POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

An improved cross-entropy method for the water pump scheduling optimization problem

Tesi di Laurea Magistrale in
Automation and Control Engineering -
Ingegneria dell'automazione e controllo

Author: **Francesco Zeng**

Student ID: 969166
Advisor: Prof. Andrea Castelletti
Co-advisors: Prof. Jingcheng Wang
Academic Year: 2022-23

Abstract

The growing demand for urban freshwater supply often corresponds to rising electrical and labor costs for pump stations. Optimizing the operation of water pumps, ensuring they run at their most efficient working points, and improving their scheduling systems offer effective ways to mitigate costs and conserve water resources. These water pump stations and their associated systems can be mathematically modeled, enabling the use of optimization algorithms to assess optimal operating conditions. Conventional optimization methods often fall short when dealing with such complex models, underscoring the significance of investigating recent algorithms. In this study, we expand the research on the cross-entropy method and introduce an improved version to address water pump scheduling optimization. Notable enhancements include an improved update step for the multivariate normal distribution used in sampling, resulting in faster and more accurate outcomes, as well as reduced susceptibility to local minima. Moreover, our adaptive objective function takes into account various cost factors and penalty functions associated with complex constraints, with the primary aim of further minimizing power consumption costs. To validate our approach, we conducted a series of experiments within a simulated environment and on a real-life-inspired scenario modeled after a water pump station located in Shanghai. The results of these experiments demonstrate the adaptability of our optimization method to complex models and constraints and highlight its satisfactory performance outcomes.

Keywords: Cross-Entropy method; WDS; water pump scheduling; optimization

Abstract in lingua italiana

La crescente domanda di approvvigionamento di acqua dolce nelle aree urbane spesso corrisponde all'aumento dei costi elettrici e del lavoro per le stazioni di pompaggio. Ottimizzare il funzionamento delle pompe dell'acqua, garantendo che operino nei punti di lavoro più efficienti e migliorando i loro sistemi di pianificazione, offre modi efficaci per ridurre i costi e preservare le risorse idriche. Queste stazioni di pompaggio dell'acqua e i relativi sistemi possono essere modellati matematicamente, consentendo l'uso di algoritmi di ottimizzazione per valutare le condizioni operative ottimali. I metodi di ottimizzazione convenzionali spesso non sono sufficienti quando si tratta di modelli così complessi, sottolineando l'importanza di indagare su algoritmi recenti. In questo studio, ampliamo la ricerca sul metodo cross-entropy e ne presentiamo una versione migliorata per affrontare l'ottimizzazione della pianificazione delle pompe dell'acqua. Miglioramenti significativi includono una fase di aggiornamento migliorata per la distribuzione normale multivariata utilizzata nel campionamento, che porta a risultati più rapidi e precisi, oltre a una minore suscettibilità ai minimi locali. Inoltre, la nostra funzione obiettivo adattiva tiene conto di vari fattori di costo e delle funzioni di penalizzazione associate a vincoli complessi, con l'obiettivo principale di ridurre ulteriormente i costi di consumo energetico. Per convalidare il nostro approccio, abbiamo condotto una serie di esperimenti in un ambiente simulato e su uno scenario ispirato alla realtà, modellato su una stazione di pompaggio d'acqua situata a Shanghai. I risultati di questi esperimenti dimostrano l'adattabilità del nostro metodo di ottimizzazione a modelli e vincoli complessi, evidenziando i suoi risultati soddisfacenti in termini di prestazioni.

Parole chiave: Metodo Cross-Entropy; WDS; pianificazione delle pompe d'acqua, ottimizzazione

POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

An improved cross-entropy method for the water pump scheduling optimization problem

Laurea Magistrale in Automation and Control Engineering - Ingegneria dell'Automazione e Controllo

Author: Francesco Zeng

Advisor: Prof. Andrea Castelletti

Co-advisor: Prof. Jingcheng Wang

Academic year: 2022-2023

1. Introduction

In urban areas facing rising demand due to population growth, water is a crucial resource. While expanding water distribution systems is necessary, the challenge of high-power consumption and energy waste in water pump stations must be addressed. Research highlights that a substantial portion of industrial electricity usage is due to electric motors [1], with water pumps accounting for a significant fraction [2]. Efficient water pump scheduling is vital for optimizing distribution systems. Traditionally, manual methods for pump activation often result in suboptimal scheduling, leading to increased costs and negative impacts on resource management and water quality. Researchers aim to find optimal operating conditions for pumps, focusing on minimizing power consumption and maximizing system efficiency. However, achieving consistently optimal solutions in complex scenarios remains a challenge for existing methodologies.
In the context of water pump scheduling, the Cross-Entropy (CE) method is a niche approach. Introduced by Rubinstein in 1997 [3], the CE method iteratively refines the probability distribution of solutions, ultimately leading to optimal or near-optimal solutions. A key strength of the CE method is its capacity to analyze solutions without relying on derivative functions, distinguishing it from traditional optimization methods. However, since the classic version of the CE method is better suited for system models with relatively simple constraints, an improved version of the CE method proves more suitable for addressing the problem of water pump station optimization.
This work makes three primary contributions. First, it expands the research on the CE method applied to the optimization of water pump scheduling. Second, it introduces the use of an asymmetric smoothed updating step to enhance the algorithm's performance, with the ultimate aim of finding the most efficient operating point for the pumps and thus minimizing pump station costs. Third, it empirically demonstrates the practical effectiveness of the improved optimization method in reducing costs for a real-life water pump station, achieved through the implementation of an adaptive cost function.

2. The pump station model

The model considers the simultaneous operation of multiple pump types to meet system requirements while minimizing operational costs. It considers the total power consumption for each pump, hourly electricity price variations, and introduces penalty functions to address complex constraints. Ultimately, by solving the model, we can determine the optimal states of the pumps, including their head, flow, power, and speed.

2.1. The affinity laws

The affinity laws govern the behavior of water pumps. They describe the fundamental relationships between the pump's head, flow, power, and speed.
Considering the following measures:

- Flow, Q (m^3/h): liquid volume.
- Head, H (m): liquid force measured.
- Speed, S (rpm): shaft speed.
- Power, P (kW): energy needed to pump a liquid.

the affinity laws state that:

$$\frac{Q_1}{Q_2} = \frac{S_1}{S_2}; \quad \frac{H_1}{H_2} = (\frac{S_1}{S_2})^2; \quad \frac{P_1}{P_2} = (\frac{S_1}{S_2})^3 \tag{1}$$

Eqs. (1) show how a 3% increase in speed results in a significant 9% surge in power consumption. In the industrial sector, this translates to a substantial rise in costs, highlighting the importance of optimal solutions.

2.2. The system and pumps curve

In water distribution systems, water pumps respond to the system's operational requirements to meet the supply demands. Changes in the system's characteristics, such as an increased flow demand affect the total dynamic head (THD). The THD is a parameter associated with the system's requirements and is composed by the static head (Hs), friction head (Hf), velocity head (Hv), and pressure head (Hp).

$$TDH = Hs + Hp + Hf + Hv \tag{2}$$

Eq. (2) characterizes the dynamics of the system and plays a fundamental role in establishing the best efficiency point (BEP) of water pumps, which is located by intersecting the system curve with the pump family curve.

