water

MDPI

*Article*

# Genetic-Algorithm Optimization of PRV Placement for DeePC-Based Pressure Control in Water Distribution Networks

**Jason Davda and Avi Ostfeld ***

Faculty of Civil and Environmental Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel; jason.davda@campus.technion.ac.il
* Correspondence: ostfeld@cv.technion.ac.il

**Abstract**

Pressure management in water distribution systems depends not only on valve operation but also on the placement and number of pressure-reducing valves (PRVs). This study develops a controller-in-the-loop framework coupling a Genetic Algorithm (GA) with Data-Enabled Predictive Control (DeePC) to optimize internal PRV placement according to closed-loop performance. The GA searches the feasible space of valve-location configurations, and each candidate is evaluated through hydraulic simulation and closed-loop DeePC control according to four performance criteria: tracking accuracy, spatial pressure variability, pressure-bound violations, and valve actuation effort. The framework was tested on the Fossolo and Modena networks. In Fossolo, the GA identified a three-PRV configuration that achieved both the lowest mean absolute error and the best combined-objective score. In Modena, configurations with two to six added internal PRVs were compared using a consistently normalized global objective. The two-PRV configuration achieved the best overall score of 0.547, although the lowest tracking error and spatial variability were obtained with four and five PRVs, respectively. These results show that adding more controllable valves does not necessarily improve overall performance and that PRV placement should be assessed jointly with controller behavior and multiple operational criteria. The proposed GA-DeePC framework demonstrates a proof-of-concept controller-in-the-loop approach for integrating PRV placement with closed-loop pressure-control performance.

**Keywords:** water distribution networks; pressure management; pressure-reducing valves; Data-Enabled Predictive Control; genetic algorithm; valve placement; design-for-control

## 1. Introduction

Water distribution systems (WDSs) are critical urban infrastructures whose operation must provide adequate water quantity, pressure, and quality under continuously changing hydraulic conditions. Their management is complicated by aging assets, spatially distributed demands, topographic variability, leakage, and the interaction among pipes, pumps, tanks, and control valves. Among the available leakage-control strategies, pressure management is widely regarded as one of the most effective approaches because excessive pressure contributes to background leakage, pipe bursts, water loss, and deterioration of network assets [1–3]. Pressure management must nevertheless maintain sufficient service pressure throughout the network, making it an operational balancing problem rather than a simple minimization of pressure. Although pressure management can also affect leakage and water losses, the present study focuses specifically on PRV placement for closed-loop



Academic Editors: Matteo Nicolini and Daniele Goi

Received: 3 August 2026
Revised: 29 August 2026
Accepted: 1 September 2026
Published: 3 September 2026

pressure-control performance; leakage reduction is not explicitly modeled or evaluated as an objective.

Pressure-reducing valves (PRVs) are commonly used to dissipate excess hydraulic head and regulate pressures in selected network areas. Their effectiveness depends on both their operating settings and their physical locations. A valve installed on one pipe can alter downstream flows and pressures over a substantial part of the network, whereas an unsuitable placement may provide limited pressure-control authority or create undesirable hydraulic conditions [3–5]. Jointly determining PRV number, placement, and operating settings therefore leads to a nonlinear, non-convex, combinatorial optimization problem that couples discrete location decisions with continuous control variables and hydraulic constraints [4–6]. A recent review of metaheuristic approaches for WDS pressure management shows that PRV location and setting have been addressed using a broad range of methods, including genetic algorithms, particle swarm optimization, harmony search, multi-objective evolutionary algorithms, and hybrid stochastic-deterministic approaches [7]. Other studies have employed graph-based screening, clustering, and mathematical optimization, illustrating the diversity of available solution strategies and the absence of a single universally adopted approach [3–6]. Across these formulations, the hydraulic benefit of added internal PRVs depends strongly on their locations and settings and cannot be inferred from valve count alone. The increasing availability of network sensors, remotely operated valves, and communication infrastructure has enabled a transition from fixed or locally modulated valve settings toward real-time control (RTC). In RTC, valve actions are updated using measured network conditions to maintain pressures near selected targets despite variations in demand and source conditions. Recent optimization-based studies have similarly focused on adapting PRV settings to changing operating conditions. Dai developed a sequential-convex-programming framework for near-real-time computation of PRV pressure settings in response to changing demand patterns, demonstrating the continuing transition from fixed valve settings toward adaptive pressure regulation [8]. However, WDS pressure control is complicated by nonlinear behavior, operating-point dependence, time delays, and uncertain system characteristics, all of which may affect control performance, robustness, and closed-loop stability [9–11]. Model Predictive Control (MPC) provides a systematic framework for constrained, multivariable pressure regulation by repeatedly predicting future behavior and optimizing control actions over a receding horizon. Its implementation, however, generally depends on a sufficiently accurate and computationally tractable model, whose identification, calibration, and maintenance may be difficult for large and uncertain WDSs [10–12].

Data-Enabled Predictive Control (DeePC) offers a direct data-driven alternative. Instead of first identifying an explicit parametric model, DeePC uses previously measured input-output trajectories to represent and predict system behavior. The method is founded on the behavioral framework and the Fundamental Lemma, according to which persistently exciting data from a controllable linear time-invariant system can span its admissible finite-length trajectories [13,14]. These trajectories are organized in Hankel matrices and incorporated directly into a receding-horizon optimization problem. For deterministic linear time-invariant systems, DeePC has been shown to be equivalent to classical MPC, while regularized and robust formulations have been developed to address noisy data, uncertainty, and model mismatch [13,15,16]. DeePC has also been demonstrated experimentally on nonlinear systems and has recently been applied to pressure management and chlorine scheduling in WDSs, where it avoids the need for an explicitly calibrated dynamic model of the controlled network [12,17].

Although DeePC determines how available actuators should be operated, it does not determine where those actuators should be installed. This distinction is particularly important for pressure control because the attainable closed-loop response depends on the hydraulic influence of the selected PRV locations. The preceding DeePC sensitivity study by Davda and Ostfeld showed that adding an internal PRV could either improve or degrade pressure-tracking performance depending on its location. In Fossolo, some locations reduced the mean absolute error relative to the inlet-only configuration, whereas poorly selected locations produced substantially worse performance. In the larger Modena network, the best individual location improved tracking only moderately, and pipes directly adjacent to the monitored node did not necessarily provide the best result [18]. These findings indicate that simple proximity-based rules are insufficient and motivate explicit optimization of PRV placement together with spatial performance assessment beyond a single controlled node [18].

A conventional hydraulic simulation can identify locations or zones experiencing excessive pressure, but this information alone does not determine the PRV locations that provide the most favorable closed-loop DeePC control performance. In a looped WDS, installing a PRV modifies the hydraulic response beyond its immediate location, and a high-pressure pipe or a pipe close to the monitored node does not necessarily provide the greatest control authority over that node. The optimization is therefore used to search for PRV configurations according to the actual closed-loop performance achieved by DeePC, while also considering spatial pressure behavior at selected high-elevation nodes.

Genetic Algorithms (GAs) provide a suitable search mechanism for such placement problems because they can represent discrete combinations of network components and evaluate them without requiring differentiable objective functions. GAs and related evolutionary algorithms have been widely used in WDS design, rehabilitation, pressure management, and multi-objective optimization, particularly where hydraulic simulation is required to evaluate each candidate solution [7,19]. Their population-based operators allow exploration of large and non-convex search spaces, while crossover, mutation, and selection progressively concentrate the search around promising regions. Nevertheless, evolutionary methods do not generally guarantee the global optimum, and their computational efficiency may deteriorate as the network size, number of decision variables, and cost of fitness evaluation increase [19,20]. Consequently, the practical objective is commonly to identify high-quality or near-optimal solutions within an available computational budget rather than to prove global optimality. The GA was deliberately selected in this study as a transparent and computationally straightforward baseline evolutionary optimizer, rather than with the intention of claiming superiority over more recent optimization algorithms. The primary objective was to examine whether a discrete outer-search layer could be effectively coupled with DeePC when each fitness evaluation requires a hydraulic simulation, a system-identification phase, and a closed-loop control experiment. In this setting, GA provides several practical advantages: it naturally handles discrete combinatorial placement decisions, does not require gradient information or an explicit analytical relationship between PRV location and closed-loop performance, and can use the resulting DeePC response directly as a black-box fitness measure [19,20]. Although more advanced evolutionary and hybrid variants may improve search efficiency, diversity preservation, or exploration of large search spaces, they do not generally provide a practical guarantee of identifying the global optimum within a finite computational budget for a black-box combinatorial problem of this type. The standard GA was therefore adopted intentionally to keep the outer optimization layer simple and interpretable while first examining the feasibility of the controller-in-the-loop concept. Having demonstrated this feasibility, the same framework can subsequently be extended to more advanced evolutionary, multi-objective, surrogate-assisted, or hybrid optimization methods [21–23].

These considerations are closely related to the emerging concept of design-for-control, in which infrastructure decisions and operational control are evaluated jointly rather than sequentially. Sequential infrastructure design followed by controller tuning can introduce unnecessary restrictions because the achievable operating performance is determined by the physical configuration made available to the controller [24]. In WDS applications, design-for-control formulations have therefore jointly considered discrete network modifications and continuous valve-control settings [21,22]. Recent studies have further extended this concept to explicitly multi-objective formulations. Jenks et al. jointly optimized the locations and operational settings of pressure-control and automatic flushing valves within a bi-objective framework integrating pressure management and self-cleaning performance [25]. Ulusoy and Stoianov subsequently formulated a multi-objective design-for-control problem in which valve and pipe installation and valve-control settings were optimized jointly with respect to pressure-induced leakage, network resilience, and design cost [26]. These developments demonstrate the value of considering infrastructure design; operational evaluations rely on hydraulic or mathematical optimization rather than on a data-driven closed-loop controller.

Within the literature reviewed here, PRV placement and operation have been optimized for objectives such as leakage, excess pressure, resilience, water age, installation cost, and spatial pressure variability [3–6,21,22], while DeePC has been investigated as a data-driven controller for WDS operation [12,18]. However, the reviewed studies do not integrate these two directions by using closed-loop DeePC performance itself to guide an evolutionary search for internal PRV locations. The resulting research gap is therefore not merely the optimization of valve placement or the application of DeePC in isolation, but the development of a controller-in-the-loop framework in which physical placement decisions are evaluated according to the pressure-control behavior that the data-driven controller can actually achieve.

Accordingly, this study develops a coupled GA-DeePC framework for optimizing the placement of internal PRVs in WDSs. The GA acts as an outer design layer that generates candidate valve-location configurations, whereas DeePC acts as the inner operational layer that performs a closed-loop pressure-control experiment for each candidate. The study first uses the smaller Fossolo network to validate the coupled search and to examine the relationship between monitored-node tracking and broader spatial and operational metrics. It then applies the framework to the larger Modena network for different numbers of added internal PRVs. Candidate configurations are assessed using monitored-node mean absolute error, spatial pressure variability, pressure-bound violations, and normalized valve actuation effort. A consistently normalized global objective is subsequently used to compare the leading solutions obtained for different valve counts.

## 2. Methodology

This study integrates an external Genetic Algorithm (GA) with Data-Enabled Predictive Control (DeePC) to optimize the placement of internal pressure-reducing valves (PRVs) in water distribution networks. The GA was used to search over candidate PRV-location configurations, while DeePC was used to evaluate the closed-loop pressure-control performance of each candidate configuration. This outer-search/inner-evaluation structure is consistent with the broader design-for-control concept, in which discrete infrastructure decisions are assessed together with the operational performance achievable under a corresponding control strategy [21,22]. The proposed framework, which extends the placement-sensitivity analysis reported in the preceding study [18], is summarized in Figure 1.

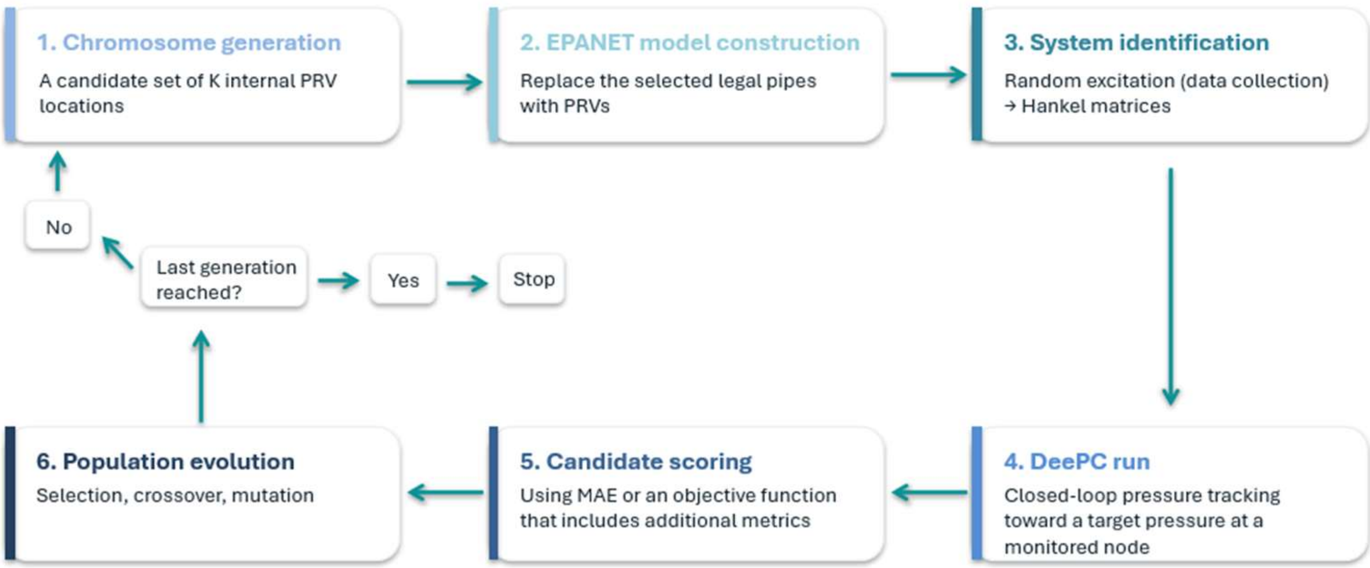



**Figure 1.** Proposed GA-DeePC framework. The Genetic Algorithm generates candidate internal PRV locations, the EPANET model is updated accordingly, DeePC performs a closed-loop pressure-control experiment, and each candidate is scored according to operational performance metrics. The score is then used by the GA to evolve the population toward improved valve-location configurations.

All computations were implemented in Python 3.11. Hydraulic simulations and candidate-network modifications were performed using EPyT version 2.3.5.0, which provides a Python interface to the EPANET 2.2 (U.S. Environmental Protection Agency, Cincinnati, OH, USA) hydraulic engine [27,28]. The DeePC optimization problems were solved using the Gurobi Optimizer (version 12.0.3) through its Python API [29]. The hydraulic demand formulation was retained from the corresponding benchmark-network files. The Fossolo network was simulated using pressure-driven analysis (PDA), with a minimum pressure of 0 m, a required pressure of 10 m, and a pressure exponent of 0.5. The Modena network was simulated using demand-driven analysis (DDA), corresponding to the default EPANET demand formulation. No changes to the demand-model formulation were introduced as part of the GA-DeePC framework.

