Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PIML-Driven Spatial Optimization and Active Telemetry Control of Integrated Grey-Green Systems to Minimize Modernization Capital Investment in US Combined Sewer Systems

Naim Sheikh

sheikhnayeem375@gmail.com

Northern University Business and Technology Khulna https://orcid.org/0009-0007-2123-7655

Research Article

Keywords: Physics-Informed Machine Learning, Combined Sewer Overflow, Green Infrastructure, Real-Time Control, Hydraulic Surrogate Modelling, Multi-Objective Optimization, Capital-Cost Optimization, Urban Drainage Systems

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11760144/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

PIML-Driven Spatial Optimization and Active Telemetry Control of Integrated Grey-Green Systems to Minimize Modernization Capital Investment in US Combined Sewer Systems

Naim Sheikh
Email: sheikhnayeem375@gmail.com
Bachelor of Computer Science and Engineering,
Northern University Business and Technology Khulna

Abstract
Combined sewer overflow (CSO) mitigation requires reliable hydraulic prediction and cost-effective infrastructure planning. This study develops a physics-informed machine learning (PIML) framework integrating surrogate modelling, real-time control (RTC), and spatial green-grey infrastructure optimization to minimize capital investment in combined sewer systems. The framework was evaluated using the 11.7-ha EPA SWMM Example 8 benchmark, NOAA Atlas 14 design storms, and 20 years of hourly rainfall data from Philadelphia. Data-driven and physics-informed surrogates were compared through one-step prediction, closed-loop rollout, and extreme-storm evaluation. Rule-based control (RBC) and model predictive control (MPC) were assessed across 15 design storms, while NSGA-II optimized infrastructure configurations and capital costs. PIML-soft achieved a CSO-flow rollout NSE of 0.87 under test conditions and 0.93 for withheld 25-100-year storms, outperforming the data-only MLP in extended prediction. RBC reduced CSO volume by 5-29% relative to passive operation and outperformed the tested MPC configurations. The selected GI-led design, comprising 38 bioretention units and 6,000 gallons of residual storage, reduced modal capital cost from $0.87 million to $0.49 million compared with grey-only storage. These findings demonstrate the potential of integrated PIML and spatial optimization for benchmark-specific, cost-effective CSO management.

Keywords: Physics-Informed Machine Learning, Combined Sewer Overflow, Green Infrastructure, Real-Time Control, Hydraulic Surrogate Modelling, Multi-Objective Optimization, Capital-Cost Optimization, Urban Drainage Systems

I. Introduction

Combined sewer systems are important components of urban wastewater infrastructure, conveying sanitary wastewater and stormwater through shared collection networks. During rainfall events, the hydraulic capacity of these systems can become insufficient, resulting in combined sewer overflows (CSOs) that discharge untreated or partially treated wastewater into receiving water bodies. Increasing urban development, changing rainfall patterns, aging drainage infrastructure, and limited municipal investment capacity make the management of these systems an important engineering and environmental challenge. Conventional approaches based primarily on expanding underground storage and conveyance infrastructure can require substantial capital investment, while operational interventions and green infrastructure offer alternative opportunities to improve system performance. However, identifying an effective combination of hydraulic control, spatial infrastructure allocation, and investment decisions requires reliable system modelling and computationally efficient decision-support methods. Recent developments in physics-informed machine learning (PIML), surrogate modelling, real-time control (RTC), and multi-objective optimization provide opportunities to improve the planning and operation of combined sewer systems. PIML can incorporate physical relationships into data-driven hydraulic prediction, potentially improving the consistency of surrogate models during extended simulations. RTC can regulate flow distribution and available storage through operational adjustments, while green infrastructure (GI), including bioretention systems, can reduce runoff entering combined sewer networks. Grey infrastructure, such as underground storage tanks and flow-control structures, can provide additional hydraulic capacity when source-control interventions alone are insufficient. Nevertheless, these approaches must be evaluated together because improvements in prediction accuracy do not necessarily produce better control outcomes, and hydraulic benefits do not automatically translate into lower capital investment. This study addresses these challenges by developing a PIML-driven spatial optimization and active telemetry control framework for integrated green–grey infrastructure planning in US combined sewer systems. The framework connects hydraulic simulation, physics-informed surrogate development, control-system evaluation, spatial infrastructure optimization,

and capital-cost assessment. EPA SWMM is retained as the hydraulic evaluation model, while the surrogate model supports computationally efficient prediction within the control process. The proposed approach is investigated using a reproducible combined sewer benchmark, allowing alternative control strategies and infrastructure configurations to be compared under consistent rainfall and hydraulic conditions.

A. Background and Motivation

Combined sewer systems operate under changing hydraulic conditions because rainfall-derived inflows interact with sanitary wastewater, available pipe capacity, regulator configurations, interceptor capacity, and downstream treatment constraints. During relatively frequent rainfall events, existing conveyance and storage capacity may accommodate much of the additional runoff. However, more intense rainfall can exceed the available hydraulic capacity, producing overflow events and potentially increasing environmental and public-health risks. Managing these conditions requires more than simply increasing storage capacity. Infrastructure planners must determine where interventions should be installed, how operational structures should respond to changing flow conditions, and how limited investment resources can be allocated to achieve the required level of CSO control. Traditional grey infrastructure interventions, including underground storage tanks, conveyance improvements, and regulator modifications, can provide additional hydraulic capacity. However, their implementation may involve substantial construction costs, land requirements, engineering complexity, and disruption to existing urban infrastructure. Green infrastructure provides a complementary approach by managing stormwater closer to its source. Bioretention systems, rain gardens, and other distributed interventions can reduce runoff entering combined sewer networks and potentially reduce the amount of additional grey storage required. The relative effectiveness of these interventions depends on their location, treatment capacity, contributing drainage area, rainfall characteristics, and interaction with the existing sewer network. Consequently, selecting GI locations without considering the hydraulic characteristics of the wider drainage system may result in inefficient investment or limited CSO reduction. Real-time control provides another opportunity to improve the utilization of existing drainage infrastructure. By adjusting regulator gates and other controllable structures in response to observed hydraulic conditions, RTC can influence flow distribution and temporarily retain or release water according to available system capacity. Rule-based control (RBC) generally applies predefined operational thresholds, whereas model predictive control (MPC) uses predictions of future system behaviour to determine control actions over a specified horizon. MPC can provide a systematic mechanism for considering hydraulic constraints and competing operational objectives. However, its performance depends on the accuracy of the internal prediction model, the availability and quality of telemetry, rainfall-forecast reliability, and the optimization procedure used to determine control actions. The increasing computational requirements of repeated hydraulic simulation create an additional motivation for surrogate modelling. Detailed hydraulic models such as SWMM can represent complex drainage-system behaviour, but repeatedly executing them within optimization and control procedures may increase computational demands. Machine-learning-based surrogates offer an alternative by approximating the relationship between hydraulic states, inflows, control settings, and subsequent system responses. Nevertheless, a data-driven surrogate may produce physically inconsistent predictions or accumulate errors when its outputs are repeatedly fed back into subsequent prediction steps. Physics-informed machine learning offers a potential solution by incorporating hydraulic relationships, such as continuity equations and hydraulic structure constraints, into the model-training process. The motivation for this study therefore arises from the need to connect four complementary engineering decisions: how to predict hydraulic behaviour reliably, how to operate controllable structures under changing conditions, where to allocate green and grey infrastructure, and how to minimize the capital investment required to achieve CSO-control objectives. Rather than evaluating these decisions independently, the proposed research considers them within a coordinated computational framework. This is particularly relevant for municipal infrastructure planning, where hydraulic effectiveness, computational feasibility, operational reliability, and investment affordability must be considered together.

Combined sewer overflow and stormwater discharge

Distributed green infrastructure and runoff management



Fig. 1. Problem and Solution.

B. Problem Statement

Despite advances in urban drainage modelling and infrastructure planning, managing CSOs remains a complex engineering problem involving interconnected hydraulic, operational, spatial, and economic constraints. A central difficulty is that conventional infrastructure planning often evaluates storage expansion, green infrastructure installation, and operational control as separate interventions. Such an approach may overlook interactions between runoff reduction, available in-line storage, regulator operation, and downstream conveyance capacity. As a result, a technically effective intervention may not necessarily represent the most economically efficient solution when evaluated across the entire drainage network. A second problem concerns the reliability of hydraulic prediction during operational control. MPC requires an internal model capable of predicting future system states and flow responses over a specified prediction horizon. When the internal model is inaccurate, the controller may select gate operations that fail to use available storage effectively or produce undesirable hydraulic conditions. Although data-driven surrogate models can reduce computational requirements, high one-step prediction accuracy does not necessarily demonstrate reliable performance during complete-event simulation. Small prediction errors can accumulate during autonomous rollout, potentially affecting predicted storage, regulator depth, interceptor flow, and overflow volume. Therefore, a surrogate model intended for control applications must be evaluated not only through conventional prediction metrics but also through physical consistency, closed-loop stability, and performance under rainfall conditions outside its training range. A third problem involves the operational limitations of real-time control. Control performance depends on the available hydraulic storage, gate configuration, telemetry observations, rainfall forecasts, prediction horizon, and control strategy. A sophisticated MPC algorithm may not consistently outperform a well-designed rule-based controller, particularly when the drainage network has limited in-line storage or when the control horizon is insufficient to represent the system's filling and drainage dynamics. Telemetry noise and rainfall-forecast uncertainty can further influence operational decisions. Consequently, the effectiveness of active control must be established through direct hydraulic evaluation rather than inferred solely from the theoretical capabilities of the optimization algorithm. A fourth problem concerns the spatial and economic allocation of infrastructure. GI interventions can reduce runoff entering combined sewer networks, whereas grey storage and RTC installations can increase the ability of existing networks to manage excess flow. However, their costs and hydraulic benefits differ considerably according to location, installation requirements, available land, existing network characteristics, and operational constraints. Selecting GI and RTC sites without evaluating their combined effects may lead to unnecessary capital expenditure. Furthermore, an infrastructure alternative that reduces CSO volume may not necessarily satisfy a specified annual overflow-frequency criterion. The planning problem must therefore identify infrastructure configurations that achieve the required hydraulic performance while minimizing total investment, including any residual storage needed to satisfy compliance objectives. The central research problem can consequently be formulated as follows: How can physics-informed hydraulic surrogate modelling, active telemetry-based control, and multi-objective spatial optimization be integrated to identify reliable and economically efficient combinations of green and grey infrastructure for reducing CSOs in combined sewer systems? Addressing this problem requires a framework that maintains the hydraulic credibility of the evaluation process, distinguishes surrogate prediction from final hydraulic assessment, compares predictive and rule-based control, and evaluates infrastructure portfolios according to both hydraulic performance and capital cost.

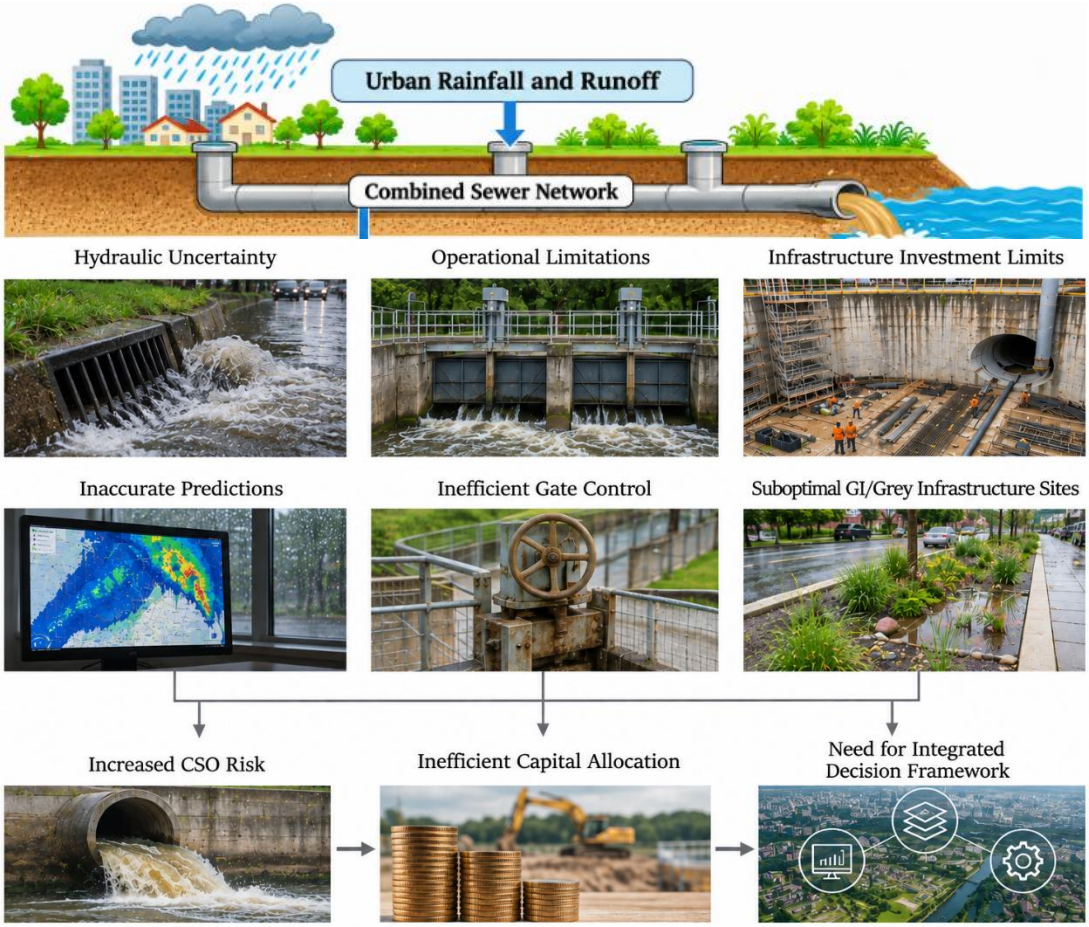


Fig. 2. Problem framework: interconnected limitations in conventional CSO management.

C. Proposed Solution

To address the identified problems, this study proposes an integrated PIML-driven framework that connects hydraulic simulation, surrogate development, active telemetry control, spatial infrastructure optimization, and capital-investment evaluation. The central design principle is to retain a physically detailed hydraulic model for generating training data and evaluating final system performance while using a computationally efficient surrogate model within the predictive control process. This separation allows the framework to benefit from machine-learning-based computational acceleration without relying exclusively on surrogate predictions when reporting final hydraulic outcomes.

The first component of the proposed framework involves generating a hydraulic dataset using EPA SWMM under historical rainfall events and design-storm conditions. The benchmark drainage network is simulated under different GI configurations and gate-operating policies to capture variations in regulator states, interceptor conditions, storage levels, lateral inflows, and overflow responses. The resulting hydraulic trajectories provide the training and evaluation data for multiple surrogate configurations, including data-only and physics-informed models. The PIML models incorporate control-volume continuity, simplified hydraulic momentum relationships, and regulator structure equations into the training objective. This approach is intended to improve the physical consistency and closed-loop reliability of hydraulic predictions rather than relying exclusively on statistical fitting. The second component develops a surrogate-assisted active control architecture. The selected surrogate is embedded within an MPC optimization process that evaluates alternative gate settings over a defined prediction horizon. Telemetry observations provide updated information about the current hydraulic state, allowing control decisions to be recalculated as system conditions change. The proposed evaluation also considers observation noise, rainfall-forecast uncertainty, and fail-safe operating logic. Rule-based control is retained as a comparison strategy to determine whether the additional computational complexity of predictive control produces meaningful hydraulic benefits under the evaluated network

conditions. SWMM remains the hydraulic plant for all final control comparisons. The third component addresses the spatial allocation of green and grey infrastructure. A multi-objective optimization procedure is used to evaluate alternative GI allocations and RTC installation locations. The decision variables represent the number of bioretention units assigned to individual sub-catchments and the selection of candidate RTC sites. The optimization process considers capital investment, CSO volume, and ponding-related traffic disruption as competing objectives. The resulting Pareto-efficient alternatives provide a set of infrastructure portfolios with different combinations of hydraulic benefits and investment requirements.