Fixed-frequency pumps maintain a constant speed ratio. Their performance curves for head and flow, as well as power and flow, can be defined with:

$$\begin{aligned} H &= h_1 + h_2 Q + h_3 Q^2 \\ P &= p_1 + p_2 Q + p_3 Q^2 \end{aligned} \tag{3}$$

Where $h_1, h_2, h_3, p_1, p_2, p_3$ are fitting parameters. Combining the pump affinity laws and the equations (3), the expressions for the variable frequency pumps are found.

$$\begin{aligned} H &= s^2 h_1 + h_2 Q + h_3 Q^2 \\ P &= s^3 p_1 + s^2 p_2 Q + s^3 p_3 Q^2 \end{aligned} \tag{4}$$

Where s is the speed ratio of the pump.

2.3. The objective function

The objective function used for optimizing the water pump scheduling problem takes various factors into account. The total power consumption is calculated as the sum of contributions from each operational pump.

$$J = min\left\{ \sum_{i=1}^{m} \omega_i (s_i^3 p_1 + s_i^2 p_2 Q_i + s_i p_3 Q_i^2) + \sum_{i=m+1}^{m+n} \omega_i (p_1 + p_2 Q_i + p_3 Q_i^2) \right\} \tag{5}$$

The status coefficients ω_i represent the on-off status of each pump and are pivotal in determining the optimal combination of pumps within the scheduling solution.

2.4. The constraints

Constraints establish the boundaries and limitations within which a solution must operate.

2.4.1 Speed ratio constraint

Variable speed pumps can be controlled to change their speed ratio in the interval $s \in [0, 1]$. However, to increase efficiency and prevent issues like cavitation and shortened life cycles, the speed ratio of each pump has been constrained.

$$s_{min,i} \leq s_i \leq 1, \quad \text{for } i = 1, 2, ..., m+n \tag{6}$$

Where m, n are, respectively, the number of variable speed pumps and fixed speed pumps.

2.4.2 Flow balance equation

The balance equation for the water pump flow states that the output flow of the system Q_s must be equal to the sum of the input flow generated by the pump station.

$$Q_s = \sum_{i=1}^{m} Q_i + \sum_{j=m+1}^{m+n} Q_j \tag{7}$$

2.4.3 Parallel operation of pumps

Pumps operated in parallel will increase the flow but not the head. When this requirement is not met, there is the risk of water flowing backward toward the pump with lower head, potentially causing significant damage. As a result, this condition is imposed as a constraint.

$$H_s = H_1 = H_2 = ... = H_{m+n} \tag{8}$$

2.4.4 High-efficiency area

It is crucial to operate the pumps such that the control variables maximize the machine's efficiency. For fixed frequency pumps, this corresponds to an interval defined by a maximum value, $Q_{max,i}$, and a minimum value, $Q_{min,i}$, of pump flow.

$$Q_{min,i} \leq Q_i \leq Q_{max,i}, \quad \text{for } i = 1, 2, ..., n \tag{9}$$

In the case of variable frequency pumps, the speed parameter extends the interval to encompass a 'best efficiency area' (BEA). The BEA is delimited by flow and speed constraints, which are indicated by the curves connecting the four points A, B, E, and F.
This region determines the area where the pump efficiency, η is equal to or greater than 85%. The curves AB and EF are the head-flow curves of the pump, respectively, at the maximum speed ratio and the minimum speed ratio. The parabolic curve EA corresponds to an efficiency of $\eta = 85\%$, while the FB curve to an efficiency of $\eta = 89.2\%$.
The two boundary curves OA and OB, where O is the origin and which contain the curves EA and FB, are described by the equations:

$$\begin{aligned} H_{OA} &= k_1 Q^2 \\ H_{OB} &= k_2 Q^2 \end{aligned} \tag{10}$$

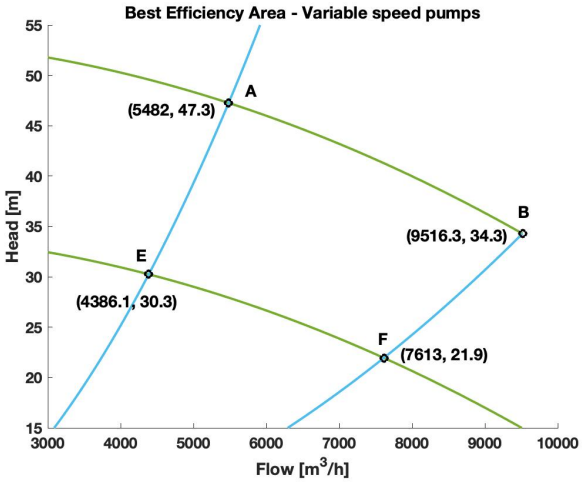


Figure 1: BEA delimited by A, B, E, F.

Where k_1, k_2 are fitting parameters.
Finally, the maximum and minimum flow values are found using equations (1), (9) and (10).

$$\begin{aligned} Q_{max,i} &= \begin{cases} \sqrt{\frac{h_1 - H_s}{h_3}} Q_A & H_s \geq H_B \\ \sqrt{\frac{H_s}{H_B}} Q_B & H_s < H_B \end{cases} \\ Q_{min,i} &= \begin{cases} \sqrt{\frac{H_s}{H_A}} Q_A & H_s \geq H_E \\ \sqrt{\frac{h_1 s_{min}^2 - H_s}{h_3}} & H_s < H_E \end{cases} \\ &\text{for } i = 1, 2, ..., m+n \end{aligned} \tag{11}$$

where H_S is the system's operating head value.

2.5. Summary

The general model is characterized as a non-linear, multi-objective, multi-variable combinatorial problem, encompassing both equality and inequality non-linear constraints.

$$J = min\left\{ \sum_{i=1}^{m} \omega_i (s_i^3 p_1 + s_i^2 p_2 Q_i + s_i p_3 Q_i^2) + \sum_{i=m+1}^{m+n} \omega_i (p_1 + p_2 Q_i + p_3 Q_i^2) \right\}$$

such that

$$\begin{aligned} & Q_s = \sum_{i=1}^{m} Q_i + \sum_{j=m+1}^{m+n} Q_j \\ & H_s = H_1 = H_2 = ... = H_{m+n} \\ & s_{min,i} \leq s_i \leq 1, && i = 1, ..., m+n \\ & Q_{min,i} \leq Q_i \leq Q_{max,i}, && i = 1, ..., m+n \end{aligned}$$

Therefore, in addressing such a problem, the CE method proves to be suitable, whereas classical methods frequently falter when faced with intricate models.

3. The improved cross-entropy method

The Cross-Entropy (CE) method derives from the Monte Carlo method and was introduced by Rubinstein in 1997 [3]. Its central concept involves iteratively refining the probability distribution of solutions, starting from an initial parameter distribution. This refinement process directs the focus toward the most promising regions within the solution space, ultimately leading to convergence towards an optimal or near-optimal solution.

3.1. The population

The initial step in the CE method is to define the probability density function (pdf) that characterizes the set of solutions and from which the initial population is sampled. A multivariate normal distribution is employed to sample the initial population.

3.2. Multivariate normal distribution

We can define each population member with a k-dimensional random vector $\boldsymbol{X} = (x_1, x_2)^T$ sampled from the initial multivariate normal distribution N.

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{12}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \bar{s} \\ \bar{h} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \left(\frac{s_{max} - \bar{s}}{3}\right)^2 & 0 \\ 0 & \left(\frac{head_{max} - \bar{h}}{3}\right)^2 \end{bmatrix}$$

$\bar{s}, \bar{h}$ are the mean values between the upper and lower bounds of the speed and head limitations given by the constraints. This ensures the proper coverage of the search region.