The temporal demand representation was also retained from the benchmark-network files. In Fossolo, nodal base demands were multiplied by a predefined deterministic, time-varying hourly demand pattern contained in the EPANET model. In Modena, no temporal demand pattern was specified, and the nodal base demands therefore remained constant throughout the simulations. No stochastic perturbations were added to consumer demands in either case; the randomized excitation used during system identification was applied only to the PRV settings. These prescribed demand conditions were kept identical across candidate PRV configurations to provide a controlled comparison of valve-placement effects. The demand trajectories were part of the hydraulic simulation and were not supplied explicitly to the DeePC controller. Leakage was not explicitly modeled in either benchmark network.

DeePC is a data-driven predictive control method that uses measured input-output trajectories directly, rather than relying on an explicitly calibrated state-space model of the controlled system [12,13]. In the present study, the control inputs are the PRV pressure settings, and the measured output is the pressure at a monitored node. The DeePC formulation follows the same general setup used in the previous DeePC sensitivity study [18], but is summarized here to clarify how it is embedded within the GA search.

Let $u^d$ and $y^d$ denote the input and output data collected during the identification phase, where $u^d$ contains the PRV settings and $y^d$ contains the pressure response at the monitored node. From these data, Hankel matrices of depth $T_{\text{ini}} + N$ are constructed and partitioned into past and future blocks [13,14]:

$$\begin{bmatrix} U_p \\ U_f \end{bmatrix} = H_{T_{\text{ini}}+N}\left(u^d\right), \begin{bmatrix} Y_p \\ Y_f \end{bmatrix} = H_{T_{\text{ini}}+N}\left(y^d\right) \tag{1}$$

where $T_{\text{ini}}$ is the initialization window and $N$ is the prediction horizon. The past blocks $U_p$ and $Y_p$ are used to match the most recent measured input-output history, while the future blocks $U_f$ and $Y_f$ are used to predict the future input-output trajectory.

At each control step, the most recent input-output measurements are denoted by $u_{\text{ini}}$ and $y_{\text{ini}}$. DeePC searches for a vector $g$ that represents the future trajectory as a linear combination of columns of the Hankel matrices [13,16]. For the pressure-tracking problem considered here, the regularized DeePC problem can be written in the following form [13,16]:

$$\min_{g,u,y,\sigma_u,\sigma_y} \| y - y_{\text{ref}} \|_2^2 + \lambda_g \| g \|_2^2 + \lambda_u \| \sigma_u \|_2^2 + \lambda_y \| \sigma_y \|_2^2 \tag{2}$$

subject to

$$\begin{bmatrix} U_p \\ Y_p \\ U_f \\ Y_f \end{bmatrix} g = \begin{bmatrix} u_{\text{ini}} \\ y_{\text{ini}} \\ u \\ y \end{bmatrix} + \begin{bmatrix} \sigma_u \\ \sigma_y \\ 0 \\ 0 \end{bmatrix} \tag{3}$$

$$u_{\min} \leq u \leq u_{\max}$$

where $y_{\text{ref}}$ is the pressure reference, $u$ is the predicted PRV-setting sequence, and $y$ is the predicted monitored-node pressure over the prediction horizon. The slack variables $\sigma_u$ and $\sigma_y$ relax the past input-output consistency constraints, and the regularization weights $\lambda_g$, $\lambda_u$, and $\lambda_y$ improve robustness to noise, model mismatch, and nonlinear behavior [16,17].

The DeePC controller was implemented in a receding-horizon manner. At each control step, the optimization problem was solved; only the first control action was applied to the hydraulic simulation, the most recent input-output history was updated, and the optimization was solved again at the next time step. The baseline DeePC parameters were adopted from Perelman and Ostfeld [12] and the subsequent sensitivity analysis of Davda and Ostfeld [18]. In these studies, the regularization weights were examined using logarithmic grid-based tuning, and the latter analysis showed only a modest influence on the monitored-node MAE within the tested ranges, although valve actuation effort was more sensitive. Therefore, the present study retained these previously tested baseline values rather than re-tuning DeePC for every candidate PRV configuration, allowing the analysis to focus on the effect of valve placement. These values should be regarded as a consistent baseline for the tested conditions rather than universally optimal parameters. Specifically, the identification phase contained 600 hourly samples, $T_{\text{ini}} = 48$ h, $N = 12$ h, and the control step was 1 h. The PRV-setting bounds were $20 \leq u \leq 70$ m, and the regularization weights were set to $\lambda_g = 10^{-2}$, $\lambda_y = 10^1$, and $\lambda_u = 10^{-2}$. In the present study, the closed-loop control phase was set to 72 hourly samples, rather than the 168 h evaluation window used in the preceding study, to limit the computational cost associated with the repeated DeePC evaluations required during the GA search. The same 72 h control window was applied consistently to all candidate configurations to preserve comparability across the GA evaluations. The hydraulic and DeePC settings used in the two case studies are summarized in Table A1.

The GA was used as an external optimization layer around the DeePC controller [19,22]. Each GA individual represented a candidate set of K internal pipes on which added internal

PRVs were installed. The value of K was fixed within each GA run. For each individual, the following evaluation procedure was performed:

1. A chromosome was generated as a set of K candidate internal PRV locations.
2. The EPANET input file was copied and modified by replacing the selected legal pipes with PRVs.
3. A randomized identification phase was simulated, in which PRV settings were excited uniformly within the allowed range of 20–70 m.
4. Hankel matrices were constructed from the resulting input-output data.
5. DeePC was run in closed loop for the control horizon, using the monitored-node pressure as the controlled output.
6. The candidate solution was scored using either the monitored-node MAE or a multi-metric objective function.
7. The GA population was evolved using selection, crossover, mutation, and elitism [19].

This structure means that the GA does not solve the hydraulic equations directly. Instead, for each candidate PRV-location configuration, the hydraulic simulation and DeePC controller generate a closed-loop performance score. That score is then used by the GA to evolve the population toward stronger PRV-location configurations.

A genome is written as

$$G = \{\ell_1, \ell_2, \ldots, \ell_K\} \tag{4}$$

where each $\ell_i$ is the pipe ID selected for the installation of one added internal PRV. The fixed source PRVs were retained in the network and were included as controllable inputs in the DeePC simulations. The fixed source PRVs and the added internal PRVs were treated consistently as controllable DeePC inputs. For all controlled PRVs, the allowable pressure-setting range was 20–70 m, and the valve settings were updated once per 1 h control interval according to the receding-horizon DeePC solution. No separate valve dynamic model or explicit rate-of-change constraint was imposed. Instead, valve movement was quantified through the actuation-effort metrics $E_{L1}$ and $E_{\text{avg}}$, as defined in Equations (8) and (9). The added internal PRVs were assigned new valve IDs and inserted into the candidate EPANET model by replacing the selected pipes.

In the present implementation, each selected pipe was treated as a candidate link to be converted into a pressure-controlled link, and was therefore replaced by a PRV connecting the same end nodes. This representation was adopted to provide a consistent and computationally simple network-modification procedure across the large number of candidate configurations evaluated by the GA. Accordingly, the study compares alternative PRV-placement configurations under a common modeling representation rather than reproducing the hydraulic effects of installing a PRV in series with the complete original pipe. Consequently, the distributed frictional head loss of the replaced pipe is not explicitly retained, and this should be regarded as a simplifying assumption of the present modeling framework.

Candidate PRV locations were constrained in both case studies in order to ensure physically meaningful and reproducible DeePC evaluations. In general, added internal PRVs were installed only on pipe links, and each genome was repaired or generated so that the same pipe could not appear more than once in a single chromosome. During model construction, each selected pipe was replaced by an added PRV with a unique valve ID, while the fixed source PRVs were retained as part of the controlled system.

For the larger Modena case study, additional feasibility rules were applied because the search space was larger and the network contained several fixed source PRVs. Candidate pipes were required to connect two junction nodes and were excluded if they shared a node with an existing valve. In addition, the genome-repair procedure prevented selecting two added internal PRVs on pipes sharing the same end node. These rules reduced the risk of creating locally clustered or hydraulically ill-conditioned valve configurations.

For both networks, PRV orientation was determined from the simulated flow direction rather than being fixed arbitrarily from the network topology. For each candidate placement, the flow sign through the selected link was inspected over the simulation period. If the resulting flow remained zero or negative without a sign change, the PRV orientation was reversed and the case was re-evaluated. Because a physical PRV has a fixed orientation, candidate locations exhibiting flow reversal during the simulated period were not considered suitable for PRV installation. This procedure prevented a valve from being installed against the prevailing flow direction and artificially restricting the link [3,18].

The scoring logic differed between the two case studies. In Fossolo, the GA was driven by the monitored-node MAE, and the final population was then re-evaluated using the additional spatial and operational performance metrics. In Modena, the multi-metric objective function was used directly during the GA search, because the larger network and the comparison between different numbers of added internal PRVs required a broader evaluation criterion from the beginning.

The GA used tournament selection, uniform crossover, mutation, and elitism. Tournament selection was applied with a tournament size of 3 for the Fossolo case and for the Modena runs with $K = 2,\ 3,\ 4$, and 5, whereas a tournament size of 4 was used for the Modena run with $K = 6$.

Crossover was applied with probability $P_c = 0.9$. Mutation was applied independently to each gene with a prescribed mutation probability, and mutation replaced a selected pipe with a nearby candidate pipe in the ordered candidate list. Elitism was used to preserve the best-performing individuals from one generation to the next [19].

For Fossolo, the GA was executed with $K = 3$ added internal PRVs, a population size of 40 individuals, and 25 generations. The GA seed was set to 123. The mutation probability was 0.30 per gene, the mutation step was 6, and one elite individual was preserved in each generation.

For Modena, separate GA runs were performed for different values of K, representing the number of added internal PRVs. The fixed source PRVs in Modena were retained, and the GA searched for additional internal PRV locations. For the representative runs shown in the results, the following search budgets were used: $K = 2$ was run for 25 generations with 40 individuals per generation; $K = 3$ was run for 30 generations with 50 individuals per generation; $K = 4$ was run for 90 generations with 70 individuals per generation; $K = 5$ was run for 120 generations with 90 individuals per generation; and $K = 6$ was run for 160 generations with 110 individuals per generation. The GA seed was set to 123. The mutation and elitism settings were adjusted according to the scale of each run, in order to balance exploration with preservation of strong candidate solutions. The complete GA settings and computational budgets used in the Fossolo and Modena runs are summarized in Table A2.

The nominal candidate-evaluation budgets, defined as population size multiplied by the number of generations, were 1000 for Fossolo and 1000, 1500, 6300, 10,800, and 17,600 for the Modena runs with K = 2, 3, 4, 5 and 6, respectively. Each fresh candidate evaluation involved 72 receding-horizon DeePC optimizations over the control phase. Previously evaluated genomes were retrieved from the evaluation cache and therefore did not require a new DeePC evaluation. In the computational runs performed for this study, a fresh candidate evaluation typically required approximately 40 s, with modest computer-dependent variation. Exact cumulative wall-clock times and standardized hardware benchmarks were not recorded during the original optimization runs.

For reproducibility, independent fixed random seeds were used for the two stochastic components of the framework: the GA operators were initialized with a seed of 123, whereas the randomized PRV excitation used during the DeePC identification phase was generated with a seed of 42. The reported optimization results therefore correspond to reproducible single-seed realizations of the coupled GA-DeePC procedure rather than to ensemble statistics over multiple independent stochastic realizations.

Each candidate PRV configuration was evaluated over the closed-loop control phase only. Let $T_c$ denote the number of control samples, let $p_m(t)$ be the pressure at the monitored node at time $t$, and let $p_{\text{ref}}$ be the pressure reference. The monitored-node tracking error was evaluated using the mean absolute error:

$$\text{MAE} = \frac{1}{T_c}\sum_{t=1}^{T_c} \mid p_m(t) - p_{\text{ref}} \mid \tag{5}$$

Spatial and multi-node pressure indicators have previously been used in pressure-management studies to evaluate network-wide behavior beyond a single monitored location [30]. To account for broader spatial performance, a set of high-elevation nodes $S$ was also evaluated. For each node $i \in S$, the node-wise MAE was calculated (5), and the spatial variability term was defined as the standard deviation of these node-wise MAE values.

$$\text{STD} = \text{std}_{i \in S}(\text{MAE}_i) \tag{6}$$

Pressure-bound violations were calculated as the percentage of pressure samples outside the allowable range:

$$\text{Viol\%} = 100 \cdot \frac{1}{T_c \mid S \mid}\sum_{t=1}^{T_c}\sum_{i \in S} \mathbb{I}[p_i(t) < p_{\min} \text{ or } p_i(t) > p_{\max}] \tag{7}$$

where $\mathbb{I}[\,\blacksquare\,]$ is an indicator function that equals 1 when the stated pressure-bound violation occurs and 0 otherwise, $p_{\min} = 25$ m, and $p_{\max} = 60$ m.

The 25–60 m range was used as a soft performance criterion rather than as a hard feasibility constraint. Candidate configurations were therefore not discarded when individual pressure samples exceeded these bounds; instead, the frequency of such excursions was quantified through Viol% and penalized within the multi-metric objective. This formulation allows the optimization to expose trade-offs between pressure compliance, tracking performance, spatial behavior, and actuation effort.

Valve actuation effort was evaluated from the PRV setting trajectories during the control phase. Let $u_j(t)$ be the setting of PRV $j$ at time $t$, and let $n_u$ be the number of controlled PRVs. The total $L_1$-type valve movement was defined as

$$E_{L1} = \sum_{t=2}^{T_c}\sum_{j=1}^{n_u}\left|u_j(t) - u_j(t-1)\right| \tag{8}$$

For cross-$K$ comparisons, where the number of controlled PRVs differs between cases, the normalized average actuation effort was calculated as

$$E_{\text{avg}} = \frac{E_{L1}}{(T_c - 1)n_u} \tag{9}$$

This metric expresses the average valve-setting change per valve and per control step, making it more suitable for comparing configurations with different numbers of added internal PRVs [18].

In Fossolo, the spatial performance metrics were evaluated at five selected high-elevation nodes, identified in the network model by node IDs 7, 24, 8, 9, and 36. In Modena, the corresponding metrics were evaluated at seven selected high-elevation nodes, with node IDs 74, 73, 75, 15, 217, 76, and 72. These high-elevation nodes were selected as representative hydraulically challenging service locations, particularly with respect to maintaining adequate pressure, rather than as an exhaustive representation of all network nodes.

The monitored nodes were retained from the preceding DeePC studies to maintain methodological continuity and direct comparability: Node 21 was used in Fossolo and Node 201 in Modena [12,18]. These nodes serve as the controlled outputs of DeePC and are not intended to represent the entire network. In the preceding Fossolo sensitivity study, repeating the PRV-location analysis with a different monitored node (Node 15) produced the same qualitative location-sensitivity behavior [18]. In the present study, broader network performance is additionally assessed through spatial pressure variability and pressure-bound violations at selected high-elevation nodes. Nevertheless, the selected monitored node may influence the optimal PRV configuration, and systematic evaluation of multiple monitored nodes is left for future work.

Because the performance components have different physical units and numerical magnitudes, a weighted normalized objective function was used when combining several metrics. The general form of the objective function is

$$J = w_{\text{MAE}} \frac{\text{MAE}}{N_{\text{MAE}}} + w_{\text{STD}} \frac{\text{STD}}{N_{\text{STD}}} + w_{\text{Viol}} \frac{\text{Viol\%}}{N_{\text{Viol}}} + w_E \frac{E}{N_E} \tag{10}$$

where $E$ is either $E_{L1}$ or $E_{\text{avg}}$, depending on the comparison being performed. The terms $N_{\text{MAE}}$, $N_{\text{STD}}$, $N_{\text{Viol}}$, and $N_E$ are normalization constants, and the weights $w_i$ define the relative importance of the metrics.