The fourth component evaluates the capital investment required to achieve the selected CSO-control criterion. Candidate infrastructure portfolios are evaluated through direct SWMM simulations under a representative rainfall year. Where the selected GI and RTC configuration does not independently satisfy the target, residual offline storage is added and sized according to the compliance requirement. Capital costs are then estimated using documented unit-cost information and uncertainty analysis. This process allows the economic performance of GI-led, grey-only, and RTC-based alternatives to be compared under consistent assumptions. The overall solution is designed to determine whether physically informed surrogate modelling can support computationally efficient control, whether active control provides additional hydraulic benefits over rule-based operation, and whether spatially distributed green infrastructure can reduce the need for expensive grey infrastructure. The framework also establishes an explicit separation between simulated hydraulic results, documented cost information, and assumed operational or traffic parameters. Accordingly, its conclusions are interpreted within the boundaries of the benchmark system rather than treated as direct evidence of field performance across all US combined sewer networks.

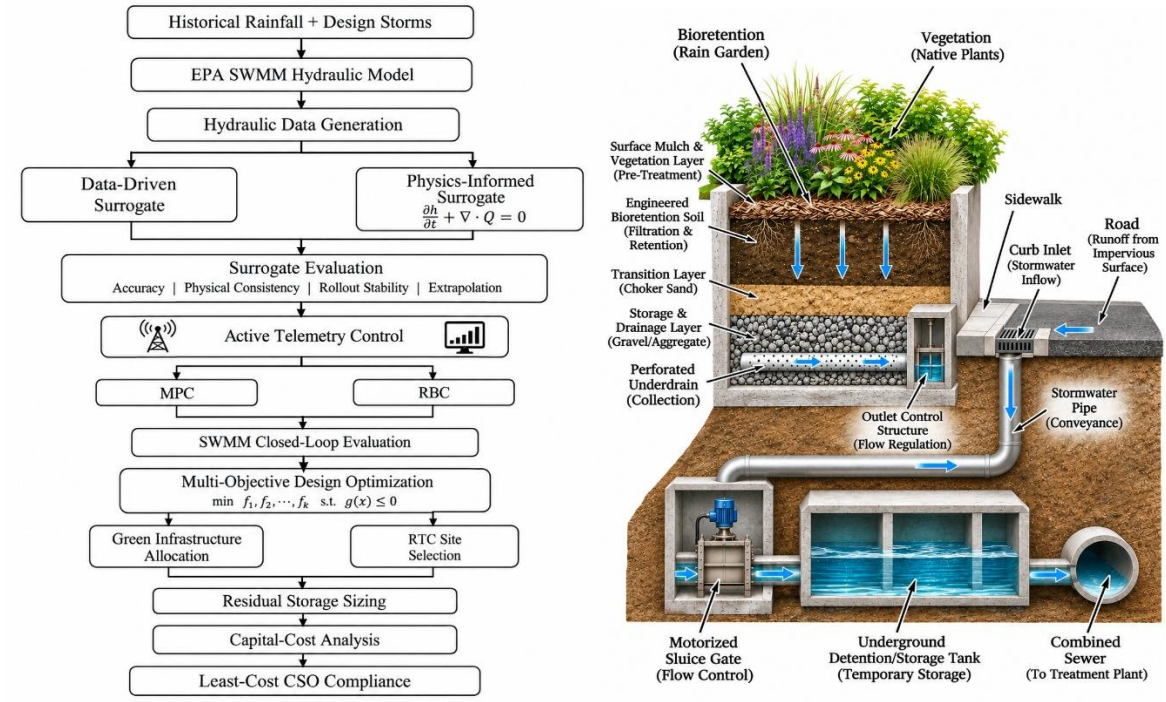


Fig. 3. Proposed integrated PIML-driven solution framework, Integrated Green Infrastructure and Underground Telemetry-Controlled Drainage Network.

D. Contributions

The principal contribution of this research is the development of an integrated computational approach that links physically informed hydraulic prediction, active control, and spatial infrastructure investment within a single CSO-management framework. Rather than treating machine learning, operational control, and infrastructure planning as independent research activities, the study connects these components through a reproducible simulation and evaluation process. The framework uses SWMM as the hydraulic reference model and employs surrogate modelling to support predictive control, thereby distinguishing computational acceleration from the assessment of final hydraulic

performance. A second contribution is the development and comparative evaluation of data-driven and physics-informed hydraulic surrogates under both one-step prediction and closed-loop rollout conditions. By incorporating hydraulic continuity and simplified flow-structure relationships into model training, the study examines whether physical constraints improve prediction stability and generalization under withheld extreme rainfall events. The analysis also distinguishes the effect of physics-informed constraints from the influence of training procedures and loss-weight selection, providing a more critical assessment of surrogate reliability. A third contribution is the comparative assessment of active predictive control and rule-based control using the same hydraulic simulation environment. The study examines the extent to which surrogate-assisted MPC can improve CSO management relative to passive and rule-based operation, while considering telemetry noise, rainfall-forecast uncertainty, and fail-safe logic. This comparison provides an operational perspective on whether the additional complexity of predictive control is justified by measurable improvements in the evaluated drainage network. A fourth contribution is the integration of spatial GI allocation and RTC site selection within a multi-objective optimization framework. The optimization considers hydraulic performance, infrastructure investment, and traffic-related disruption to identify alternative green–grey infrastructure portfolios. The resulting alternatives are subsequently evaluated under a representative rainfall year, with residual offline storage included where necessary to satisfy the selected CSO-control criterion. A fifth contribution is the comparative evaluation of capital-investment requirements across GI-led, grey-only, and RTC-based infrastructure alternatives. By combining direct hydraulic simulation with documented unit-cost information and Monte Carlo cost uncertainty analysis, the study examines the economic implications of alternative infrastructure configurations. This provides a transparent basis for identifying the conditions under which distributed green infrastructure may offer a more cost-effective compliance pathway than additional grey storage or new RTC installations.

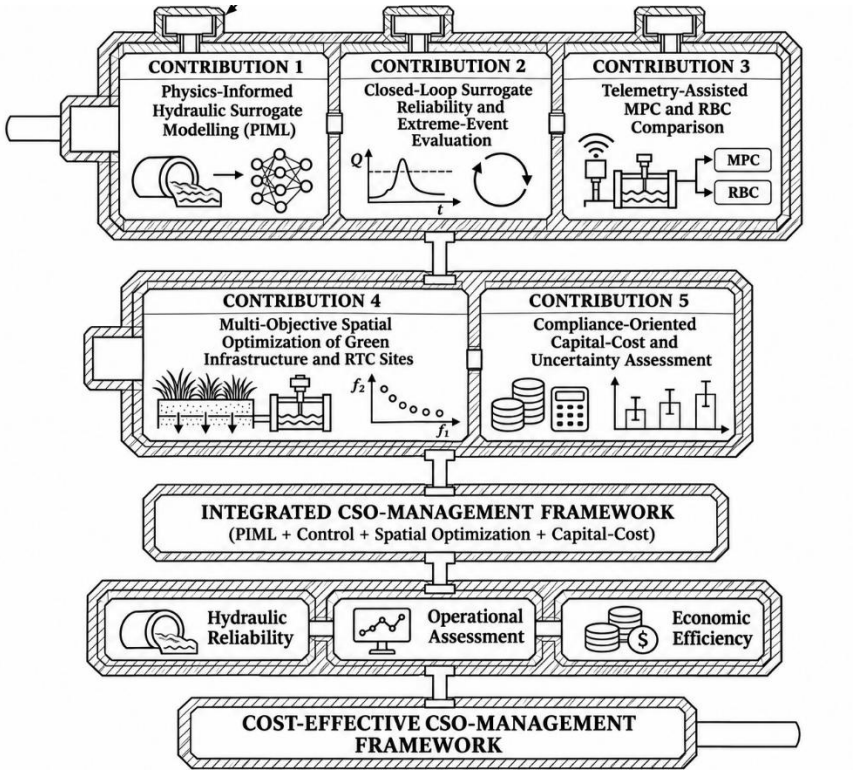


Fig. 4. Research contributions and their interrelationships in developing an integrated CSO-management framework.

E. Paper Organization

The remainder of this paper is organized as follows. Section II presents the related work, reviewing previous research on hydraulic modelling, physics-informed machine learning, urban drainage surrogate models, combined sewer overflow mitigation, real-time control, integrated green–grey infrastructure, and multi-objective optimization. It also identifies the research gaps that motivate the proposed framework. Section III describes the research methodology,

including the benchmark drainage network, rainfall forcing, hydraulic dataset generation, surrogate-model development, physical consistency constraints, real-time control formulation, spatial optimization procedure, and capital-cost assessment. Section IV presents the experimental results and discussion, including surrogate accuracy, closed-loop prediction performance, physical consistency, computational cost, control performance, robustness to telemetry and forecast uncertainty, Pareto-efficient infrastructure configurations, and comparative capital-investment requirements. Section V concludes the study by summarizing the principal findings, discussing the engineering implications, acknowledging the methodological limitations, and identifying priorities for future research and practical implementation.

II. Related Work

Related research on combined sewer overflow (CSO) mitigation has increasingly integrated hydraulic simulation, physics-informed machine learning (PIML), real-time control (RTC), and green–grey infrastructure optimization. This section reviews the relevant developments, evaluates their contributions to urban drainage management, and identifies the research gaps addressed by the proposed framework.

A. Hydraulic Modelling and Computational Platforms for Urban Drainage Systems

Hydraulic modelling provides the foundation for evaluating the performance of combined sewer systems under varying rainfall conditions, flow regimes, and operational strategies. The Storm Water Management Model (SWMM) is widely used to represent urban drainage networks and simulate runoff generation, hydraulic routing, storage, and combined sewer overflow processes. Its application framework provides a basis for constructing reproducible drainage simulations and assessing alternative infrastructure configurations [1]. The development of simulation environments specifically designed for stormwater control research has further strengthened the evaluation of drainage management strategies. The pystorms framework provides a simulation sandbox for developing and comparing stormwater control algorithms, allowing researchers to investigate alternative control policies under consistent hydraulic conditions [2]. Similarly, PySWMM provides a Python interface to SWMM, supporting programmatic model execution and integration with external computational methods, optimization algorithms, and control systems [3]. Despite these capabilities, conventional hydraulic simulation can become computationally demanding when repeated model evaluations are required for surrogate training, real-time optimization, and multi-objective infrastructure planning. This motivates the development of computationally efficient modelling approaches that preserve important hydraulic relationships while reducing the computational burden.

B. Physics-Informed Machine Learning and Surrogate Modelling

Physics-informed machine learning has emerged as an approach for combining data-driven prediction with governing physical relationships. Physics-informed neural networks incorporate physical equations and constraints into model training, enabling neural networks to represent systems governed by differential equations while reducing dependence on purely empirical relationships [4]. The broader development of physics-informed machine learning demonstrates its potential for integrating physical knowledge into learning architectures across scientific and engineering applications [5]. In urban drainage applications, physics-guided machine learning has been investigated as a means of accelerating hydrodynamic simulations while retaining important characteristics of the underlying hydraulic system. Palmitessa et al. demonstrated the application of physics-guided machine learning to accelerate urban drainage simulations, providing a relevant foundation for developing computationally efficient hydraulic surrogates [6]. Garzón et al. reviewed machine-learning-based surrogate modelling for urban water networks and identified methodological developments and future research directions, highlighting the importance of evaluating surrogate models beyond conventional prediction accuracy [7]. More recent research has explored network-aware learning architectures for representing the spatial and hydraulic dependencies of drainage systems. Zhang et al. investigated graph neural network-based surrogate modelling for real-time hydraulic prediction in urban drainage networks, illustrating the relevance of graph-based approaches for capturing network connectivity and hydraulic interactions [8]. Luo et al. examined machine-learning-based surrogate modelling within stochastic model predictive control of urban drainage systems, demonstrating the connection between computationally efficient prediction and control-oriented decision-making [9]. Although these studies establish the potential of machine learning for hydraulic prediction and computational acceleration, surrogate accuracy under individual prediction steps does not necessarily guarantee stable performance during extended closed-loop simulations. This distinction is particularly important when predicted storage, flow, and overflow states are repeatedly used by an optimization or control algorithm. Consequently, the integration of physical consistency constraints and closed-loop evaluation remains important for reliable surrogate-assisted drainage management.

C. Integrated Green Grey Infrastructure and Combined Sewer Overflow Mitigation

The integration of green infrastructure (GI), grey infrastructure, and real-time control has received increasing attention as a strategy for reducing CSO volumes while managing infrastructure investment. Jean et al. investigated the optimization of real-time control alongside green and grey infrastructure design, demonstrating the relevance of jointly considering operational control and infrastructure configuration when developing cost-effective CSO mitigation strategies [10]. Their subsequent research compared model predictive control and rule-based control in combination with green infrastructure, providing a basis for evaluating alternative operational strategies within integrated drainage management [11]. Further research has examined the spatial distribution of green and grey infrastructure in conjunction with real-time sewer control. Jean et al. investigated the optimal distribution of these infrastructure types as an adaptation strategy for controlling CSOs under climate-related challenges, emphasizing the importance of coordinated spatial and operational decisions [12]. These contributions indicate that the effectiveness of CSO mitigation depends not only on the selection of individual infrastructure technologies but also on their spatial allocation and operational coordination. However, infrastructure planning must also consider the computational requirements of evaluating alternative configurations and the financial implications of implementing different combinations of green and grey interventions.

D. Real-Time Control and Model Predictive Control in Urban Drainage Networks

Real-time control provides an operational approach for managing water levels, flow distribution, and available storage within urban drainage systems. Model predictive control (MPC) has attracted particular attention because it can use system predictions and operational constraints to determine control actions over a specified prediction horizon. Lund et al. reviewed MPC applications in urban drainage systems and discussed its potential contribution to smart real-time water management [13]. The performance of RTC systems is also influenced by uncertainty in system observations, rainfall forecasts, and hydraulic responses. Van der Werf et al. examined risks associated with uncertainties in real-time control of combined sewer systems, highlighting the need to account for uncertain operating conditions when assessing control performance [14]. García et al. reviewed modelling and real-time control approaches for urban drainage systems, providing a broader overview of the relationship between hydraulic modelling and operational control strategies [15]. These studies establish the importance of predictive and rule-based operational strategies for drainage management. Nevertheless, control performance depends on the quality of the internal system representation, the available storage capacity, the prediction horizon, and the reliability of telemetry and forecast information. Consequently, evaluating control algorithms within a hydraulic simulation environment is necessary to determine whether improvements in predictive accuracy translate into measurable reductions in CSO volume.

E. Green Infrastructure Planning, Spatial Allocation, and Investment Considerations

Green infrastructure planning involves selecting appropriate intervention types and allocating them across drainage catchments to reduce runoff entering combined sewer networks. Fu et al. developed a scenario-based stormwater management planning support system for reducing CSOs, demonstrating the relevance of scenario evaluation in infrastructure planning and decision support [16]. Joshi et al. examined how different sustainable drainage system approaches influence CSO outcomes, indicating that the selection and configuration of source-control interventions can affect drainage performance [17]. The comparison between green and grey infrastructure is also important for understanding the resilience and performance of urban drainage systems. Dong et al. investigated future resilience in urban drainage systems through a comparison of green and grey infrastructure, contributing to the discussion of alternative infrastructure-based adaptation strategies [18]. Botturi et al. reviewed established practices and innovative solutions for mitigating the environmental and public-health impacts of CSOs, emphasizing the broader importance of effective overflow management [19]. Infrastructure investment decisions are influenced not only by hydraulic performance but also by governance and implementation conditions. Hopkins et al. examined the influence of governance structures on green stormwater infrastructure investment, illustrating the relevance of institutional arrangements to infrastructure deployment [20]. In addition, Pregnolato et al. developed a depth-disruption function for assessing the impact of flooding on road transport, providing a methodological basis for representing the relationship between flood depth and transport disruption [21]. Economic comparisons of infrastructure alternatives are essential when selecting practical CSO mitigation strategies. Cohen et al. compared conventional grey CSO-control infrastructure with combined green–grey alternatives, directly connecting infrastructure configuration with cost considerations [22]. Collectively, these studies support the need to evaluate hydraulic performance, spatial allocation, and investment requirements together rather than treating infrastructure selection as an isolated engineering decision.