3.3. Asymmetric smoothed update

In the CE method, the update step is responsible for retaining a record of the best solutions found and for the refinement of the pdf used in the subsequent iteration.
In the classic version of the CE method this step is straightforward. The parameter vector $\hat{v}_t$, which comprises the mean and covariance matrices, is updated at each iteration according to $\hat{v}_{t-1} = \hat{w}_t$, where t is the iteration and $\hat{w}_t$ is the new parameter vector that describes the mean and standard deviation of the elite samples. This update method shows to be inefficient and renders the algorithm vulnerable to local minima.
Decomposing the parameters vector into its its mean and covariance components and introducing a second parameter β, the previous issue is addressed by providing an asymmetric smoothed update.

$$\begin{aligned} \hat{v}_{\mu,t} &= \alpha\hat{w}_{\mu,t} + (1-\alpha)\hat{v}_{\mu,t-1} \quad 0.4 \leq \alpha \leq 0.9 \\ \hat{v}_{\sigma,t} &= \beta\hat{w}_{\sigma,t} + (1-\beta)\hat{v}_{\sigma,t-1} \quad 0.3 \leq \beta \leq 0.7 \end{aligned}$$

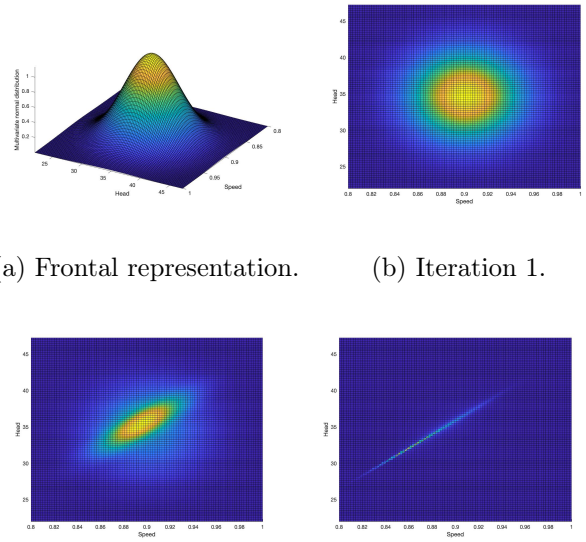
In optimization problems, the factor $(1-\alpha)$ is commonly employed in smoothed update steps. In this context, our focus extends beyond the introduction of a second parameter β, which enables an asymmetric smoothed update. The CE method samples from a Gaussian pdf, which, in 2 dimensions, forms a bell-shaped curve (Fig. 2). When using a single parameter α, in each update step, the bell-shaped curve 'moves towards the best solution' as the mean matrix updates and 'reduces in size' when the covariance matrix updates. Unfortunately, this simultaneous movement and scaling can lead the algorithm to converge into local minima, compromising accuracy. By applying two parameters instead, we enable the Gaussian pdf to 'move' and 'scale' in an asymmetric manner, as shown in Fig. 2. This results in faster convergence, reduced susceptibility to local minima, and an improved ability to explore better solutions, particularly when specific directions offer a higher probability of yielding superior results.

3.4. Adaptive objective function

To achieve a faster and more accurate optimization a different parametrization has been applied to the expression (5) by reformulating the flow values Q_i in terms of H_s and s_i.

$$Q_i = \sqrt{\frac{h_1 s_i^2 - H_s}{h_3}}$$

The constraints and the choice of their mathematical representation can significantly impact the feasibility and quality of the solution. Therefore, two penalty functions Pe_1 and Pe_2 that take into account the constraints expressed by the equations (7) and (9) are introduced in the model.
The penalty function Pe_1 represents the BEA



(a) Frontal representation. (b) Iteration 1.

(c) Iteration 5. (d) Iteration 15.

Figure 2: 'Moving' and 'resizing' in the update step.

constraint.

$$Pe_1 = \sum_{i=1}^{n+m} (\Delta Q_i)^2$$
$$= \begin{cases} \sum_{i=1}^{n+m} (Q_i - Q_{min,i})^2, & Q_i < Q_{min,i} \\ 0, & Q_{min,i} < Q_i < Q_{max,i} \\ \sum_{i=1}^{n+m} (Q_i - Q_{max,i})^2, & Q_i > Q_{max,i} \end{cases}$$

The penalty function Pe_2 represents the flow balance constraint.

$$Pe_2 = (\sum_{i=1}^{n+m} \omega_i \sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} - Q_s)^2$$

Therefore, the problem's mathematical formulation to be solved with the CE method is summarized in:
Where σ is a weight coefficient that decreases with the iterations, giving less importance to the penalty functions as the solutions converge to the optima. It is characterized by an initial value σ_0 and a cooling parameter γ.

$$\sigma = \sigma_0 \frac{1}{\gamma T}, \quad \gamma \in [0, 1]$$

4. Comparative analysis of performance

The improved CE method has introduces enhancements that improve best existing solutions.

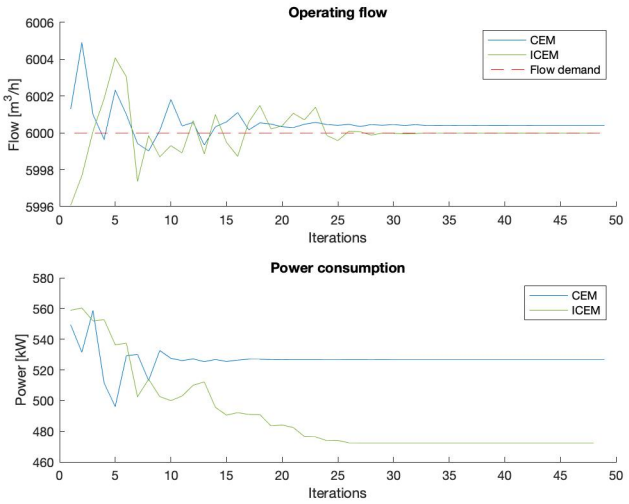


Figure 3: Comparison of the improved CE method in a single pump case.

Table 1: Comparative analysis

	CEM	ICEM
Flow demand (m^3/h)	6000	6000
Computed flow (m^3/h)	6004.0	6000
Power (kW)	526.77	472.43
Iterations	30	32

Its performance is thoroughly analyzed from various perspectives to highlight its contributions and improvements. Fig. 3 shows the comparison between the classic CE method (CEM) and improved CE method(ICEM). Furthermore, Tab. I shows the their operational parameters under the condition of system flow demand equal to $6000m^3/h$.
The improved CE method performs better compared to the classic CE method, as shown in Tab. 1. The improved CE method is able to converge to the global minima in power consumption, achieving a remarkable reduction of approximately 11%, in the $6000m^3/h$ case. Furthermore, it demonstrated superior effectiveness in refining its search for the optimal working point, resulting in a very accurate flow value. The differences in convergence observed can be directly attributed to the fundamental principles underlying the improved CE method. By smoothing the update of the multivariate normal distribution, the algorithm doesn't rapidly converge toward the best-found solution but al-

lows room for exploration in the proximity of the current optimal point. This approach enhances the algorithm's capability to search for global minima and improves its overall accuracy.

5. Practical implementations

The model has been adapted to optimize a real-life-inspired scenario featuring the water pump station of a megalopolis equipped with four pumps, including two variable frequency pumps (pump 1, 2) and two fixed frequency pumps (pump A, B). Their real performance curves have been incorporated, and the operation of the water pump station over several months has served as the foundation for the database in use. With the application of the improved CE method, effective working points for these set of pumps have been defined, ensuring compliance with all constraints, especially their operation within the BEA (Fig. 2). Additionally, the method has provided insights into the pumps' statuses (on-off), indicating the optimal combination of pumps that should be operated to minimize the costs of the pump station.
Since the system flow demand is constantly varying, a key flow demand value of $25000m^3/h$ have been chosen. The respective performance values are detailed in Table II and the optimal operating condition for the variable-speed pumps are indicated with a star symbol in Fig. 4. The efficiency of the variable-speed pumps consistently exceeds 85%, and the head constraint is respected. The implementation of the resulting scheduling system effectively reduces the operational and maintenance costs of the water pump station.
In conclusion, the application of the improved CE method results in the successful operation of the pumps within their best efficiency area, while satisfying the water system constraints and requirements of head and flow.

