Article

Improving Water and Energy Resource Management: A Comparative Study of Solution Representations for the Pump Scheduling Optimization Problem

Sergio A. Silva-Rubio [1] , **Yamisleydi Salgueiro** [2] , **Daniel Mora-Meliá** [3,4,*]
and **Jimmy H. Gutiérrez-Bahamondes** [5]

1   Doctorado en Sistemas de Ingeniería, Facultad de Ingeniería, Universidad de Talca, Camino Los Niches Km 1, Curico 3340000, Chile; sergio.silva@utalca.cl
2   Departamento de Ingeniería Industrial, Facultad de Ingeniería, Universidad de Talca, Camino Los Niches Km 1, Curico 3340000, Chile; ysalgueiro@utalca.cl
3   Departamento de Ingeniería y Gestión de la Construcción, Facultad de Ingeniería, Universidad de Talca, Camino Los Niches Km 1, Curico 3340000, Chile
4   Departamento de Ingeniería Hidráulica y Medio Ambiente, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
5   Departamento de Ciencias de la Computación, Facultad de Ingeniería, Universidad de Talca, Camino Los Niches Km 1, Curico 3340000, Chile; jgutierrezb@utalca.cl
*   Correspondence: damora@utalca.cl, damome@upv.edu.es or damora@ut; Tel.: +56-9-51265044

Abstract: Water distribution networks (WDNs) are vital for communities, facing threats like climate change and aging infrastructure. Optimizing WDNs for energy and water savings is challenging due to their complexity. In particular, pump scheduling stands out as a fundamental tool for optimizing both resources. Metaheuristics such as evolutionary algorithms (EAs) offer promising solutions, yet encounter limitations in robustness, parameterization, and applicability to real-sized networks. The encoding of decision variables significantly influences algorithm efficiency, an aspect frequently overlooked in the literature. This study addresses this gap by comparing solution representations for a multiobjective pump scheduling problem. By assessing metrics such as execution time, convergence, and diversity, it identifies effective representations. Embracing a multiobjective approach enhances comprehension and solution robustness. Through empirical validation across case studies, this research contributes insights for the more efficient optimization of WDNs, tackling critical challenges in water and energy management. The results demonstrate significant variations in the performance of different solution representations used in the literature. In conclusion, this study not only provides perspectives on effective pump scheduling strategies but also aims to guide future researchers in selecting the most suitable representation for optimization problems.

Keywords: optimization; solution representation; evolutionary algorithms; multiobjective problem; NSGA-II; pump scheduling; water distribution networks; EPANET

MSC: 90C26; 90C29; 90C31; 90C90; 49Q12; 76D55

1. Introduction

Water distribution networks (WDNs) and their associated infrastructure are considered critical structures due to their vital importance to the well-being of communities. It is forecasted that by 2050, the demand for water systems will intensify as the worldwide population is expected to increase to between 9.4 and 10.2 billion individuals [1]. WDNs face various threats that jeopardize their effective operations. One of these threats is climate change, which leads to extreme weather events such as droughts and floods that impact the availability and quality of water. Additionally, old pipes and treatment systems can experience breakdowns, resulting in leaks, water losses, and supply interruptions. Moreover,

rising energy costs also impact WDNs, as water extraction, treatment, and distribution require significant energy [2]. Electric power is one of the dominant costs for water utilities, consuming approximately 5–7% of the total energy produced worldwide [3]. Consequently, reducing energy consumption levels and conserving available natural resources such as water are among society's challenges [4] and the Water Europe Strategic Research Agenda (SIRA) [5].

The discrete (binary) nature of some variables and the size of the solution space are among the main difficulties encountered when optimizing water- and energy-saving systems. In this context, the use of metaheuristics has proven to be suitable for numerous WDN problems [6]. Researchers have effectively employed evolutionary algorithms (EAs) [7] and multiobjective EAs (MOEAs) [8] to address various WDN problems. These algorithms have demonstrated versatility in handling challenges concerning water resources, including leak detection [9], optimal pipe sizing [10], and water quality optimization [11]. Despite their utility, EAs and MOEAs present notable limitations, such as issues related to robustness, accuracy, and parameterization [12]. Various factors, such as objective functions, problem constraints, algorithmic parameters, and initial conditions, along with the solution representation, can influence the performance of these algorithms [13]. Moreover, many studies have focused primarily on academic cases, which lack practical relevance to real-world complexities, especially concerning the exponential growth in the problem size. Consequently, enhancements and refinements to optimization methods for addressing these shortcomings are pressingly needed, emphasizing the imperative for the development of methodologies and guidelines aimed at enhancing algorithmic efficiency and reducing the search space, which remain significant research challenges in this field. Within the realm of metaheuristics, a highly efficacious strategy for enhancing computational efficiency is to modify the representations of decision variables [14,15]. Properly configuring such representations simplifies the model formulation process, enabling the achievement of solutions that are both more precise and optimized. However, the literature lacks in-depth investigations into how the representations of solutions directly influence the efficiency of particular EAs.

The electricity consumption linked to the operation of a WDN is primarily attributed to the operational costs of its water pumps. Approximately 80% of the energy usage is dedicated to running motors for pumping [16], which is essential for moving water from collection points to consumers in a manner that satisfies their requirements. In this context, previous research efforts have been directed towards identifying the optimal pumping operation [17]. Pump operations are optimized to enable pumps to consume a minimal amount of energy. Historically, the pump operation process has been formulated as either an implicit or explicit control problem [18]. In the implicit formulation, the decision variables of the optimal control problem are pump flows [19], tank water trigger levels [20], or pump speeds for variable-speed pumps [21]. In the explicit formulation, the decision variables are the numbers of times that pumps operate (pump scheduling).

The pump scheduling problem has historically been preferred by researchers. Explicit pump programming is an approach that is used to switch pumps on and off [22] based on predefined time intervals. Most studies adopt this approach, and its state representations during each time interval have been studied using different encodings that employ binary and integer variables with various levels of discretization [15,18]. Pump scheduling optimization has proven to be a practical and highly effective method for reducing energy consumption without affecting the actual infrastructure of the whole system [23]. Commonly, pump scheduling can be specified by on/off pump switches during predefined equal time intervals [24,25]. The difficulty encountered in this approach is mainly caused by the large number of decision variables required for real WDNs with numerous pump stations, which causes an exponential increase in the size of the solution space. Other pump scheduling methods [26] reduce the number of variables and the search space, leading to better computational efficiency and increasing the chance of identifying higher-quality solutions.

Given the nondeterministic polynomial-time (NP)-hard nature of pump scheduling problems [27], various optimization models using both mono-objective and multi-objective approaches have been proposed. Notably, these studies often select binary or integer formats for solution representation [28,29] but lack detailed justification for these choices and assessments of their impact on model efficacy. To address this gap, our study performs a comparative analysis of various solution representation methods applied to a multiobjective pump scheduling problem with the goals of:

- Minimizing pumping operation costs.
- Enhancing water quality.

This presents a crucial opportunity to thoroughly investigate how different solution representations impact the performance of optimization models. The performance of each representation is evaluated using the following metrics:

- Execution time.
- Convergence.
- Pareto front coverage.
- Diversity.
- Sensitivity.

Experiments are conducted in three case studies using representative datasets; the obtained outcomes are analyzed and compared across different representations to draw conclusions and provide recommendations on which representations might be most suitable for the target problem. Adopting a multiobjective approach allows us to consider multiple objectives simultaneously, providing a holistic evaluation of solutions and facilitating the development of robust strategies. This approach not only enhances understanding but also underscores the significant contributions of our study to the field of optimization.

The remainder of this paper is organized as follows: Section 2 introduces the proposed methodology, detailing the problem statement and exploring the influence of data structure representations on the performance of optimization algorithms when applied to pump scheduling problems. Section 3 presents the formulation of the comparison method, the computational environment, and the case studies examined. Section 4 delivers the results and discussion, initially focusing on the efficiency of the nondominated sorting genetic algorithm (NSGA)-II across the reviewed case studies, comparing the tested solution representations through their approximate Pareto fronts, and assessing these representations in terms of diversity and convergence according to various metrics. Finally, Section 5 offers conclusions, summarizing the key findings and their implications for future research.

2. Model Outline and Encoding Decision Variables in the Pump Scheduling Problem

2.1. Model Outline and Mathematical Notations

Pump programming is often formulated as a cost optimization problem [30,31], which aims to minimize the operational costs involved in transporting potable water and have the pumps consume a minimal amount of energy. This work proposes a multiobjective optimization method that simultaneously considers energy costs and water quality. The goal is to find the best pumping program in a typical operating cycle, minimizing the total operational costs while ensuring competent network service without compromising the quality of the water supply.