F. Multi-Objective Optimization and Computational Decision Methods

Multi-objective optimization provides a systematic approach for evaluating infrastructure alternatives under competing objectives, including capital investment, hydraulic performance, and operational impacts. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) is an established evolutionary optimization method designed to identify Pareto-efficient solutions in multi-objective problems [23]. The pymoo framework provides a computational implementation of multi-objective optimization methods, supporting the development and evaluation of optimization workflows for engineering applications [24]. For complex decision problems involving uncertain or stochastic evaluations, the cross-entropy method provides an optimization approach based on iterative sampling and updating of candidate solution distributions. Its theoretical and computational foundations have been presented in a comprehensive tutorial, supporting its application to simulation-based optimization problems [25]. MPC has also been investigated as a means of optimizing the operation of urban drainage systems. Abou Rjeily et al. examined MPC for optimizing urban drainage operations, contributing to the development of predictive operational strategies for drainage management [26]. Svensen et al. investigated MPC for urban drainage systems while considering uncertainty, further highlighting the importance of uncertainty-aware operational decision-making [27]. The economic effectiveness of RTC is another important consideration when comparing operational improvements with infrastructure investment. Dirckx et al. examined the cost-efficiency of real-time control for mitigating CSO impacts, demonstrating the importance of assessing the financial implications of control implementation alongside its environmental benefits [28].

G. Hydraulic Uncertainty and Spatial Design of Real-Time Control Systems

Uncertainty in urban drainage modelling can influence the reliability of hydraulic predictions and the interpretation of infrastructure performance. Deletic et al. assessed uncertainties in urban drainage models, establishing the importance of considering modelling uncertainty when evaluating drainage-system behavior and management alternatives [29]. The spatial placement of flow-control infrastructure also affects the performance and feasibility of RTC systems. Eulogi et al. investigated the hydraulic optimization of multiple flow-control locations for designing local real-time control systems, demonstrating the importance of selecting control locations in relation to the hydraulic characteristics of the drainage network [30]. These findings support the consideration of control-site selection as part of a broader infrastructure-planning problem rather than as a purely operational decision.

H. Research Gaps

Despite substantial progress in hydraulic simulation, physics-informed machine learning, real-time control, and green grey infrastructure planning, several methodological and practical gaps remain relevant to integrated CSO management.

i. Limited integration of physics-informed surrogate modelling and infrastructure optimization

Existing studies have investigated physics-guided hydraulic surrogates and machine-learning-assisted control, while other research has examined the optimization of green–grey infrastructure configurations. However, the coordinated use of physically constrained surrogate models, operational control, spatial infrastructure allocation, and capital-investment assessment within one reproducible framework remains an important research opportunity [6]–[12], [16]–[18].

ii. Insufficient emphasis on closed-loop surrogate reliability

Conventional surrogate evaluation may prioritize prediction accuracy under one-step or independently sampled conditions. However, control applications repeatedly use predicted states, potentially allowing prediction errors to accumulate over time. The need to assess surrogate reliability during autonomous rollout, physical consistency, and extreme-event extrapolation remains important for dependable surrogate-assisted control [4]–[9].

iii. Limited comparative assessment of predictive and rule-based control within integrated infrastructure planning

MPC and rule-based control have both been investigated for urban drainage management, including applications involving green infrastructure. Nevertheless, their relative performance should be evaluated under consistent hydraulic conditions and in relation to available storage, control-horizon selection, and operational constraints. Improved surrogate prediction should not automatically be interpreted as improved control performance [10]–[15], [26], [27].

iv. Incomplete coordination of spatial infrastructure allocation and capital-cost minimization

Previous research has considered green–grey infrastructure combinations, CSO mitigation, and infrastructure investment. However, selecting GI locations, determining RTC deployment sites, and sizing residual storage require

coordinated evaluation of hydraulic performance and capital investment. This creates a need for optimization frameworks that explicitly compare alternative infrastructure portfolios under consistent cost assumptions [16]–[22], [28], [30].

v. Insufficient consideration of uncertainty across hydraulic modelling and operational control
Rainfall forecasts, telemetry observations, and hydraulic model representations can introduce uncertainty into drainage-system decisions. Although uncertainty has been examined in urban drainage modelling and RTC, its combined influence on surrogate-assisted control and infrastructure selection warrants further investigation [14], [27], [29].

vi. Limited integration of operational and construction-related impacts into infrastructure decision-making
Infrastructure alternatives can affect both hydraulic performance and surrounding urban activities. Flood-related road disruption and governance-related investment conditions have been examined separately, indicating the importance of incorporating relevant operational and implementation considerations into integrated infrastructure assessments [19]–[21].

III. Methodology

The methodology integrates EPA SWMM-based hydraulic simulation, physics-informed machine learning (PIML) surrogate development, and telemetry-based real-time control to predict and manage combined sewer overflows. It then applies NSGA-II multi-objective spatial optimization and capital-cost analysis to identify the most cost-effective combination of green and grey infrastructure for achieving CSO reduction targets.

A. Research design

The study couples four steps (Fig. 1): (i) generation of a hydraulic dataset with EPA SWMM 5.2.4 under historical and design rainfall, randomized green infrastructure (GI) layouts and randomized gate operation; (ii) training of data-only and physics-informed surrogates of the regulator-interceptor dynamics; (iii) closed-loop model predictive control (MPC) of the regulator gates with the surrogate as the internal model and SWMM as the plant; and (iv) multi-objective GI and real-time control (RTC) siting followed by least-cost sizing of residual storage to meet the US CSO Control Policy. Costs and traffic disruption are evaluated after the hydraulic simulations, so that uncertain economic inputs do not alter the hydraulic results.

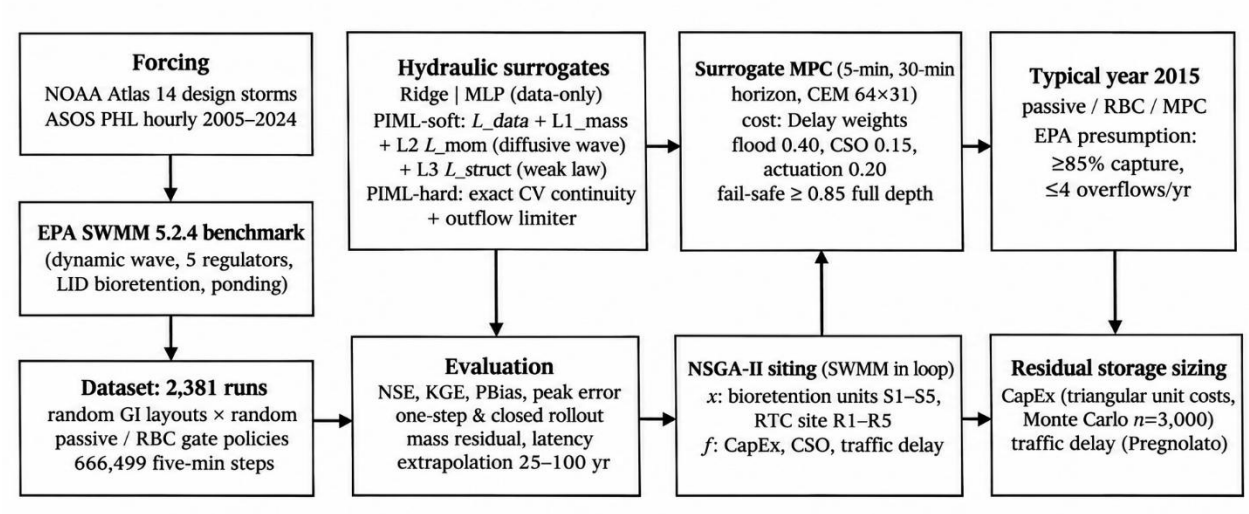


Fig. 5. Integrated Research Workflow for PIML-Driven Hydraulic Surrogate Modeling, Active CSO Control, Spatial Infrastructure Optimization, and Capital-Cost Evaluation.

Because the draft did not identify a catchment and no utility data were available, the analysis uses a public benchmark so that every result can be reproduced. The network is the combined sewer system of Example 8 of the EPA SWMM Applications Manual [1, 2], distributed as scenario 'alpha' in the pystorms control sandbox [3] and run through PySWMM [4]. It drains seven sub-catchments with a total area of 11.7 ha (28.9 ac). Sub-catchments S1–S5 drain to

five combined-sewer regulators (R1–R5); each regulator diverts dry-weather and part of the wet-weather flow through an orifice (Or1–Or5) to an interceptor that conveys flow to the treatment plant (WWTP) and spills the excess over a transverse weir (W1–W5) to the receiving creek (Fig. 6). Junction J1 also spills through a relief conduit (C1a). Sub-catchments S6 and S7 drain to the creek and are not part of the combined system. The benchmark is small compared with a city-wide system, and the cost results in Section IV must be read at that scale.

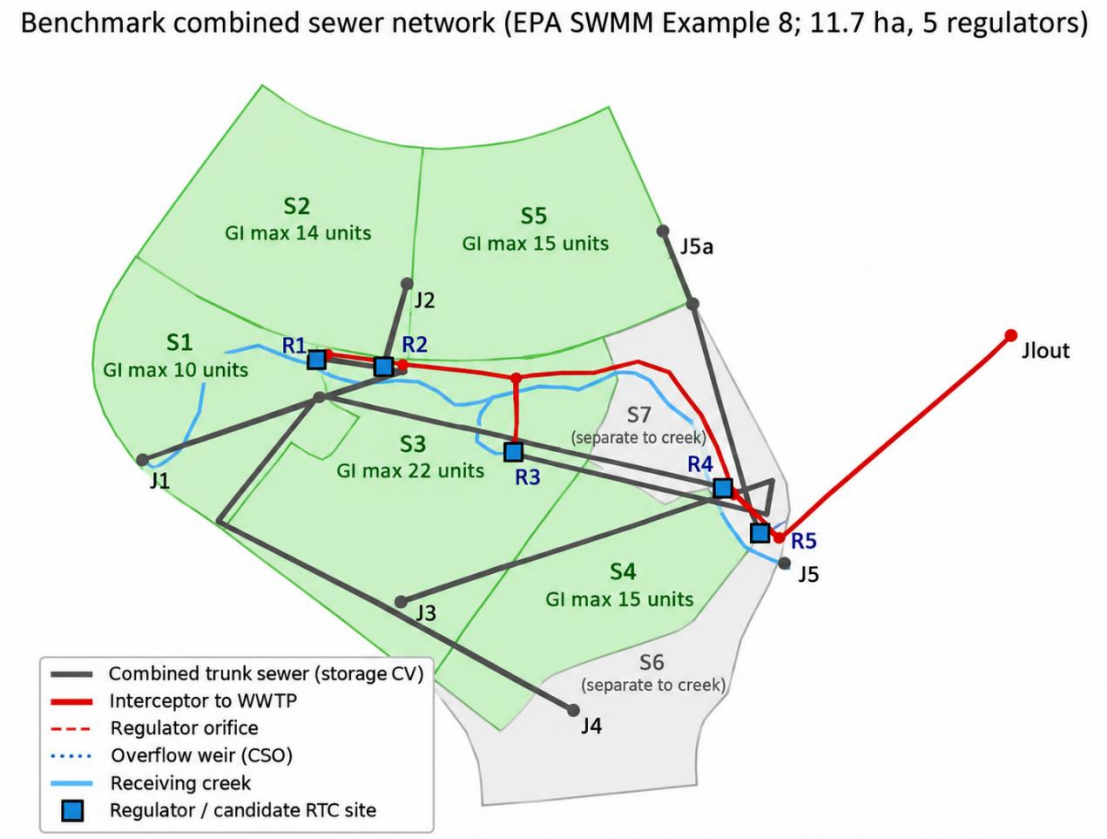


Fig. 6. Benchmark combined sewer network (EPA SWMM Applications Manual Example 8 / pystorms 'alpha'): sub-catchments, regulators R1–R5 with orifice and weir outlets, interceptor and WWTP outfall.

Table 1. *Benchmark Combined Sewer Network Characteristics.*

Parameter	Description
Hydraulic model	EPA SWMM 5.2.4
Benchmark network	EPA SWMM Applications Manual Example 8
Benchmark scenario	pystorms – alpha
Total catchment area	11.7 ha (28.9 ac)
Number of subcatchments	7
Combined-system subcatchments	S1–S5
Separate creek-draining subcatchments	S6–S7
Number of combined sewer regulators	5 (R1–R5)
Flow-control structures	Orifices and transverse weirs
Receiving water body	Receiving creek
Downstream treatment facility	Wastewater Treatment Plant (WWTP)

B. Rainfall forcing and dataset generation

Historical forcing consists of hourly precipitation at Philadelphia International Airport (ASOS station PHL) for 2005–2024, obtained from the Iowa Environmental Mesonet [5]. Rainfall events were separated with a 6-h minimum inter-event time and a 0.10-in (2.5-mm) minimum depth, which gives 1286 events. Design storms were built as alternating-block hyetographs of 1-, 2- and 6-h duration from the NOAA Atlas 14 partial-duration depths at the same location [6] for return periods of 1–100 years (Fig. 7). Calendar year 2015 (43.8 in, 62 events) was chosen as the typical year for compliance checks because its depth and event count are closest to the 2005–2024 means (44.2 in, 64 events).

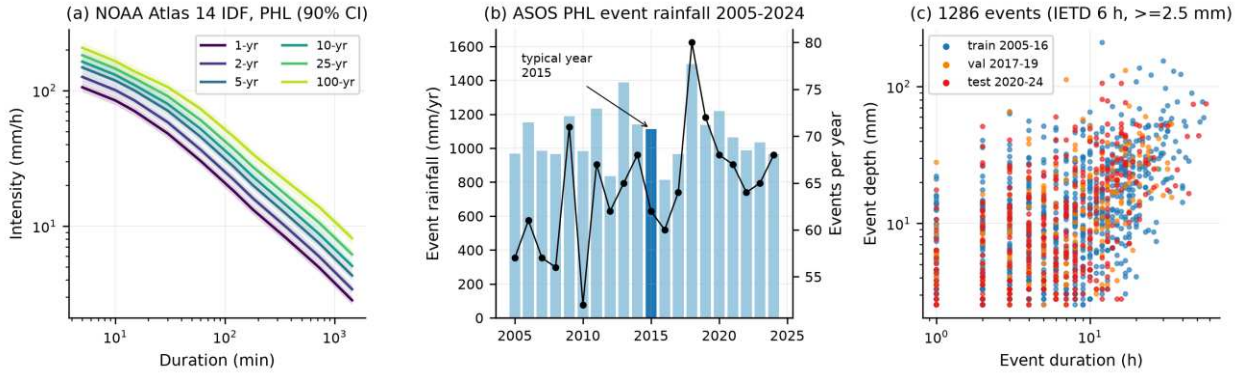


Fig. 7. Rainfall forcing: NOAA Atlas 14 depth-duration–frequency at PHL, alternating-block design hyetographs, and the 2005–2024 ASOS event record with the typical year (2015) highlighted.

Each SWMM run combined one rainfall event with a randomly drawn GI layout (bioretention units in S1-S5, each 400 ft², treating 15 times its area of impervious surface, up to the feasible maximum per sub-catchment) and one gate policy: random settings held for random durations, a passive policy (gates fixed open), or a rule-based policy with random thresholds. SWMM was run with dynamic-wave routing, and states and step-mean flows were recorded every 5 min, which is the control interval. Splits are chronological so that test events are never seen in training (Table 2). The 25-, 50- and 100-yr design storms were withheld from training to test extrapolation beyond the trained rainfall range. In total 2381 runs and 666,499 five-minute steps were generated (simulated).