6. Conclusions

The optimization of the water pump station scheduling was explored using an improved CE method. The algorithm was developed and simulated in MATLAB, and its performance was tested in a real-life-inspired scenario, representing a megalopolis water pump station. The results of this study were highly satisfactory, with the system and its components operating

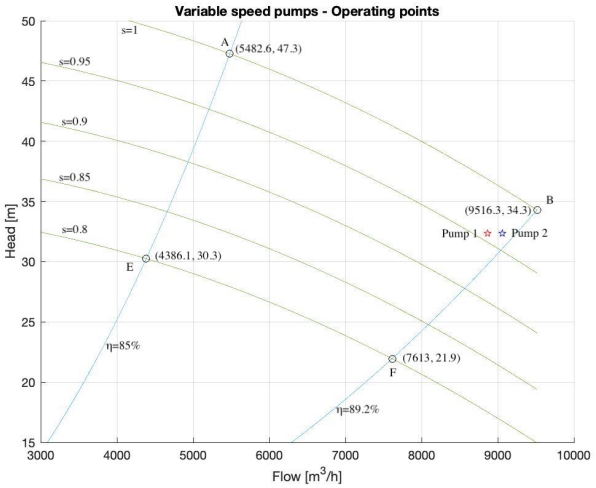


Figure 4: Working points when $Q_s = 25000m^3/h$.

Table 2: Pump station parameters when $Q_s = 25000m^3/h$.

Pump	**A**	**B**	**1**	**2**
Status	On	On	On	On
Output flow	24969.7 (m^3/h)			
Power	2394.05 (kW)			
Head	32.33 (m)			
Flow (m^3/h)	4196	3902	8299	8570
Speed ratio	0.92	0.87	0.87	0.89
Efficiency (%)	75.83	75.83	90.95	90.51

at their most efficient points while adhering to all specified constraints.

References

[1] V.K. Shankar, S. Umashankar, S. Paramasivam, N. Hanigovszki. "A comprehensive review on energy efficiency enhancement initiatives in centrifugal pumping system". Applied Energy, Nov. 2016, pp. 495-513.

[2] European Commission. Study on improving the energy efficiency of pumps; 2001.

[3] Rubinstein, Reuven Y. "The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning." 1997.

Contents

Introduction

Water is an essential resource in contemporary society. It holds the utmost importance for the well-being of urban inhabitants and for its influence over their daily lives and activities. The exponential rise in urban populations has prompted a matching escalation in water demands. In the context of China, boasting approximately 2.8 trillion cubic meters of freshwater resources, constituting 6% of the global total, the challenges are particularly pronounced. The demographic surge in China, with a population of approximately 1.4 billion in 2019 and an anticipated surge in urban population of roughly 600 million by 2050, underscores the pressing need for the efficient management of urban fresh water distribution systems as an integral aspect of urban planning endeavors.

A water distribution system (WDS), is a set of hydraulic components efficiently linked to convey water from sources to end consumers. This intricate network comprises an array of electromechanical constituents, including pipes, pumps, valves, and storage tanks. Typically, such systems are modeled as graphs, wherein nodes signify sources and consumers, while links represent the connecting components encompassing pipes, pumps, and valves. Its behavior is governed by three main aspects: the physical laws that describe flow and pressure distributions; the consumers' demands; and the system layout.

A WDS has to adhere to the established system requirements and specifications, such as continuity of supply, minimum delivery pressure, compliance with regulations, and response to changing demand. Reliability and risk considerations are integral to all aspects of WDS management. Therefore, to solve this crucial optimization problem, engineers have to resolve to holistic optimization approaches.

In the last decades, optimization methods garnered significant attention from researchers. They have been employed in various fields, including engineering, logistics, machine learning, and WDSs, to make informed decisions and improve the efficiency of processes. Traditional approaches such as linear programming, non-linear programming, and dynamic programming, which were once considered effective, now are seen as dated and the focus has shifted to more intuitive and high-performing algorithms like metaheuristic methods. Metaheuristic algorithms are versatile and heuristic-driven problem-solving strategies used

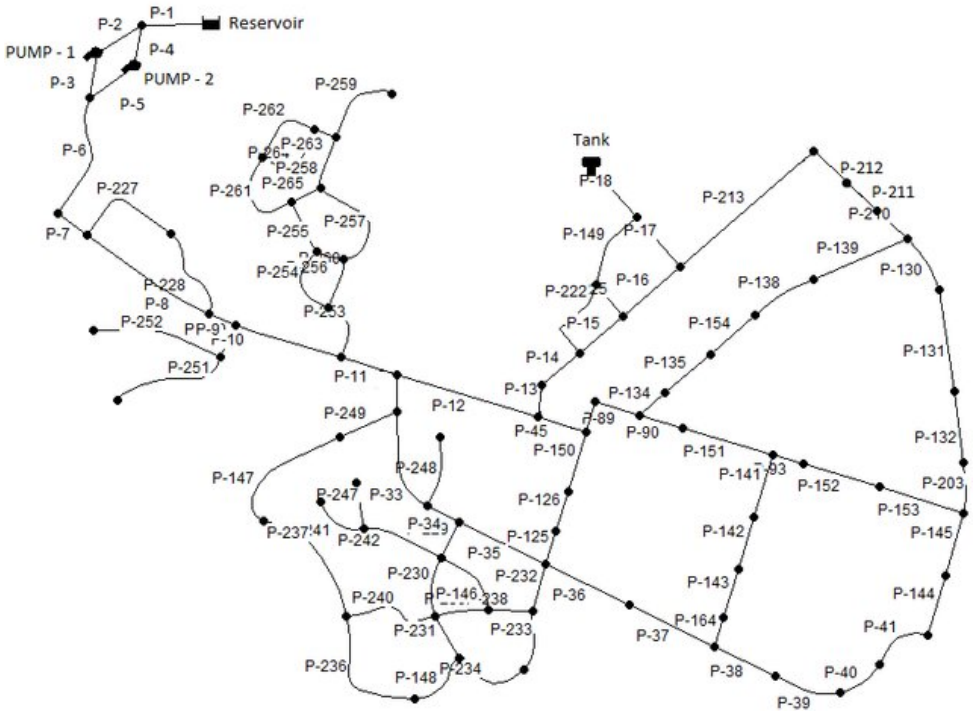


Figure 1: A water distribution system model.

to tackle complex optimization problems where finding an exact solution is impractical due to large search spaces or the absence of a known mathematical model. Two notable paradigms were developed during the first half of the last century and became important branches of metaheuristics: Evolutionary Computation (EC) and Swarm Intelligence (SI). These nature-inspired techniques derive from observing the analogous collective behaviors that flocks of birds, swarms of ants or bees, and schools of fish have when engaged in group activities, whether it's searching for food or safeguarding their species. This phenomenon, related to their innate mechanisms of natural selection and genetics, act as an active force that favors the selection of the most adept individuals within their respective groups. Consequently, these intricate systems evolve over time, adapting and enhancing their abilities to increase their prospects of survival. The researchers, have sought to comprehensively study, analyze, and model these mechanisms, and proposed such techniques to offer effective ways to solve complex and constrained mathematical problems.