The normalization constants were selected to place the performance components on comparable numerical scales based on the characteristic magnitudes and operational ranges considered in each comparison. Within each comparison, they were applied consistently and were not tuned to favor a particular configuration or value of K.

The MAE term represents local pressure-tracking accuracy at the monitored node and was assigned the largest weight. The STD term represents spatial pressure uniformity across selected high-elevation nodes. The violation term penalizes pressure values outside the allowable operating range. The actuation-effort term penalizes excessive PRV movement, which is relevant for operability and valve wear. The objective-function weights were prescribed a priori to reflect the focus of this study on DeePC pressure-tracking performance. Accordingly, the monitored-node MAE was assigned the largest weight, while the remaining weight was distributed among the spatial and operational criteria.

In the Fossolo case study, the GA search itself was driven by the monitored-node MAE only. After the GA run was completed, the final population was re-evaluated using the additional performance metrics defined above: spatial STD, pressure-bound violations, and normalized average valve actuation effort, $E_{\text{avg}}$. This post-run evaluation allowed the final population to be ranked not only according to local pressure tracking, but also according to broader operational behavior.

For the final Fossolo population, the combined objective had the same weighted-normalized structure described above, but no $K$-penalty was included because all Fossolo candidates contained the same number of added internal PRVs, $\text{K} = 3$:

$$J_{\text{Fossolo,final}} = 0.45 \cdot \frac{\text{MAE}}{5} + \frac{0.55}{3} \cdot \frac{\text{STD}}{4} + \frac{0.55}{3} \cdot \frac{\text{Viol\%}}{15} + \frac{0.55}{3} \cdot \frac{E_{\text{avg}}}{10} \tag{11}$$

For the Fossolo final-population evaluation, the monitored-node MAE was assigned a weight of 0.45, while the remaining weight was divided equally among spatial pressure variability, pressure-bound violations, and normalized average valve actuation effort. The corresponding normalization constants were 5 m for MAE, 4 m for STD, 15% for pressure-bound violations, and 10 m for $E_{\text{avg}}$. These values were applied consistently to all individuals in the final Fossolo population.

In the Modena case study, the objective function was used already during the GA search, rather than only after the final generation.

For each fixed value of K, the GA minimized a local multi-metric objective:

$$J_{\text{local}} = 0.50 \cdot \frac{\text{MAE}}{5} + \frac{1}{6} \cdot \frac{\text{STD}}{1} + \frac{1}{6} \cdot \frac{\text{Viol\%}}{25} + \frac{1}{6} \cdot \frac{E_{L1}}{N_{L1,K}} \tag{12}$$

where $N_{L1,\text{K}}$ is the actuation-effort normalization used in the corresponding fixed-K run. The local objective values were used only to rank candidate solutions within the same GA run. The local-objective weights were prescribed a priori, with MAE assigned half of the total weight and the remaining half divided equally among spatial variability, pressure-bound violations, and actuation effort. Because these weights directly governed the GA evolution, changing them could alter both the explored populations and the selected fixed-K winners. Consequently, they were not varied in the post-run sensitivity analysis, since a like-for-like assessment of alternative local weights would require new GA searches. They were not used for direct comparison across different values of K, because the actuation term and the dimensionality of the control problem depend on the number of valves. The actuation-effort normalization constant $N_{L1,\text{K}}$ was scaled according to the total number of controlled PRVs:

$$N_{L1,\text{K}} = 8000 \cdot \left( \frac{4 + \text{K}}{6} \right) \tag{13}$$

where four represents the fixed source PRVs in the Modena network and K represents the number of added internal PRVs. The denominator of six corresponds to the $\text{K} = 2$ reference case. The resulting values were rounded to the nearest integer before implementation and are summarized in Table A2.

After the best solution was selected for each fixed-K run, the leading solutions were re-scored using a global objective function with consistent normalization. This enabled a fair comparison between configurations with different numbers of added internal PRVs. The global objective was defined as

$$J_{\text{global}} = 0.40 \cdot \frac{\text{MAE}}{5} + 0.175 \cdot \frac{\text{STD}}{1} + 0.175 \cdot \frac{\text{Viol\%}}{25} + 0.175 \cdot \frac{E_{\text{avg}}}{20} + 0.075 \cdot \frac{\text{K}}{6} \tag{14}$$

The use of $E_{\text{avg}}$, rather than total $E_{L1}$, allows actuation effort to be compared across cases with different numbers of controlled valves. The final term penalizes the number of added internal PRVs and reflects the operational preference for simpler configurations when performance is otherwise comparable [22].

Unlike the local-objective weights, the global-objective weights were applied only after the fixed-K winners had already been selected. Their influence on the final cross-K ranking could therefore be examined post hoc without repeating the GA or DeePC simulations. In addition to the baseline weighting in Equation (14), the five selected solutions were re-scored under three illustrative preference structures: balanced, tracking-focused, and spatial/service-focused. The normalization constants and the valve-count penalty $w_{\text{K}} = 0.075$ were retained in all cases so that only the relative emphasis among the four performance metrics was varied. The tested weight structures and resulting rankings are summarized in Table 1.

**Table 1.** Sensitivity of the Modena cross-K ranking to alternative global-objective weighting structures. The selected fixed-K winners were re-scored using the normalization constants of Equation (14). The valve-count penalty was fixed at $w_K = 0.075$ in all scenarios. Rankings are ordered from the lowest (preferred) to the highest re-scored global-objective value.

| Weighting Scenario | $w_{MAE}$ | $w_{STD}$ | $w_{Viol}$ | $w_{E_{avg}}$ | Ranking of (Best → Worst) | Preferred |
|---|---|---|---|---|---|---|
| Baseline | 0.400 | 0.175 | 0.175 | 0.175 | 2, 3, 4, 5, 6 | 2 |
| Balanced | 0.231 | 0.231 | 0.231 | 0.231 | 2, 3, 4, 5, 6 | 2 |
| Tracking-focused | 0.625 | 0.100 | 0.100 | 0.100 | 4, 3, 2, 6, 5 | 4 |
| Spatial/service-focused | 0.250 | 0.275 | 0.275 | 0.125 | 2, 3, 4, 5, 6 | 2 |

Thus, the Modena evaluation consisted of two stages. First, each fixed-$K$ GA run selected a local winner according to the local objective. Second, the local winners from $K = 2, 3, 4, 5, 6$ were compared using the global objective function. This separation avoided direct comparison of local objective values across different K values and provided a consistent basis for selecting the overall preferred configuration.

## 3. Case Studies

The proposed GA-DeePC framework was evaluated on the publicly available Fossolo and Modena benchmark water distribution networks obtained from the WaterBenchmarkHub [31]. The case studies were designed to assess whether an external GA can identify internal PRV locations that improve DeePC-based pressure control at the monitored node while also accounting for broader spatial and operational performance. This follows directly from the previous DeePC sensitivity study, which showed that PRV placement can strongly affect tracking performance and that, particularly in larger networks, manual or exhaustive location testing is not practical.

In both networks, the GA was used as an outer optimization layer, while DeePC remained the controller used to evaluate each candidate valve configuration through closed-loop simulation. The network setup, monitored nodes, pressure references, and fixed source PRVs were kept consistent with the previous study as much as possible, so that the new optimization results could be interpreted as a direct extension of the earlier sensitivity analysis. The Fossolo case study was used first as a smaller validation case for the GA-DeePC loop, whereas the Modena case study was used as a larger and more complex test case for comparing different numbers of added internal PRVs.

### 3.1. Fossolo Case Study

The Fossolo network was selected as the first case study because of its relatively small size and because it had already been analyzed in the previous DeePC sensitivity study. This made it suitable for testing whether the GA could reproduce and extend the earlier observation that internal PRV placement has a strong influence on control performance. The same monitored node and pressure reference were used as in the previous analysis: Node 21 was controlled toward a 30 m pressure reference, with the fixed inlet PRV retained as part of the control configuration.

In the Fossolo GA run, each individual represented a set of three added internal PRV locations. During the evolutionary search, the GA was driven by the MAE at the monitored node, allowing it to identify valve-location combinations that improve local pressure tracking. The GA was executed for 25 generations, with a population size of 40 individuals in each generation.

To move beyond a purely local tracking criterion, the final GA population was re-evaluated using the broader objective function defined in Equation (11). This objective function combines the monitored-node MAE with spatial pressure variability, pressure-bound violations, and normalized average valve actuation effort, $E_{\mathrm{avg}}$. Therefore, the Fossolo case study includes both the MAE-driven GA search and the post-run multi-metric evaluation of the final population.

Figure 2 presents the Fossolo network and highlights the monitored node, the fixed inlet PRV, and two representative internal PRV configurations. The green configuration corresponds to the winning solution, with added internal PRVs located on Pipes 45, 35, and 33. This configuration achieved both the minimum monitored-node MAE and the best combined-objective score. The red configuration represents a poor-performing placement.

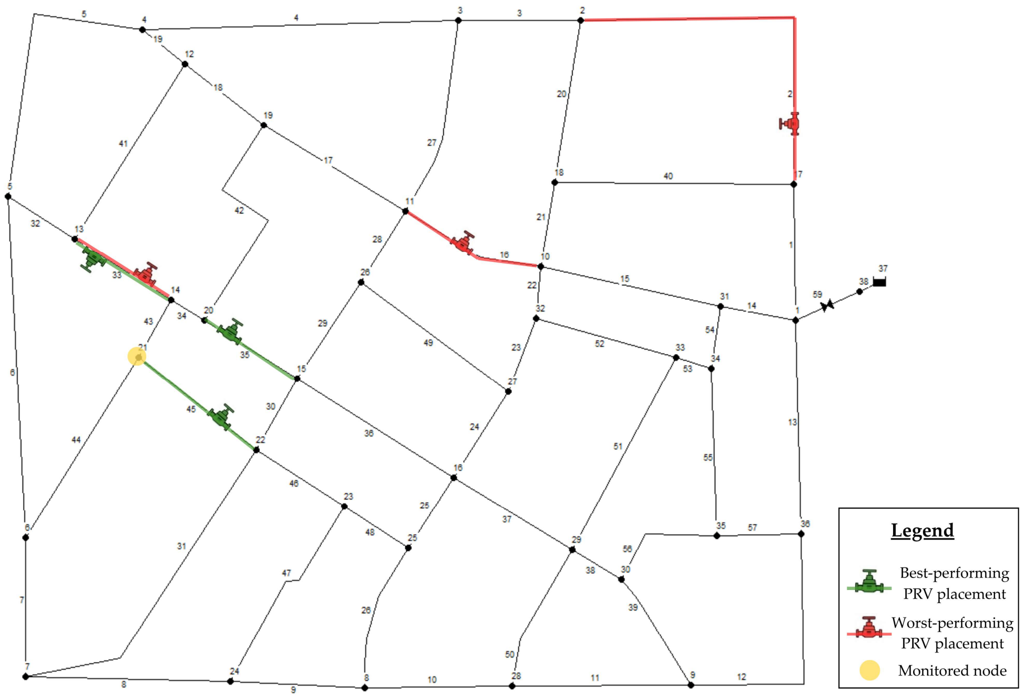



**Figure 2.** Fossolo network used for the GA-DeePC validation case study. The figure shows the monitored node, the fixed inlet PRV, the winning internal PRV configuration, and a poor-performing configuration. The green markers indicate the winning placement on Pipes 45, 35, and 33, while the red markers indicate the poor-performing placement.

Figure 3 presents the convergence of the GA in the Fossolo case study. The blue curve shows the best MAE obtained in each generation, while the orange curve shows the median MAE of the population. The best MAE decreased during the first generations and then remained nearly constant for most of the run, reaching a final value of approximately 2.37 m. The median MAE also decreased relative to its initial value, while showing fluctuations across generations.

Figure 4 illustrates the population diversity throughout the Fossolo GA run, expressed as the number of unique genomes in the population at each generation. In this implementation, duplicate individuals were allowed to remain in the population rather than being explicitly rejected and replaced. The diversity decreased from 40 unique genomes at the beginning of the run, reached a minimum of 27 unique genomes during the later generations, and ended with 32 unique genomes in the final generation.

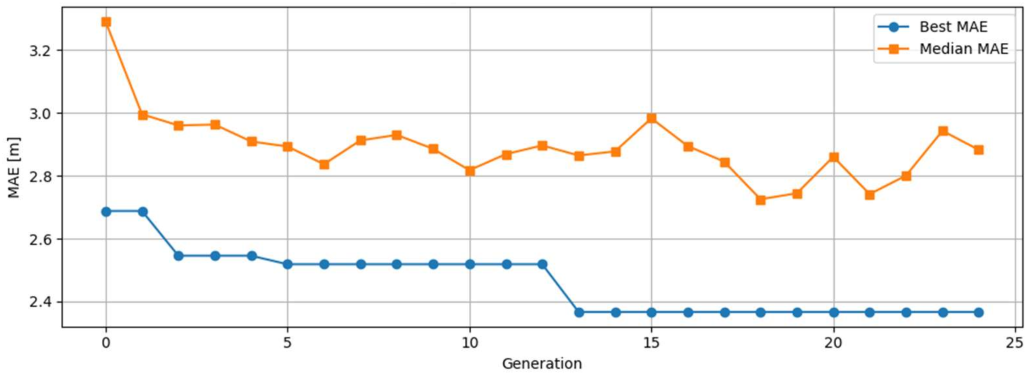



**Figure 3.** GA convergence in the Fossolo case study. The plot shows the best MAE (in blue) and median MAE (in orange) over generations. The rapid early improvement followed by stabilization indicates convergence toward improved internal PRV configurations.

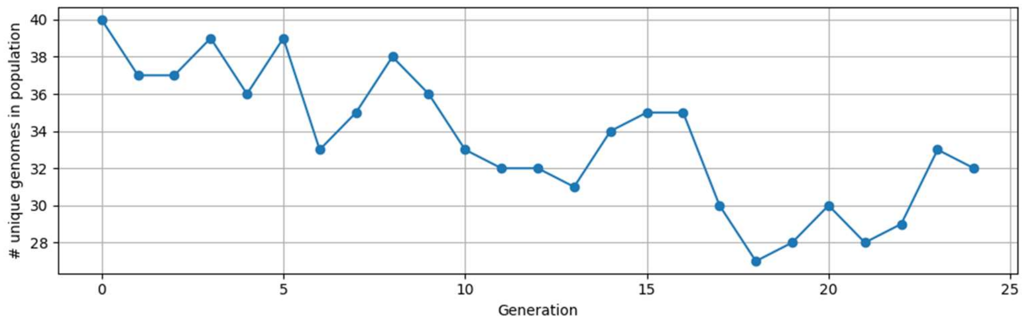



**Figure 4.** Population diversity over generations in the Fossolo GA run. The decrease in diversity indicates that the GA gradually converged as better candidate solutions became more dominant in the population.

Figure 5 presents a component-wise evaluation of the final GA population in the Fossolo case study. For each individual in the final population, four performance components were evaluated during the closed-loop control phase. The first component is the MAE at the monitored Node 21 relative to the 30 m pressure reference. The second component is the spatial pressure-error variability, represented by the standard deviation of the node-wise MAE values across five selected high-elevation nodes: Nodes 7, 24, 8, 9, and 36. The third component is the percentage of pressure-bound violations outside the prescribed pressure range of 25–60 m at these nodes. The fourth component is the total valve actuation effort, denoted by $E_{L1}$, which represents the accumulated valve-setting fluctuations during the control phase.