Table 2. *Surrogate dataset (all values with SWMM 5.2.4).*

Split	Runs	5-min steps	Gate policies	Max. event depth (in)
Training (ASOS ≤2016)	1,070	311,266	Random	8.26
Training (Atlas 14, 1–10 yr)	96	22,944	Random	4.80
Validation (ASOS 2017–2019)	465	131,259	Passive, random, RBC	5.16
Test (ASOS 2020–2024)	642	175,218	Passive, random, RBC	4.16
Extrapolation (Atlas 14, 25–100 yr)	108	25,812	Passive, random, RBC	7.69

C. Decision weights

The MPC objective (Section III-F) combines surface flooding, CSO volume, gate actuation and WWTP surcharge with weights 0.40, 0.15, 0.20 and 0.05. These weights are the values listed in the submitted draft. In this revision they are treated as author-specified preferences, because no expert elicitation, questionnaire or interview data were collected for this study and no ethics approval exists for such work. If the authors wish to retain a Delphi procedure, the panel composition, rounds, consensus criterion, response data and the approval of an institutional review board must be reported; otherwise the weights should be described exactly as here. Because the design and cost results in Sections IV-G and IV-H are computed with SWMM and fixed rule-based or MPC operation, they do not depend on these weights except through the MPC runs.

D. Surrogate model formulation

The surrogate is a one-step model of the regulator–interceptor system at the 5-min control interval. The input vector x_k(43 values) contains the state s_k(28 values: trunk-sewer storage V_i, regulator depth $h_{R,i}$and upstream-junction depth $h_{J,i}$at the five regulators, interceptor storage V_{int}, six interceptor node depths $h_I$and six interceptor link flows Q_I), the step-mean lateral inflows $Q_{lat,i}$of the five combined subcatchments, and the ten gate settings (orifice and weir, 0 = closed/crest raised, 1 = open). The output contains twelve step-mean flows (five orifice flows $Q_{or,i}$, five weir overflows $Q_{w,i}$, the relief flow $Q_{relief}$and the WWTP inflow Q_{WWTP}) and the next state s_{k+1}. All neural models share one architecture: a multilayer perceptron with three hidden layers of 128 SiLU units (43,816 parameters) implemented in PyTorch. A ridge regression model on the same inputs serves as a linear baseline. Physics-informed machine learning follows the residual-penalty approach used in physics-informed neural networks [4] and its application to urban drainage modelling [6]; here, the physics terms are formulated for the control volumes of this network rather than for a continuous partial differential equation.

Three physical relations are used. First, the finite-volume continuity equation for each regulator control volume and for the interceptor:

$$V_i(k+1) - V_i(k) = \Delta t\left[Q_{lat,i} - Q_{or,i} - Q_{w,i} - Q_{relief,i}\right],\ V_{int}(k+1) - V_{int}(k) = \Delta t\left[\sum_i Q_{or,i} - Q_{WWTP}\right] \quad (1)$$

with $\Delta t = 300$s (Q_{relief} only at R1). Second, the momentum equation on the six interceptor conduits in diffusive-wave form, in which the inertial terms of the Saint-Venant momentum equation are neglected and the friction slope equals the water-surface slope:

$$Q_{I,j} = \text{sign}\left(S_{f,j}\right)\frac{1.49}{n}A_j R_j^{2/3} \mid S_{f,j} \mid^{1/2},\ S_{f,j} = \frac{H_{up,j} - H_{dn,j}}{L_j} \quad (2)$$

with Manning $n = 0.016$, flow area $A$and hydraulic radius $R$of the circular section evaluated at the mean end depth, and $H$the hydraulic head. This is a simplification of the full Saint-Venant system and is described as such throughout. Third, the structure laws at the regulators: a transverse weir whose crest is raised by the gate and an orifice whose open height is set by the gate:

$$Q_{w,i} = K_i C_w L_{w,i}\left[h_{R,i} - z_{c,i} - \left(1 - \theta_{w,i}\right)H_{g,i}\right]_+^{1.5},\ Q_{or,i} = C_d A_o\left(\theta_{or,i}\right)\left[2g\left(h_{R,i} - a_i/2\right)\right]^{1/2} \quad (3)$$

where $C_w = 3.3$, $C_d = 0.65$, $z_c$the fixed crest, $H_g$the gate height, $\theta$the gate setting and $a$the open height. The weir relation is applied only in the free-flow regime (head ≤ 1 ft), because SWMM switches to submerged and surcharged formulations above that. The effective coefficients K_i(0.47–0.69) were fitted to the training data in that regime. The training loss is

$$L = L_{data} + \lambda_m L_{mass} + \lambda_q L_{mom} + \lambda_s L_{struct} + \lambda_+ L_{neg} \quad (4)$$

where $L_{data}$is the mean squared error of standardised outputs, L_{mass}, $L_{mom}$and $L_{struct}$are the mean squared residuals of Eqs. (1)–(3) normalised by their scale in the training data, and L_{neg}penalises negative flows. Two physics-informed variants were trained. PIML-soft uses $\lambda = (0.1, 0.02, 0.02)$. PIML-hard enforces Eq. (1) exactly by construction: flows pass through a soft plus, outflows are limited by the available storage, and the next storage is computed from Eq. (1) rather than predicted; its storage outputs are therefore removed from $L_{data}$and $\lambda = (0, 0.02, 0.02)$. The data-only MLP uses $\lambda = 0$.

Training has two stages. Stage 1 fits the one-step map for 50 epochs (Adam W, learning rate $2 \times 10^{-3}$with cosine decay, batch 1024) on 334,210 training transitions. Stage 2 fine-tunes each network for 12 epochs on 12-step (1-h) closed-loop sequences with back-propagation through time (learning rate 5×10^{-4}, batch 256), so that the model is trained on its own predicted states. Stage 2 was introduced because all stage-1 networks drifted when rolled out over complete events. Model selection used the validation years only. All training ran on one CPU thread.

E. Surrogate evaluation

Accuracy is reported with the Nash–Sutcliffe efficiency (NSE), the Kling–Gupta efficiency (KGE), percent bias and peak error for five aggregated targets: total CSO flow, WWTP inflow, total trunk storage, mean upstream-junction depth and mean interceptor depth. Two modes are distinguished. One-step skill uses the true SWMM state at every step, whereas closed-loop rollout feeds the surrogate its own predictions over the whole event, which is the condition faced inside a controller. Physical consistency is measured by the absolute continuity residual of Eq. (1) as a percentage of inflow volume (one-step, and median over events in rollout) and by the root-mean-square residuals of Eqs. (2) and (3). The same residuals were computed for the SWMM records themselves. SWMM's cumulative continuity error is about 0.03%, but the 5-min aggregation of its internal time step leaves a non-zero residual against Eqs. (1)–(3); this value is reported as the reference floor. Latency was timed per call with batch sizes 1 and 64 and compared with the wall time of SWMM per 5-min step.

F. Real-time control

Three operating modes are compared, always with SWMM as the plant. In passive operation all gates stay open. Rule-based control (RBC), written as native SWMM control rules, raises the weir crest at a regulator while its upstream junction is below 50% of full depth and the WWTP interceptor node is below 75% of full depth, and lowers it otherwise. MPC re-optimizes the ten gate settings every 5 min over a 30-min horizon (six steps, two move blocks)

using the cross-entropy method with 64 samples, 3 iterations and 8 elite samples; the first move is applied to SWMM. The stage cost iskpidepot

$$J = \sum_{h} \left[0.15 \frac{V_{CSO,h}}{100} + 0.40 \cdot 50\Phi_{flood,h} + 0.05 \cdot 50\Phi_{WWTP,h} \right] + 0.20 \cdot 0.5 \sum_{h} \| \Delta\theta_h \|^2 \qquad (5)$$

where $V_{CSO}$is the predicted overflow volume (ft³), $\Phi_{flood}$is a quadratic penalty on junction depths above 80% of full depth and on storage above 95% of capacity, Φ_{WWTP}penalizes interceptor depth above 75% of full depth and $\Delta\theta$is the change of gate settings. Lateral inflows over the horizon are taken from a passive SWMM run of the same event, which represents a perfect runoff forecast. Forecast error is then added as multiplicative log-normal noise. A fail-safe rule overrides the optimiser and opens the gates when any upstream-junction depth exceeds 85% of full depth. Telemetry noise is simulated as multiplicative Gaussian error on all observed states, smoothed by an exponential filter ($\alpha = 0.6$); this is a simple filter, not a Kalman filter. MPC was run with the PIML-soft surrogate (selected on validation rollout skill) and, for comparison, with the data-only MLP. Robustness was tested on the 2-yr 6-h and 10-yr 2-h storms for telemetry noise of 0, 5, 10 and 20% and forecast error of 0, 25 and 50% (three random seeds each), and the fail-safe was switched off for one combination (10% noise, 25% forecast error) to measure its effect. These perturbation levels were selected as scenario-based sensitivity ranges to evaluate the controller's response to telemetry noise and rainfall-forecast uncertainty.

G. Spatial optimization of GI and RTC and compliance sizing

The siting problem has ten integer decision variables: the number of bioretention units in S1–S5 (0 to the feasible maximum of 10, 14, 22, 15 and 15) and a binary RTC flag for each regulator. Three objectives were minimized with NSGA-II [23] in pymoo [24] (population 40, 30 generations, 1,162 SWMM evaluations): (f1) capital cost of GI and RTC at modal unit costs; (f2) total CSO volume over the 1-yr 2-h and 2-yr 6-h storms with RBC at the selected sites; and (f3) ponding-induced traffic delay for the 10-yr 1-h storm. Every evaluation used SWMM directly, so the Pareto front does not depend on surrogate error.

Each Pareto design, a grey-only reference and RTC-only references were then simulated for the 62 events of the typical year. The US CSO Control Policy accepts, under the presumption approach, no more than four overflow events per year on average or capture of at least 85 % of wet-weather combined flow. The passive network already captures 89.8 % of wet-weather volume (simulated) but overflows in 13 events, so the four-event criterion is the binding one. Because GI removes runoff from the combined system, it lowers the denominator of the capture ratio and can reduce the percent capture even when it reduces overflow volume; the percent-capture criterion is therefore not used to rank designs. For each design, off-line storage tanks at the regulators were sized greedily in 2,000-gal steps until no more than four events per year overflowed by more than 1 m³. The least-cost compliant design is reported as the selected design.

H. Capital cost model

The capital cost model estimates the investment required for GI installation, RTC retrofits, and residual offline storage using documented unit-cost ranges expressed in 2024 USD. Triangular cost distributions, a 30% design and contingency markup, and 5,000 Monte Carlo simulations are applied to quantify investment requirements and uncertainty across alternative infrastructure configurations.

Table 3. *Unit costs (2024 USD) used as triangular distributions; a 30 % design and contingency markup is applied to all items.*

Item	Low	Mode	High	Basis (documented)
Bioretention / curb-extension rain garden ($/ft²; 400 ft² per unit)	10	20	40	Camden GI handbook; EPA bioretention fact sheet and Opti-Tool cost data [23, 24]
RTC retrofit per regulator (gate, actuator, RTU, communications)	$0.62 M	$0.85 M	$3.24 M	EPA in-line storage fact sheet, $264k (ENR 5750) escalated to 2024 ; Jean et al. ≈1.15 M CAD per site ; Buffalo, two RTC sites $6.48 M
Level sensor (2 per RTC site)	$3,300	$7,500	$12,900	EPA fact sheet (not escalated)
Off-line storage tank ($/gal)	5.6	12.0	17.7	Elizabeth NJ LTCP, 12.5 MG for $69.9 M; Alexandria VA LTCP, 3 MG for $35–53 M

Capital cost is evaluated at modal values and by Monte Carlo sampling (5,000 draws from the triangular distributions, independent between items, common to all designs) to obtain the distribution of savings relative to grey-only storage. Operation and maintenance costs, land costs and life-cycle discounting are not included. The RTC break-even analysis in Section IV-H evaluates the site-level retrofit cost at which RTC would become economically comparable with the

grey-only alternative under the assumptions of this benchmark. The draft's own unit costs ($15k per gate, $2.5k per sensor) were also evaluated for transparency.

I. Traffic disruption model

No traffic data exist for the benchmark, so traffic disruption is an assumed, illustrative quantity. Each street manhole was assigned a road class (arterial, collector or local) with assumed average daily traffic of 18,000, 7,000 and 1,500 vehicles and free-flow speeds of 50, 40 and 30 km/h, and represents a 100-m segment. Ponded depth from SWMM is converted to a speed limit with the depth–disruption function of Pregnolato et al. [21], $v(d) = 0.0009d^2 - 0.5529d + 86.9448$ (km/h, d in mm), with closure above 300 mm and a 0.5-km detour at 30 km/h. Construction disruption is computed from assumed lane-closure durations (120 days plus 60 days per million gallons for a tank, 5 days per four GI units, 20 days per RTC site) and a 25 % speed reduction over a 300-m work zone on an arterial. Neither quantity should be cited as a forecast of real traffic losses.

J. Software and reproducibility

All steps are scripted in Python 3.14 with PySWMM 2.2.0 (SWMM 5.2.4), PyTorch (CPU), pymoo 0.6.2 and matplotlib. The package lists the run order, random seeds (2026), raw rainfall files, the SWMM input file, trained model weights and all result tables, so that every figure and table in Section IV can be regenerated. Dataset generation took about 14 min and each surrogate training stage 4–6 min on two CPU cores.

IV. Results and Discussion

Section IV presents the experimental findings and evaluates the proposed framework in terms of surrogate prediction accuracy, physical consistency, computational efficiency, real-time control performance, and robustness under uncertainty. It further examines spatial GI–RTC optimization, capital investment requirements, and traffic disruption to identify the most hydraulically effective and economically viable CSO-mitigation strategy.

A. Surrogate accuracy

Fig. 8 shows the validation losses. In stage 1 the data-only MLP reached the lowest data loss but the highest mass and weir-law residuals, whereas the hard-constrained model held the continuity residual at numerical precision (about 10^{-13}). The jump at the start of stage 2 is expected, because the loss is now computed on the models' own 12-step predictions; all models then decreased again.

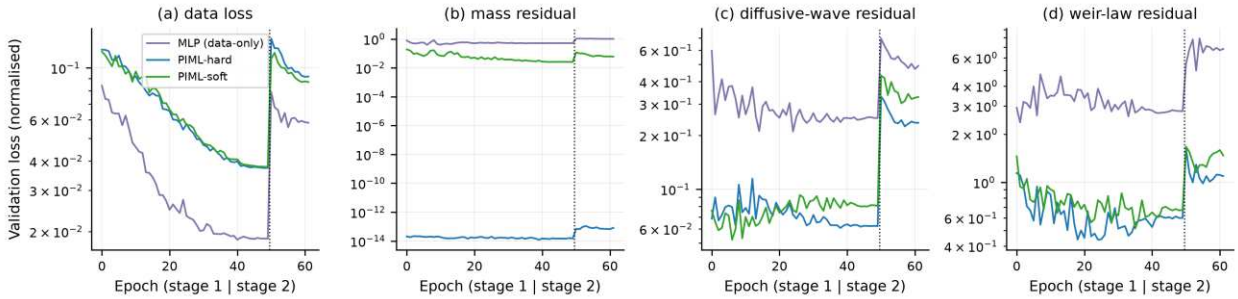


Fig. 8. Validation loss components during stage 1 (one-step, epochs 0-49) and stage 2 (12-step rollout fine-tuning, epochs 50–61). Values are normalized by their training-data scale.