This research aims to investigate and implement the CE method, contributing to the existing literature by applying this algorithm to the specific scenario of the water pump scheduling problem. The study introduces an innovative approach: employing an asymmetric smoothed updating step to enhance the algorithm's performance. The primary goal is to identify the most efficient operating point for the pumps, thereby minimizing costs associated with pump station operations. To validate the algorithm's performance,

this study proposes a comparative analysis between the CE method, Genetic Algorithm (GA), and Ant Colony Optimization (ACO) methods. By undertaking this comparative analysis, the aim is to highlight the strengths and potential advantages of the CE method in this context. Moreover, the research demonstrates the practical efficacy of the improved optimization method by reducing costs in a real-life water pump station. This is achieved through the implementation of an adaptive cost function, providing empirical evidence of the algorithm's effectiveness in practical settings. This comprehensive approach not only explores the application of the CE method to the water pump scheduling problem but also aims to enhance its performance and validate its efficiency through comparative analysis and real-life implementation.

1 | Background

Energy wastage is a common problem in water pump stations. A comprehensive research conducted by Shankar et al. [1] found that almost 70% of industrial electricity usage is attributed to electric motors (see Fig. 1.1), with water pumps alone accounting for about one-fifth of the energy consumed by electric motors worldwide, according to the European Commission [2]. This energy imbalance is a common problem in water pump stations. Among the numerous approaches available for optimizing system operation and reducing both electric and maintenance costs, the most impactful one is the optimization of the water pump scheduling system. Optimizing pump scheduling has emerged as a practical and highly effective strategy for reducing operational, electrical, and maintenance costs without necessitating extensive modifications to the existing infrastructure. Pumping stations function with multiple pumps working in tandem to fulfill the required pressure or water flow output. As a consequence, some pumps remain active while others remain inactive at any given moment. The challenge in scheduling pump operations revolves around identifying the most optimal combination of pumps to operate during specific time intervals within a scheduling period. A pump schedule includes the selection of pump combinations designated for each time interval within this scheduling horizon. Traditionally, pump stations have relied on empirical methods, where personnel manually activate pumps based on their experience. However, this conventional approach often yields suboptimal and inefficient scheduling, leading to escalated costs and detrimental impacts on resource management and water quality.

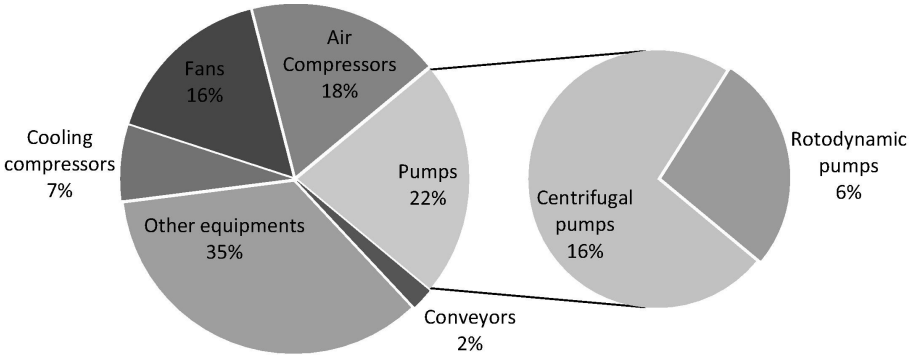


Figure 1.1: Energy consumption of pumping systems.

1.1. Metaheuristic methods

Starting from the 1960s, both domestic and international scholars embarked on research activities concerning the optimal regulation of urban water pumping unit scheduling. Over time, diverse methodologies have been employed to address this problem. These methodologies encompass linear approaches [3], like in Joewitt et al. [4], nonlinear approaches [5], as well as metaheuristic techniques, which started to gain popularity during the last decades. These algorithms explore the solution space iteratively, guided by a set of rules or principles, with the goal of finding near-optimal solutions within a reasonable time frame. Unlike traditional algorithms, they do not guarantee global optimality but offer a trade-off between solution quality and computational resources, making them invaluable tools for addressing real-world optimization challenges.

Within the landscape of EC and SI methodologies, Genetic Algorithm (GA) [6] and Ant Colony Optimization (ACO) [7, 8] stood out for their effectiveness. The genetic algorithm is a type of optimization and search algorithm inspired by the process of natural selection and genetics. It is used to find approximate solutions to optimization and search problems by mimicking the process of evolution within a population of candidate solutions. It utilizes mechanisms such as selection, crossover, and mutation to iteratively refine solutions with the hope that over time, the population converges toward better results. GA is particularly useful in optimization problems where the search space is large, complex, and poorly understood, or when gradient-based optimization techniques are not applicable. ACO is a computational optimization technique that draws inspiration from the foraging behavior of ants. It replicates the way real ants find the shortest path between their nest and a food source by laying pheromone trails and making informed decisions based on these trails. In the realm of computational problem-solving, ACO mimics this behavior by utilizing artificial ants to explore solution spaces and discover optimal or near-optimal solutions.

To establish a comprehensive foundation, it is worthwhile to consider certain prominent sources that offer an extensive background. We can mention Goldberg [9], who developed a comprehensive review of how to solve combinatorial optimization with GA. Hajela et al. [10] discuss how GA can be employed to find optimal designs for complex structural systems, Fonseca [11] introduces the concept of multi-objective optimization GA and several authors developed hybrid or improved versions of this methodology to address specific systems and problems [12–14].Similarly, Torregrossa et al. [15] proposed an innovative way to solve the scheduling problem with a similar nature-inspired metaheuristic. Mambretti, Villacampa, and their team employed GAs to optimize the scheduling of as many

as 31 pumping stations within the city of Milan. They further demonstrated the validity and reliability of this methodology through rigorous testing in a chosen set of field pumping stations [16]. In the field of SI instead, Jang et. al [17] implemented a novel binary ACO to solve the unit commitment problem in power systems. Several applications of ACO to the optimal control of water pumps in water distribution networks can be seen in [18, 19]. Zheng et al. [20] used an adaptive convergence-trajectory controlled ACO and applied it to WDSs. Improved and hybrid versions have been proposed as well, such as in López-Ibáñez et al. [21], where the optimal control of water pumps in a water distribution network was achieved using a hybrid ACO.

The random search method plays a key role within the domain of metaheuristic algorithms. It imparts a crucial element of randomness to the search process to further progress, making the methods less susceptible to modeling inaccuracies and affording them the capacity to break free from local optima, thus converging towards global optima. It is shared among a broad spectrum of metaheuristic algorithms such as GA, ACO and the Cross-Entropy method.

Introduced by Rubinstein in 1997 [22], the cross-entropy (CE) method iteratively refines the probability distribution of solutions, ultimately leading to optimal or near-optimal solutions. It has been used in a variety of fields, such as signal detection by Liu et al. [23]; vehicle routing optimization with stochastic demands by Chepuri and Homem-de-Mello [24]; power system combinatorial optimization problems by Ernst et al. [25]; multidimensional independent component analysis by Szab'o et al. [26]. Additionally, it has been successfully applied by Busoniu et al. [27] to optimal policy search; by Kothari and Kroese [28] to mixed integer nonlinear programming; by Kroese et al. [29] to continuous multi-extremal optimization. In the context of water pump scheduling, the CE method is a niche approach. It excels in handling, multivariable cost functions and a key strength of the CE method is its capacity to analyze solutions without relying on derivative functions, distinguishing it from traditional optimization methods. However, it's important to acknowledge that the initial version of the CE method is better suited for system models with relatively simple constraints. Consequently, an improved version of the CE method proves more suitable for addressing the problem of water pump station optimization.