The figure presents the results for all 40 individuals in the final population. As also reflected in the population-diversity analysis, the final generation contained 32 unique genomes, meaning that some individuals corresponded to repeated PRV-location configurations. To allow the four metrics to be shown in a single figure, separate vertical scales were used. The blue bars represent the monitored-node MAE, and the orange bars represent the spatial STD term; both are pressure-based metrics and are shown in meters. The green bars represent the percentage of pressure-bound violations. The red bars show the cumulative valve actuation effort, $E_{L1}$. For calculation of the combined objective, this quantity was converted to the normalized average effort $E_{\text{avg}}$ according to Equation (9).

Figure 6 provides a time-domain view of the winning Fossolo solution. The upper plot shows the pressure at Node 21 relative to the 30 m reference, while the lower plot shows the setting trajectories of the fixed inlet PRV and the three added internal PRVs located on Pipes 45, 35, and 33. The grey region represents the randomized identification phase, and the white region represents the closed-loop DeePC control phase. In the final population, this configuration achieved both the minimum MAE and the best score according to the combined objective function.

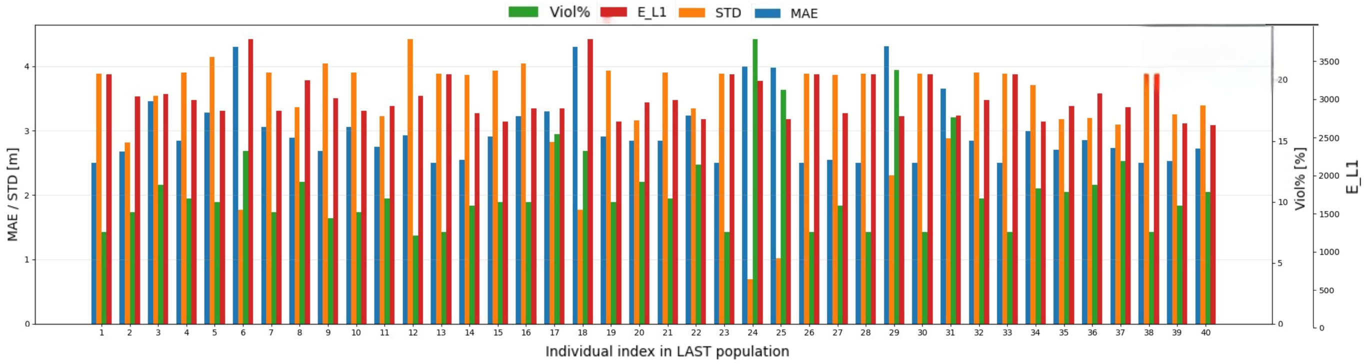



**Figure 5.** Multi-metric comparison of the final Fossolo GA population. The figure summarizes the monitored-node MAE, spatial pressure variability, pressure-bound violations, and cumulative valve actuation effort, $E_{L1}$. For calculation of the combined objective, $E_{L1}$ was converted to the normalized average valve actuation effort, $E_{\text{avg}}$, according to Equation (9). In this case, the configuration with the minimum MAE also achieved the best combined-objective score.

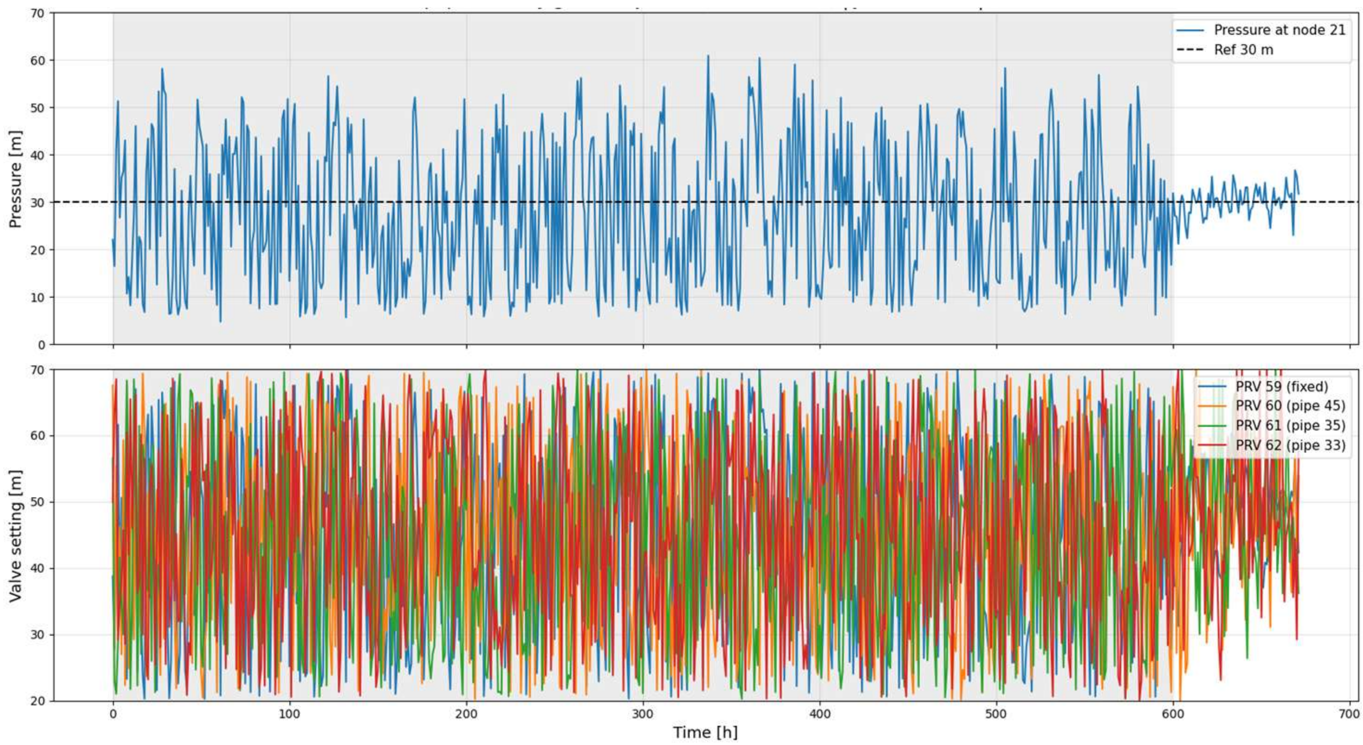



**Figure 6.** Closed-loop DeePC behavior for the winning Fossolo solution, which achieved both the minimum MAE and the best combined-objective score in the final GA population. The upper plot shows the pressure at Node 21 relative to the 30 m reference, while the lower plot shows the setting trajectories of the fixed inlet PRV and the three added internal PRVs located on Pipes 45, 35, and 33. The grey region represents the identification phase with randomized valve excitation, and the white region represents the closed-loop DeePC control phase. In the final population, this configuration achieved both the minimum MAE and the best score according to the combined objective function, indicating that in this case the strongest local tracking solution also provided the best overall operational performance. The trajectories shown correspond to the fixed-seed realization used in the reported optimization and are not ensemble averages.

### 3.2. Modena Case Study

The Modena network was used as the larger-scale case study for evaluating the GA-DeePC framework. Compared with Fossolo, Modena contains a larger number of pipes, several source PRVs, and a more complex hydraulic structure. Therefore, it provides a more suitable test case for examining whether a genetic algorithm can support PRV-location selection when direct or exhaustive search becomes impractical.

The main purpose of this case study was not only to identify favorable internal PRV locations, but also to examine how the number of added internal PRVs affects DeePC-based pressure control. As the number of valves increases, the number of possible placement combinations grows rapidly, making a full enumeration of all candidate configurations computationally infeasible. The GA was therefore used as an efficient search tool that can explore the combinatorial design space and identify promising solutions without testing all possible valve combinations. For $N_c$ legal candidate pipes, an exhaustive fixed-K search can require up to $\binom{N_c}{K}$ placement evaluations before accounting for additional pairwise feasibility constraints. The GA therefore reduces the required DeePC-in-the-loop evaluations from combinatorial enumeration to the finite search budgets reported in Table A2. However, the results are interpreted as the best solutions found within the performed searches, rather than as guaranteed global optima.

The hydraulic setup was kept consistent with the previous DeePC sensitivity study as much as possible, including the Modena network configuration, the fixed PRVs located near the sources, the monitored node, and the baseline reservoir-head conditions. This allows the present results to be interpreted as a direct extension of the earlier sensitivity analysis, moving from the evaluation of individual PRV locations to the optimization of multi-valve configurations.

In contrast to the Fossolo case, where the GA was initially driven by the monitored-node MAE and the final population was later re-evaluated using the combined objective function, the Modena runs used the multi-metric objective function during the GA search for each fixed number of added internal PRVs. Separate GA runs were performed for different values of K, representing the number of added internal PRVs. After the best solution was selected within each fixed-K run, the leading solutions were re-evaluated using a global objective function with consistent normalization.

Figure 7 shows the Modena network and summarizes the spatial outcome of the GA searches. The fixed source PRVs, the monitored Node 201, and the selected internal PRV locations are marked on the network. Each color represents the winning configuration for a different number of added internal PRVs.

Figure 8 presents the GA convergence behavior in the Modena case study for three representative numbers of added internal PRVs: K = 2, K = 4, and K = 6. In each panel, the blue curve represents the best objective value in each generation, while the orange curve represents the median objective value of the population. The objective values in these plots should be interpreted only within each fixed-K run, because each run used a local objective formulation and normalization defined for that specific number of added internal PRVs.

The computational search budget was increased as the number of added internal PRVs increased. For K = 2, the GA was executed for 25 generations with a population size of 40 individuals. For K = 4, the search was extended to 90 generations with 70 individuals per generation, and for K = 6, it was further extended to 160 generations with 110 individuals per generation. In the K = 2 run, the best objective value decreased during the first generations and then remained almost constant until the end of the search. In the K = 4 run, the best objective value improved in several discrete steps. The K = 6 run shows a rapid early reduction in the best objective value, followed by a long period of stabilization and a smaller later improvement.

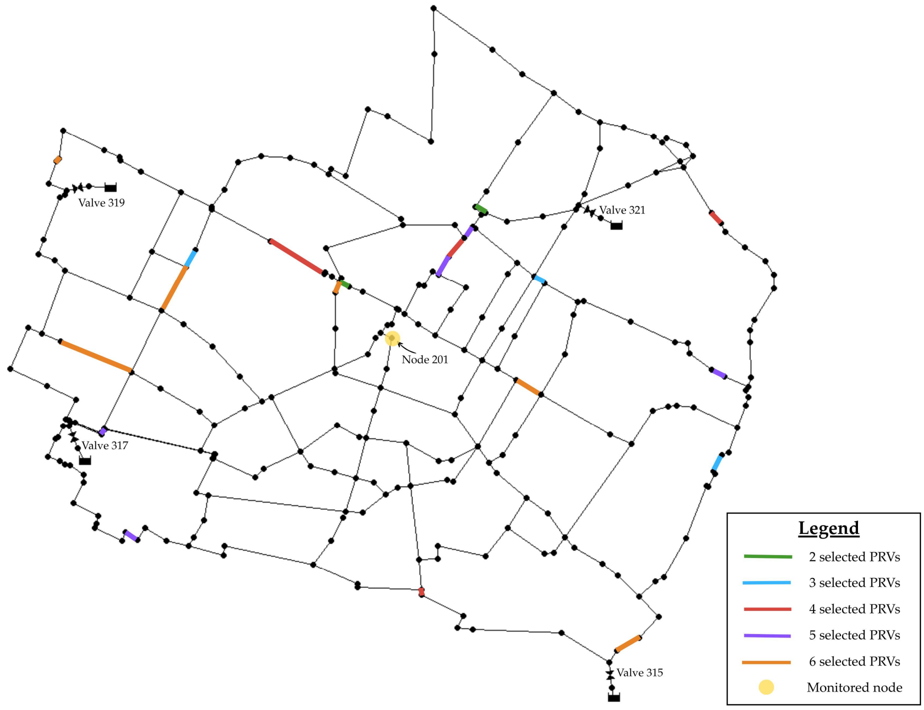
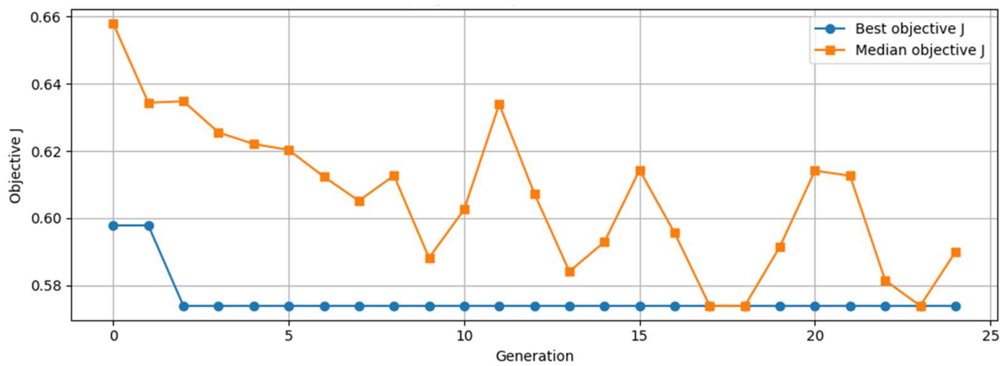
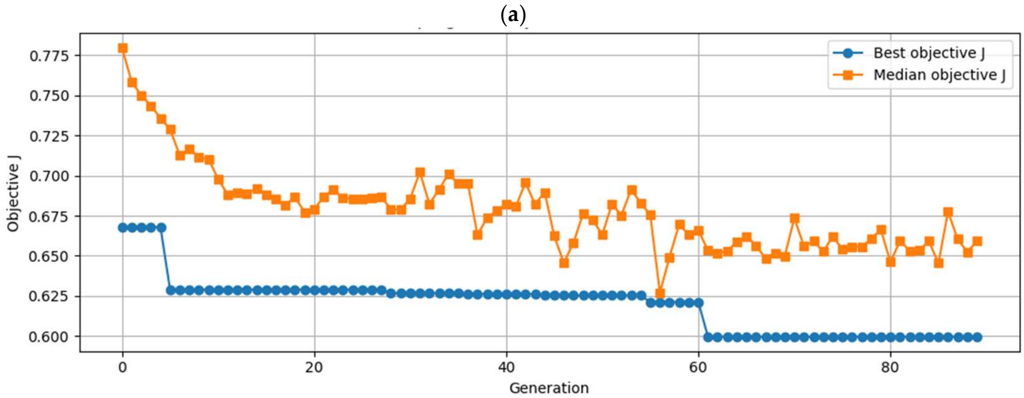



**Figure 7.** Modena network used for the larger-scale GA-DeePC case study. The figure shows the monitored node, the fixed source PRVs, and the winning internal PRV locations identified in the different GA runs for the selected number of PRVs, with each run marked by a different color. For each fixed value of K, the GA searched for the best set of K internal PRV locations. The objective function included MAE, spatial pressure variability, pressure-bound violations, and normalized actuation effort over the control window. This allowed the evaluation to reflect both local pressure tracking and broader network behavior.