Table 4. *Surrogate skill (NSE) on the test years (2020–2024, 642 events) and on the withheld 25–100-yr Atlas 14 storms (108 runs). One-step uses SWMM states as inputs; rollout feeds the model its own predictions over the full event.*

Model	Test one-step CSO	Test rollout CSO	Test rollout WWTP	Test rollout storage	Test rollout junction depth	25–100 yr rollout CSO	25–100 yr rollout storage	Test rollout CSO PBIAS (%)
Ridge	0.71	-0.30	0.51	-0.32	0.66	0.87	0.84	-11
MLP (data-only)	0.91	-1.04	0.67	-0.86	0.10	0.06	0.85	618
PIML-soft	0.89	0.87	0.97	0.58	0.91	0.93	0.93	-11
PIML-hard	0.84	0.75	0.96	-20	-1.34	0.87	-0.90	-17

One-step skill was high for all neural models (CSO NSE 0.84-0.91; Table 4, Fig. 9a), and on this measure the data-only MLP was best. The ranking reversed in closed-loop rollout. The MLP's rollout NSE for CSO flow fell to -1.04 on the test years with a bias of +618 %, because small errors in storage accumulated into spurious overflows in events that SWMM simulates without any spill. PIML-soft kept a rollout NSE of 0.87 for CSO flow, 0.97 for WWTP inflow and 0.91 for junction depth. Storage was the weakest target (NSE 0.58). PIML-hard reproduced CSO and WWTP flows reasonably (NSE 0.75 and 0.96) but its storage drifted strongly (NSE -20): exact continuity guarantees that mass is conserved, not that it is routed to the right place, and storage that is over-predicted early in an event is drained only slowly (Fig. 10).

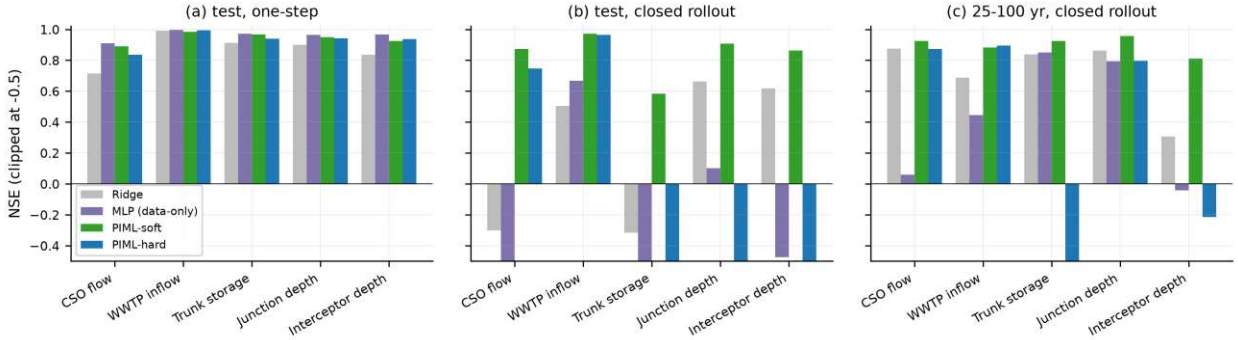


Fig. 9. NSE of the four surrogates for five targets: (a) one-step on the test years, (b) closed-loop rollout on the test years, (c) closed-loop rollout on the 25–100-yr storms. Bars are clipped at −0.5.

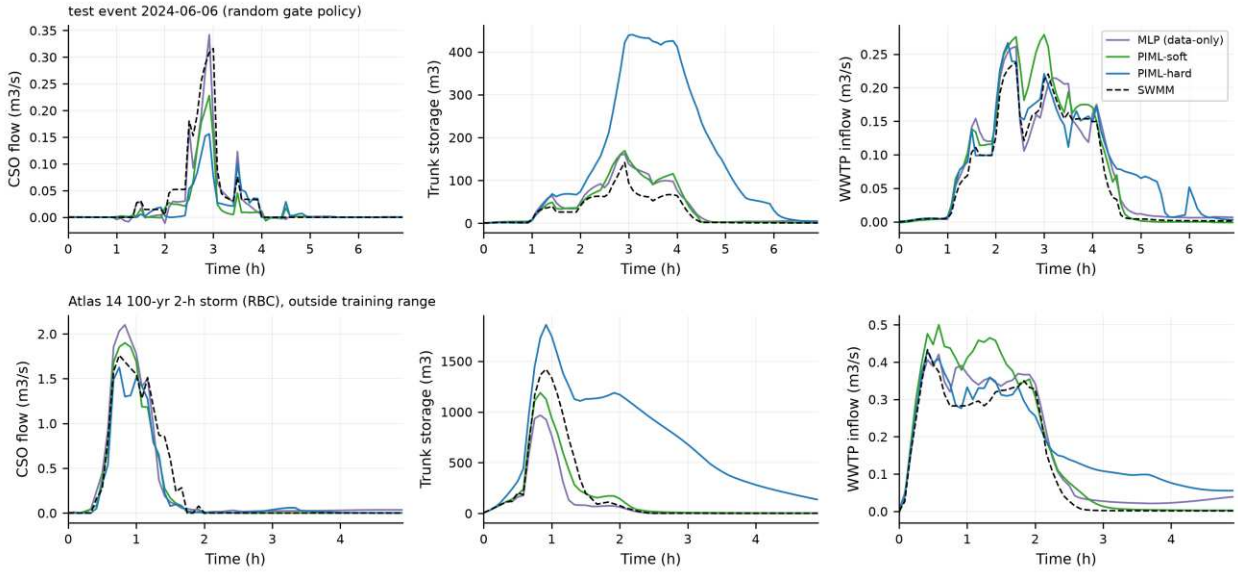


Fig. 10. Closed-loop rollout hydrographs against SWMM for a test event under random gate operation (top) and the 100-yr 2-h storm under RBC, whose peak intensity exceeds that of any training event (bottom).

The event-scale picture is consistent (Fig. 11). For the test events, PIML-soft and PIML-hard event volumes lie close to the 1:1 line above about 100 m³, while the MLP produces large false overflows for many events with little or no simulated spill. On the 25–100-yr storms, whose short-duration intensities exceed those of every training event (the largest training event peaked at 2.56 in/h), PIML-soft reached a rollout CSO NSE of 0.93 and storage NSE of 0.93, against 0.06 and 0.85 for the MLP. The linear ridge model also extrapolated well in this regime (CSO NSE 0.87), because large storms drive the regulators into a nearly linear overflow regime, but it failed on ordinary events (test rollout NSE -0.30). These results support the use of physics terms for rollout stability and extrapolation; they do not support the draft's single NSE of 0.952, which mixes the two evaluation modes.

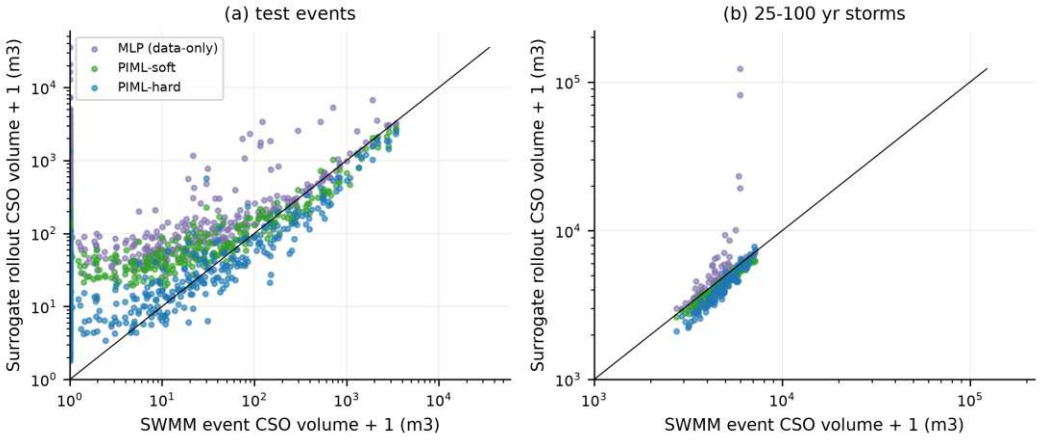


Fig. 11. Event CSO volume from closed-loop surrogate rollout versus SWMM, (a) test events and (b) 25–100-yr storms (log axes, +1 m³ offset).

B. Physical consistency

Physical consistency is evaluated by measuring how accurately the surrogate models satisfy hydraulic continuity, diffusive-wave momentum relationships, and regulator weir-flow equations. The analysis compares one-step and closed-loop residuals against the SWMM reference to assess mass conservation, hydraulic validity, and physically realistic flow predictions under both normal and extreme rainfall conditions.

Table 5. *Residuals of the physical relations, test years unless stated. Mass residual is the absolute continuity error of Eq. (1) as a percentage of lateral inflow volume. The SWMM row is the residual of SWMM's own 5-min records against the same relations.*

Model	One-step mass residual (%)	Rollout mass residual, median (%)	Rollout mass residual, 25–100 yr (%)	Diffusive-wave RMSE (L/s)	Weir-law RMSE (L/s)	Negative flows (% of outputs)
SWMM reference (5-min records)	2.34	0.06	0.94	3.77	2.28	0.0
Ridge	5.65	9.32	0.90	3.79	3.26	25.1
MLP (data-only)	15.20	44.79	8.32	3.93	3.24	7.9
PIML-soft	7.53	14.03	3.02	3.45	1.93	9.3
PIML-hard	0.00	0.00	0.00	3.41	1.59	0.0

Table 5 quantifies how far each model departs from the physical relations. SWMM's own 5-min records leave a one-step continuity residual of 2.3 % because the relations are evaluated on step means of a model that integrates at a few seconds; a surrogate should not be expected to do better than this floor. The data-only MLP had a residual of 15.2 % in one-step mode and a median of 45 % over test-event rollouts, and 8 % of its flow outputs were negative. PIML-soft halved the one-step residual (7.5 %) and reduced the rollout residual to 14 %, and PIML-hard closed continuity exactly with no negative flows. Residuals of the diffusive-wave and weir relations were also smallest for the physics-informed models, but the differences were small for the diffusive-wave term, which SWMM's dynamic-wave solution itself satisfies only approximately.

C. Ablation of loss weights and training-data size

This subsection examines how physics-loss weighting and training-data availability affect surrogate-model stability and closed-loop prediction accuracy. It compares alternative combinations of mass, momentum, and structural loss weights and evaluates model performance across different training-data proportions using test events and withheld extreme storms.

Table 6. *Loss-weight ablation at reduced budget (stage-2 fine-tuning for 4 epochs on 50 % of training sequences from a common data-only initialization). Rollout NSE on 150 test events and on the 25–100-yr storms.*

(λ_mass, λ_mom, λ_struct)	Test rollout CSO NSE	25–100 yr rollout CSO NSE	Test rollout junction-depth NSE	Test rollout mass residual (%)
(0, 0, 0)	-39	0.88	-3.49	65
(0.01, 0.002, 0.002)	-57	0.76	-220	105
(0.1, 0.02, 0.02)	0.79	0.92	0.24	49
(1, 0.2, 0.2)	-7.80	-2.41	-3.67	41
(0.1, 0, 0)	-7,361,951	0.91	-1,227,476	141
(0, 0.02, 0.02)	0.59	0.63	0.41	49

The loss-weight ablation (Table 6, Fig. 12b) shows that the benefit of physics terms depends strongly on their weight. Without physics terms, or with very small weights, rollout on the test events diverged (CSO NSE far below zero) even though the extreme storms were still reproduced. The weights used for PIML-soft, (0.1, 0.02, 0.02), gave the best test rollout (NSE 0.79). Raising them tenfold to (1, 0.2, 0.2), close to the weights used in an earlier version of this work, degraded all targets, because the physics residuals of Eqs. (2) and (3) are themselves approximate and start to compete with the SWMM data. Using the mass term alone was unstable, and the structure terms alone were stable but less accurate. The training-size ablation (Fig. 12a) was run with a shorter schedule than the full models (20 + 4 epochs). In that reduced setting the data-only MLP was more accurate in rollout than PIML-soft at 10, 25 and 50 % of the data (test CSO NSE 0.69–0.89 against 0.05–0.60), and it was also better than the fully trained MLP. The rollout failure of the full-budget MLP in Table 4 is therefore partly a consequence of its longer one-step training, which fits one-step targets closely at the expense of stability, and not only of the missing physics terms. The claim that physics terms reduce data needs is not supported by this experiment, and the advantage of PIML-soft in Table 4 should be presented as specific to the training protocol used.

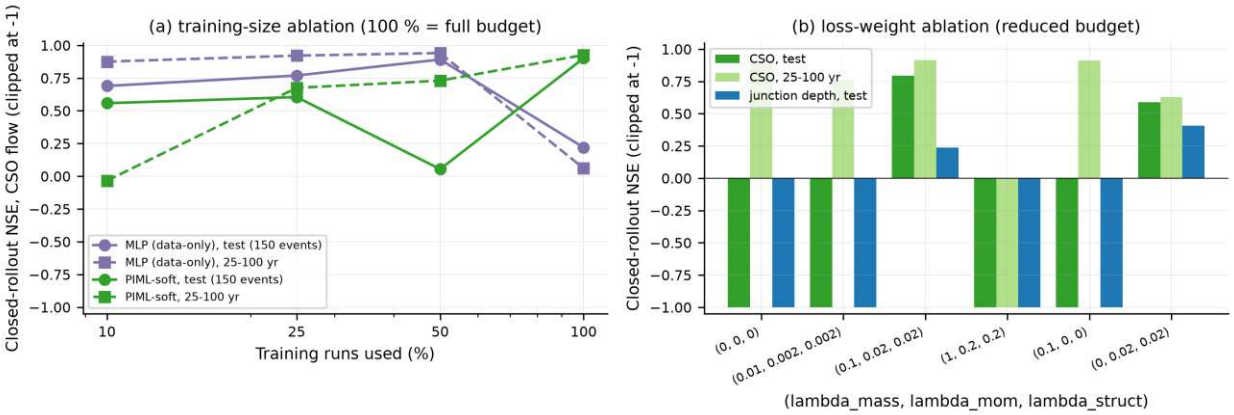


Fig. 12. (a) Closed-loop CSO NSE versus share of training runs, on 150 sampled test events and the 25–100-yr storms (reduced training budget for 10–50 %; full models at 100 %). (b) Loss-weight ablation. Values clipped at −1.

D. Computational cost

The proposed PIML-soft surrogate requires only 0.070 ms per prediction call, approximately 27 times faster than SWMM hydraulic simulation (1.89 ms per 5-min step). The complete MPC optimization takes an average of 7.8 ms, demonstrating that surrogate-assisted prediction and control can operate efficiently within the 300-s control interval while reducing computational demands for repeated hydraulic evaluations.

Table 7. *Latency on one CPU thread.*

Model	ms per call, batch 1	ms per call, batch 64	Parameters
Ridge	0.080	0.118	1,760
MLP (data-only)	0.070	0.155	43,816
PIML-soft	0.070	0.155	43,816
PIML-hard	0.199	0.350	43,816
SWMM 5.2.4, one 5-min step (dynamic wave)	1.89	–	–
MPC solve with PIML-soft (64 × 3 candidates × 6 steps)	7.8 (mean)	10.4 (p95)	–

A single surrogate step took 0.070 ms, about 27 times less than SWMM needs to advance this small network by 5 min (Table 7). A full MPC decision, which evaluates 192 candidate trajectories over six steps, took 7.8 ms on average. Both values are far below the 300-s control interval. The draft compared a 0.28–0.32-ms surrogate call with a 12–18-min SWMM run; those figures refer to different operations and cannot be compared. For a network of this size SWMM itself is fast enough for MPC, so the speed advantage matters mainly for larger networks or for sampling-based optimizer such as the one used here.

E. Closed-loop control under design storms

Closed-loop control performance was evaluated under 15 NOAA Atlas 14 design storms by comparing passive operation, rule-based control (RBC), and surrogate-assisted model predictive control (MPC) using SWMM as the

hydraulic plant. RBC consistently outperformed both MPC configurations, reducing CSO volume by 5–29%, while MPC effectiveness declined under higher-return-period storms because of limited in-line storage and the short prediction horizon.