When formulating the mathematical optimization problem, we define decision variables that correspond to the operational decisions of the pumps, specifically their on/off statuses. The number of variables is dependent on both the number of pumps and the number of time intervals. For instance, in a pump program denoted as S, when indicating which pumps will be operational during each time interval, the decision variables are represented as $S(n, t)$, where n refers to the number of pumps and t refers to the time interval.

Two objective functions are defined. The first function, represented by Equation (1), aims to minimize energy costs [30,31], which consist of the sum of the energies consumed

by different pumps during each time interval. Additionally, this cost is influenced by the energy price set by electricity tariffs.

$$Minimize\ C_E = \sum_{n=1}^{NP} \sum_{t=0}^{NT} (P_c(n,t)E_c(n,t)S(n,t)) \tag{1}$$

where NP is the total number of pumps, NT is the number of time intervals (usually in hours), $P_c(n,t)$ is the energy consumption tariff imposed on pump n for each interval t, $E_c(n,t)$ is the energy consumption of pump n in interval t, and $S(n,t)$ is the binary state indicating whether pump n is operating during interval t.

The second objective function is related to improving water quality [17,32,33]. Deterioration in quality is associated with water age (WANET). Consequently, the aim is to minimize the water retention time. The water age is assessed as the weighted average water age based on demand (Equation (2)) and represents the average of the calculated ages, assigning a weight equal to the requested demand at each time step to each node.

$$Minimize\ WANET = \frac{\sum_{n=1}^{ND} \sum_{t=0}^{NT} WA_{n,t} Q_{n,t}}{\sum_{n=1}^{ND} \sum_{t=0}^{NT} Q_{n,t}} \tag{2}$$

where $WA_{n,t}$ is the water age at the nth node at time t, ND is the number of demand nodes in the network, NT represents the number of time intervals, and $Q_{n,t}$ is the water demand requested by node n at time step t.

These two objective functions are in conflict because, while we aim to minimize the energy cost, doing so can decrease pump operations, thereby increasing the water retention time.

In general, two different types of constraints are presented in this problem. The first type consists of hydraulic constraints such as mass and energy conservation constraints, which define the hydraulic equilibrium state of the system. They are presented in Equations (3) and (4), respectively.

$$\sum q_{in} - \sum q_{out} = C_j \tag{3}$$

$$\sum h_f - \sum E_p = 0 \tag{4}$$

Pipe head losses are estimated here using the Hazen–Williams equation (Equation (5)).

$$h_f = \frac{10.67 L_q^{1.85}}{CH^{1.85} D^{4.87}} \tag{5}$$

where q_{in} and q_{out} are the inflows and outflows at a node, respectively, C_j is the consumption at node j, h_f is the head loss due to friction, CH is the Hazen–Williams coefficient, L is the length of the pipe, and D is the diameter of the pipe.

The second type includes limit restrictions and represents system performance criteria. For example, maintaining a minimum pressure level at nodes (Equation (6)) is essential for ensuring optimal water flow and guaranteeing sufficient access, and this is supported by adequate pressure. This constraint is established to ensure that the pressure at each node does not drop below a predetermined value.

$$h_{i,t} \geq h_i^{min} \tag{6}$$

where $h_{i,t}$ is the pressure at node i during time t and h_i^{min} is the minimum pressure at node i.

Moreover, limiting the maximum flow rate of the pumps is crucial for preventing the distribution system from becoming overloaded and avoiding pump damage due to excessive operation. This contributes to maintaining safe and efficient pump operations.

$$Q_{i,t} \leq Q_i^{max} \tag{7}$$

where $Q_{i,t}$ is the flow rate of pump i during time t and Q_i^{max} is the maximum flow rate supported by pump i.

Additionally, when analyzing the water levels in the tanks after the optimization period, they must be at least the same as those at the beginning of the process (Equation (8)). Finally, maintaining appropriate water levels in the tanks is necessary to ensure constant availability and prevent both emptying and overflowing (Equation (9)), thus contributing to the stability of the system.

$$TL_{i,NT} \geq TL_{i,0} \tag{8}$$

$$TL_{i,min} \leq TL_{i,NT} \leq TL_{i,max} \tag{9}$$

where $TL_{i,NT}$ is the level of tank i during period NT, $TL_{i,0}$ is the level of tank i during the 0 period, and $TL_{i,NT}$ is the level of tank i during the final period. $TL_{i,min}$ and $TL_{i,max}$ are the minimum and maximum levels, respectively, for tank i.

2.2. Encoding Decision Variables

The representation of a solution in the formulation of an optimization problem refers to how the information describing a potential solution to the problem is structured [34]. Each decision variable is encoded so that it can be manipulated and evaluated by the employed optimization algorithm.

Common representations include binary, integer, and real-valued encodings [35]. Furthermore, the chosen representations determine how parameters and operators are used, which in turn affects the performance of the search process. The choice of a suitable representation type may have a considerable impact on the effectiveness and efficiency of the optimization algorithm used, significantly affecting its ability to effectively explore and exploit the search space.

In the context of the pump scheduling optimization problem, the use of binary or integer encodings varies significantly depending on the approach utilized and the desired solution type. The number of decision variables in an evolutionary algorithm can influence its performance, though it is not necessarily directly correlated with better outcomes. A higher number of decision variables can increase problem complexity, requiring more computational resources and time, potentially impeding algorithm convergence and optimal solution finding. Conversely, a lower number of variables may lead to an underspecified problem, limiting the algorithm's effectiveness in finding solutions. Striking a balance between an appropriate number of variables and efficient representation is often crucial for optimal performance in an evolutionary algorithm. Historically, binary encoding has been predominantly favored in situations where decisions involve representing on/off states directly, such as in the context of operating pumps at various time intervals. Therefore, an abundance of research has been conducted on binary encoding in this specific context due to its intuitive nature and efficiency. On the other hand, integer encoding is better suited for more complex contexts where decisions cover multiple levels or states in the operations of the pumps or when more detailed time or sequencing considerations are needed. Despite the extensive literature on this subject, empirical evidence that definitively indicates which of the two encodings is superior in terms of efficiency, accuracy, and applicability in different pump scheduling contexts is notably lacking. A detailed comparative study would not only fill this gap in the existing research but also provide practical guidance for those involved in optimizing pumping systems, helping them to choose the most appropriate encoding strategy based on the specific characteristics of their problem.

In this context, this article compares five types of representations that have been used in the literature for the pump scheduling problem.

- Binary Representation (bin) [24–26,31]: This strategy is used to represent the pump states observed in each time interval using accepted values s, where $s \in [0, 1]$. Here, 0 and 1 are representative of the off and on states, respectively. The size of the solution vector is determined by the number of time intervals (NT) and the number of

pumps (NP), and it is calculated as $NT \cdot NP$. Additionally, the search space for this representation type equals $2^{(NT \cdot NP)}$.

- Integer Representation (int) [36]: The operations of pumps are represented by integers, which are in the range $\{x \in \mathbb{Z} \mid 0 \leq x < 2^{NP}\}$, where NP is the number of pumps. Once the valid values are defined, a conversion from each integer to its binary equivalent is performed to represent the state of each pump. For example, if we have a time interval NT_i and obtain the corresponding integer value, we convert it to a binary number. Each bit of the binary number is used to define the state of each pump in the interval NT_i. In this representation, the size of the solution space is the same as that in the binary representation, i.e., $2^{(NT \cdot NP)}$, and the size of the solution vector is equal to NT.

- Restricted Formulation (int_r) [18,37]: In this variant, the decision variables represent the start and end times of pump operations, and they are bounded between 0 (pump off) and Δt (the duration of the time interval). To determine the number of decision variables, the ranges of the time intervals (NTR) are defined. For example, if we consider a total of 24 h and define NTR as 4, we have a total of 6 decision variables for each set of pumps (NP). In general, the formula for calculating the total number of decision variables is $(NT/NTR) \cdot NP$, and the total search space is $(NTR \cdot 2 + 1)^{((NT/NTR) \cdot NP)}$, where $NTR \equiv 0 \ (mod \ NT)$.

- Absolute Time-Controlled Triggers (int_at) [15]: In this representation strategy, the decision variables are absolute times, meaning that each decision variable represents the time elapsed from the start of the scheduling period until the point at which the status of a pump changes. A pair of decision variables represents the operating interval during which the associated pump is active. This representation approach allows for scheduling the turning on and turning off of pumps at specific times, and a maximum change limit (SW) must be defined. The total number of decision variables is $(SW \cdot 2) \cdot NP$, and the size of the search space is $(NT)^{((SW \cdot 2) \cdot NP)}$.

- Relative Time-Controlled Triggers (int_rt) [15]: For decision variables that represent relative time intervals, each pair signifies the duration from the beginning of the scheduling period to the first state change exhibited by the corresponding pump. In other words, they denote the periods of inactivity and activity for a pump, respectively. The number of decision variables is $(SW \cdot 2) \cdot NP$, and $\sum_{i=1}^{(SW \cdot 2) \cdot NP} x_i \leq 24$ must be satisfied, where x_i represents each decision variable in the vector.

