. Model. Software 37 (2012), 103–109. https://doi.org/10.1016/j.envsoft.2012.03.008.
Schütze, M., Willems, P., and Vaes, G.: Integrated Simulation of Urban Wastewater Systems – How Many Rainfall Data Do We Need?, in: Global Solutions for Urban Drainage, American Society of Civil Engineers, Lloyd Center Doubletree Hotel, Portland, Oregon, United States, 1–11, Doi: 10.1061/40644(2002)244, 2002.
Shin, M.-J., Guillaume, J.HA., Croke, B.FW., Jakeman, A.J., 2013. Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. J. Hydrol. 503, 135–152.
Singh, S.K., Bárdossy, A., 2012. Calibration of hydrological models on hydrologically unusual events. Adv. Water Resour. 38, 81–91.
Sivapalan, Murugesu, K. Takeuchi, S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, X. Liang et al. "IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences." *Hydrological sciences journal* 48, no. 6 (2003): 857-880.
Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Chonggang, Xu., 2015. Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. J. Hydrol. 523, 739–757.

Sorooshian, S., Gupta, V.K., 1983. Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness. Water Resour. Res. 19 (1), 260–268.
Tamm, O., Kokkonen, T., Warsta, L., Dubovik, M., Koivusalo, H., 2023. Modelling urban stormwater management changes using SWMM and convection-permitting climate simulations in cold areas. J. Hydrol. 622, 129656.
Todeschini, S., Manenti, S., Creaco, E., 2019. Testing an innovative first flush identification methodology against field data from an Italian catchment. J. Environ. Manage. 246, 418–425. https://doi.org/10.1016/j.jenvman.2019.06.007.
Todeschini, S., Papiri, S., Ciaponi, C., 2018. Placement strategies and cumulative effects of wet-weather control practices for intermunicipal sewerage systems. Water Resour. Manag. 32 (8), 2885–2900. https://doi.org/10.1007/s11269-018-1964-y.
Tscheikner-Gratl, F., Zeisl, P., Kinzel, C., Leimgruber, J., Ertl, T., Rauch, W., Kleidorfer, M., 2016. Lost in calibration: why people still do not calibrate their models, and why they still should – a case study from urban drainage modelling. Water Sci. Technol. 74, 2337–2348. https://doi.org/10.2166/wst.2016.395.
Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrol. Earth Syst. Sci. 5 (1), 13–26.
Wagener, T., Hoshin, V.G., 2005. Model identification for hydrological forecasting under uncertainty. Stochastic Environ. Res. Risk Assess. 19, 378–387.
Walsh, C.J., Roy, A.H., Feminella, J.W., Cottingham, P.D., Groffman, P.M., Morgan, R.P., 2005. The urban stream syndrome: current knowledge and the search for a cure. J. N. Am. Benthol. Soc. 24 (3), 706–723.
Wang, K.-H., Altunkaynak, A., 2012. Comparative case study of rainfall-runoff modeling between SWMM and fuzzy logic approach. J. Hydrol. Eng. 17 (2), 283–291.
Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. Int. J. Climatol. 32 (13), 2088–2094.
Wu, Y., Dunxian She, Xia, J., Zhang, Y., Lei, Z., 2024. Evaluation of the number of events' influence on model performance and uncertainty in urban data-scarce areas based on behavioral parameter ranking method. J. Hydrol. 636, 131298.
Wu, Y., Liu, S., 2014. A suggestion for computing objective function in model calibration. Eco. Inform. 24, 107–111.
Yang, B., Zhang, T., Li, J., Feng, P., Miao, Y., 2023. Optimal designs of LID based on LID experiments and SWMM for a small-scale community in Tianjin, north China. J. Environ. Manage. 334, 117442.
Zhan, C.-S., Song, X.-M., Xia, J., Tong, C., 2013. An efficient integrated approach for global sensitivity analysis of hydrological model parameters. Environ. Model. Software 41, 39–52.
Zhu, Z., Chen, Z., Chen, X., Guo, Yu., 2019. An assessment of the hydrologic effectiveness of low impact development (LID) practices for managing runoff with different objectives. J. Environ. Manage. 231, 504–514.

# Conclusion

This thesis set out to improve the robustness and reliability of stormwater modelling in urban catchments, focusing on both water quantity and water quality within the EPA-SWMM framework. Building on the theoretical background provided in Chapters 2 and 3, the research was developed through original studies that examined hydrological formulation choices, calibration design strategies, and optimization methods for pollutant build-up and wash-off. Collectively, the results contribute to a deeper understanding of how modelling decisions shape predictive reliability and provide practical guidance for urban water management.

For stormwater quantity, studies demonstrated that model formulation and calibration strategy exert a strong influence on predictive outcomes. Chapter 4 showed that the choice between non-linear reservoir and unit hydrograph formulations shapes calibration effort and the accuracy of hydrograph metrics, with the non-linear reservoir approach generally providing more robust results. Chapter 6 extended this by showing that calibration outcomes also depend on the selection of storm events and the choice of objective function. It showed that rainfall depth-based calibration improves volume prediction, while high-intensity and hyetograph-center events improve peak estimation. Furthermore, transformed and index-based objective functions (e.g., NSE-sqrt, IoA) offered more robust performance across diverse storms than traditional squared-error criteria. Together, these studies highlight that calibration design, event selection, parameter sensitivity screening, and performance criteria, are as critical as model structure in determining the quality of urban runoff simulations.

For stormwater quality, chapter 5 built directly on the hydrological foundation established in chapter 4, recognizing that pollutant wash-off predictions require accurate flow simulations. This study introduced two methodological innovations: the extension of Mat-SWMM to allow direct calibration of quality parameters, and the development of an integrated event calibration strategy. By comparing concentration-based and global-metric objectives, it was shown that calibrations targeting total pollutant mass and peak pollutant fluxes yield more physically plausible, efficient, and generalizable results than those based solely on instantaneous concentrations.

Taken together, the thesis demonstrates that methodological choices: including model formulation, event selection, parameter sensitivity screening, and objective functions are as critical as the model itself. It provides evidence-based guidance for both researchers and practitioners, emphasizing that hydrological and water quality modelling are inseparable and that the effect of decisions made in one domain propagate into the other. While the work was applied to a single catchment and focused primarily on TSS under event-based conditions, it opens pathways for future research on multi-pollutant applications, continuous simulation, and hybrid modelling approaches.

In conclusion, the thesis contributes methodological innovations and practical recommendations that advance urban stormwater modelling. By clarifying how model building shape outcomes, it strengthens the scientific foundation for stormwater management and supports more reliable and resilient design and operation of urban drainage systems.