Table 8. *Event CSO volume (m^3) with SWMM as the plant, and reduction relative to passive operation.*

Atlas 14 storm	Passive	RBC	MPC (PIML-soft)	MPC (MLP)
1-yr 1-h	1049	741 (29%)	809 (23%)	802 (24%)
1-yr 2-h	1146	871 (24%)	895 (22%)	896 (22%)
1-yr 6-h	1332	1020 (23%)	1091 (18%)	1067 (20%)
2-yr 1-h	1554	1198 (23%)	1347 (13%)	1311 (16%)
2-yr 2-h	1745	1339 (23%)	1534 (12%)	1452 (17%)
2-yr 6-h	1982	1714 (14%)	1759 (11%)	1769 (11%)
5-yr 1-h	2311	2040 (12%)	2189 (5%)	2241 (3%)
5-yr 2-h	2650	2239 (16%)	2459 (7%)	2411 (9%)
5-yr 6-h	2945	2544 (14%)	2842 (3%)	2838 (4%)
10-yr 1-h	2888	2691 (7%)	2890 (-0.1%)	2873 (1%)
10-yr 2-h	3369	3064 (9%)	3152 (6%)	3172 (6%)
10-yr 6-h	3727	3265 (12%)	3683 (1%)	3707 (1%)
25-yr 1-h	3636	3422 (6%)	3606 (1%)	3712 (-2%)
25-yr 2-h	4325	4090 (5%)	4153 (4%)	4212 (3%)
25-yr 6-h	4810	4481 (7%)	4550 (5%)	4712 (2%)

RBC reduced CSO volume relative to passive operation for every storm, and it outperformed both MPC variants on all 15 storms (Table 8, Fig. 13). MPC with PIML-soft reduced overflow on most storms but was marginally worse than passive operation for the 10-yr 1-h storm (CSO volume 0.1 % higher), and MPC with the MLP surrogate was worse than passive for the 25-yr 1-h storm (2.1 % higher). Reductions were largest for frequent storms, where unused trunk-sewer volume is available: RBC reduced overflow by 29 % for the 1-yr 1-h storm and MPC with PIML-soft by 23 %. For 10- and 25-yr storms the reductions were 5–12 % for RBC and -0.1 to 6 % for MPC, because the in-line storage is small (the trunk sewers hold a few hundred cubic meters) compared with the event runoff. No strategy eliminated overflows for any storm, in contrast to the 100 % reduction stated in the draft.

Three reasons for the advantage of RBC over MPC are evident from the runs. The 30-min horizon is short compared with the time it takes storage to fill and drain, so MPC releases water that RBC holds. The surrogate's storage prediction is its weakest output (Table 4), and the cost function relies on it. The fail-safe was triggered 5–15 times per event and cut MPC's storage use further. MPC with the PIML-soft and MLP surrogates performed similarly, which shows that the closed-loop setting with re-observation every 5 min hides much of the MLP's open-loop drift. The results agree with comparisons in which well-tuned rules capture most of the benefit of MPC on small systems [11], and they mean that the value of the physics-informed surrogate in this study lies in its open-loop fidelity and robustness, not in a better control outcome.

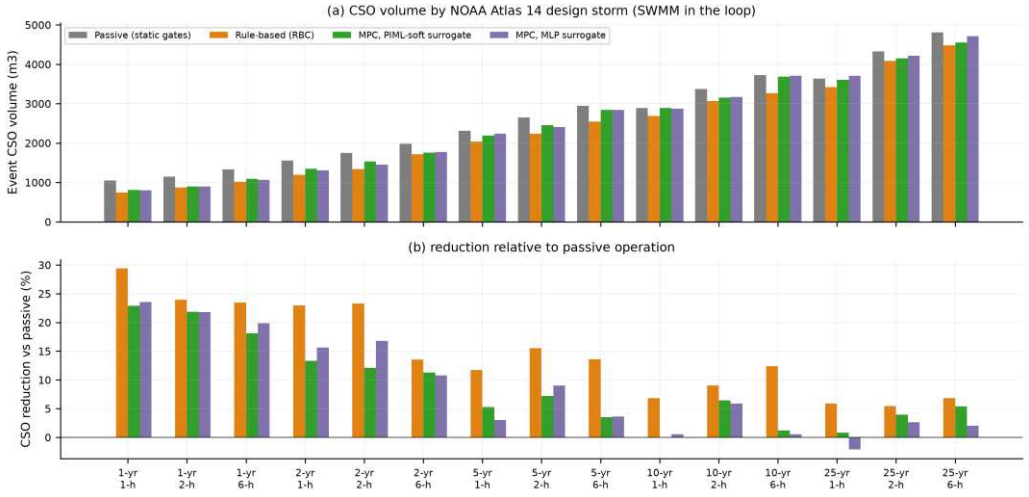


Fig. 13. (a) Event CSO volume for 15 NOAA Atlas 14 design storms under passive operation, RBC and MPC with the PIML-soft and MLP surrogates, and (b) reduction relative to passive operation (simulated, SWMM as plant).

F. Robustness to telemetry noise, forecast error and fail-safe logic

MPC performance was insensitive to the tested uncertainty (Fig. 14a,b). Across all combinations of up to 20 % telemetry noise and 50 % forecast error, the mean CSO volume differed from the noise-free case by -0.5 % to +2.7 %, which is of the same order as the variation between random seeds. Switching off the fail-safe (10 % noise, 25 % forecast error) increased CSO volume slightly (1790 to 1834 m³ for the 2-yr 6-h storm and 3199 to 3253 m³ for the 10-yr 2-h storm) and changed the surface flooding duration by less than 0.2 h (Fig. 14c). The insensitivity reflects the small benefit of control on this network: when the controller's room for improvement is 5–20 %, moderate observation and forecast errors cannot remove much of it. Larger systems with more storage are known to be more exposed to such errors [17].

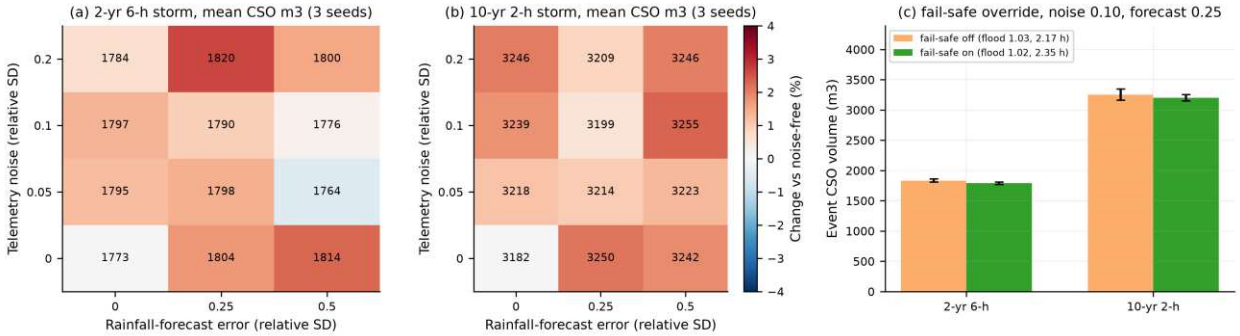


Fig. 14. Mean event CSO volume under MPC (PIML-soft) for combinations of telemetry noise and rainfall-forecast error, (a) 2-yr 6-h and (b) 10-yr 2-h storm, color showing change from the noise-free case; (c) effect of the fail-safe override (3 seeds, error bars ±1 SD).

G. Spatial GI-RTC design

The NSGA-II front contains 40 non-dominated designs (Fig. 15a). It splits into a GI-only branch below $1 M and branches with one, two or more RTC sites, each roughly $0.9 M more expensive because the gate retrofit cost dominates. Within the GI-only branch CSO volume on the two design storms fell from about 2,800 to 1,700 m³ as GI increased. R4 was the most frequently selected RTC site (30 % of front designs), followed by R5 (18 %); R3 was never selected because its weir crest is already high and its trunk sewer offers little storage. GI was allocated mainly to S4 and S5 and never to S3 (Fig. 15b), which is consistent with the larger flows and overflow volumes at R4 and R5. The traffic objective varied little (0.25–0.64 veh-h for the 10-yr 1-h storm) and did not drive the selection.

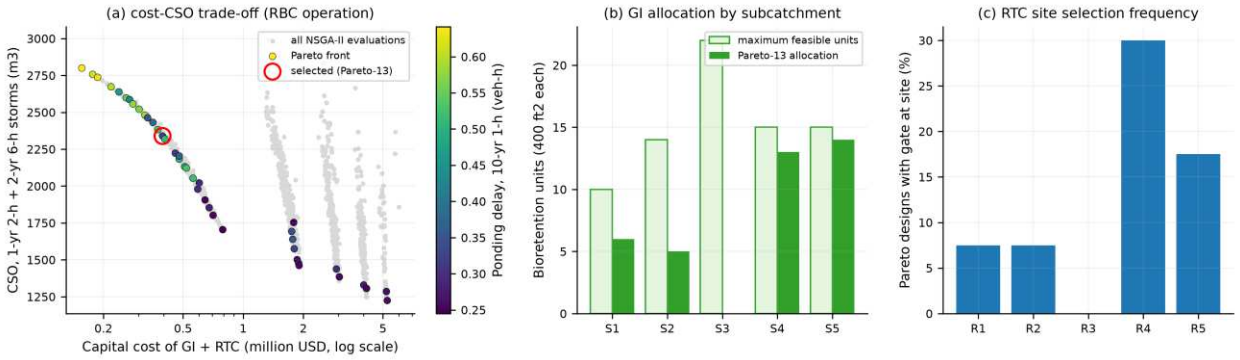


Fig. 15. (a) NSGA-II evaluations and Pareto front (color: ponding delay for the 10-yr 1-h storm), selected design circled; (b) GI allocation of the selected design (Pareto-13) against the feasible maximum; (c) share of Pareto designs that place a gate at each regulator.

H. Capital investment for CSO Control Policy compliance

The capital investment assessment compares GI-led, grey-only, and RTC-based infrastructure configurations to identify the least-cost solution for satisfying the US CSO Control Policy's annual overflow criterion. The analysis integrates residual storage requirements, documented infrastructure unit costs, and Monte Carlo uncertainty assessment to evaluate the economic feasibility of alternative compliance strategies.

Table 9. *Typical-year (2015) performance and least-cost compliance with ≤4 overflow events per year. Hydraulics simulated; costs from documented unit costs (Table 3); capital cost in million USD (2024). Ctrl = gate operation; Capt. = volumetric capture; P(< grey) = Monte Carlo probability of costing less than grey-only.*

Alternative	Ctrl	Events no tank	Capt. (%)	Tank (gal)	Events with tank	GI total ($M)	RTC total ($M)	Tank total ($M)	Total ($M)	P5–P95 ($M)	P(< grey)
Grey-only (passive network)	Passive	13	89.8	56,000	4	0.00	0.00	0.87	0.87	0.56–1.16	0.00
Grey + RTC all sites (RBC)	RBC	2	92.9	0	2	0.00	5.62	0.00	5.62	5.30–17.42	0.00
Pareto-13	Passive	7	89.6	6,000	3	0.40	0.00	0.09	0.49	0.36–0.78	0.93
Grey + RTC (MPC)	MPC	3	93.8	0	3	0.00	5.62	0.00	5.62	5.30–17.42	0.00
Pareto-13 GI + RTC (MPC)	MPC	2	91.8	0	2	0.40	5.62	0.00	6.02	5.77–17.88	

In the typical year the passive network overflowed in 13 events with 7,279 m³ of CSO . Meeting the four-event criterion with storage alone required 56,000 gal of off-line tanks at R1, R2, R4 and R5, costing $0.87 M at modal unit costs (Table 9). The least-cost compliant design was Pareto-13, comprising 38 bioretention units (6, 5, 0, 13 and 14 in S1–S5) without RTC. This configuration initially produced 7 overflow events per year, which were further reduced to 3 after adding 6,000 gal of residual offline storage, thereby satisfying the four-event annual overflow criterion. Its capital cost was $0.49 M, 44 % less than grey-only storage. Under the Monte Carlo cost uncertainty the median saving was 36 % (5th–95th percentile -4 % to 59 %), and the GI design was cheaper than grey storage in 93 % of draws (Fig. 16b). RTC reduced overflow events most effectively: RBC at all five regulators reached two events per year without any tank, and MPC at all sites reached three events with the lowest CSO volume (4,441 m³, capture 93.8 %). At documented costs, however, five gate retrofits cost $5.62 M, more than six times the grey option. The break-even analysis (Fig. 16c) shows that RTC at all sites would match the grey option only if a complete site retrofit cost about $60k, which is roughly one tenth of the lowest documented cost. With the draft's unit costs ($15k per gate, $2.5k per sensor), RTC at all sites would cost $0.13 M; these unit costs are not supported by the cited sources. The draft's statement that RTC enables an 82–92 % capital saving is therefore not reproduced. In this network the saving comes from source control by GI, and RTC is justified only where gate hardware already exists or can be installed at a small fraction of documented costs. This is consistent with previous research examining the cost-effective integration of real-time control with green and grey infrastructure, which highlights the importance of evaluating operational control alongside infrastructure configuration and available storage capacity [10–12].

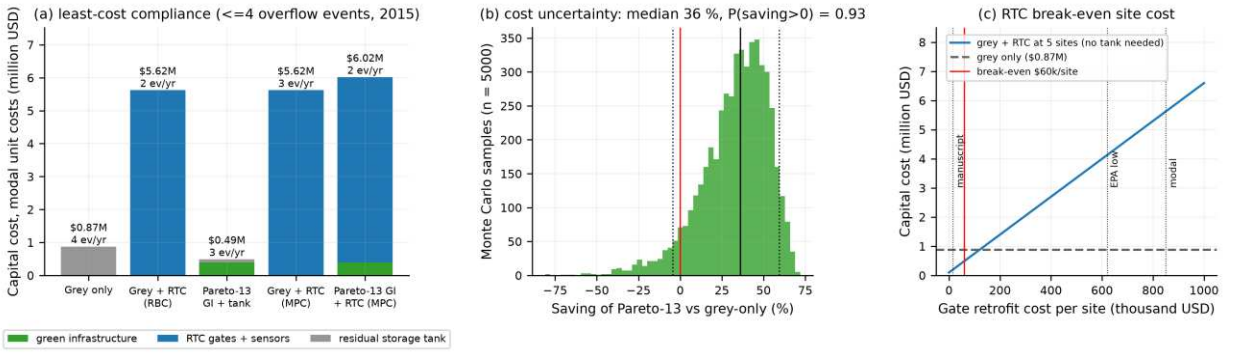


Fig. 16. (a) Capital cost of compliant alternatives at modal unit costs, split into GI, RTC and residual tank, with overflow events per year after sizing; (b) Monte Carlo distribution of the saving of the selected GI design over grey-only storage (5,000 draws; lines at P5, median, P95 and zero); (c) total cost of RTC at five sites versus site cost, with the break-even point.

The absolute costs are small (below $1 M for the compliant options) because the benchmark drains only 11.7 ha. They should not be scaled to city-wide programs by area: storage costs per gallon fall with size, GI siting is limited by available right-of-way, and the cost of RTC depends on the number and condition of existing regulators. Governance structures and institutional arrangements also influence green stormwater infrastructure investment across utilities [20].

I. Traffic disruption

Operational traffic delay from street ponding (assumed traffic data) was zero for storms up to the 2-yr return period and reached about 14 veh-h for the 25-yr 6-h storm (Fig. 17a). The control strategy changed delay by less than 1 veh-h, because surface flooding in this network occurs upstream of the regulators and is governed by the capacity of the collection pipes rather than by gate settings. The draft's reduction from 1,450 to 85 veh-h could not be reproduced and has no basis in observed traffic data. Construction disruption, computed from assumed lane-closure durations, was 483 lane-closure days for the grey option and 288 days for the selected GI design (Fig. 17b). GI is built in many small work zones, so its disruption is spread out rather than avoided; the draft's claim of zero construction traffic losses is not supported.