Additional details on how to calculate the number of decision variables in each representation, along with specific examples demonstrating the process, are available in the Supplementary Materials.

3. Materials and Methods

3.1. Solution Representation Comparison Methodology

The aim of this study is to conduct a comprehensive comparison between different solution representations for a given multiobjective optimization problem. A multiobjective optimization problem can be mathematically defined as:

$$\begin{aligned} Minimize \ F(\mathrm{x}) &= (f_1(\mathrm{x}), \dots, f_m(\mathrm{x}))^T \\ subject \ \mathrm{x} &\in \Omega \end{aligned} \tag{10}$$

where Ω is the (non-empty) decision space and $\mathrm{x} \in \Omega$ is the decision vector. $F(\mathrm{x})$ consists of $m \geq 2$ conflicting objective functions $f_i : \Omega \to \mathbb{R}, i = 1, \dots, m$ where $\mathbb{R}_m$ is the objective space.

To achieve this goal, a rigorous methodology was developed and applied to assess the performance of each representation strategy through a comprehensive set of performance indicators.

The first step is to run the optimization algorithm. To ensure the reliability of our findings, thirty independent experiments were conducted. This approach enables us to assess the consistency of the outcomes achieved with each solution version across various quality indicators. The consistency of the results is crucial for confirming the reliability and replicability of this study. The data acquired from these trials are presented in the tables in the Results section, highlighting the median values obtained for each quality indicator associated with the different solution representations. This allows for an effective comparison between the different representation variants and a determination of their stability throughout the thirty experiments.

Then, the second step of the methodology is to define a real or reference Pareto front. In our optimization problem (pump scheduling), the true Pareto front was unknown. Consequently, this analysis was carried out by constructing an approximate Pareto front derived from the amalgamation of all solutions from all simulations conducted under the various representations considered. Within such a set, some solutions improve one or more objectives yet exhibit not-so-great values in the rest of the objectives. These solutions can be formalized using the following definitions:

Definition 1. *A vector $u = (u_1, \ldots, u_m)^T$ strongly dominates another vector $v = (v_1, \ldots, v_m)^T$, denoted as $u \prec v$, iff $\forall i \in \{1, \ldots, m\}$, $u_i < v_i$.*

Definition 2. *A vector $u = (u_1, \ldots, u_m)^T$ weakly dominates another vector $v = (v_1, \ldots, v_m)^T$, denoted as $u \preccurlyeq v$, iff $\forall i \in \{1, \ldots, m\}$, $u_i \leq v_i$ and $\exists j \in \{1, \ldots, m\}$ such that $u_j < v_j$.*

Definition 3. *A feasible solution $x^* \in \Omega$ of Equation (7) is called a Pareto optimal solution, if $\nexists y \in \Omega$ such that $y \preccurlyeq x^*$. The set of all the Pareto optimal solutions is called the Pareto set (PS), denoted as: $PS = \{x^* \in \Omega \mid \nexists y \in \Omega, \, y \preccurlyeq x^*\}$.*

Definition 4. *The image of the Pareto set in the objective space is called the Pareto front (PF): $PF = \{F(x) \mid x \in PS\}$.*

Figure 1 exemplifies the process of constructing an approximate Pareto front. This procedure enables us to inspect a wide range of efficient solutions and systematically compare them.
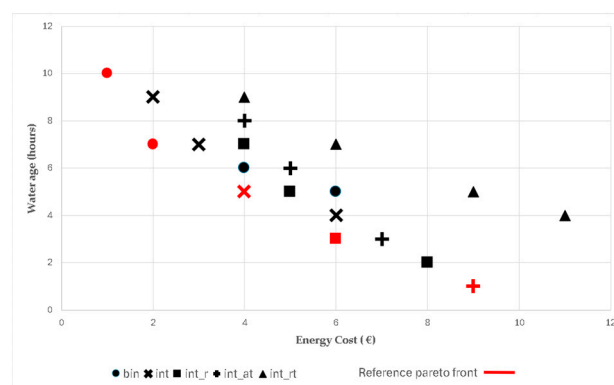

Figure 1. Construction of an approximate Pareto front.

Initially, all non-dominated solutions generated by each representation used in the case studies were collected. These points are represented in black in the figure. Among all non-dominated solutions, those that are not dominated by any other solutions within the set are selected to form the reference Pareto front. These points are highlighted in red in the figure. It is important to note that some representations may contribute one or more points to the reference Pareto front, while others may not contribute any. This is due to the variability in the performance of different representations in exploring the solution space.

Finally, the last stage involves calculating the performance indicators. In the realm of multiobjective optimization, unlike single-objective optimization, it is impractical to rely on a single performance evaluation metric due to the inherently complex nature of comparing solutions that simultaneously address multiple objectives. This complexity stems from the need to evaluate various dimensions concurrently. Therefore, the performance indicators selected for this evaluation focus on three fundamental areas: the computational time, diversity, and convergence of the algorithm.

On the one hand, computational time is a fundamental indicator that evaluates the practical efficiency of each solution representation in terms of the resources consumed during the search process. Its importance is especially highlighted in the context of large-scale problems, where time and computational resource constraints can have substantial impacts on the feasibility of proposed approaches. In this study, all experiments were conducted using a computer system equipped with two 2.00-GHz Intel Xeon Gold 6330 CPUs, each with 28 cores and 56 execution threads. This environment contains 256 GB of RAM and offers a storage capacity of 2.7 TB on an HDD and 894 GB on an SSD. To achieve enhanced performance, a multiprocessing parallelization module was implemented to facilitate the creation of secondary processes for parallel task execution, fully leveraging the processing resources available in the multiple cores and threads of each processor. The results section addresses the practical efficiency of each solution representation in terms of the resources consumed during the search process, focusing on execution time as a key indicator.

On the other hand, evaluating diversity and convergence is essential for ensuring a wide range of alternatives. Diversity is crucial for preventing premature convergence towards locally optimal areas of the solution space, allowing for more thorough explorations. It is considered an indispensable metric when determining the effectiveness of the tested algorithms. In the multiobjective optimization literature, numerous metrics have been designed to assess the diversity of the solutions generated by algorithms. The choice of the most appropriate metric depends on the specific study objectives and the characteristics of the problem under consideration, with several metrics commonly used to obtain more robust and comprehensive analyses.

In our study, we not only aimed for solutions that exhibit variability but also strived to make them as close as possible to the reference Pareto front. Therefore, we adopted the inverted generational distance plus (IGD+) [38] metric, which measures diversity relative to a reference set, thereby providing a comprehensive perspective of the dispersion of the observed solutions along the reference Pareto front. The mathematical formulation of IGD+ is based on calculating the distance between the set of solutions obtained by a multiobjective optimization algorithm and a reference Pareto front. The results of the IGD+ index offer deep insights into the performance of multiobjective optimization algorithms, where low IGD+ values indicate notable proximity between the generated solution front and the reference Pareto front, thus demonstrating high solution quality in terms of convergence and diversity. Conversely, high IGD+ values suggest a significant discrepancy between both fronts, which can be interpreted as an indication that the tested algorithm is not converging effectively or that the generated solutions lack the necessary diversity.

Convergence evaluates the proximity of the obtained solutions to the real or reference Pareto front. This indicator is essential for determining the degree of effectiveness with which a method can approximate the optimal solutions to multiobjective problems. In the context of pump scheduling, convergence ensures that the solutions found are optimal in terms of their energy costs and water quality levels. Typically, the Epsilon [39] and hypervolume [40] metrics serve as relevant measures of convergence in multiobjective optimization experiments.

The Epsilon metric provides a measure of how much a set of solutions needs to improve to reach another reference set, or the true Pareto front, quantifying the distance of the solutions from the Pareto front. It is fundamental for assessing how close the solutions found are to the Pareto front, indicating the quality of the approximation. Epsilon is

interpreted by considering its numerical value in relation to the reference Pareto front. An Epsilon value equal to zero signifies that the generated solution front is identical to the reference Pareto front; hence, the closeness of Epsilon to zero implies better convergence and solution quality.

The hypervolume (HV) indicator captures the size of the objective space dominated by the solutions, enabling an assessment of the coverage extension within the objective space. Indirectly, it also provides diversity information by calculating the volume occupied by the proposed solutions. Notably, while Epsilon focuses more on the proximity of the solution set to the Pareto front, HV captures both the proximity and dispersion of solutions; the latter is highly valuable because of its ability to simultaneously evaluate multiple aspects of algorithmic performance. A higher HV indicates better coverage of the Pareto space, implying closer convergence to the ideal Pareto front and a greater diversity of solutions in the set. Conversely, a lower HV reflects less coverage of the objective space, suggesting that the solutions may be more dispersed or not converging efficiently towards the optimal Pareto front.