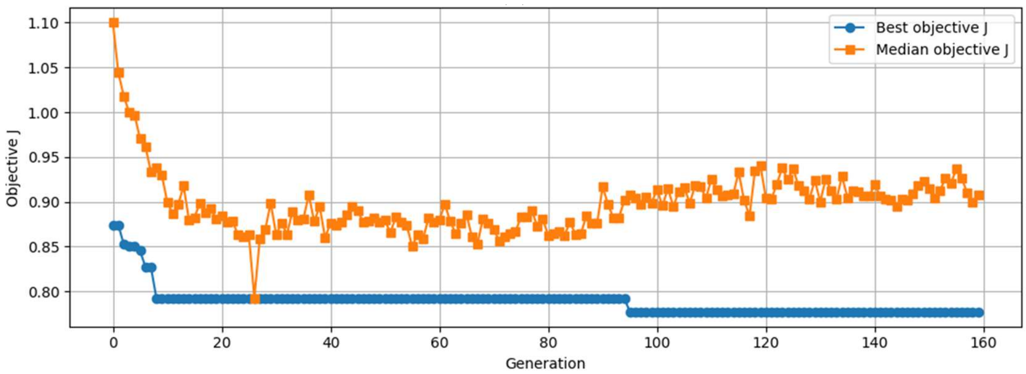



(**a**)




(**b**)

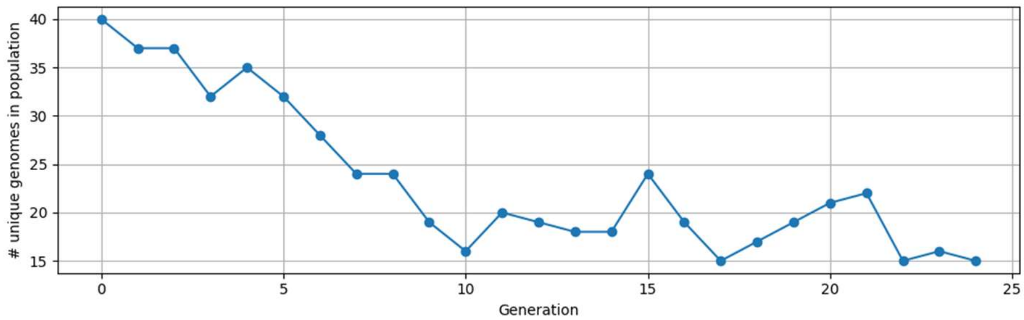
**Figure 8.** *Cont.*



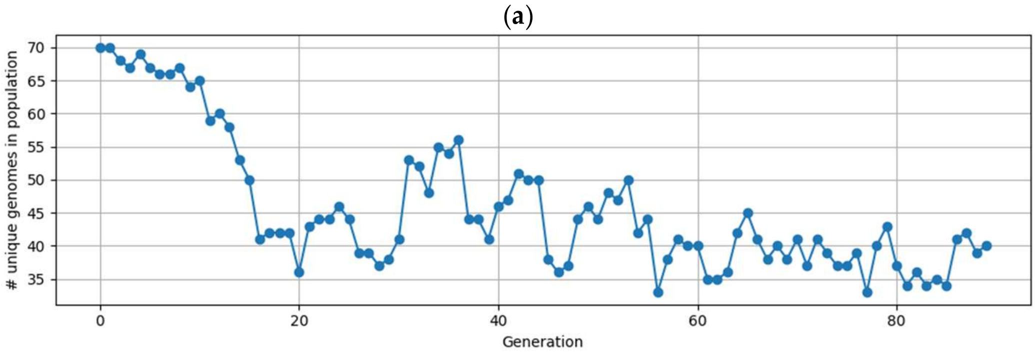

(**c**)

**Figure 8.** GA convergence in the Modena case study for different numbers of added internal PRVs. In each panel, the blue curve shows the best local objective value in each generation, while the orange curve shows the median objective value of the population: (**a**) K = 2 added internal PRVs; (**b**) K = 4 added internal PRVs; and (**c**) K = 6 added internal PRVs. These plots demonstrate the convergence behavior within each fixed-K run. The objective values should be compared directly only within the same run; cross-K comparisons are performed using the consistently normalized global objective.

Figure 9 presents the population diversity during the Modena GA runs for K = 2, K = 4, and K = 6. As in the Fossolo case study, diversity is measured as the number of unique genomes in the population at each generation. For K = 2, the diversity decreases from 40 unique genomes at the beginning of the run to about 15 unique genomes by the final generation. For K = 4, the diversity decreases from nearly 70 unique genomes at the beginning to approximately 40 unique genomes toward the end of the run. For K = 6, the diversity decreases rapidly from nearly 110 unique genomes to roughly 60 unique genomes during the first part of the search and then remains variable.

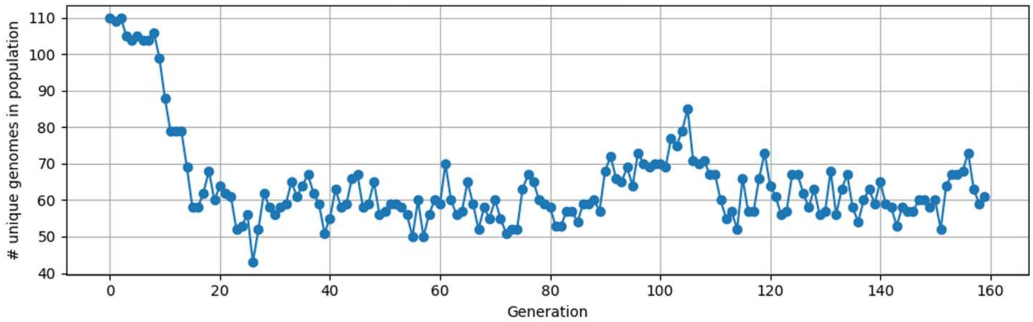



(**a**)




(**b**)

**Figure 9.** *Cont.*




(**c**)

**Figure 9.** Population diversity over generations in the Modena GA runs for different numbers of added internal PRVs. Diversity is measured as the number of unique genomes in the population at each generation, where higher values indicate a broader exploration of different valve-location combinations: (**a**) K = 2 added internal PRVs; (**b**) K = 4 added internal PRVs; (**c**) K = 6 added internal PRVs. Overall, the decrease in diversity supports the interpretation that the GA progressively concentrated around stronger candidate valve configurations rather than sampling random solutions throughout the run.

The corresponding convergence and population-diversity results for the intermediate K = 3 and K = 5 Modena cases are provided in Figure A1. Figure 10 summarizes the global comparison between the leading Modena solutions obtained for different numbers of added internal PRVs. Unlike the convergence plots in Figure 8, which describe the progress of each fixed-K GA run using its own local objective formulation, this figure presents the re-evaluated results using the global objective function defined in Equation (14). This step allows the selected solutions for different values of K to be compared using a consistent normalization framework.

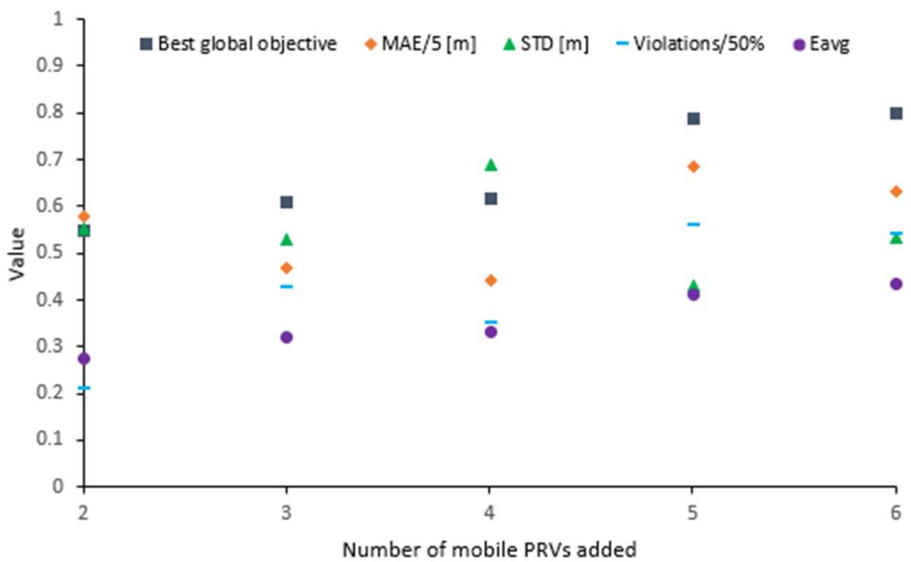



**Figure 10.** Global comparison of the best Modena solution for each number of added internal PRVs. The figure compares the final global objective values after consistent normalization across K values. The individual metric components were rescaled only for visualization using the denominators indicated in the legend; these plotting scales were not used in calculating $J_{\text{global}}$. The plotted values correspond to the same single-seed realizations reported in Table 2.

Figure 10 includes both the final global objective value and the individual performance components used to interpret the result. The plotted components are shown with the normalization indicated in the figure. For visualization only, the individual metric components in Figure 10 were rescaled using the denominators indicated in the legend so that all components could be displayed on a common vertical scale. These plotting normalizations are independent of the normalization constants and weights used to calculate $J_{\text{global}}$. All global-objective values were computed exclusively according to Equation (14). The actuation-related term is represented by $E_{\text{avg}}$, which expresses the average valve-setting change per valve and per control step. The best global objective value was obtained for K = 2.

**Table 2.** Global comparison of the best Modena solution obtained for each number of added internal PRVs. For each fixed value of K, the table reports the re-evaluated global objective value and its main performance components: mean absolute error (MAE) at the monitored node, spatial pressure variability (STD), percentage of pressure-bound violations (Viol%), and normalized average valve actuation effort ($E_{avg}$). Global-objective values were calculated using the unrounded metric values, whereas the individual performance components shown in the table were rounded for presentation. The reported values correspond to the single reproducible seeded realization used for each fixed-K optimization and should not be interpreted as ensemble means.

| Number of Added PRVs | Global Objective | MAE [m] | STD [m] | Viol% | $E_{avg}$ [m] |
|---|---|---|---|---|---|
| 2 | 0.547 | 2.891 | 0.553 | 10.516 | 13.814 |
| 3 | 0.608 | 2.348 | 0.529 | 21.429 | 16.016 |
| 4 | 0.615 | 2.210 | 0.688 | 17.460 | 16.627 |
| 5 | 0.788 | 3.433 | 0.429 | 27.976 | 20.557 |
| 6 | 0.800 | 3.159 | 0.532 | 26.984 | 21.706 |

To assess the sensitivity of the final cross-K selection to the prescribed global-objective weights, the five selected fixed-K solutions were re-scored under three alternative illustrative preference structures while retaining the normalization constants of Equation (14) and the valve-count penalty $w_K = 0.075$. Table 1 summarizes the tested weights and the resulting ranking of the five configurations.

The baseline ranking was unchanged under both the balanced and spatial/service-focused weighting structures, with K = 2 remaining the preferred configuration. Under the deliberately tracking-focused scenario, however, K = 4 became the preferred solution, consistent with its lowest monitored-node MAE in Table 2. This result shows that the preference for K = 2 is robust to the tested balanced and spatial/service-oriented weighting structures, but is not universal: when monitored-node tracking is assigned dominant importance, the preferred valve count can change. The global-objective weights should therefore be interpreted as an explicit engineering preference structure rather than as uniquely optimal values.

Table 2 provides the numerical values underlying the global comparison shown in Figure 10. The table reports the exact values obtained for each number of added internal PRVs after re-evaluation with consistent normalization. The K = 2 configuration achieved the best global objective value, 0.547. This configuration also had the lowest pressure-violation percentage, 10.516%, and the lowest normalized valve actuation effort, $E_{\text{avg}} = 13.814$. The lowest MAE was obtained for K = 4, with 2.210 m, and the lowest STD was obtained for K = 5, with 0.429 m.

The pressure-bound violation metric requires additional interpretation from an operational perspective. Even the preferred K = 2 configuration exhibited a Viol% of 10.516%, indicating that the 25–60 m evaluation range was not maintained at all evaluated node-time samples. Consequently, this result should not be interpreted as demonstrating full operational compliance with a hard pressure constraint. Rather, Viol% was intentionally treated as a penalized performance indicator so that the GA could quantify and compare the severity of pressure excursions together with the other objectives. The lower violation rate of K = 2 relative to the other tested configurations contributes to its selection as the best overall compromise, but a field implementation requiring strict pressure compliance would need to impose the applicable utility-specific pressure limits as hard constraints or use a stronger violation penalty.

The selected internal PRV locations for the leading solution obtained in each fixed-*K* Modena run are listed in Table A3. Figure 11 provides a time-domain view of the selected Modena solution with K = 2 added internal PRVs, which achieved the best global objective value among the tested cases. Similar to the Fossolo case, the figure is divided into an identification phase and a closed-loop control phase. The grey region represents the randomized identification stage, during which valve movements were used to generate the input-output data required for DeePC, while the white region represents the subsequent closed-loop control stage.

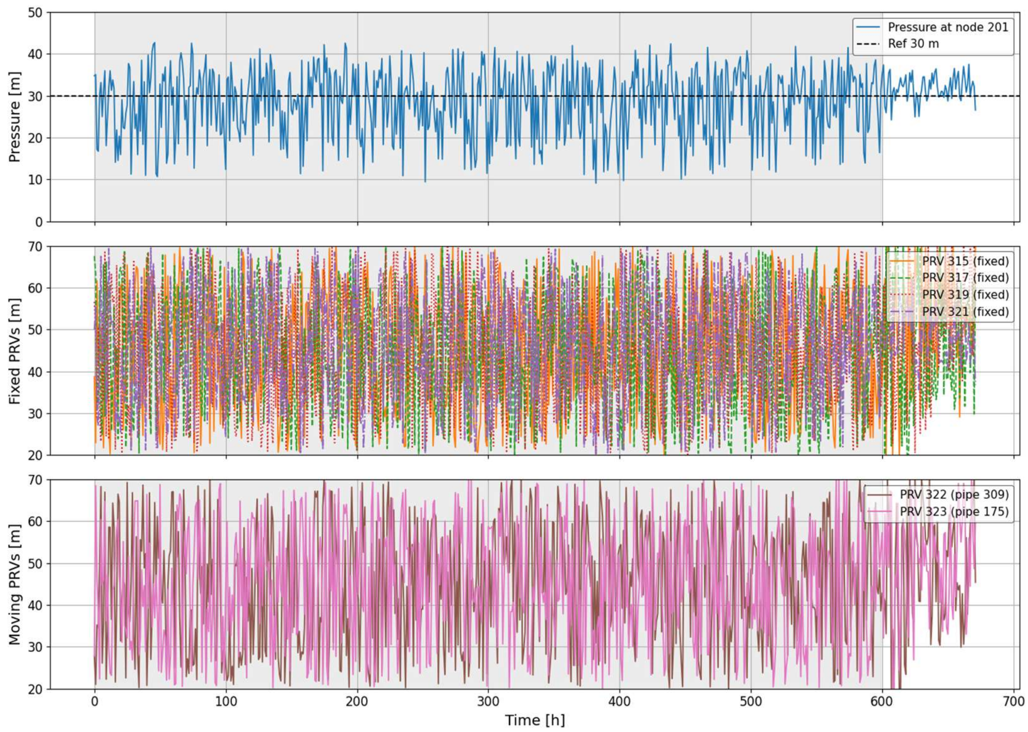



**Figure 11.** Closed-loop DeePC response for the selected Modena configuration with K = 2 added internal PRVs, which achieved the best global objective value among the tested cases. The grey region represents the randomized identification phase, while the white region represents the closed-loop control phase. The upper panel shows the pressure at the monitored Node 201 relative to the 30 m reference. The middle panel shows the setting trajectories of the four fixed source PRVs, and the lower panel shows the two added internal PRVs located on Pipes 309 and 175. Different line styles are used in the middle panel, in addition to color, to improve distinction between the overlapping fixed-PRV trajectories. The trajectories shown correspond to the fixed-seed realization used in the reported optimization and are not ensemble averages.