1.2. The water pump station

A water pump station, also known as a pumping station or water pumping facility, is a critical infrastructure component designed to efficiently transport and distribute water

from its source to various points of use or storage within a municipal or industrial water supply system. These facilities play a vital role in ensuring a reliable supply of clean and potable water to homes, businesses, and other consumers. Water pump stations are strategically located within the water distribution network to maintain adequate pressure, overcome elevation differences, and facilitate the movement of water across vast distances.

In essence, water pump stations are the workhorses of water supply systems, as they are responsible for boosting the pressure of water, often originating from reservoirs, wells, or treatment plants, to ensure it reaches consumers at a sufficient flow rate and pressure. They come in various sizes and capacities, ranging from small, locally situated stations serving neighborhoods to massive facilities that serve entire cities or regions.

The urban raw water supply system primarily draws water from sources such as rivers, lakes, and groundwater, which often exhibit complex natural conditions marked by numerous uncertain factors. This complexity adds a layer of intricacy to the water supply process, necessitating passage through several key facilities, such as the intake pumping station, booster pumping station, and raw water plant. These crucial stations work together to transform raw water into a clean and safe product for consumers. Raw water, at its source, typically contains a variety of impurities, including organic decomposition byproducts, sand particles, algae, iron, bacteria, and viruses. The fundamental phases of a raw water treatment system include:

- Introduction of chemical agents to aid in the flocculation or coagulation of suspended solids.
- Employment of a clarifier to effectively remove larger solid particles.
- Implementation of filtration mechanisms to eliminate smaller particulate matter.

The comprehensive treatment process encompasses several key stages to ensure the purification of raw water for urban consumption:

1. Raw water intake: at the initial stage, untreated water is sourced from natural bodies like rivers, oceans, lakes, or groundwater. Large debris and objects are typically removed at this point to initiate the treatment process.
2. Coagulation: coagulation commences by using mixing reactors and introducing specific chemical agents capable of aggregating smaller particles into larger, more manageable particles that can be easily separated. Slight adjustments in water pH may aid in this coagulation process.
3. Flocculation: after coagulation, the treated water enters a flocculation chamber,

where longer-chain polymers facilitate the gathering of larger particles. This process results in the formation of visible, snowflake-like particles that readily settle.

4. Sedimentation: the sedimentation device, often a large circular structure, directs the water flow into the chamber, allowing it to circulate. This circulation enables the solid particles to settle to the bottom of the clarifier, forming a sludge blanket. The clarified water is then separated from the sludge through a filtering process.

5. Filtration: filtration involves the use of substantial filtration areas, typically filled with two to four feet of sand. The processed water passes through these filters, effectively trapping any remaining particles. In more recent water treatment methods, membranes are employed as a cutting-edge technology for efficient filtration.

6. Disinfection: the disinfection stage serves to cleanse and eliminate any bacteria present in the water, ensuring it is safe for consumption.

7. Distribution: the treated water is pumped into a holding tank and subsequently distributed throughout the urban area through a network, creating a continuous supply loop that serves the needs of the city's residents.

A schematic example of a urban freshwater distribution network in Shanghai is shown in Fig. 1.2.

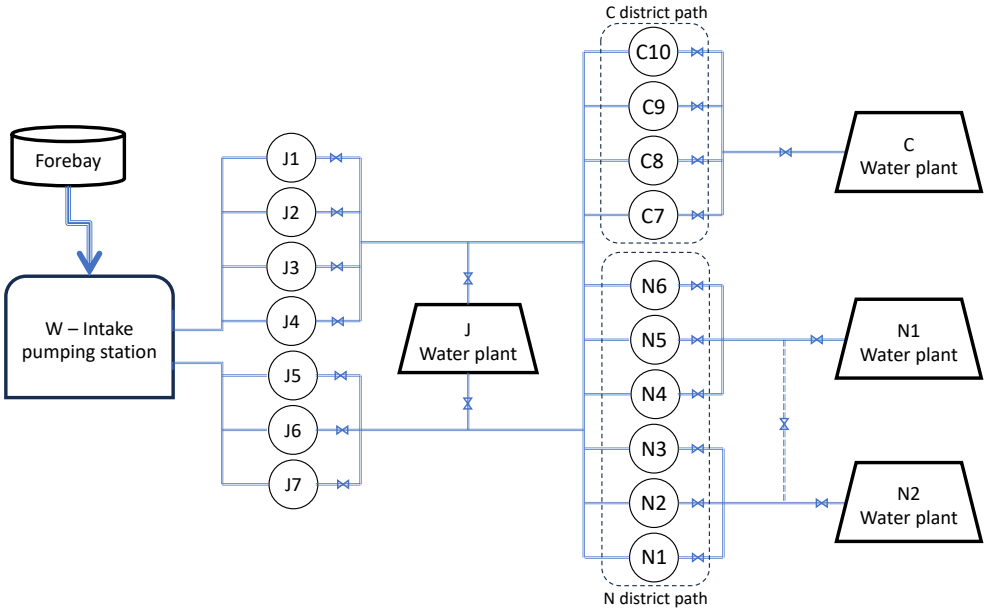


Figure 1.2: Scheme of a WDS located in Shanghai, China.

The raw water system comprises an intake pumping station that conveys raw water from the forebay to the pressurized pumping station, W, and the water treatment plant, J.

The intake station's main role is to acquire raw water, and it involves multiple pumps, a lengthy pipeline, and complex water dispatching. Due to the vast network and distance to downstream water plants, relying solely on the intake station to increase system pressure can be challenging and costly. To address this, booster pumping stations are strategically placed to ensure adequate pressure at each plant, considering the network's characteristics and plant demands. The booster pumps are grouped into two sets, pumps N (N1 through N6) and pumps C (C7 through C10). These pumps are responsible for directing water to their designated districts. The raw water treatment plant is essential for converting raw water into clean water through clarification, sedimentation, and filtration. Regular maintenance is necessary, and the design includes various types of reservoirs with different functionalities.

1.3. The water pump principles

A centrifugal pump is a mechanical device that uses the principles of fluid dynamics to transport liquids from low-pressure zones to high-pressure zones. It may also accelerate liquids through pipes or move them from a low elevation into a higher elevation. A pump works by converting the mechanical energy from a rotating impeller into kinetic energy in the fluid, which then causes the fluid to be displaced and pumped through a system. The central component of a centrifugal pump is the impeller, which is a rotating wheel with curved blades or vanes. The impeller is typically mounted on a shaft connected to a motor. As the impeller rotates, it generates a high-velocity flow of fluid. The pump has an inlet or suction side where the liquid to be pumped enters the pump. This fluid is drawn into the pump by the rotating impeller due to the low-pressure area created at the center of the impeller. The rotation of the impeller imparts centrifugal force to the liquid particles, causing them to move outward from the center of the impeller. The directional movement is illustrated by the arrows depicted in Fig. 1.3, available for reference in [30]. The pressurized liquid exits the pump through a discharge nozzle. The pressure at the discharge is higher than the pressure at the inlet due to the conversion of kinetic energy into pressure energy. Finally, as the impeller continues to rotate, a constant flow of fluid is drawn in from the inlet and discharged from the outlet achieving the pump's main operation, which is to provide a continuous, consistent flow rate and pressure.