As a final step, the Wilcoxon statistical test is applied to analyze the values of the three quality indicators (IGD+, Epsilon, and HV). The Wilcoxon test is a widely recognized and utilized nonparametric statistical tool for comparing two related or paired samples. This test provides a robust statistical framework for determining whether the observed differences in quality indicators are statistically significant or simply a result of the inherent variability of the data. This last step not only adds an additional layer of methodological rigor to the study but also ensures the reliability and validity of the findings by providing solid statistical evidence regarding the performance differences among the various solution representations in the context of multiobjective optimization.

3.2. Computational Environment

To solve the optimization model, a computational method is needed. Specifically, this work uses an elitist NSGA-II [41]. NSGA-II ranks the given population by considering the dominance relationship of each solution. Solutions are organized into different groups, facilitating the identification and selection of higher-quality solutions during the evolutionary process. Nondominated solutions are classified into the first group, and these are followed by those that are surpassed by at least one individual from the previous group, and so on. The choice of NSGA-II is justified because it is one of the most popular algorithms for solving multiobjective optimization problems [42,43]. Additionally, it stands out due to its ability to adapt to the different representation forms considered in this study, facilitating the evaluation of how different encodings influence the resulting performance.

The implementation of NSGA-II is carried out through jMetalPy, an open-source Python library dedicated to addressing both single-objective and multiobjective optimization problems. Drawing inspiration from the Java-based jMetal library, jMetalPy showcases a suite of evolutionary algorithms, local search techniques, and hybrid approaches for various optimization tasks [44]. This work specifically harnesses Python 3.10 as its programming language. The objective function call is implemented according to the guidelines described in Section 3.1. This framework is capable of performing extensive simulations, working in conjunction with a hydraulic network solver for comprehensive analysis purposes. In this regard, hydraulic simulations are carried out using the programmer toolkit of the EPANET software [45]. This is widely used open-source software developed by the U.S. Environmental Protection Agency (EPA) for modeling water distribution systems. The EPANET 2.2 library is efficiently integrated into this parallelization environment; this step is facilitated by the Python library owa-epanet, which serves as a wrapper for the EPANET hydraulic toolkit. This software provides tools for modeling, simulating, and analyzing the hydraulic behavior of RDAs with different pump activation and deactivation patterns.

To ensure a minimum level of statistical confidence in the results, 30 experiments were performed and analyzed. Each independent experiment begins with the prior selection of a solution representation. Subsequently, NSGA-II is implemented, starting with an initial

population of individuals whose quantity is predefined and remains constant throughout the process. This initial population is affected by the operators in each generation and concludes when the established criterion is satisfied, which in this study involves reaching a total number of evaluations. Upon completion of the algorithmic execution, a final population of solutions is obtained. Importantly, this final population may contain repeated individuals, as well as feasible and infeasible solutions. The observed diversity is attributed to both the implementation of the algorithm and the influence of the representation used, reflecting the dynamic process of the solution search.

Regarding the NSGA-II operators, a version of the SBX crossover [46] is used for integer representations, and the SPX crossover [47] is used for binary representations. These operators combine the characteristics of the selected individuals to generate offspring that inherit information from both parents. In terms of mutation, the polynomial mutation [46] is applied in its integer version, and the Bit Flip mutation [47] is applied for integer and binary representations. These operators induce random changes in individuals to explore new solutions in the search space.

Importantly, in the int_at representation, it is noted, following the original article [15], that the crossover and mutation operators have the potential to produce invalid solutions by breaking the ascending order of values. To address this issue, an additional process is introduced to sort the resulting values in ascending order. This measure is implemented to ensure the consistency and validity of the generated sequence. In the same manner, the int_rt representation poses an additional challenge wherein the solutions generated by the operators can violate representation constraints by exceeding the NT scheduling period. In response to this observation, the value of a randomly selected time interval is iteratively reduced by one unit of time. This repair strategy is designed to incrementally adjust the solution until the total sum complies with the NT constraint. Both operator adjustments were proposed and justified in the original article [15] as effective solutions for enhancing the consistency and validity of different representations.

Finally, the behavior of algorithms is conditioned by properly adjusting their parameters. In the case of NSGA-II, population size (P), crossover frequency (Pc), and mutation frequency (Pm) are the main parameters affecting the performance of the algorithm. It is important to highlight that operators such as crossover and mutation must be specifically designed for the chosen representation. For instance, in a binary representation, mutation involves changing one or more bits in the chain. In contrast, in an integer representation, this process entails exchanging integer values between two solutions [35]. Furthermore, in the context of EAs, the number of evaluations (Ne) determines how much time the employed algorithm has to converge towards a solution. An insufficient number of evaluations may result in inadequate exploration of the solution space, whereas a limit that is too high might allow the algorithm to continue iterating without significant improvements, thereby wasting resources. For this study, we utilized values that are commonly recommended in the literature as baselines. To determine the optimal parameter combination, we conducted a grid search. This method assesses a matrix of values for each parameter to identify the set that yields the best performance. The optimal values determined for the different parameters are as follows: P = 300, Pc = 0.9, Pm = 0.05 and Ne = 45,000.

3.3. Case Studies

In EPANET, various key elements serve specific functions: junctions represent connection points in the WDN, such as tanks, reservoirs, and pipe intersections, facilitating water flow and distribution. Pipes connect nodes and represent conduits through which water flows in the network, enabling effective water transportation. Pumps symbolize devices that propel water through the network, crucial for maintaining adequate flow and pressure. Valves are markers representing control devices regulating water flow in specific network sections, allowing precise adjustments in the distribution system. Tanks are symbols depicting water storage tanks in the network, utilized for strategic reserves and pressure stabilization. The reservoirs indicate water sources in the network, like water wells

or external supplies, serving as crucial entry points that feed the water distribution system. These elements play vital roles in modeling and simulating water distribution networks in EPANET, enabling detailed and efficient analysis of hydraulic network operation.

To apply the methodology described above, three case studies were conducted. On the one hand, the Anytown and Anytown Modified WDNs [26] are well-known benchmarking networks and have been tested in some previous works. On the other hand, the Curicó network [48] is a real WDN located in the city of Curicó (Chile). The hydraulic analysis was conducted for one day, dividing the duration into one-hour periods. However, for the int_r representation, the time intervals were adjusted to half-hour periods, as this approach allows such flexibility by definition. Consequently, the demand patterns were also adjusted to align with the half-hour time intervals. The patterns used to characterize the time and cost variations in demand and information about the nodes and pipelines can be found in the Supplementary Materials. A brief description of each case study is provided below.

The Anytown (AT) case study is composed of an infrastructure that includes 41 pipelines, 16 nodes with demand, 1 tank, a supply source and 1 pumping station equipped with 4 pumps. Figure 2 shows the topology of the Anytown WDN called the AT network.
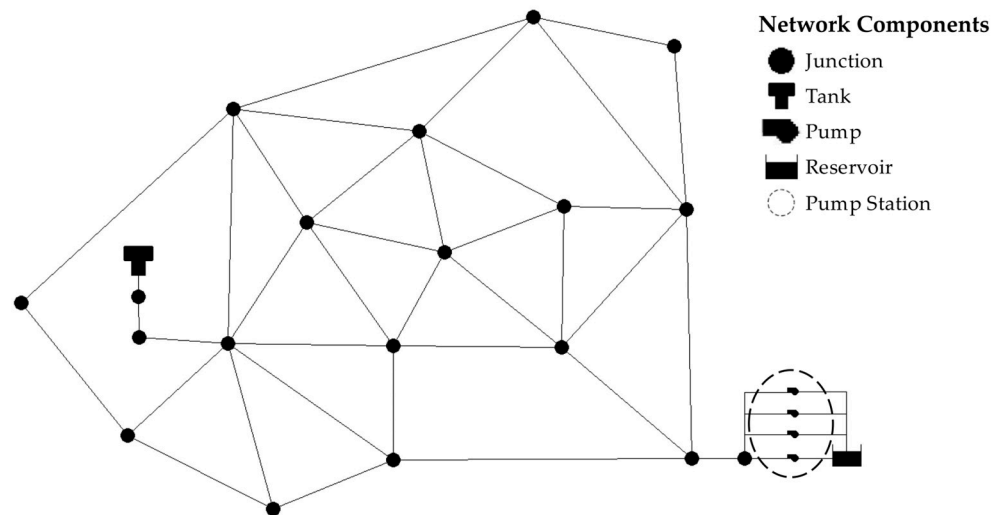

Figure 2. Anytown network.