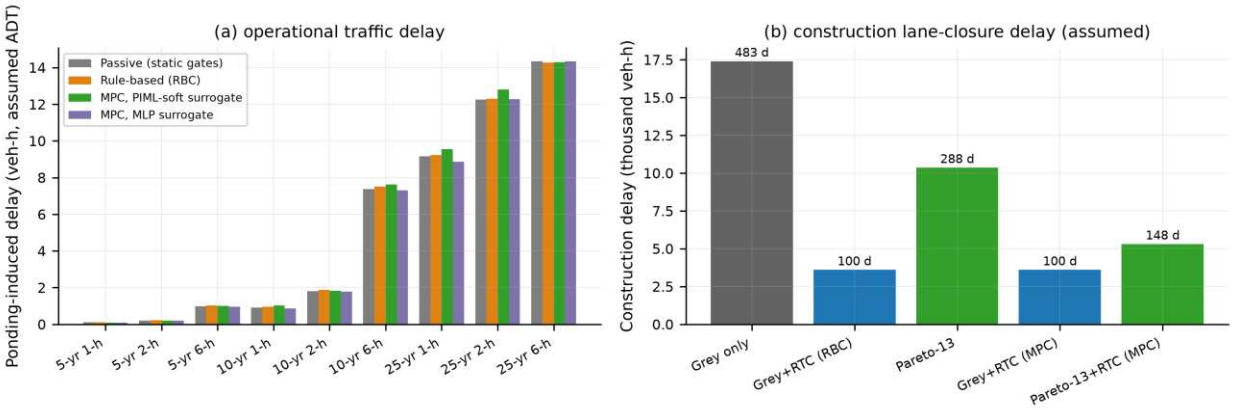


Fig. 17. (a) Ponding-induced traffic delay for 5–25-yr storms under each operating mode; (b) construction lane-closure delay of the compliant alternatives, with lane-closure days.

B. Discussion

The findings demonstrate that the effectiveness of physics-informed machine learning in combined sewer system modelling depends not only on its ability to reproduce individual hydraulic observations but also on its stability during extended simulation and its suitability for control-oriented applications. The comparative evaluation of Ridge regression, the data-only multilayer perceptron (MLP), PIML-soft, and PIML-hard reveals important differences between one-step prediction accuracy, closed-loop rollout performance, and physical consistency. Although the data-only MLP achieved the highest one-step CSO-flow prediction accuracy, with a Nash–Sutcliffe efficiency (NSE) of 0.91, its performance deteriorated substantially during autonomous rollout. Its test-event CSO-flow rollout NSE decreased to −1.04, accompanied by a positive bias of 618%. This deterioration indicates that high one-step prediction accuracy does not necessarily ensure reliable representation of hydraulic behavior when predicted states are repeatedly used as inputs. In contrast, PIML-soft maintained a test-event CSO-flow rollout NSE of 0.87, together with an NSE of 0.97 for wastewater treatment plant (WWTP) inflow and 0.91 for upstream-junction depth. Its CSO-flow rollout NSE further increased to 0.93 under withheld 25–100-year design storms. These results indicate that incorporating appropriately weighted hydraulic relationships can improve surrogate stability under the evaluated full-event and extreme-storm conditions. The comparison between PIML-soft and PIML-hard provides an additional insight into the relationship between physical conservation and hydraulic prediction accuracy. PIML-hard enforced control-volume mass continuity exactly, producing zero continuity residual and eliminating negative predicted flow outputs. However, its storage-state prediction remained substantially inaccurate, with a test-event storage NSE of −20. This result demonstrates that exact mass conservation alone does not guarantee correct hydraulic routing or accurate prediction of the distribution of water within a drainage network. A model may conserve total volume while incorrectly predicting where water accumulates and how quickly it moves through the system. PIML-soft provided a more balanced performance by reducing the median rollout mass residual from approximately 45% for the data-only MLP to approximately 14%, while retaining considerably stronger storage and flow predictions. Nevertheless, its residual remained above the SWMM reference floor, indicating that physical consistency was improved rather than completely achieved. The loss-weight and training-data ablation experiments further demonstrate that the benefits of physics-informed learning are conditional on the formulation and training procedure. The selected PIML-soft weights, (0.1, 0.02, 0.02), produced the best test rollout performance among the reported loss-weight configurations, whereas increasing the physics weights tenfold degraded the prediction performance. This deterioration is associated with the approximate nature of the diffusive-wave momentum and regulators structure relationships embedded in the surrogate. When these approximate relationships dominate the data-fitting objective, the model may be encouraged to reproduce

simplified physical relationships that do not fully correspond to the dynamic-wave behavior represented by SWMM. Furthermore, the reduced-training experiment showed that data-only MLPs trained with a shorter schedule could outperform PIML-soft at certain training-data fractions. Therefore, the results do not establish that physics-informed constraints universally outperform data-driven models or reduce data requirements. Rather, they demonstrate that appropriately calibrated physical constraints can improve closed-loop reliability under the specific training protocol and hydraulic benchmark evaluated in this study. From a computational perspective, the surrogate models substantially reduced the time required for individual prediction steps. The PIML-soft surrogate required approximately 0.070 ms per call, compared with approximately 1.89 ms for a five-minute SWMM simulation step. The complete MPC optimization process required an average of 7.8 ms per control decision, with a 95th-percentile execution time of 10.4 ms. These computational requirements were considerably shorter than the 300-second control interval. The results demonstrate that surrogate-assisted prediction can provide computationally efficient hydraulic state estimation and candidate evaluation. However, the practical significance of this speed advantage should be interpreted in relation to the size of the benchmark network. For the evaluated 11.7-ha system, SWMM itself was sufficiently fast for the tested control configuration. Therefore, the principal computational advantage of surrogate modelling is expected to become more relevant in larger networks or optimization procedures requiring substantially more candidate simulations. An important finding concerns the relationship between surrogate accuracy and operational control performance. Despite the improved closed-loop prediction performance of PIML-soft, rule-based control (RBC) consistently outperformed both tested MPC configurations across all 15 NOAA Atlas 14 design storms. RBC reduced CSO volume by approximately 5-29% relative to passive operation, whereas MPC using PIML-soft produced reductions ranging from a slight increase of 0.1% to approximately 23%. The largest control benefits occurred under relatively frequent rainfall events, when unused in-line storage was available to retain additional flow. Under more severe storms, the limited storage capacity of the benchmark network restricted the potential benefits of both control approaches. The results therefore demonstrate that improved surrogate fidelity does not automatically translate into superior control performance. The operational benefit of predictive control depends on the interaction between the internal model, available storage, control horizon, optimization procedure, and hydraulic characteristics of the drainage network. The weaker performance of MPC can be explained by several characteristics of the evaluated system and controller configuration. The 30-minute prediction horizon was relatively short compared with the time required for the sewer network to fill and drain, limiting the controller's ability to anticipate the longer-term consequences of gate adjustments. In addition, the sampling-based optimization procedure and conservative fail-safe operation restricted the extent to which available storage could be utilized. The surrogate's comparatively weak storage prediction also affected the MPC objective because predicted storage was used to evaluate alternative gate settings. The fail-safe mechanism, which opened gates when upstream-junction depth exceeded the specified threshold, was triggered repeatedly during individual events and further constrained the use of in-line storage. These findings indicate that the benchmark's hydraulic characteristics and the selected controller configuration were important determinants of the observed performance difference between RBC and MPC. Accordingly, the results should not be interpreted as evidence that rule-based control is universally superior to predictive control. They demonstrate that a relatively simple control strategy can be more effective than a computationally sophisticated alternative when the available storage and prediction horizon provide limited opportunities for predictive optimization. The uncertainty experiments indicate that MPC performance was comparatively stable under the tested telemetry and rainfall-forecast perturbations. Across combinations of telemetry noise up to 20% and rainfall-forecast error up to 50%, the mean CSO volume differed from the noise-free baseline by approximately −0.5% to +2.7%. Disabling the fail-safe mechanism under the specified uncertainty condition increased CSO volume from 1,790 to 1,834 m^3 for the two-year, six-hour storm and from 3,199 to 3,253 m^3 for the ten-year, two-hour storm. However, these changes were relatively small compared with the overall hydraulic response of the benchmark. This apparent robustness should be interpreted in the context of the limited control benefit available in the network. Since the controller could achieve only modest improvements under several severe rainfall conditions, moderate telemetry and forecast perturbations had limited scope to eliminate an otherwise substantial control advantage. The results therefore establish scenario-specific robustness within the tested uncertainty ranges rather than demonstrating operational reliability under all real-world telemetry and forecasting conditions. The spatial optimization results provide further evidence that infrastructure location and hydraulic effectiveness are closely connected. The NSGA-II procedure identified 40 non-dominated infrastructure configurations through 1,162 direct SWMM evaluations. The resulting Pareto set contained a GI-dominated group of alternatives below approximately $1 million, alongside configurations involving one or more RTC installations with substantially higher capital requirements. Within the GI-only alternatives, increasing the allocation of bioretention units reduced the combined CSO volume across the selected design storms from approximately 2,800 to 1,700 m^3. The optimization also concentrated GI deployment in sub-catchments S4 and S5, while sub-catchment S3 received no bioretention units in the selected configuration. This spatial distribution corresponds to the hydraulic characteristics of the benchmark,

where greater flows and overflow contributions were associated with the downstream regulators. The results suggest that allocating GI according to network-specific runoff and overflow characteristics can be more effective than distributing interventions uniformly across all sub-catchments. The RTC site-selection results further indicate that the hydraulic value of a control installation depends on its position within the drainage network. Regulator R4 appeared in 30% of the non-dominated designs, followed by R5 in 18%, whereas R3 was not selected. The absence of R3 from the Pareto selections is consistent with its existing weir configuration and comparatively limited available trunk-sewer storage. These results indicate that the potential benefits of RTC are not uniform across regulators. Consequently, infrastructure planning should consider the hydraulic contribution of each candidate control location rather than assuming that installing automated control equipment at every regulator will produce the most economically efficient outcome. The traffic-delay objective varied only between approximately 0.25 and 0.64 vehicle-hours across the Pareto set and had limited influence on the final infrastructure selection. The capital-investment analysis provides the most direct evidence concerning the economic implications of integrated green–grey infrastructure planning. Under the 2015 typical-year rainfall assessment, passive operation produced 13 overflow events and a total CSO volume of 7,279 m^3. Achieving the selected four-event annual overflow criterion through grey-only infrastructure required 56,000 gallons of residual offline storage, with an estimated modal capital cost of $0.87 million. In comparison, the selected Pareto-13 configuration combined 38 bioretention units with only 6,000 gallons of residual storage. This configuration reduced the number of annual overflow events to three and achieved an estimated modal capital cost of $0.49 million. The resulting 44% modal cost reduction indicates that source-control GI can reduce the additional grey storage required to meet the selected compliance criterion within the evaluated benchmark. The Monte Carlo cost assessment provides a more cautious interpretation of this economic advantage. Across 5,000 cost simulations, the Pareto-13 configuration achieved a median saving of 36% relative to grey-only storage, with a fifth-to-ninety-fifth percentile range of -4% to 59%. The GI-led alternative was less expensive than the grey-only alternative in 93% of the simulated cost draws. Although these results support the economic potential of GI-led infrastructure planning, the negative lower percentile indicates that the cost advantage is not guaranteed under all sampled cost conditions. The findings also show that the economic role of RTC differs from its hydraulic effectiveness. RTC installations reduced overflow events more substantially under some operating configurations; however, retrofitting all five regulators resulted in an estimated capital cost of $5.62 million. The break-even analysis indicated that the complete RTC retrofit cost would need to decrease to approximately $60,000 per site to become economically comparable with the grey-only alternative under the study's assumptions. Therefore, the economic advantage identified in this benchmark is primarily associated with distributed source-control infrastructure and limited residual storage, rather than the installation of new RTC hardware at every regulator. The traffic-disruption assessment provides an additional planning indicator but does not change the principal hydraulic and economic conclusions. Operational traffic delay associated with surface ponding was estimated to be zero for storms up to the two-year return period and approximately 14 vehicle-hours for the 25-year, six-hour storm. The differences between control strategies were less than one vehicle-hour because ponding in the benchmark was primarily associated with upstream collection-pipe capacity rather than regulator operation. Construction-related disruption was estimated at 483 lane-closure days for the grey-only alternative and 288 days for Pareto-13. However, these estimates depend on assumed road classifications, traffic volumes, construction durations, and work-zone characteristics. They should therefore be interpreted as illustrative planning quantities rather than measured traffic benefits.

Overall, the findings demonstrate that integrating PIML, active control, and spatial GI optimization provides a useful computational approach for evaluating alternative CSO-management strategies. However, the results also reveal that the individual components contribute differently to the final engineering outcome. PIML-soft primarily improves surrogate reliability during closed-loop prediction and extreme-event extrapolation; RBC provides the strongest control performance in the evaluated benchmark; and spatially optimized GI with limited residual storage provides the lowest-cost compliance configuration. The central engineering implication is that the value of an integrated framework lies not in assuming that every advanced technology must be deployed simultaneously, but in identifying which combination of prediction, control, and infrastructure interventions provides the most appropriate balance between hydraulic performance and capital investment.

C. Limitations and Future Work

The findings of this study should be interpreted in relation to several methodological, hydraulic, operational, and economic limitations. The first limitation concerns the use of a small public benchmark rather than a calibrated municipal combined sewer catchment. The evaluated EPA SWMM Example 8 network covers approximately 11.7 ha and contains five combined-sewer regulators. Its hydraulic behavior, available in-line storage, regulator configuration, and infrastructure cost requirements may differ considerably from those of larger municipal networks. All hydraulic

performance results were generated through SWMM simulations or surrogate predictions, and no field measurements, utility telemetry records, or observed CSO event data were available for independent calibration and validation. Consequently, the reported hydraulic and economic outcomes demonstrate the behavior of the selected computational benchmark rather than verified performance in an operational US sewer system. The second limitation concerns the scope and physical representation of the surrogate model. The developed surrogate represents the regulator–interceptor subsystem, while lateral runoff inflows are supplied directly by SWMM. Therefore, the model does not independently learn runoff generation, surface infiltration, or the complete hydrological and hydraulic behavior of the catchment. Furthermore, the physics-informed formulation uses control-volume continuity, diffusive-wave momentum relationships, and simplified regulator structure equations rather than a complete representation of the dynamic-wave equations implemented in SWMM. The substantial storage-state drift observed in PIML-hard also demonstrates that exact mass conservation is insufficient to ensure accurate hydraulic-state prediction. Although PIML-soft achieved more balanced closed-loop performance, its results remain dependent on the selected physical constraints, training protocol, and loss-weight configuration. A further limitation concerns the temporal coverage of the compliance assessment. The infrastructure alternatives were evaluated using a single representative rainfall year, 2015, selected from the historical rainfall record. Although this year was chosen because its rainfall depth and event count were close to the long-term averages, a single-year assessment cannot fully represent the variability of long-term rainfall sequences or the frequency of rare hydrological conditions. Continuous multi-year simulation is therefore necessary to establish whether the selected GI–grey infrastructure configuration can maintain the specified overflow-frequency criterion under a broader range of rainfall conditions. The operational evaluation also contains forecast-related limitations. The MPC forecast baseline was generated using a passive SWMM simulation of the same rainfall event and therefore represented a perfect runoff forecast before controlled perturbations were introduced. This approach allowed the effects of telemetry noise and forecast error to be examined systematically, but it does not reproduce the full uncertainty associated with operational weather forecasts, radar nowcasting, rainfall-runoff prediction, or changing catchment conditions. The tested telemetry and forecast-error ranges are scenario assumptions, while the MPC objective weights were specified by the authors rather than derived through expert elicitation. Accordingly, the control results should be interpreted as controlled simulation experiments and not as evidence of performance under a fully operational forecasting and telemetry environment. The economic analysis is subject to additional limitations because the capital-cost model excludes several components of actual infrastructure expenditure. Operation and maintenance costs, land acquisition, financing, equipment replacement, and life-cycle discounting were not included in the cost comparison. The cost of RTC retrofits is particularly sensitive to the condition of existing gate hardware, actuator requirements, communication systems, and installation conditions. Therefore, the reported break-even RTC installation cost and relative economic ranking may change when utility-specific asset conditions and life-cycle expenditure are considered. Similarly, the traffic-disruption assessment relies on assumed road classes, average daily traffic, free-flow speeds, detour lengths, and construction durations. Neither operational ponding delay nor construction-related disruption was calibrated against observed traffic data. These estimates are consequently illustrative planning indicators rather than measured co-benefits. Future research should first transfer the proposed framework to a named municipal combined sewer catchment using calibrated SWMM parameters, continuous multi-year rainfall records, observed flow and water-level telemetry, and verified CSO event records. This would allow the proposed surrogate models, control strategies, and infrastructure configurations to be evaluated against actual drainage-system behavior. Field-informed calibration and independent validation would also help determine whether the performance differences observed in the public benchmark remain consistent under more complex network configurations and operational conditions. Second, the surrogate modelling framework should be extended beyond the regulator–interceptor subsystem to represent runoff generation and the full coupled drainage network. Graph-based neural architectures, hybrid state-space models, and more comprehensive hydraulic constraints could be investigated to capture the spatial relationships between sub-catchments, conduits, regulators, storage structures, and downstream treatment facilities. Further research should also investigate how physical constraints can be formulated to remain consistent with the dynamic-wave hydraulic model while avoiding the storage-state errors observed in the hard-constrained formulation. Such developments would provide a stronger foundation for using surrogate models in large-scale urban drainage applications. Third, the MPC formulation should be improved by investigating longer prediction horizons, terminal storage penalties, alternative optimization algorithms, and explicit actuator constraints. Operational rainfall forecasts and radar-based nowcasts should replace the same-event passive SWMM forecast baseline to provide a more realistic assessment of predictive control under uncertain rainfall conditions. Further experiments should examine whether gradient-based optimization through differentiable surrogate models can improve control decisions and whether more advanced telemetry filtering and state-estimation approaches can improve the reliability of active control. These investigations would help determine whether MPC can provide measurable benefits over RBC in networks with greater available storage and longer hydraulic travel times. Fourth, future infrastructure optimization