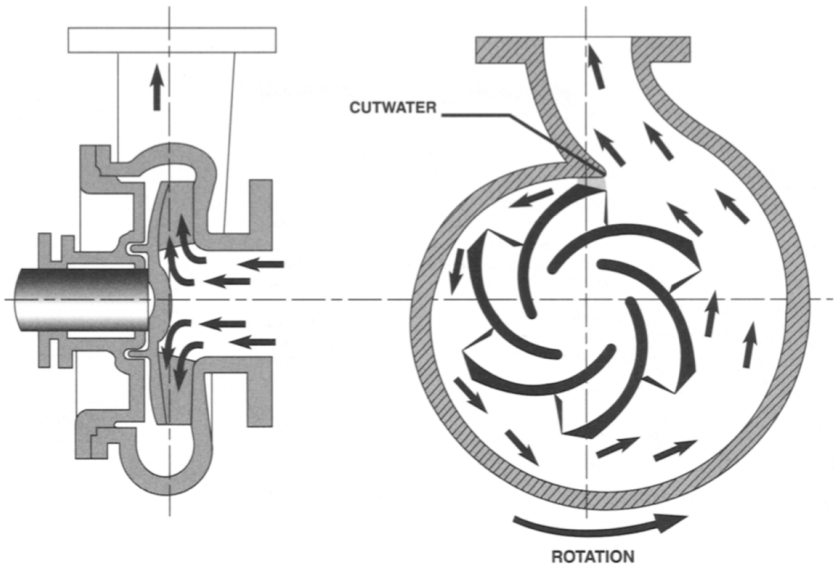


Figure 1.3: Impeller and casing of a centrifugal pump.

1.3.1. The affinity laws

The affinity laws are a set of rules that govern the behavior of water pumps [31]. They served, in the past, to guess the pump's working point when the exact pump parameters were not available, such as when working with different electricity standards in foreign countries. In recent decades, with the development of the variable speed electric motor and the Variable Frequency Drive (VFD), the affinity laws became increasingly important. They describe the fundamental relationship between the pump's head, flow, power and speed.

Considering the following operational characteristics:

- Flow, Q: the liquid volume, measured in cubic meters per hour (m^3/h).
- Head, H: the liquid force, measured in meters (m).
- Speed, S: the shaft speed, measured in revolutions per minute (rpm).
- Power, P: the energy needed to pump a liquid, measured in kilowatts (kW).

the affinity laws state that:

- The flow, Q, is directly proportional to the velocity variation.
- The head, H, is directly proportional to the square of the velocity variation.
- The power, P, is directly proportional to the cube of the velocity variation.

which can be expressed by the their equation form:

$$\begin{aligned} \frac{Q_1}{Q_2} &= \frac{S_1}{S_2} \\ \frac{H_1}{H_2} &= (\frac{S_1}{S_2})^2 \\ \frac{P_1}{P_2} &= (\frac{S_1}{S_2})^3 \end{aligned} \tag{1.1}$$

Where the subscrips '1' and '2' denote the initial and new values of the parameters. It can be noted from equation (1.1) that a simple 3% increase in speed results in a significant 9% surge in power consumption. In the industrial sector, this translates to a substantial rise in costs, highlighting the importance of optimal solutions.

The affinity laws also describe the relationships between the pump's impeller diameter, flow, head and power. Considering a constant operational velocity, a variation in the impeller size would affect the pump such that:

- The flow, Q, is directly proportional to the diameter variation.
- The head, H, is directly proportional to the square of the diameter variation.
- The power, P, is directly proportional to the cube of the diameter variation.

which, in their equation form are expressed by:

$$\begin{aligned} \frac{Q_1}{Q_2} &= \frac{D_1}{D_2} \\ \frac{H_1}{H_2} &= (\frac{D_1}{D_2})^2 \\ \frac{P_1}{P_2} &= (\frac{D_1}{D_2})^3 \end{aligned} \tag{1.2}$$

The clear implication is that adjusting the flow and pressure by altering the impeller diameter can significantly save kilowatts of energy. For example, a 10% reduction in impeller diameter could lead to a 30% decrease in power consumption. This reduction in energy usage, eventually, not only offsets the expenses associated with additional labor and the maintenance of multiple impeller pumps but also makes it a cost-effective solution.

1.3.2. The system curve

The Affinity Laws provide intuitive insights and a practical understanding of the operation of centrifugal pumps. However, to develop a useful and optimized model, it is essential

to understand the system curve.

Water pumps serve as reactive elements in WDSs, responding to the system's operational requirements to meet the supply demands. For instance, an increased flow demand would affect the pressure head (Hp), while the addition of pipes to the network would impact the friction head (Hf). Deteriorated or undersized pipes influence both flow and velocity head (Hv), and elevated water levels in storage tanks introduce static head (Hs). The parameter that sums up the requirements of each pumping system and characterizes its dynamics is known as the Total Dynamic Head (THD).

$$TDH = Hp + Hf + Hv + Hs \tag{1.3}$$

The TDH is composed of the following four head values:

1. Hp — The pressure head: describes the pressure changes across the system. It is expressed in meters of head. In case there is no pressure change, then this term doesn't exist. Consider a water recirculation filter pump, commonly utilized in swimming pools to both drain and replenish water within the same container as an example.

$$Hp = \frac{\Delta psi \times 2.31}{sp.gr.} \tag{1.4}$$

   Where Δpsi is the pressure difference and 'sp. gr.' is an abbreviation for specific gravity or relative density.

2. Hf — The friction head: describes the system head losses caused by the friction between the internal walls of the pipes, valves, and the water. It is expressed in meters of head. It is an important factor to take in consideration during the design of the systems since the energy losses in the system can not be avoided.

$$Hf = \frac{Kf \times L}{100} \tag{1.5}$$

   In equation (1.5) Kf is the friction constant for every 38.48m of pipe according to tables, and L is the actual length of the pipe.

3. Hv — The velocity head: describes the energy lost within the system as a result of the fluid's motion through the pipelines.

$$Hv = \frac{V^2}{2g} \tag{1.6}$$

Where V is the velocity of the fluid moving through the pipe.

While its magnitude may be relatively small, Hv plays a significant role as it is necessary for calculating the friction head.

$$Hf = K \times Hv \tag{1.7}$$

with K a constant derived from tables.

4. Hs — The static head: describes the variation in liquid elevation throughout the system. It is expressed in meters of elevation change. It represents the difference in liquid surface levels between the point of origin, typically at the suction source, and the point where the pump discharges the liquid.

The system curve plays a fundamental role in establishing the best efficency point (BEP) at which the pumps are expected to operate. The BEP is the working point of the pump in which the operational parameters have values such that the efficiency is maximized. It is located by intersecting the system curve with the pump performance family curve, as shown in the following graph.

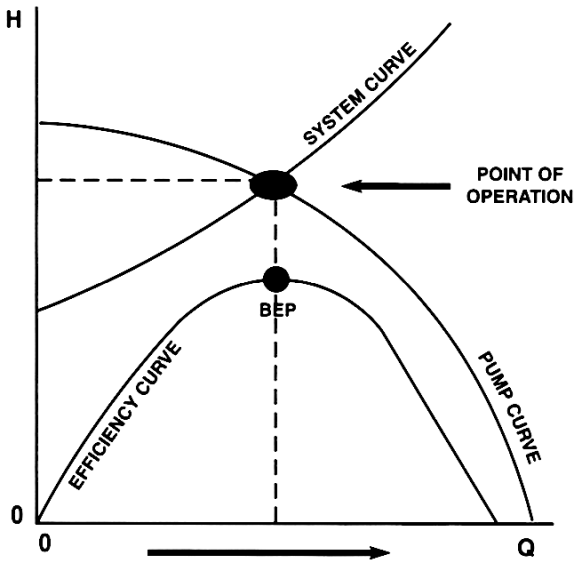


Figure 1.4: The best operational point (BEP).