The modified Anytown (Figure 3) case study consists of 41 pipelines, 19 nodes with demand, 3 tanks, a supply source and 1 pumping station with 3 identical parallel pumps.
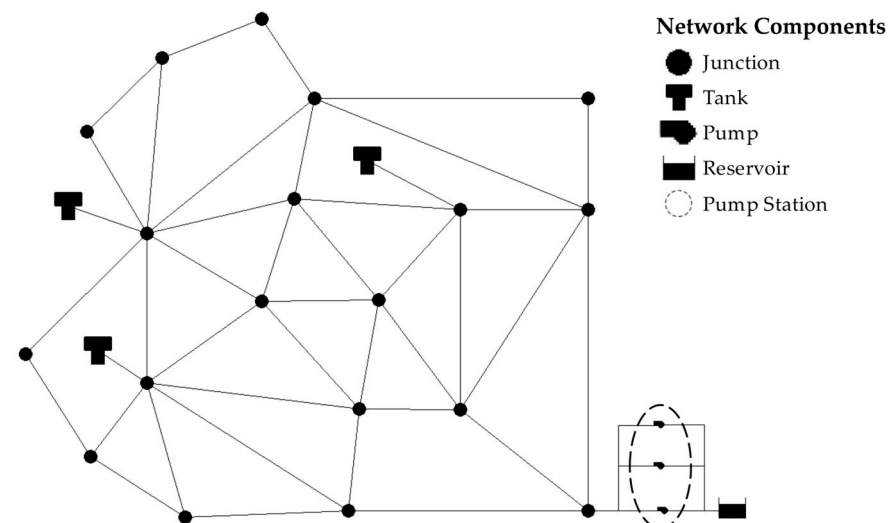

Figure 3. Anytown modified network.

Finally, the Curico WDN is part of the water supply network of the city of Curico (Chile). The network is depicted in Figure 4 and comprises 5217 pipes, 3983 nodes with water demand, 4 tanks, a supply source and 1 pumping station equipped with 5 identical pumps.
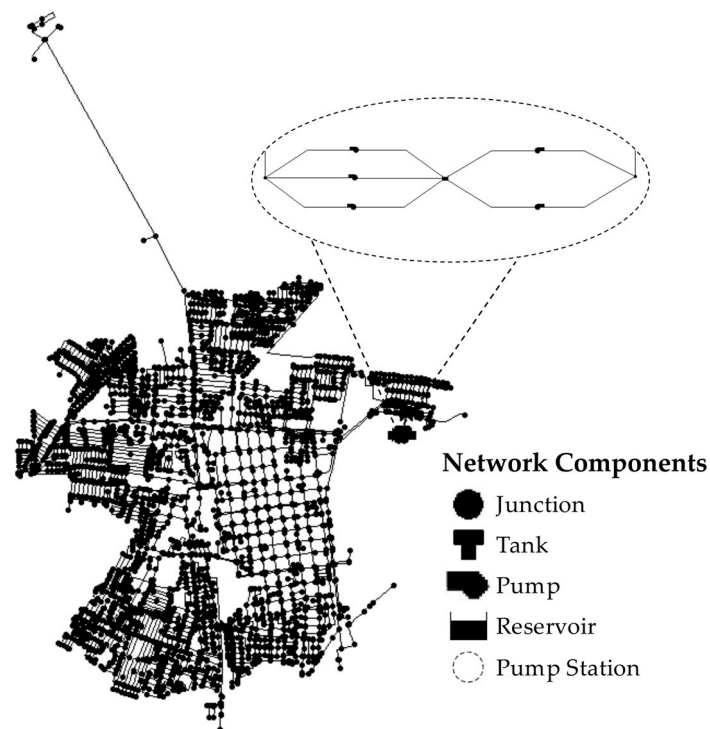

Figure 4. Curico network.

4. Results and Discussion

This section is divided into two subsections. First, Section 4.1 emphasizes the results achieved through NSGA-II, analyzing the numbers of unique, feasible, and nondominated solutions obtained with each solution representation in the three analyzed case studies. Next, an overview of the approximate Pareto diagrams constructed from the set of all solutions obtained for all the utilized representations is provided. This enables us to compare each representation individually with the approximate Pareto front. Section 4.2 delves into solution diversity and convergence evaluations through metrics such as HV, Epsilon (EP), and IGD+. These metrics are crucial for assessing the quality of different representations in various case studies, providing insights into how well these representations perform in addressing the optimization problems at hand. Additionally, a Wilcoxon test is employed to determine the differences between the datasets, shedding light on the statistical significance of the outcomes. Alongside these analyses, a focus on computational efficiency draws attention to the execution times of different representations across diverse networks, offering valuable insights into the trade-offs between speed and solution quality.

4.1. Overview of the Simulation Results

Table 1 encapsulates the findings obtained from the 30 experiments conducted for each representation strategy. It is crucial to note that every experiment comprises 300 independent solutions, resulting in 9000 cumulative simulations per row in the table. The outcomes are presented both as aggregate totals and as averages per experiment. These include the number of unique, feasible, and nondominated solutions for each representation and case study examined. Unique solutions refer to those not replicated within the population and serve as a measure of diversity among the explored solutions. Feasible solutions satisfy all the stipulated constraints imposed on the problem, while nondominated solutions are those

that cannot be simultaneously improved upon across all objectives by any other solution within the experiment.

Table 1. Summary of solutions for each case study.

Network	Representation	Total Solutions			Average Unique Solutions per Experiment (%)	Feasibility (%)	Contribution to the Reference Front (%)
		Unique	Feasible	Nondominated			
Anytown	bin	7743	7173	11	93.7	93.0	0.0
	int	2372	2099	13	27.0	88.6	58.8
	int_r	253	245	9	2.8	96.8	41.2
	int_at	476	465	16	5.4	97.7	0.0
	int_rt	297	297	12	3.5	100.0	0.0
Anytown Modified	bin	5961	5961	13	67.3	100.0	0.0
	int	538	538	23	6.3	100.0	9.7
	int_r	424	424	33	6.7	100.0	90.3
	int_at	245	245	19	3.3	100.0	0.0
	int_rt	69	69	12	2.8	100.0	0.0
Curicó	bin	8125	470	7	97.3	5.7	0.0
	int	153	153	6	1.7	100.0	20.0
	int_r	214	214	9	2.4	100.0	60.0
	int_at	235	229	11	2.6	97.4	0.0
	int_rt	131	131	8	1.5	100.0	20.0

The results demonstrate how the binary representation strategy excels in terms of the number of unique solutions generated, which can arguably indicate that this representation is particularly effective at extensively exploring the solution space. Regarding feasibility, the solutions generated through binary representation exhibit feasibility levels of 93% and 100% for the benchmarking cases of Anytown and Anytown Modified, respectively. In contrast, in the more complex Curicó network, the feasibility drastically decreases to 5.7%. This contrast is not observed in the integer representations, where the feasibility levels reach 100% in most experiments.

The dominance assessment reveals that the nondominated solutions, which are defined as those that cannot be outperformed with respect to simultaneously optimizing all objectives by other solutions, exhibit comparable quantities across the different representations. This suggests that, regardless of the employed representation, a comparable set of solutions is obtained in terms of their dominance. However, the similarity among the numbers of nondominated solutions obtained for the different representations does not necessarily imply that these solutions are clustered near the approximate Pareto front.

The last column of Table 1 shows that only two representations (int and int_r) contribute to the approximate Pareto diagram in the Anytown and Anytown Modified networks. However, in the case of the Curicó network, three representations contribute to this Pareto front (int, int_r, and int_rt). Overall, the int_r representation contributes most significantly. Figures 5–7 highlight this comparison by showing the approximate Pareto front (marked in red) along with the specific Pareto fronts obtained for each representation and case study, emphasizing the differences near the theoretical optimum.