should incorporate utility-specific asset inventories, land availability, construction constraints, and complete life-cycle cost information. The economic comparison should include operation and maintenance, land acquisition, financing, equipment replacement, and long-term discounting. Traffic-disruption estimates should also be developed using measured traffic volumes, verified road-network information, and realistic construction schedules. These improvements would allow the optimization framework to evaluate infrastructure alternatives using more representative financial and operational conditions. Finally, independent replication across multiple combined sewer catchments is necessary before generalizing the reported cost savings or identifying the framework as a transferable city-scale modernization strategy. Different catchments may exhibit substantially different regulator configurations, in-line storage capacities, rainfall characteristics, GI installation opportunities, and RTC retrofit requirements. Evaluating the framework across several municipal systems would establish the extent to which its predictive, operational, and economic findings remain consistent under different hydraulic and institutional conditions. Field demonstrations would provide the additional evidence required to assess the practical feasibility and long-term effectiveness of the proposed approach.

V. Conclusion

This study developed and evaluated a reproducible computational framework integrating physics-informed machine learning, real-time control, spatial green-infrastructure allocation, and capital-cost optimization for combined sewer overflow management. The framework connected rainfall-driven EPA SWMM simulations, hydraulic surrogate development, closed-loop control evaluation, multi-objective infrastructure optimization, and compliance-oriented residual storage sizing. A central methodological feature was the separation of the surrogate model's role in computational prediction from SWMM's role as the hydraulic evaluation model, ensuring that the final infrastructure and control outcomes were assessed through direct hydraulic simulations. The surrogate evaluation demonstrated that one-step prediction accuracy alone was insufficient to establish reliability during extended hydraulic simulation. Although the data-only MLP achieved the highest one-step CSO-flow NSE of 0.91, its test-event closed-loop rollout NSE deteriorated to −1.04, with a positive bias of 618%. PIML-soft achieved a more balanced performance, maintaining a test-event CSO-flow rollout NSE of 0.87 and an NSE of 0.93 under withheld 25–100-year design storms. PIML-hard enforced mass continuity exactly but experienced substantial storage-state drift, demonstrating that conservation alone does not guarantee accurate hydraulic routing. These findings establish the importance of evaluating physics-informed surrogate models through closed-loop performance, physical consistency, and extreme-event extrapolation rather than relying exclusively on one-step prediction metrics. The control experiments demonstrated that RBC outperformed the tested MPC configurations across all 15 NOAA Atlas 14 design storms. RBC achieved CSO-volume reductions of approximately 5–29% relative to passive operation, whereas MPC using PIML-soft produced reductions ranging from −0.1% to 23%. The limited in-line storage, short prediction horizon, storage-prediction uncertainty, and conservative fail-safe operation constrained the benefits of predictive control in the evaluated network. Although the surrogate-assisted MPC demonstrated computationally efficient execution and comparatively stable performance under the tested telemetry and forecast-error scenarios, improved surrogate prediction did not translate into superior control performance in this particular benchmark. The spatial optimization and capital-cost analysis identified a GI-led infrastructure configuration as the least-cost pathway to the selected annual CSO-control criterion. The selected Pareto-13 design allocated 38 bioretention units across five sub-catchments and required only 6,000 gallons of residual offline storage. Its estimated modal capital cost was $0.49 million, compared with $0.87 million for the grey-only alternative requiring 56,000 gallons of storage. The GI-led configuration therefore achieved a 44% reduction in modal capital investment, while Monte Carlo analysis indicated a median saving of 36%, with a fifth-to-ninety-fifth percentile range of −4% to 59%. The design was less expensive than the grey-only alternative in 93% of the simulated cost draws. In contrast, installing RTC at all five regulators resulted in an estimated capital cost of $5.62 million, indicating that new RTC hardware was not economically competitive under the documented retrofit-cost assumptions used in this benchmark. The principal conclusion is that, within the evaluated 11.7-ha public benchmark, spatially distributed green infrastructure combined with limited residual grey storage provided a more cost-effective compliance pathway than either grey-only storage or comprehensive RTC retrofitting. The results also demonstrate that physics-informed surrogate modelling, predictive control, and infrastructure optimization should be evaluated according to their individual contributions rather than assuming that their simultaneous deployment necessarily produces the best outcome. The framework provides a reproducible computational basis for investigating these interactions and identifying infrastructure alternatives that balance hydraulic performance, operational requirements, and capital investment. However, the findings remain conditional on the benchmark network, simulated hydraulic conditions, documented unit costs, and assumed operational and traffic parameters. They do not establish field performance, universal cost savings, or city-wide

applicability across US combined sewer systems. Practical implementation requires calibration against utility-specific hydraulic data, operational telemetry, multi-year rainfall records, verified project costs, and independent field validation. Extending the framework to multiple municipal catchments and more comprehensive hydraulic and economic representations is therefore the necessary next step toward evaluating its suitability for real-world combined sewer modernization.

References

[1] G. Gironás, L. A. Roesner, L. A. Rossman, and J. Davis, "A new applications manual for the Storm Water Management Model (SWMM)," Environmental Modelling & Software, vol. 25, pp. 813–814, 2010. https://doi.org/10.1016/j.envsoft.2009.11.009

[2] S. P. Rimer, A. Mullapudi, S. C. Troutman, G. Ewing, B. D. Bowes, A. A. Akin, J. Sadler, R. Kertesz, B. McDonnell, L. Montestruque, J. Hathaway, J. L. Goodall, J. Norton, and B. Kerkez, "pystorms: A simulation sandbox for the development and evaluation of stormwater control algorithms," Environmental Modelling & Software, vol. 162, Art. no. 105635, 2023. https://doi.org/10.1016/j.envsoft.2023.105635

[3] B. E. McDonnell, K. Ratliff, M. E. Tryby, J. J. X. Wu, and A. Mullapudi, "PySWMM: The Python interface to Stormwater Management Model (SWMM)," Journal of Open Source Software, vol. 5, no. 52, Art. no. 2292, 2020. https://doi.org/10.21105/joss.02292

[4] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, vol. 378, pp. 686–707, 2019. https://doi.org/10.1016/j.jcp.2018.10.045

[5] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nature Reviews Physics, vol. 3, pp. 422–440, 2021. https://doi.org/10.1038/s42254-021-00314-5

[6] R. Palmitessa, M. Grum, A. P. Engsig-Karup, and R. Löwe, "Accelerating hydrodynamic simulations of urban drainage systems with physics-guided machine learning," Water Research, vol. 223, Art. no. 118972, 2022. https://doi.org/10.1016/j.watres.2022.118972

[7] A. Garzón, Z. Kapelan, J. Langeveld, and R. Taormina, "Machine learning-based surrogate modeling for urban water networks: Review and future research directions," Water Resources Research, vol. 58, no. 5, Art. no. e2021WR031808, 2022. https://doi.org/10.1029/2021WR031808

[8] Z. Zhang, W. Tian, C. Lu, Z. Liao, and Z. Yuan, "Graph neural network-based surrogate modelling for real-time hydraulic prediction of urban drainage networks," Water Research, vol. 263, Art. no. 122142, 2024. https://doi.org/10.1016/j.watres.2024.122142

[9] X. Luo, P. Liu, Q. Xia, Q. Cheng, W. Liu, Y. Mai, C. Zhou, Y. Zheng, and D. Wang, "Machine learning-based surrogate model assisting stochastic model predictive control of urban drainage systems," Journal of Environmental Management, vol. 346, Art. no. 118974, 2023. https://doi.org/10.1016/j.jenvman.2023.118974

[10] M.-È. Jean, C. Morin, S. Duchesne, G. Pelletier, and M. Pleau, "Optimization of real-time control with green and gray infrastructure design for a cost-effective mitigation of combined sewer overflows," Water Resources Research, vol. 57, no. 12, Art. no. e2021WR030282, 2021. https://doi.org/10.1029/2021WR030282

[11] M.-È. Jean, C. Morin, S. Duchesne, G. Pelletier, and M. Pleau, "Real-time model predictive and rule-based control with green infrastructures to reduce combined sewer overflows," Water Research, vol. 221, Art. no. 118753, 2022. https://doi.org/10.1016/j.watres.2022.118753

[12] M.-È. Jean, C. Morin, J. E. Ossa Ossa, S. Duchesne, G. Pelletier, and M. Pleau, "Optimal distribution of green and grey infrastructures coupled with real time control of the sewer for combined sewer overflows control as an adaptation measure to climate change," Urban Water Journal, vol. 21, no. 4, pp. 419–435, 2024. https://doi.org/10.1080/1573062X.2024.2312497

[13] N. S. V. Lund, A. K. V. Falk, M. Borup, H. Madsen, and P. S. Mikkelsen, "Model predictive control of urban drainage systems: A review and perspective towards smart real-time water management," Critical Reviews in Environmental Science and Technology, vol. 48, no. 3, pp. 279–339, 2018. https://doi.org/10.1080/10643389.2018.1455484

[14] J. A. van der Werf, Z. Kapelan, and J. Langeveld, "Real-time control of combined sewer systems: Risks associated with uncertainties," Journal of Hydrology, vol. 617, Art. no. 128900, 2023. https://doi.org/10.1016/j.jhydrol.2022.128900

[15] L. García, J. Barreiro-Gomez, E. Escobar, D. Téllez, N. Quijano, and C. Ocampo-Martinez, "Modeling and real-time control of urban drainage systems: A review," Advances in Water Resources, vol. 85, pp. 120–132, 2015. https://doi.org/10.1016/j.advwatres.2015.08.007

[16] X. Fu, H. Goddard, X. Wang, and M. E. Hopton, "Development of a scenario-based stormwater management planning support system for reducing combined sewer overflows (CSOs)," Journal of Environmental Management, vol. 236, pp. 571–580, 2019. https://doi.org/10.1016/j.jenvman.2018.12.089

[17] P. Joshi, J. P. Leitão, M. Maurer, and P. M. Bach, "Not all SuDS are created equal: Impact of different approaches on combined sewer overflows," Water Research, vol. 191, Art. no. 116780, 2021. https://doi.org/10.1016/j.watres.2020.116780

[18] X. Dong, H. Guo, and S. Zeng, "Enhancing future resilience in urban drainage system: Green versus grey infrastructure," Water Research, vol. 124, pp. 280–289, 2017. https://doi.org/10.1016/j.watres.2017.07.038

[19] A. Botturi, E. G. Ozbayram, K. Tondera, N. I. Gilbert, P. Rouault, N. Caradot, O. Gutierrez, S. Daneshgar, N. Frison, C. Akyol, A. Foglia, A. L. Eusebi, and F. Fatone, "Combined sewer overflows: A critical review on best practice and innovative solutions to mitigate impacts on environment and human health," Critical Reviews in Environmental Science and Technology, vol. 51, no. 15, pp. 1585–1618, 2021. https://doi.org/10.1080/10643389.2020.1757957

[20] K. G. Hopkins, N. B. Grimm, and A. M. York, "Influence of governance structure on green stormwater infrastructure investment," Environmental Science & Policy, vol. 84, pp. 124–133, 2018. https://doi.org/10.1016/j.envsci.2018.03.008

[21] M. Pregnolato, A. Ford, S. M. Wilkinson, and R. J. Dawson, "The impact of flooding on road transport: A depth-disruption function," Transportation Research Part D: Transport and Environment, vol. 55, pp. 67–81, 2017. https://doi.org/10.1016/j.trd.2017.06.020

[22] J. P. Cohen, R. Field, A. N. Tafuri, and M. A. Ports, "Cost comparison of conventional gray combined sewer overflow control infrastructure versus a green/gray combination," Journal of Irrigation and Drainage Engineering, vol. 138, no. 6, pp. 534–540, 2012. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000432

[23] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182–197, 2002. https://doi.org/10.1109/4235.996017

[24] J. Blank and K. Deb, "pymoo: Multi-objective optimization in Python," IEEE Access, vol. 8, pp. 89497–89509, 2020. https://doi.org/10.1109/ACCESS.2020.2990567

[25] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," Annals of Operations Research, vol. 134, no. 1, pp. 19–67, 2005. https://doi.org/10.1007/s10479-005-5724-z

[26] Y. Abou Rjeily, O. Abbas, M. Sadek, I. Shahrour, and F. Hage Chehade, “Model predictive control for optimising the operation of urban drainage systems,” Journal of Hydrology, vol. 566, pp. 558–565, 2018. https://doi.org/10.1016/j.jhydrol.2018.09.044
[27] J. L. Svensen, C. Sun, G. Cembrano, and V. Puig, “Model predictive control of urban drainage systems considering uncertainty,” IEEE Transactions on Control Systems Technology, vol. 31, no. 6, pp. 2968–2975, 2023. https://doi.org/10.1109/TCST.2023.3286648
[28] G. Dirckx, M. Schütze, S. Kroll, C. H. Thoeye, G. De Gueldre, and B. Van De Steene, “Cost-efficiency of RTC for CSO impact mitigation,” Urban Water Journal, vol. 8, no. 6, pp. 367–377, 2011. https://doi.org/10.1080/1573062X.2011.630092
[29] A. Deletic, C. B. S. Dotto, D. T. McCarthy, M. Kleidorfer, G. Freni, G. Mannina, M. Uhl, M. Henrichs, T. D. Fletcher, W. Rauch, J.-L. Bertrand-Krajewski, and S. Tait, “Assessing uncertainties in urban drainage models,” Physics and Chemistry of the Earth, Parts A/B/C, vol. 42–44, pp. 3–10, 2012. https://doi.org/10.1016/j.pce.2011.04.007
[30] M. Eulogi, S. Ostojin, P. Skipworth, J. D. Shucksmith, and A. Schellart, “Hydraulic optimisation of multiple flow control locations for the design of local real time control systems,” Urban Water Journal, vol. 18, no. 2, pp. 91–100, 2021. https://doi.org/10.1080/1573062X.2020.1860238