The upper panel shows the pressure response at the monitored Node 201 relative to the 30 m reference pressure. The middle panel shows the setting trajectories of the four fixed source PRVs, while the lower panel shows the trajectories of the two added internal PRVs located on Pipes 309 and 175.

Figure 12 provides a direct visual comparison of the pressure responses at the monitored node for the leading Modena solutions obtained with K = 2, K = 4, and K = 6 added internal PRVs. Each panel corresponds to the upper pressure-response panel of the full closed-loop DeePC figures for the respective case. The grey region represents the randomized identification phase and the white region represents the closed-loop DeePC control phase. The dashed horizontal line marks the 30 m pressure reference.

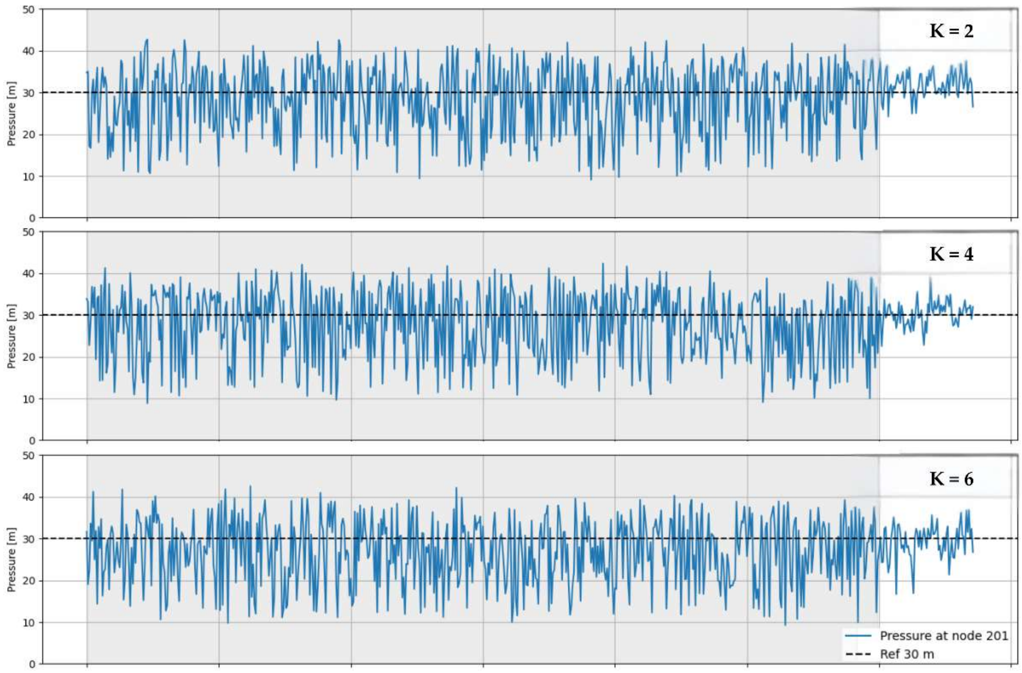



**Figure 12.** Pressure response at the monitored node in the Modena network for the leading GA solutions obtained with different numbers of added internal PRVs. The grey region represents the randomized identification phase, while the white region represents the closed-loop DeePC control phase. The dashed line marks the 30 m pressure reference. The upper panel shows the selected solution for K = 2 added internal PRVs, corresponding to the best global objective value among the tested cases. The middle panel shows the selected solution for K = 4 added internal PRVs. The lower panel shows the selected solution for K = 6 added internal PRVs. The corresponding monitored-node MAE values are 2.891 m, 2.210 m, and 3.159 m for K = 2, K = 4, and K = 6, respectively. The trajectories shown correspond to the fixed-seed realization used in the reported optimization and are not ensemble averages.

The corresponding monitored-node pressure responses for the intermediate K = 3 and K = 5 solutions are presented in Figure A2.

# 4. Discussion

## 4.1. Fossolo Case Study

The Fossolo case study served as a validation case for the GA-DeePC framework. The preceding DeePC sensitivity study showed that the placement of an added internal PRV can either improve or degrade pressure-tracking performance, even when the controller and the remaining hydraulic conditions are unchanged [18]. The present results extend that observation from individual valve locations to combinations of three internal PRVs selected through an evolutionary search. This outcome is also consistent with previous PRV-placement studies, which have shown that valve effectiveness depends on the interaction between location, operating settings, and the hydraulic response of the surrounding network rather than on location alone [3,4,32].

The convergence results show that the GA was able to improve the monitored-node MAE during the evolutionary search. The rapid early reduction in the best MAE indicates that promising valve-location combinations were identified within the first generations, while the later stabilization of the best value suggests convergence toward a strong candidate solution within the performed search budget. The decrease in median MAE further indicates that the improvement was not limited to a single isolated individual, but that the population as a whole shifted toward better-performing configurations. This behavior is consistent with the general operation of population-based genetic algorithms in WDS optimization, in which selection progressively favors stronger individuals while crossover and mutation continue to explore alternative regions of the search space [19]. The stabilization of the best and median values therefore provides evidence of convergence within the performed search, but it does not establish that the global optimum was reached.

The population-diversity results provide a complementary view of the GA behavior. The reduction in the number of unique genomes indicates that repeated or similar PRV-location configurations became more dominant as the search progressed. At the same time, the diversity did not collapse to a single solution, and fluctuations were observed during the run. This behavior is consistent with a typical GA process in which selection and elitism preserve strong solutions, while mutation and crossover continue to introduce new candidate configurations and maintain exploration [19,23].

The final-population analysis demonstrates why it is useful to evaluate more than the monitored-node MAE. Previous pressure-management studies have considered network-wide or multi-criteria measures because improvement at one location does not necessarily imply uniform or satisfactory pressure behavior throughout the network [5,30].

In the present study, the Fossolo population was therefore re-evaluated using spatial pressure variability, pressure-bound violations, and valve actuation effort in addition to local tracking accuracy. In this particular case, the configuration that achieved the minimum MAE also achieved the best combined-objective score, indicating that the local tracking term was closely aligned with the broader operational criteria for the selected network, nodes, and objective weights.

However, this result should not be interpreted as evidence that MAE alone is always sufficient. Rather, it shows that in this relatively small validation network, the broader evaluation confirmed the same leading solution. This may be related to the simpler hydraulic structure of Fossolo and to the fact that the GA search itself was driven by MAE from the beginning. Therefore, the Fossolo results support the validity of the GA-DeePC loop, but they also motivate testing the framework in a larger and more complex network where local and spatial performance criteria may diverge.

### *4.2. Modena Case Study*

The Modena case study extended the analysis to a larger and more complex hydraulic network. The objective was not only to identify favorable internal PRV locations, but also to compare different numbers of added internal PRVs. Valve-placement problems combine discrete installation decisions, continuous hydraulic and operational variables, and nonlinear network equations, and their complexity increases rapidly with network size and the number of candidate valves [4,22,32]. In the present framework, increasing K also increased the number of controllable inputs available to DeePC and the number of possible multi-valve placement combinations. Modena therefore represented a more demanding design-control search than the Fossolo validation case.

The convergence plots show that the GA reduced the objective value in all representative Modena runs. For $K = 2$, the best solution was identified early and remained nearly unchanged for most of the search. For $K = 4$, the best objective improved in several steps, suggesting a more gradual search process. For $K = 6$, a strong early improvement was followed by a long stabilization period and a smaller later improvement. These trends indicate that the GA was able to identify improved PRV-location configurations for different values of K, although the search became more challenging as the number of added internal PRVs increased. Similar scalability difficulties have been reported for evolutionary WDS optimization, where increasing the number of design variables enlarges the search space and may require larger populations, additional generations, or decomposition and hybridization strategies [19,20,22].

The diversity plots support this interpretation. In all three representative runs, the number of unique genomes decreased relative to the initial population, showing that the population progressively concentrated around stronger candidate regions. At the same time, the remaining diversity and the observed fluctuations indicate that the GA continued to explore alternative configurations throughout the search. This is especially important in the higher-dimensional cases, where premature convergence could prevent the algorithm from exploring promising multi-valve combinations. Population size, mutation, crossover, and diversity preservation are therefore important algorithmic considerations in large WDS optimization problems [19,23].

The global comparison provides the main Modena result. After selecting the leading solution within each fixed-K run, the solutions were re-evaluated using a global objective function with consistent normalization. This step was essential because the local objective values from the individual GA runs should not be compared directly across different values of K. Under the consistently normalized global objective, K = 2 achieved the lowest overall score despite not being optimal with respect to every individual performance metric.

This result illustrates the importance of a multi-metric objective. PRV-placement and pressure-management studies have previously considered competing criteria such as pressure reduction, leakage, water age, resilience, installation cost, and spatial or temporal pressure variability [5,21,22,30]. Similarly, in the present Modena analysis, the lowest MAE was obtained for K = 4, whereas the lowest spatial STD was obtained for K = 5. Neither configuration achieved the best global score because its improvement in one component was accompanied by less favorable pressure-bound violations, actuation effort, or valve-count penalty. Thus, the best solution according to a single metric was not necessarily the best operational compromise when all selected criteria were evaluated together.

The gradual increase in the global objective with increasing K should not be interpreted as a monotonic deterioration of all performance metrics. Different values of K performed best for different criteria: K = 4 achieved the lowest MAE and K = 5 the lowest spatial STD, whereas K = 2 provided the lowest pressure-violation rate and normalized actuation effort. The trend therefore reflects trade-offs among the individual criteria rather than a uniform loss of performance. In addition, the global objective includes a deliberately small penalty on K, favoring simpler configurations when other performance measures are comparable. Larger K values also substantially increase the combinatorial placement search space and the number of control degrees of freedom. Although larger populations and more generations were used for these cases, the GA does not guarantee global optimality. Accordingly, K = 2 represents the best overall compromise identified under the baseline objective formulation and computational budgets, rather than evidence that added internal PRVs are inherently detrimental. The global-weight sensitivity analysis in Table 1 further shows that this preference is retained under the tested balanced and spatial/service-focused weighting structures, but shifts to K = 4 when monitored-node tracking is assigned dominant importance.

To examine these trade-offs more directly, a descriptive pairwise Spearman rank-correlation analysis was performed on the unique genomes in the final GA population of each Modena run for K = 2 to K = 6. Duplicate genomes were removed before the analysis, leaving 15, 37, 40, 58, and 61 unique final-population solutions, respectively. Spearman correlation was selected because it evaluates monotonic relationships between ranked observations without requiring a linear dependence assumption. The resulting correlation coefficients are summarized in Figure A4.

The correlation analysis shows that no single performance metric consistently represents the others. The relationship between MAE and pressure-bound violations is positive for K = 3 to K = 6 ($\rho$ = 0.27, 0.66, 0.48, and 0.36, respectively), whereas for K = 2 it is negative ($\rho = -0.52$), demonstrating that the relationship between local tracking accuracy and spatial pressure compliance depends on the valve configuration. Relationships involving spatial STD are also configuration-dependent: the STD-violation correlation ranges from $-0.68$ for K = 2 to 0.34 for K = 5, while the MAE-STD relationship becomes weak or negligible for several higher-K cases. Correlations involving $E_{avg}$ are generally weak ($|\rho| \leq 0.36$), indicating that actuation effort contributes information that is largely complementary to the hydraulic performance indicators. Overall, the results confirm that tracking accuracy, spatial pressure variability, pressure compliance, and actuation effort describe different aspects of closed-loop performance and support their joint consideration when evaluating PRV configurations. Consistent with these configuration-dependent trade-offs, adding more internal PRVs did not automatically improve the overall performance.

Previous placement studies likewise indicate that the marginal benefit of added internal PRVs is network-dependent: in some case studies, the first valve produced most of the achievable pressure reduction, whereas subsequent valves produced smaller improvements; in other networks, a larger number of valves was justified by the hydraulic structure and service requirements [3,4]. This means that valve count should not be interpreted independently of placement, settings, installation implications, and the selected performance criteria. In the present study, additional valves increased the available control degrees of freedom, but also enlarged the search and control spaces and were associated with less favorable actuation or pressure-performance components in several cases. Under the tested search budgets, objective weights, and normalization scheme, the K = 2 configuration therefore provided the best overall operational compromise.

Figure 12 provides a qualitative time-domain comparison of the selected K = 2, K = 4, and K = 6 pressure responses, while the quantitative tracking comparison is given by the monitored-node MAE values reported in Table 2. The K = 4 configuration achieved the lowest MAE (2.210 m), compared with 2.891 m for K = 2 and 3.159 m for K = 6. Thus, the selection of K = 2 as the preferred overall configuration does not result from superior monitored-node tracking alone, but from its more favorable combined performance across the selected hydraulic and operational criteria.

The importance of PRV location can also be observed by comparing configurations containing the same number of added valves. Figure A3 presents a poorly performing K = 2 Modena configuration, with PRVs located on Pipes 158 and 229, for comparison with the selected K = 2 configuration shown in Figure 11. The poorly located configuration produced a monitored-node MAE of 4.806 m and a pressure-bound violation rate of 21.429%, compared with 2.891 m and 10.516%, respectively, for the selected configuration. Thus, even for the same number of added internal PRVs, the resulting closed-loop DeePC performance depends strongly on their hydraulic locations.

An important difference emerges when comparing Modena with Fossolo. In Fossolo, the minimum-MAE solution in the final population was also the combined-objective winner. In Modena, however, the more complex network structure and the comparison across different values of K led to a clearer separation between the best solution according to local tracking accuracy and the best solution according to the global objective. This supports the use of a broader multi-metric evaluation framework in larger networks, where relying only on the monitored-node MAE may overlook spatial pressure behavior, pressure-bound violations, and valve actuation effort.

*4.3. Overall Interpretation and Limitations*

The results of the two case studies support the use of GA as an external optimization layer for DeePC-based PRV placement. The GA does not replace the hydraulic simulator or the DeePC controller; rather, it searches over discrete infrastructure configurations, while each candidate is evaluated through a hydraulic simulation and a closed-loop control experiment. This structure follows the broader design-for-control principle, in which design decisions are assessed according to the operational performance that can be achieved under an associated control strategy [21,22,24]. The distinctive feature of the present framework is that the inner operational evaluation is performed using DeePC rather than a model-based nonlinear control optimizer.

The two benchmark case studies are intended to demonstrate the feasibility and behavior of the proposed GA-DeePC coupling under networks of different scale and complexity, rather than to establish superiority over existing PRV-placement or control methods. A direct head-to-head comparison was not performed in the present study. Such a comparison would require alternative methods to be evaluated under the same network models, candidate-location sets, hydraulic assumptions, constraints, performance objectives, and computational budgets in order to provide a like-for-like assessment. Accordingly, the present results should be interpreted as evidence of feasibility and case-specific performance of the controller-in-the-loop framework. A systematic benchmark against alternative evolutionary, surrogate-assisted, multi-objective, and model-based design-for-control approaches is an important direction for future work.