Figures 5–7 display the nondominated solutions obtained during all independent experiments conducted for each case study. On the one hand, the results show that the binary representation strategy does not contribute to the approximate Pareto front in any of the case studies. This demonstrates that despite obtaining the highest number of unique and/or feasible solutions, the majority of these solutions are of lower quality than those of the other representations. Consequently, despite offering greater solution diversity, this approach does not benefit the optimization problem studied in this work. On the other hand, this qualitative analysis illustrates the unique contributions of the int and int_r representations in shaping the approximate Pareto fronts pertaining to each problem. In particular, the int_r representation stands out for its superior generation of nondominated solutions, signifying its ability to produce higher-quality solutions than those of the other examined representations.
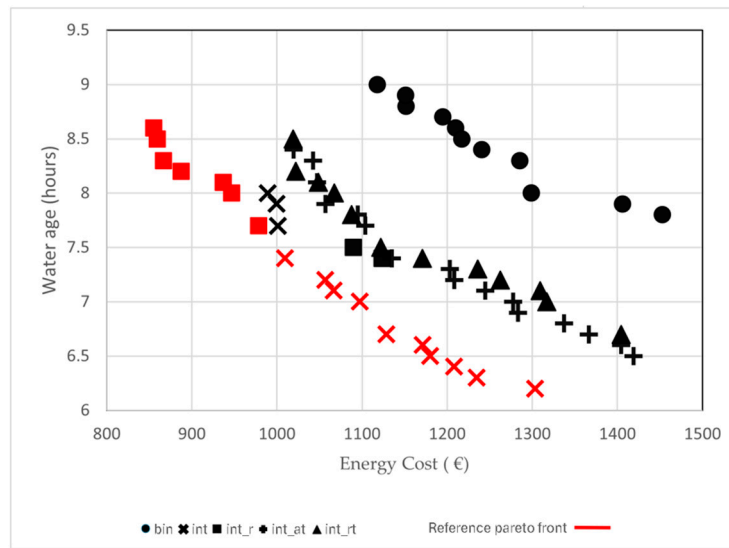
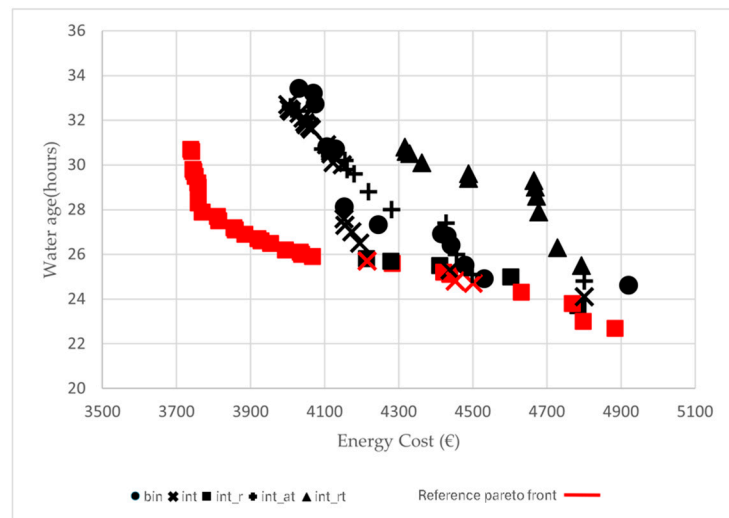**Figure 5.** Pareto front for the Anytown network.

Figure 6. Pareto front for the Anytown Modified network.

Nevertheless, it is crucial to consider that a visual assessment of the nondominated solutions may be insufficient for comparing the various representations in terms of quality and performance. Therefore, it is imperative to resort to statistical indicators and inference methods to obtain a more rigorous and objective evaluation of the superiority and performance of each representation. This approach allows for a precise identification of whether the observed disparities between the representations are statistically significant. Additionally, it provides a deeper understanding of how each representation addresses the problem objectives and determines if any of them significantly outperforms the others in terms of the quality of their generated nondominated solutions.
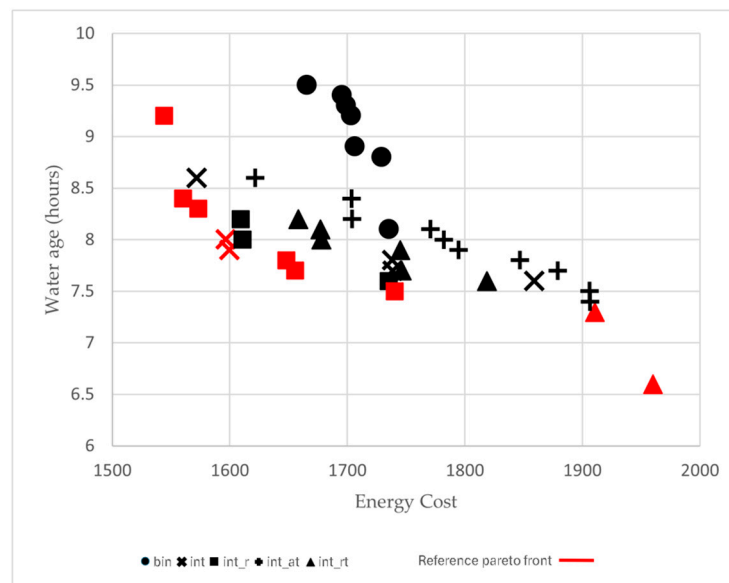
Figure 7. Pareto front for the Curicó network.

4.2. Performance Metrics

To assess the diversity and convergence of the solutions obtained for each representation, the HV, Epsilon, and IGD+ metrics are calculated for each independent experiment. Table 2 displays these metrics in terms of their medians. Given the fact that the exact data distribution is unknown, these measures are less susceptible to the influence of outliers, ensuring a more stable and reliable evaluation of the central tendencies and dispersion trends in the results. To aid in the interpretation and analysis of the tables, the representation producing the best value for each metric is highlighted in gray.

Table 2. Median metric values.

Network	Quality Indicator	bin	int	int_r	int_at	int_rt
Anytown	HV	0.65	0.70	0.73	0.70	0.70
	EP	0.09	0.05	0.01	0.06	0.05
	IGD+	0.05	0.02	0.01	0.02	0.02
Anytown Modified	HV	0.29	0.30	0.35	0.28	0.24
	EP	0.07	0.07	0.01	0.09	0.13
	IGD+	0.05	0.04	0.00	0.06	0.09
Curicó	HV	0.15	0.19	0.22	0.16	0.16
	EP	0.10	0.06	0.03	0.09	0.09
	IGD+	0.05	0.02	0.01	0.04	0.04

The HV metric measures the space covered by the solutions found, where a higher value indicates greater diversity and coverage in the objective space. In all three analyzed scenarios, the int_r representation stands out by yielding the highest values, suggesting its extensive coverage in the objective space. Regarding the Epsilon metric, which is an indicator of proximity to the optimal Pareto front, lower values indicate higher quality. Once again, the int_r representation excels by producing the lowest values in all three cases, indicating closer proximity to the reference optimum. IGD+, which assesses the distance from a reference set of solutions to those found, considers both convergence to the Pareto front and the diversity of the output solutions. This is crucial because it evaluates not only the diversity of the solutions but also their proximity to the approximate Pareto front. Across all studied cases, the int_r representation achieves the lowest IGD+ values, reflecting its superior performance in terms of convergence and diversity.

A Wilcoxon test is now introduced to provide a robust statistical framework for determining whether the observed differences among the quality metrics are statistically significant or simply the result of inherent data variability. By utilizing this analysis approach, the previous findings can be reliably validated.

The Wilcoxon test compares the differences between two related datasets. Consequently, it is necessary to choose a sample or reference representation for comparison with the other four representations. This is carried out for the five representations studied in this work. Additionally, each sample contains quality metrics (HV, EP, and IGD+ values) for each case study. To facilitate data interpretation and follow the methodology proposed by Salgueiro et al. [49], Table 3 summarizes the comparisons between each reference representation and the remaining representations for each case study. We define the Null Hypothesis (H0) as there being no significant difference in performance between the compared representations. For each of the pairwise tests, if the *p*-value is ≤ 0.05, we reject H0 and conclude that there are significant differences between the samples. This results in a total of twelve comparisons for each evaluated quality metric (4 representations compared $\times$ 3 case studies). Each cell in the table indicates the number of times a statistically significant result is observed in favor of the reference representation ('+'), in favor of the compared representation ('−'), or lacking significant differences ('=').

Table 3. Wilcoxon test results.

	bin			int			int_r			int_at			int_rt		
	HV	EP	IGD+	HV	EP	IGD+	HV	EP	IGD+	HV	EP	IGD+	HV	EP	IGD+
+	2	2	2	5	5	5	12	12	12	2	2	2	2	2	2
−	8	8	8	6	6	6	0	0	0	8	8	8	7	7	7
=	2	2	2	1	1	1	0	0	0	2	2	2	3	3	3

An analysis of the results demonstrates that the superiority of the int_r representation in terms of the studied quality metrics (HV, EP, and IGD+) is not a result of randomness; rather, statistically significant differences in favor of this representation are observed in all case studies. Consequently, this representation is the most suitable strategy for this particular optimization problem.

In contrast, the Wilcoxon test reveals that the binary representation approach not only achieves poorer numerical performance in 8 out of the 12 comparisons made for each metric (Table 3) but also that this unfavorable discrepancy is statistically significant and not a result of chance. Additionally, significantly favorable differences with respect to the binary representation are observed only in two out of the twelve comparisons for each metric. Therefore, despite being common in the literature, the binary solution representation strategy does not seem to be the most suitable technique for this type of optimization problem based on pump scheduling.

Finally, Table 4 shows data related to the average execution times of the different representations across the three case studies. For each network and representation, the average execution time is recorded.

Table 4. Computational times of the tested methods.