The present study also builds directly on the preceding DeePC sensitivity analysis [18], which included explicit baseline configurations without added internal PRVs. In Fossolo, DeePC with only the fixed inlet PRV produced a monitored-node MAE of 3.35 m, while testing one additional internal PRV across the network showed that favorable locations could improve tracking whereas poor locations could substantially degrade it. In Modena, the baseline configuration with the four fixed source PRVs and no added internal PRV produced an MAE of 2.19 m; among 54 tested single-PRV locations, the best candidate reduced the MAE to 2.02 m, although with higher maximum deviation and actuation effort. That study also examined sensitivity to the control time step, reservoir-head conditions, and DeePC regularization weights. These baseline and sensitivity results established that DeePC performance depends on both hydraulic configuration and controller settings and motivated the transition in the present study from one-at-a-time placement testing to an external GA search over multi-valve configurations. Because the present study uses a modified evaluation protocol and a broader multi-metric objective, the earlier baseline values are provided as methodological context rather than as direct numerical comparators for the current fixed-K results.

Several limitations should be considered when interpreting the results. Evolutionary algorithms provide search-based solutions rather than certificates of global optimality. Their performance depends on population size, number of generations, genetic operators, initial population, and random seed, and larger WDS optimization problems may require substantially greater computational budgets or specialized decomposition and hybridization methods [19,20,23]. Consequently, the reported configurations are the best solutions identified within the specific searches performed in this study. Repeated runs with additional seeds, larger populations, or alternative GA parameterizations could identify other strong configurations.

In addition, because each candidate fitness was obtained from a single seeded randomized identification experiment, variability associated with alternative identification realizations was not quantified. The differences reported among the fixed-K cases should therefore be interpreted as outcomes of the specific reproducible optimization runs rather than as statistically characterized differences. A multi-seed ensemble analysis would be required to quantify this source of uncertainty.

Every fitness evaluation required a randomized identification phase followed by a closed-loop DeePC simulation. This made the search computationally expensive, particularly for the higher-K Modena cases. The selected population and generation budgets therefore represented a practical compromise between computational effort and search coverage. Hybrid methods have been proposed in WDS design-for-control to reduce the cost of evolutionary exploration by combining it with deterministic optimization or refinement [21,22]. Related future work could examine parallel candidate evaluations, prescreening, surrogate models, or hybrid refinement of GA solutions, provided that the final candidate ranking remains based on full DeePC closed-loop evaluation. Repeating the complete nested optimization over multiple independent seeds would multiply this computational burden and was therefore not included in the present proof-of-concept study.

The combined-objective results depend on the selected metrics, normalization constants, and weights. The present formulation intentionally prioritized monitored-node tracking while retaining penalties for spatial pressure variability, pressure-bound violations, actuation effort, and, in the global Modena comparison, the number of added internal PRVs. Alternative utility priorities could produce a different ranking, as demonstrated by the global-weight sensitivity analysis presented in Table 1. A limitation of the present single-objective formulation is that the different hydraulic and operational criteria are aggregated into a single weighted fitness value. This simplifies both the GA search and the subsequent selection of one preferred configuration, but it necessarily embeds the prescribed engineering preferences in the optimization and may discard near-optimal solutions that offer different trade-offs among tracking accuracy, spatial pressure behavior, pressure compliance, and actuation effort. In contrast, a multi-objective evolutionary formulation could retain a set of non-dominated solutions and allow operators to select among alternative practically implementable trade-offs after the optimization. The single-objective formulation was intentionally retained here to provide a comparatively simple and interpretable outer-search framework for the initial evaluation of the GA-DeePC coupling. The local GA-search weights were not varied because changing them would alter the GA evolution and would require new optimization runs for a like-for-like assessment. Future work should therefore examine Pareto-based multi-objective formulations and compare their solution diversity and operational trade-offs with the present weighted-sum approach [21–23]. An additional modeling limitation is that each selected pipe was replaced by a PRV link rather than retaining the complete original pipe in series with the valve. Accordingly, the original pipe frictional loss was not explicitly preserved, and future implementations could examine pipe-preserving PRV insertion when detailed retrofit representation is required.

A further limitation is that pressure-dependent leakage was not explicitly included in the benchmark simulations. Because leakage can vary with pressure, its inclusion could modify the hydraulic response and potentially influence the relative performance of alternative PRV configurations. Evaluating the proposed framework under leakage-inclusive hydraulic models is therefore recommended as a direction for future work.

The present optimization was evaluated under the prescribed benchmark demand conditions, which were held fixed across candidate configurations. Although this provides a controlled basis for comparing PRV locations, alternative temporal or stochastic demand scenarios could modify the hydraulic response and potentially the preferred valve configuration. Testing the framework across multiple demand patterns and uncertain-demand realizations is therefore recommended for future work.

The closed-loop evaluation was limited to a single 72 h window to keep the computational cost of the nested GA-DeePC search tractable. Although this common window enables a consistent comparison among candidate configurations, longer evaluation periods could influence their relative performance. Extending the assessment to longer operating periods is therefore recommended for future work.

Spatial performance was evaluated at selected high-elevation nodes as a practical representation of challenging service areas rather than at every network node. Future work should examine whether the selected solutions remain favorable under network-wide pressure compliance, additional monitored nodes, and abnormal hydraulic conditions. The framework should also be tested on operational networks, where uncertain demands, sensor noise, actuator limitations, and model mismatch may affect both the identification data and closed-loop response. Robust and distributionally robust DeePC formulations provide possible directions for treating noisy or uncertain input-output data while maintaining constraint and performance guarantees [16,33].

Taken together, the case studies indicate that the preferred PRV configuration is case-specific and depends on network topology, existing source PRVs, monitored and spatial service locations, the available search budget, and the relative weighting of hydraulic and operational criteria.

## 5. Conclusions

This study developed and evaluated a coupled GA-DeePC framework for optimizing the placement of internal pressure-reducing valves in water distribution networks. The Genetic Algorithm was used as an external design layer to search over discrete PRV-location configurations, while each candidate was evaluated through a complete closed-loop DeePC pressure-control experiment. Accordingly, PRV placement was evaluated in terms of the closed-loop performance achievable by the data-driven controller rather than through hydraulic or topological criteria alone.

The Fossolo case study demonstrated the feasibility of the proposed framework in a relatively small network. The GA progressively identified stronger three-PRV configurations, while the simultaneous reduction in the best and median MAE and the decrease in population diversity indicated convergence of the search. Re-evaluation of the final population using local, spatial, and operational metrics showed that the configuration with the minimum monitored-node MAE also achieved the best combined-objective score. Thus, for this network and the selected objective weights, strong local pressure tracking was aligned with the broader operational criteria.

The Modena case study extended the framework to a larger search problem and compared configurations containing two to six added internal PRVs. The leading solution from each fixed-K search was re-evaluated using a global objective with consistent normalization. The K = 2 configuration achieved the best global objective value of 0.547 and provided the most favorable overall compromise among monitored-node tracking, spatial pressure behavior, pressure-bound violations, valve actuation effort, and configuration complexity. However, it did not achieve the best value for every individual metric: K = 4 produced the lowest MAE, whereas K = 5 produced the lowest spatial STD. This contrast demonstrates that a single performance metric may favor a different configuration from that selected through an integrated operational assessment.

Overall, the results show that PRV placement and valve count are central design variables in DeePC-based pressure management. Increasing the number of available control valves increases the controller's degrees of freedom, but it does not necessarily improve the overall performance, because it also enlarges the placement search space and may increase control coordination requirements and actuation variability. The preferred configuration therefore depends on the hydraulic topology, the existing source PRVs, the monitored and spatial service locations, and the relative importance assigned to tracking accuracy, pressure compliance, and operational effort. These findings demonstrate the feasibility of the proposed GA-DeePC framework but should not be interpreted as evidence of superiority over alternative optimization or control approaches.

The identified configurations should be interpreted as the best solutions found within the performed GA searches rather than as proven global optima. Future work should examine additional random seeds, larger search budgets, alternative local-objective weights, and network-wide pressure evaluation under different demand and boundary conditions. A true multi-objective formulation could also be used to generate Pareto-optimal trade-offs instead of relying on a single weighted score. Further developments may include parallel or surrogate-assisted candidate evaluation, hybrid refinement of GA solutions, robust DeePC formulations, and validation on operational networks with uncertain demands, measurement noise, and realistic actuator limitations. These extensions would support the development of a scalable controller-in-the-loop framework for jointly designing and operating pressure-control infrastructure in water distribution systems.

**Author Contributions:** Conceptualization, J.D.; methodology, J.D.; software, J.D.; validation, J.D.; formal analysis, J.D.; investigation, J.D.; data curation, J.D.; writing—original draft preparation, J.D.; writing—review and editing, J.D. and A.O.; visualization, J.D.; supervision, A.O.; funding acquisition, A.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Israeli Water Authority under project number 2033800, and by the Bernard M. Gordon Center for Systems Engineering at the Technion.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| DeePC | Data-Enabled Predictive Control |
| GA | Genetic Algorithm |
| MAE | Mean Absolute Error |
| MPC | Model Predictive Control |
| PRV | Pressure-Reducing Valve |
| RTC | Real-Time Control |
| WDS | Water Distribution System |
| DDA | Demand-Driven Analysis |
| PDA | Pressure-Driven Analysis |

## Appendix A. Computational Settings and Reproducibility

**Table A1.** Hydraulic and DeePC Settings Used in the Case Studies.

| Parameter | Fossolo | Modena |
|---|---|---|
| Monitored node | Node 21 | Node 201 |
| Pressure reference | 30 m | |
| Identification samples | 600 | |
| Control duration | 72 h | |
| Control time step | 1 h | |
| Tini | 48 h | |

**Table A1.** *Cont.*

| Parameter | Fossolo | Modena |
|---|---|---|
| Prediction horizon N | 12 h | |
| PRV-setting bounds (fixed and added internal PRVs) | 20–70 m | |
| λg | $10^{-2}$ | |
| λy | $10^{1}$ | |
| λu | $10^{-2}$ | |
| Pressure evaluation bounds | 25–60 m | |
| Spatial evaluation nodes | 7, 24, 8, 9, 36 | 74, 73, 75, 15, 217, 76, 72 |
| Demand model | PDA | DDA |
| Temporal demand representation | Predefined hourly time-varying pattern | Constant nodal base demands |

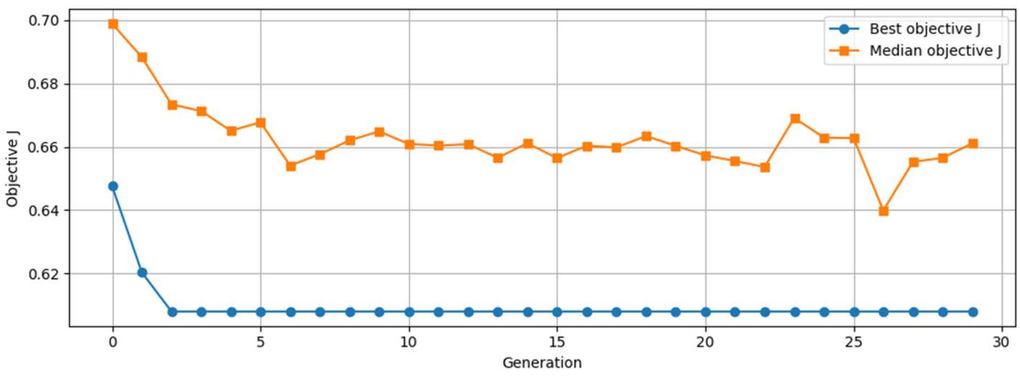
**Table A2.** Genetic-Algorithm Settings and Computational Budgets.

| Case | Added PRVs | Population Size | Generations | Tournament Size | Crossover Probability | Nominal Candidate Evaluations |
|---|---|---|---|---|---|---|
| Fossolo | 3 | 40 | 25 | 3 | 0.9 | 1000 |
| Modena | 2 | 40 | 25 | 3 | | 1000 |
| | 3 | 50 | 30 | 3 | | 1500 |
| | 4 | 70 | 90 | 3 | | 6300 |
| | 5 | 90 | 120 | 3 | | 10,800 |
| | 6 | 110 | 160 | 4 | | 17,600 |
| **Case** | **Added PRVs** | **Gene Mutation Probability** | **Mutation Step** | **Elite** | **Actuation-Effort Normalization Constant ($N_{L1,K}$)** | |
| Fossolo | 3 | 0.3 | 6 | 1 | - | |
| Modena | 2 | 0.3 | 6 | 1 | 8000 | |
| | 3 | 0.3 | 6 | 1 | 9333 | |
| | 4 | 0.2 | 5 | 2 | 10,667 | |
| | 5 | 0.18 | 5 | 3 | 12,000 | |
| | 6 | 0.15 | 5 | 3 | 13,333 | |

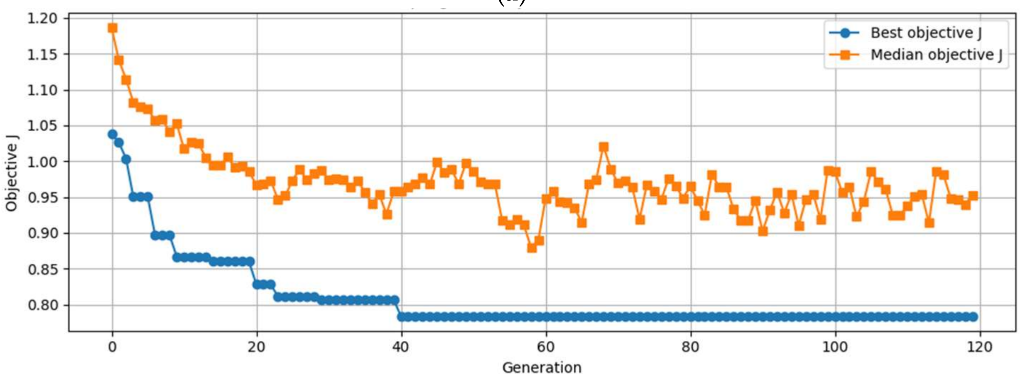
## Appendix B. Additional Modena GA Results

**Table A3.** Selected Internal PRV Locations for the Leading Modena Solutions.

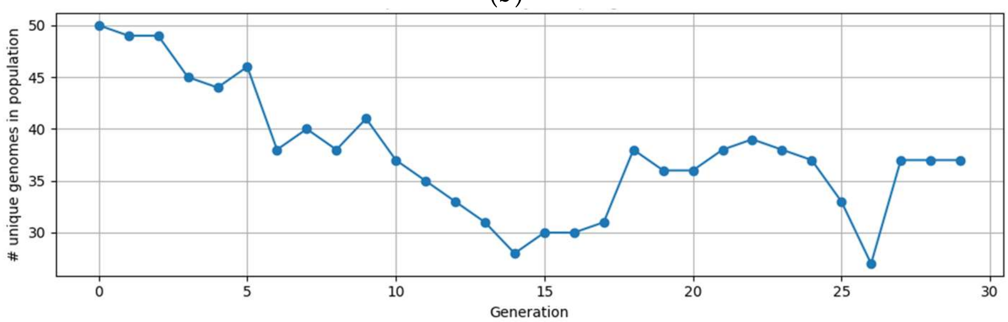
| Added PRVs | Selected Pipe IDs |
|---|---|
| 2 | 309, 175 |
| 3 | 6, 165,76 |
| 4 | 227, 273, 186, 156 |
| 5 | 163, 49, 297, 256, 20 |
| 6 | 164, 58, 149, 175, 193, 68 |



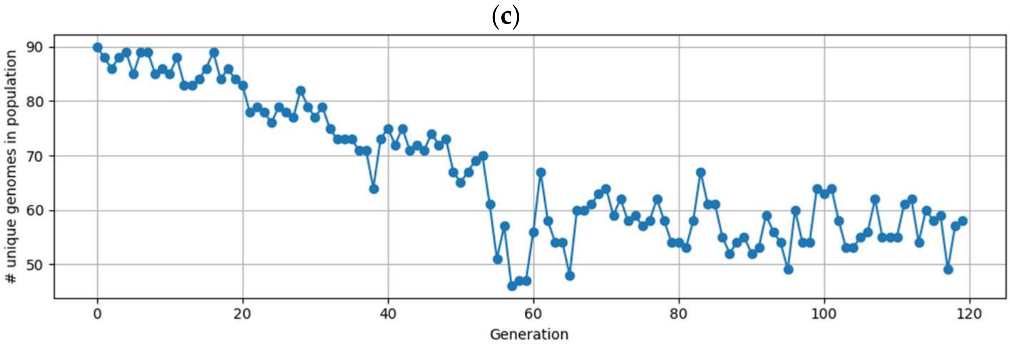

(**a**)




(**b**)

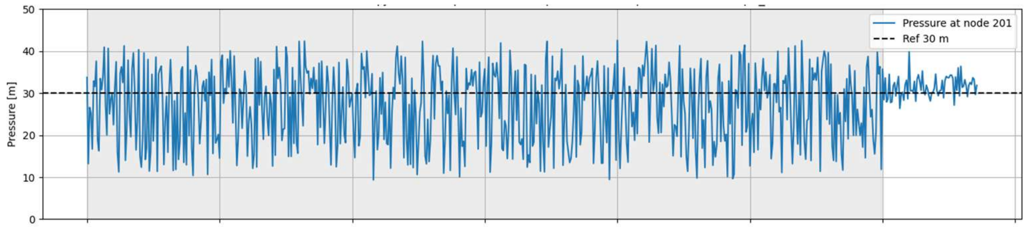



(**c**)

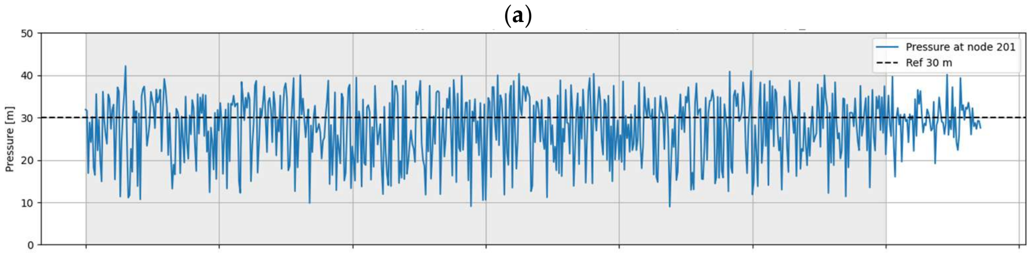



(**d**)

**Figure A1.** Genetic-algorithm convergence and population diversity for the intermediate Modena cases. Panels (**a**,**b**) show the best and median local objective values over generations for K = 3 and K = 5, respectively. Panels (**c**,**d**) show the corresponding number of unique genomes in each generation. The local objective values are intended for evaluating convergence within each fixed-K run and should not be compared directly across different values of K.




(**a**)




(**b**)

**Figure A2.** Pressure response at the monitored node in the Modena network for the leading solutions obtained with (**a**) K = 3 and (**b**) K = 5 added internal PRVs. The grey region represents the randomized identification phase, the white region represents the closed-loop control phase, and the dashed horizontal line indicates the 30 m pressure reference.

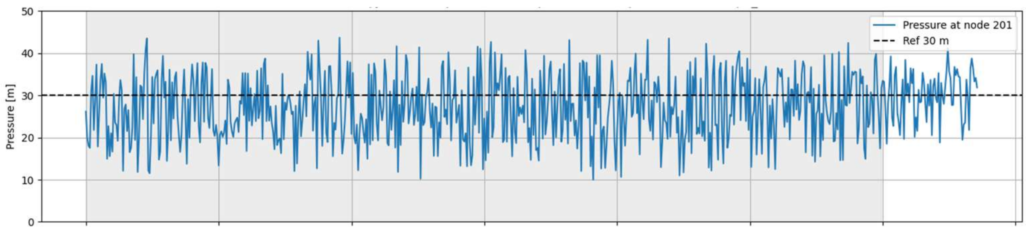



**Figure A3.** Closed-loop DeePC response for a poorly performing Modena configuration with K = 2 added internal PRVs located on Pipes 158 and 229. The grey region represents the randomized identification phase and the white region represents the closed-loop control phase. The figure shows the pressure response at monitored Node 201 relative to the 30 m reference. This configuration produced poorer closed-loop performance than the selected K = 2 configuration shown in Figure 11, with a monitored-node MAE of 4.806 m and a pressure-bound violation rate of 21.429%.

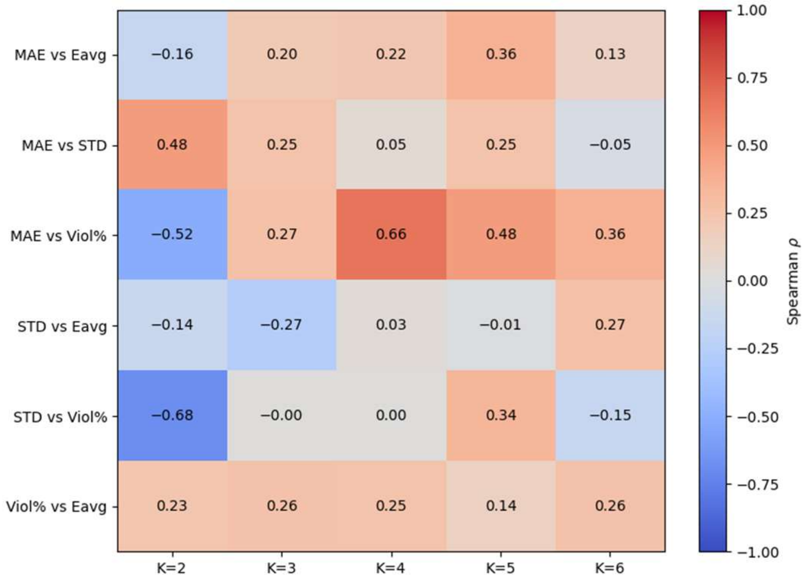



**Figure A4.** Pairwise Spearman rank-correlation coefficients between the performance metrics in the final GA populations of the Modena case for K = 2 to K = 6 added internal PRVs. Correlations were calculated separately within each fixed-K final population after removing duplicate genomes. The corresponding numbers of unique solutions were 15, 37, 40, 58, and 61 for K = 2, 3, 4, 5 and 6, respectively. Positive coefficients indicate that the ranked values of the two metrics tend to increase together, whereas negative coefficients indicate an inverse ranked relationship.

## References

1. Abd Rahman, N.; Muhammad, N.S.; Wan Mohtar, W.H.M. Evolution of Research on Water Leakage Control Strategies: Where Are We Now? *Urban Water J.* **2018**, *15*, 812–826. [CrossRef]
2. Cavazzini, G.; Pavesi, G.; Ardizzon, G. Optimal Assets Management of a Water Distribution Network for Leakage Minimization Based on an Innovative Index. *Sustain. Cities Soc.* **2020**, *54*, 101890. [CrossRef]
3. Price, E.; Abhijith, G.R.; Ostfeld, A. Pressure Management in Water Distribution Systems through PRVs Optimal Placement and Settings. *Water Res.* **2022**, *226*, 119236. [CrossRef]
4. Huzsvár, T.; Wéber, R.; Szabó, M.; Hős, C. Optimal Placement and Settings of Valves for Leakage Reduction in Real Life Water Distribution Networks. *Water Resour. Manag.* **2023**, *37*, 4949–4967. [CrossRef]
5. Korder, K.; Cao, H.; Salomons, E.; Ostfeld, A.; Li, P. Simultaneous Minimization of Water Age and Pressure in Water Distribution Systems by Pressure Reducing Valves. *Water Resour. Manag.* **2024**, *38*, 3561–3579. [CrossRef]
6. Rajakumar, A.G.; Cornelio, A.A.; Mohan Kumar, M.S. Leak Management in District Metered Areas with Internal-Pressure Reducing Valves. *Urban Water J.* **2020**, *17*, 714–722. [CrossRef]
7. Ferrarese, G.; Medoukali, D.; Mirauda, D.; Malavasi, S. Review of Metaheuristic Methodologies for Leakage Reduction and Energy Saving in Water Distribution Networks. *Water Resour. Manag.* **2024**, *38*, 3973–4001. [CrossRef]
8. Dai, P.D. A Real Time Optimization Based Sequential Convex Program for Pressure Management in Water Distribution Systems. *Water Resour. Manag.* **2023**, *37*, 4751–4768. [CrossRef]
9. Galuppini, G.; Creaco, E.; Magni, L. A Gain Scheduling Approach to Improve Pressure Control in Water Distribution Networks. *Control Eng. Pract.* **2020**, *103*, 104612. [CrossRef]
10. Galuppini, G.; Creaco, E.F.; Magni, L. Multinode Real-Time Control of Pressure in Water Distribution Networks via Model Predictive Control. *IEEE Trans. Control Syst. Technol.* **2023**, *31*, 2201–2216. [CrossRef]
11. Galuppini, G.; Creaco, E.; Magni, L. Real-Time Pressure Control in Water Distribution Networks: Stability Guarantees via Gain-Scheduled Internal Model Control. *IEEE Trans. Control Syst. Technol.* **2024**, *32*, 731–743. [CrossRef]
12. Perelman, G.; Ostfeld, A. Data Enabled Predictive Control for Water Distribution Systems Optimization. *Water Resour. Res.* **2025**, *61*, e2024WR039059. [CrossRef]
13. Coulson, J.; Lygeros, J.; Dörfler, F. Data-Enabled Predictive Control: In the Shallows of the DeePC. In *Proceedings of the 2019 18th European Control Conference (ECC), Naples, Italy, 25–28 June 2019*; IEEE: New York, NY, USA, 2019; pp. 307–312.
14. Willems, J.C.; Rapisarda, P.; Markovsky, I.; De Moor, B.L.M. A Note on Persistency of Excitation. *Syst. Control Lett.* **2005**, *54*, 325–329. [CrossRef]
15. Berberich, J.; Allgöwer, F. An Overview of Systems-Theoretic Guarantees in Data-Driven Model Predictive Control. *Annu. Rev. Control Robot. Auton. Syst.* **2025**, *8*, 77–100. [CrossRef]
16. Huang, L.; Zhen, J.; Lygeros, J.; Dörfler, F. Robust Data-Enabled Predictive Control: Tractable Formulations and Performance Guarantees. *IEEE Trans. Autom. Control* **2023**, *68*, 3163–3170. [CrossRef]
17. Elokda, E.; Coulson, J.; Beuchat, P.N.; Lygeros, J.; Dörfler, F. Data-enabled Predictive Control for Quadcopters. *Int. J. Robust Nonlinear Control* **2021**, *31*, 8916–8936. [CrossRef]
18. Davda, J.; Ostfeld, A. DeePC Sensitivity for Pressure Control with Pressure-Reducing Valves (PRVs) in Water Networks. *Water* **2026**, *18*, 253. [CrossRef]
19. Parvaze, S.; Kumar, R.; Khan, J.N.; Al-Ansari, N.; Parvaze, S.; Vishwakarma, D.K.; Elbeltagi, A.; Kuriqi, A. Optimization of Water Distribution Systems Using Genetic Algorithms: A Review. *Arch. Comput. Methods Eng.* **2023**, *30*, 4209–4244. [CrossRef]
20. Chen, W.-N.; Jia, Y.-H.; Zhao, F.; Luo, X.-N.; Jia, X.-D.; Zhang, J. A Cooperative Co-Evolutionary Approach to Large-Scale Multisource Water Distribution Network Optimization. *IEEE Trans. Evol. Comput.* **2019**, *23*, 842–857. [CrossRef]
21. Ulusoy, A.-J.; Mahmoud, H.A.; Pecci, F.; Keedwell, E.C.; Stoianov, I. Bi-Objective Design-for-Control for Improving the Pressure Management and Resilience of Water Distribution Networks. *Water Res.* **2022**, *222*, 118914. [CrossRef]
22. Ulusoy, A.-J.; Stoianov, I. Hybrid Evolutionary-Exact Optimization Method for the Bi-Objective Design-For-Control of Water Distribution Networks. *Water Resour. Res.* **2026**, *62*, e2025WR040688. [CrossRef]
23. Mohamad Shirajuddin, T.; Muhammad, N.S.; Abdullah, J. Optimization Problems in Water Distribution Systems Using Non-Dominated Sorting Genetic Algorithm II: An Overview. *Ain Shams Eng. J.* **2023**, *14*, 101932. [CrossRef]
24. Wang, Y.; Weyer, E.; Manzie, C.; Simpson, A.R.; Blinco, L. Stochastic Co-Design of Storage and Control for Water Distribution Systems. *IEEE Trans. Control Syst. Technol.* **2025**, *33*, 274–287. [CrossRef]

25. Jenks, B.; Ulusoy, A.-J.; Pecci, F.; Stoianov, I. Dynamically Adaptive Networks for Integrating Optimal Pressure Management and Self-Cleaning Controls. *Annu. Rev. Control* **2023**, *55*, 486–497. [CrossRef]
26. Ulusoy, A.-J.; Stoianov, I. Multi-Objective Branch-and-Bound for the Design-for-Control of Water Distribution Networks with Global Bounds. *Struct. Multidiscip. Optim.* **2024**, *67*, 83. [CrossRef]
27. Rossman, L.A.; Woo, H.; Tryby, M.; Shang, F.; Janke, R.; Haxton, T. *EPANET 2.2 User Manual*; U.S. Environmental Protection Agency: Washington, DC, USA, 2020.
28. Kyriakou, M.S.; Demetriades, M.; Vrachimis, S.G.; Eliades, D.G.; Polycarpou, M.M. EPyT: An EPANET-Python Toolkit for Smart Water Network Simulations. *J. Open Source Softw.* **2023**, *8*, 5947. [CrossRef]
29. Gurobi Optimizer, Version 12.0.3. Available online: https://docs.gurobi.com/projects/optimizer/en/12.0/index.html (accessed on 28 August 2026).
30. Hedaiaty Marzouny, N.; Jalili Ghazizadeh, M.; Moslehi, I.; Komeily, M. Application of Pressure Reducing and Throttle Control Valves for Optimal Pressure Management in Water Distribution Networks. *Urban Water J.* **2024**, *21*, 65–79. [CrossRef]
31. Artelt, A.; Giese, K.; Vrachimis, S.G.; Eliades, D.G.; Polycarpou, M.M.; Hammer, B. *The WaterBenchmarkHub: A Platform for Benchmarks in Water Distribution Networks*; The University of Sheffield: Sheffield, UK, 2025.
32. Morani, M.C.; Carravetta, A.; D'Ambrosio, C.; Fecarotta, O. A New Mixed Integer Non-Linear Programming Model for Optimal PAT and PRV Location in Water Distribution Networks. *Urban Water J.* **2021**, *18*, 394–409. [CrossRef]
33. Coulson, J.; Lygeros, J.; Dörfler, F. Distributionally Robust Chance Constrained Data-Enabled Predictive Control. *IEEE Trans. Autom. Control* **2022**, *67*, 3289–3304. [CrossRef]