Network	Average Times (s)				
	bin	int	int_r	int_at	int_rt
Anytown	216.46	212.58	209.72	218.08	212.88
Anytown Modified	203.15	206.00	215.18	214.52	205.32
Curicó	28,729.87	26,842.75	36,730.12	20,756.05	15,807.77

The computational efficiency metrics reveal that all representations exhibit similar execution times for the Anytown and Anytown Modified benchmarking networks, which are comparable in size and topology. Consequently, the enhanced solution quality of 'int_r' over that of the other representations does not come at the expense of its computational speed, making it the most efficient representation for this particular problem size.

In contrast, significant differences exist among the representations with respect to the Curicó network, which is known for its increased complexity and size. In this scenario, 'int_r' incurs the highest computational cost, yet it also delivers the highest solution quality, as observed previously. Interestingly, 'int_rt' proves to be the representation executing simulations most swiftly in this case, and it attains average performance in terms of solution quality, as depicted in Figure 7. The choice between one representation or another in this case depends on the specific objectives of the optimization problem of interest.

4.3. Benefits and Disadvantages of the Suggested Solution Codification

In this section, we provide a comprehensive evaluation of the benefits and disadvantages of the int_r method. This analysis is supported by the metrics discussed in Sections 4.1 and 4.2. The aim is to present a balanced view of the method's performance, highlighting its strengths in achieving convergence and diversity as well as its potential limitations in terms of problem simplification and resource consumption. By examining these aspects, we aim to offer valuable insights for researchers and practitioners in water distribution network optimization.

Benefits:

- Enhanced Convergence: The int_r method achieves superior convergence towards the Pareto front, as indicated by the lower Epsilon (EP) and IGD+ values across all networks (Table 2). The int_r representation consistently shows the closest proximity to the reference Pareto front, ensuring high-quality solutions that balance energy costs and water quality objectives.
- Improved Solution Diversity: The int_r representation excels in generating diverse solutions, as evidenced by the higher Hypervolume (HV) values in all three case studies (Table 2). This indicates extensive coverage in the objective space, which enhances the robustness and applicability of the solutions.

Disadvantages:

- Problem Simplification: The int_r representation significantly simplifies the problem by reducing the number of decision variables. While this leads to faster convergence, it also means that the method might not explore the entire search space thoroughly. Consequently, it could potentially miss some solutions that are closer to the optimal, which might be found using more detailed representations.
- Resource Consumption: The int_r method, while efficient in some scenarios, can be resource-intensive, particularly for larger and more complex networks like Curicó. As observed in Table 4, the execution time for the Curicó network is significantly higher (36,730.12 s) compared to the other representations. This indicates that the computational demands increase with the complexity and size of the network, potentially requiring substantial computational resources and longer processing times.

The detailed metrics and statistical analyses support the conclusions and demonstrate the robustness of the int_r representation in solving the pump scheduling optimization problem. However, these results should not be generalized to other kinds of mathematical optimization problems.

5. Conclusions

Enhancing energy efficiency in pumping systems is crucial for reducing operational expenses for water supply companies. Pump scheduling optimization is a common approach, utilizing different encoding techniques such as binary and integer variables. However, the complexity increases with the number of decision variables, leading to challenges in

modeling real-world water distribution networks (WDNs). In this study, five solution representations were compared for a multiobjective pump scheduling problem, aiming to minimize energy costs and improve water quality. NSGA-II was chosen for its effectiveness in solving multiobjective optimization problems. Through analysis and comparison using convergence and diversity metrics across three case studies, significant insights were gained.

- The binary representation strategy excels at generating a high number of unique solutions, implying robust exploration of the solution space. However, this advantage does not necessarily translate into higher solution quality, as indicated by the lower feasibility levels observed, especially in complex networks like Curicó. In contrast, integer representations demonstrate higher feasibility rates across most experiments, showcasing their effectiveness in meeting problem constraints. Furthermore, the analysis of nondominated solutions reveals comparable quantities across representations, suggesting that each strategy yields solutions with similar dominance characteristics.
- However, a closer examination shows that certain representations contribute more significantly to shaping the approximate Pareto front, particularly the int_r representation. The visualizations provided in Figures 5–7 highlight the unique contributions of specific representations to the Pareto fronts, emphasizing differences near the theoretical optimum.
- To quantitatively evaluate solution diversity and convergence, metrics such as HV, Epsilon, and IGD+ were calculated. The results consistently favor the int_r representation, indicating its superior performance in terms of solution diversity, proximity to the Pareto front, and convergence.
- Statistical analysis using the Wilcoxon test confirms the significance of these differences, establishing int_r as the most suitable representation strategy for the optimization problem studied. Conversely, the binary representation strategy demonstrates inferior numerical performance across comparisons, suggesting its inadequacy for this type of optimization problem.
- Regarding computational efficiency, int_r maintains competitive execution times across benchmarking networks, indicating that its enhanced solution quality does not compromise computational speed. However, in more complex scenarios like the Curicó network, where computational costs vary significantly among representations, the choice depends on specific optimization objectives.

In conclusion, the findings underscore the importance of selecting an appropriate solution representation strategy based on the problem's complexity and objectives. While binary representation may offer extensive exploration of the solution space, it does not necessarily lead to high-quality solutions. On the other hand, integer representations, particularly int_r, demonstrate superior performance in terms of solution quality, diversity, and convergence, making them more suitable for this optimization problem.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

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

References

1. Boretti, A.; Rosa, L. Reassessing the Projections of the World Water Development Report. *NPJ Clean Water* **2019**, *2*, 15. [CrossRef]
2. Giustolisi, O.; Walski, T.M. Demand Components in Water Distribution Network Analysis. *J. Water Resour. Plan Manag.* **2012**, *138*, 356–367. [CrossRef]
3. Liu, J.; Li, X.; Yang, H.; Han, G.; Liu, J.; Zheng, C.; Zheng, Y. The Water-Energy Nexus of Megacities Extends beyond Geographic: A Case of Beijing. *Environ. Eng. Sci.* **2019**, *36*, 778–788. [CrossRef] [PubMed]
4. Gilron, J. Water-Energy Nexus: Matching Sources and Uses. *Clean Technol Env. Policy* **2014**, *16*, 1471–1479. [CrossRef]
5. de Oliveira, V.; Wehn, U.; Casale, G.; Cordeiro Ortgara, A.R.; Uhlenbrook, S.; Genné, I.; Campling, P.; Amorsi, N.; Smith, D.; Delargy, O.; et al. *Water in the 2030 Agenda for Sustainable Development: How Can Europe Act?* UNESCO and Water Europe: Brussels, Belgium, 2019.
6. Kumar, V.; Yadav, S.M. A State-of-the-Art Review of Heuristic and Metaheuristic Optimization Techniques for the Management of Water Resources. *Water Supply* **2022**, *22*, 3702–3728. [CrossRef]
7. Maier, H.R.; Razavi, S.; Kapelan, Z.; Matott, L.S.; Kasprzyk, J.; Tolson, B.A. Introductory Overview: Optimization Using Evolutionary Algorithms and Other Metaheuristics. *Environ. Model. Softw.* **2019**, *114*, 195–213. [CrossRef]
8. 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]
9. Quiñones-Grueiro, M.; Ares Milián, M.; Sánchez Rivero, M.; Silva Neto, A.J.; Llanes-Santiago, O. Robust Leak Localization in Water Distribution Networks Using Computational Intelligence. *Neurocomputing* **2021**, *438*, 195–208. [CrossRef]
10. Mora-Melia, D.; Iglesias-Rey, P.L.; Martinez-Solano, F.J.; Ballesteros-Pérez, P. Efficiency of Evolutionary Algorithms in Water Network Pipe Sizing. *Water Resour. Manag.* **2015**, *29*, 4817–4831. [CrossRef]
11. Aldrees, A.; Khan, M.A.; Tariq, M.A.U.R.; Mustafa Mohamed, A.; Ng, A.W.M.; Bakheit Taha, A.T. Multi-Expression Programming (MEP): Water Quality Assessment Using Water Quality Indices. *Water* **2022**, *14*, 947. [CrossRef]
12. Niknam, A.; Zare, H.K.; Hosseininasab, H.; Mostafaeipour, A.; Herrera, M. A Critical Review of Short-Term Water Demand Forecasting Tools—What Method Should I Use? *Sustainability* **2022**, *14*, 5412. [CrossRef]
13. Li, Q.; Liu, S.-Y.; Yang, X.-S. Influence of Initialization on the Performance of Metaheuristic Optimizers. *Appl Soft Comput* **2020**, *91*, 106193. [CrossRef]
14. de Araujo Lima, S.J.; de Araújo, S.A. Genetic Algorithm Applied to the Capacitated Vehicle Routing Problem: An Analysis of the Influence of Different Encoding Schemes on the Population Behavior. *Am. Sci. Res. J. Eng. Technol. Sci.* **2020**, *73*, 96–110.
15. López-Ibáñez, M.; Prasad, T.D.; Paechter, B. Representations and Evolutionary Operators for the Scheduling of Pump Operations in Water Distribution. *Evol. Comput.* **2011**, *19*, 429–467. [CrossRef] [PubMed]
16. Copeland, C.; Carter, N.T. *Energy-Water Nexus: The Water Sector's Energy Use*; Congressional Research Service: Washington, DC, USA, 2017.
17. Marchi, A.; Salomons, E.; Ostfeld, A.; Kapelan, Z.; Simpson, A.R.; Zecchin, A.C.; Maier, H.R.; Wu, Z.Y.; Elsayed, S.M.; Song, Y.; et al. Battle of the Water Networks II. *J. Water Resour. Plan Manag.* **2014**, *140*, 4014009. [CrossRef]
18. Ormsbee, L.; Lingireddy, S.; Chase, D. Optimal Pump Scheduling for Water Distribution Systems. In Proceedings of the Multidisciplinary International Conference on Scheduling: Theory and Applications (MISTA 2009), Dublin, Ireland, 10 August 2009; pp. 341–356.
19. Bene, J.G.; Selek, I.; Hős, C. Comparison of Deterministic and Heuristic Optimization Solvers for Water Network Scheduling Problems. *Water Supply* **2013**, *13*, 1367–1376. [CrossRef]
20. Manuel, L.-I.; Devi, P.T.; Ben, P. Ant Colony Optimization for Optimal Control of Pumps in Water Distribution Networks. *J. Water Resour. Plan. Manag.* **2008**, *134*, 337–346. [CrossRef]
21. Janus, T.; Ulanicki, B.; Diao, K. Optimal Scheduling of Variable Speed Pumps Using Mixed Integer Linear Programming—Towards An Automated Approach 2023. *arXiv* **2023**, arXiv:2309.04715.
22. Candelieri, A.; Ponti, A.; Giordani, I.; Archetti, F. Lost in Optimization of Water Distribution Systems: Better Call Bayes. *Water* **2022**, *14*, 800. [CrossRef]
23. Abdelsalam, A.A.; Gabbar, H.A. Energy Saving and Management of Water Pumping Networks. *Heliyon* **2021**, *7*, e07820. [CrossRef]

24. Vieira, T.P.; Almeida, P.E.M.; Meireles, M.R.G.; Souza, M.J.F. Use of Computational Intelligence for Scheduling of Pumps in Water Distribution Systems: A Comparison between Optimization Algorithms. In Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.

25. Singhtaun, C.; Rungraksa, W. Pump Scheduling for Water Supply Production Using Mathematical Programming. In Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering, Osaka, Japan, 7–9 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 261–265.

26. Luigi, C.; Andrea, D.; Luca, C. Boosting Genetic Algorithm Performance in Pump Scheduling Problems with a Novel Decision-Variable Representation. *J. Water Resour. Plan. Manag.* **2020**, *146*, 04020023. [CrossRef]

27. Fooladivanda, D.; Taylor, J.A. Energy-Optimal Pump Scheduling and Water Flow. *IEEE Trans. Control Netw. Syst.* **2018**, *5*, 1016–1026. [CrossRef]

28. Jafari-Asl, J.; Azizyan, G.; Monfared, S.A.H.; Rashki, M.; Andrade-Campos, A.G. An Enhanced Binary Dragonfly Algorithm Based on a V-Shaped Transfer Function for Optimization of Pump Scheduling Program in Water Supply Systems (Case Study of Iran). *Eng. Fail Anal.* **2021**, *123*, 105323. [CrossRef]

29. Shao, Y.; Zhou, X.; Yu, T.; Zhang, T.; Chu, S. Pump Scheduling Optimization in Water Distribution System Based on Mixed Integer Linear Programming. *Eur. J. Oper. Res.* **2024**, *313*, 1140–1151. [CrossRef]

30. López-Ibáñez, M. Operational Optimisation of Water Distribution. Ph.D. Thesis, Edinburgh Napier University, Edinburgh, UK, 2009.

31. Makaremi, Y.; Haghighi, A.; Ghafouri, H.R. Optimization of Pump Scheduling Program in Water Supply Systems Using a Self-Adaptive NSGA-II; a Review of Theory to Real Application. *Water Resour. Manag.* **2017**, *31*, 1283–1304. [CrossRef]

32. Kurek, W.; Ostfeld, A. Multi-Objective Optimization of Water Quality, Pumps Operation, and Storage Sizing of Water Distribution Systems. *J Env. Manag.* **2013**, *115*, 189–197. [CrossRef]

33. Oscar Osvaldo Marquez Calvo Claudia Quintiliani, L.A.C.D.C.A.L.D.S.; de Marinis, G. Robust Optimization of Valve Management to Improve Water Quality in WDNs under Demand Uncertainty. *Urban Water J.* **2018**, *15*, 943–952. [CrossRef]

34. Lucasius, C.B.; Kateman, G. Understanding and Using Genetic Algorithms Part 2. Representation, Configuration and Hybridization. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 99–145. [CrossRef]

35. Talbi, E.-G. *Metaheuristics: From Design to Implementation*; Wiley Publishing: Hoboken, NJ, USA, 2009; ISBN 0470278587.

36. Barán, B.; von Lücken, C.; Sotelo, A. Multi-Objective Pump Scheduling Optimisation Using Evolutionary Strategies. *Adv. Eng. Softw.* **2005**, *36*, 39–47. [CrossRef]

37. Brás, M.; Moura, A.; Andrade-Campos, A. Cost Reduction of Water Supply Systems through Optimization Methodologies: A Comparative Study of Optimization Approaches. In Proceedings of the International Conference on Industrial Engineering and Operations Management, IEOM Society International, Lisbon, Portugal, 18–20 July 2023.

38. Ishibuchi, H.; Masuda, H.; Nojima, Y. A Study on Performance Evaluation Ability of a Modified Inverted Generational Distance Indicator. In Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, Madrid, Spain, 11–15 July 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 695–702.

39. Knowles, J.D.; Thiele, L.; Zitzler, E. *A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers*; ETH Zurich: Zurich, Switzerland, 2006.

40. Zitzler, E.; Thiele, L. Multiobjective Optimization Using Evolutionary Algorithms—A Comparative Case Study. In Proceedings of the Parallel Problem Solving from Nature PPSN V; Eiben, A.E., Bäck, T., Schoenauer, M., Schwefel, H.-P., Eds.; Springer: Berlin, Heidelberg, 1998; pp. 292–301.

41. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]

42. Gutiérrez-Bahamondes, J.H.; Salgueiro, Y.; Silva-Rubio, S.A.; Alsina, M.A.; Mora-Meliá, D.; Fuertes-Miquel, V.S. JHawanet: An Open-Source Project for the Implementation and Assessment of Multi-Objective Evolutionary Algorithms on Water Distribution Networks. *Water* **2019**, *11*, 2018. [CrossRef]

43. Zhang, K.; Yan, H.; Zeng, H.; Xin, K.; Tao, T. A Practical Multi-Objective Optimization Sectorization Method for Water Distribution Network. *Sci. Total Environ.* **2019**, *656*, 1401–1412. [CrossRef] [PubMed]

44. Benítez-Hidalgo, A.; Nebro, A.J.; García-Nieto, J.; Oregi, I.; Del Ser, J. JMetalPy: A Python Framework for Multi-Objective Optimization with Metaheuristics. *Swarm Evol. Comput.* **2019**, *51*, 100598. [CrossRef]

45. Rossman, L. *EPANET 2.0 User Manual*; Water Supply and Water Resources Division, National Risk management Laboratory, USEPA: Cincinnati, OH, USA, 2000.

46. Deb, K.; Sindhya, K.; Okabe, T. Self-Adaptive Simulated Binary Crossover for Real-Parameter Optimization. In Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, London, UK, 7–11 July 2007; Association for Computing Machinery: New York, NY, USA, 2007; pp. 1187–1194.

47. Hassanat, A.; Almohammadi, K.; Alkafaween, E.; Abunawas, E.; Hammouri, A.; Prasath, V.B.S. Choosing Mutation and Crossover Ratios for Genetic Algorithms—A Review with a New Dynamic Approach. *Information* **2019**, *10*, 390. [CrossRef]

48. Negrete Flores, M.Y. Modelación Computacional en Epanet de un Sector de la Red de Abastecimiento de Agua Potable. Bachelor's Thesis, University of Talca, Curico, Chile, 2021.

49. Salgueiro, Y.; Toro, J.L.; Bello, R.; Falcon, R. Multiobjective Variable Mesh Optimization. *Ann. Oper. Res.* **2017**, *258*, 869–893. [CrossRef]