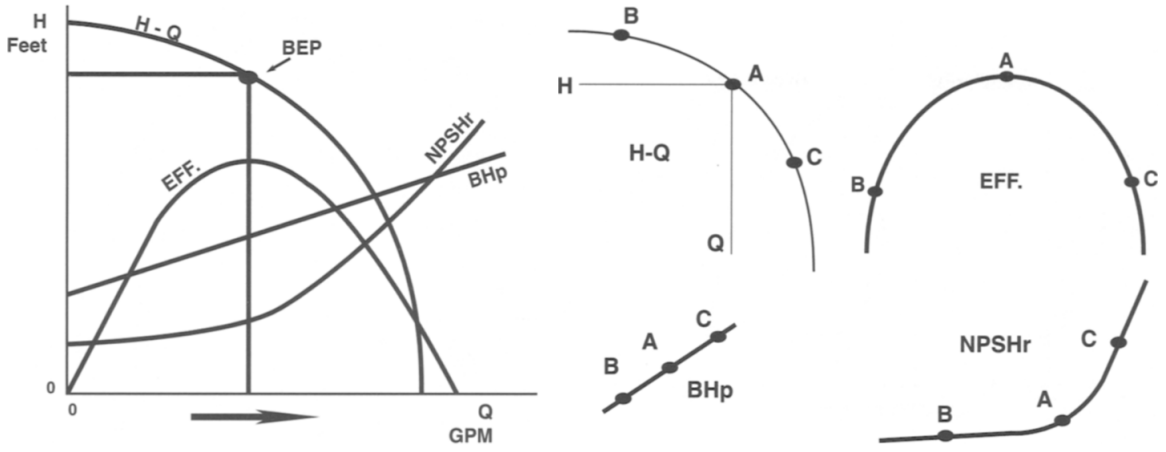
1.3.3. The performance curves

The pump performance curves provide detailed information about the performance characteristics of the pump. These curves typically display how the pump behaves under

different operating conditions. The key elements typically found on a pump curve include:

- Head (H) vs. Flow rate (Q).
- Power (BHp) vs. Flow rate (Q).
- Efficiency (η) vs. Flow rate (Q).
- NPSH vs. Flow rate (Q).

Understanding these curves is essential for selecting the right pump for a specific application and for optimizing its performance within a given system. They provide insights into how the pump will perform under various conditions and help in making informed decisions regarding pump selection and operation. In this section the performance curves of centrifugal pumps with negligible static head are presented.



(a) Summary of the key curves.

(b) Highlight of the specific points A, B, C.

Figure 1.5: The pump performance curves.

The Head and Power curves

The head-flow curve, commonly know as HQ curve, shows how the pump's total head, the energy imparted to the fluid, changes as the flow rate through the pump varies. It helps in understanding how the pump handles different flow rates and pressure requirements.

The power-flow curve shows the power consumption of the pump at different flow rates. It's essential for assessing the energy requirements of the pump under different conditions.

As variable frequency pumps see greater adoption in the industrial sector, it becomes increasingly important to discern them from conventional fixed frequency pumps. Fixed-frequency pumps maintain a constant speed ratio. Their performance curves for head and flow, as well as power and flow, can be defined with:

$$\begin{aligned} H &= h_1 + h_2 Q + h_3 Q^2 \\ P &= p_1 + p_2 Q + p_3 Q^2 \end{aligned} \tag{1.8}$$

Where $h_1, h_2, h_3, p_1, p_2, p_3$ are fitting parameters. Combining the pump affinity laws and the equations (1.8), the expressions for the variable frequency pumps are found.

$$\begin{aligned} H &= s^2 h_1 + h_2 Q + h_3 Q^2 \\ P &= s^3 p_1 + s^2 p_2 Q + s^3 p_3 Q^2 \end{aligned} \tag{1.9}$$

Where s is the speed ratio of the pump.

In variable frequency pumps, alterations in speed significantly influence the pump's operating point due to the quadratic and cubic relationships of the head and power functions with the speed. As depicted in Fig. 1.6, a decrease in operating speed causes a downward shift in both the head vs. flow curve and the power vs. flow curve. Consequently, when confronted with a constant flow demand, lowering the operating speed can result in reduced power consumption. However, it's crucial to note that this adjustment also decreases the head value, therefore, exercising careful consideration is essential when utilizing speed control to operate a pump.

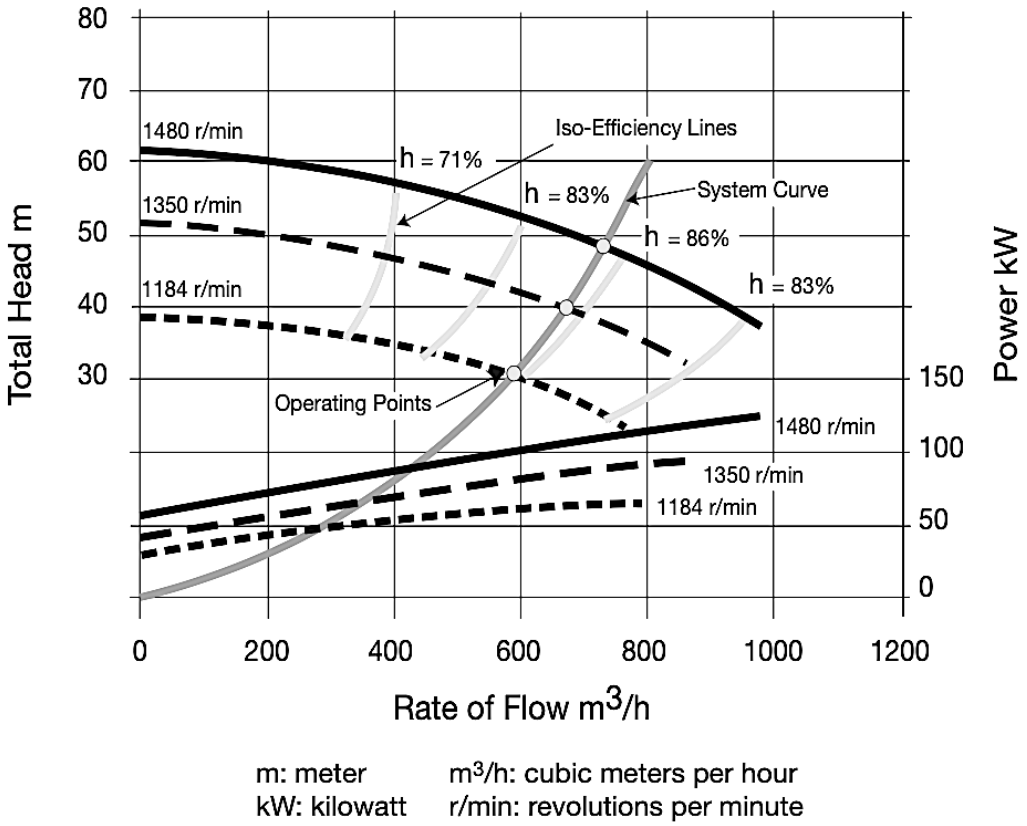


Figure 1.6: The effect of speed variation on the head and power curves.

The Efficiency and NPSH curve

The efficiency vs. flow curve is an essential component of the performance curves for water pumps. This curve shows how the pump's efficiency changes as the flow rate varies. A fundamental goal in the operation of water pumps is to maintain the pump's working point close to the BEP, which corresponds to the peak of the efficiency curve.

Unlike the head and power curves, small variations in speed do not significantly impact the pump's efficiency. For instance, when the speed is reduced, it moves the intersection point on the system curve along a line of constant efficiency, as illustrated in Fig. 1.7.

Several factors that influence efficiency also include:

- The diameter of the eye of the impeller.
- The angle of the impeller blades.
- The thickness of the impeller.

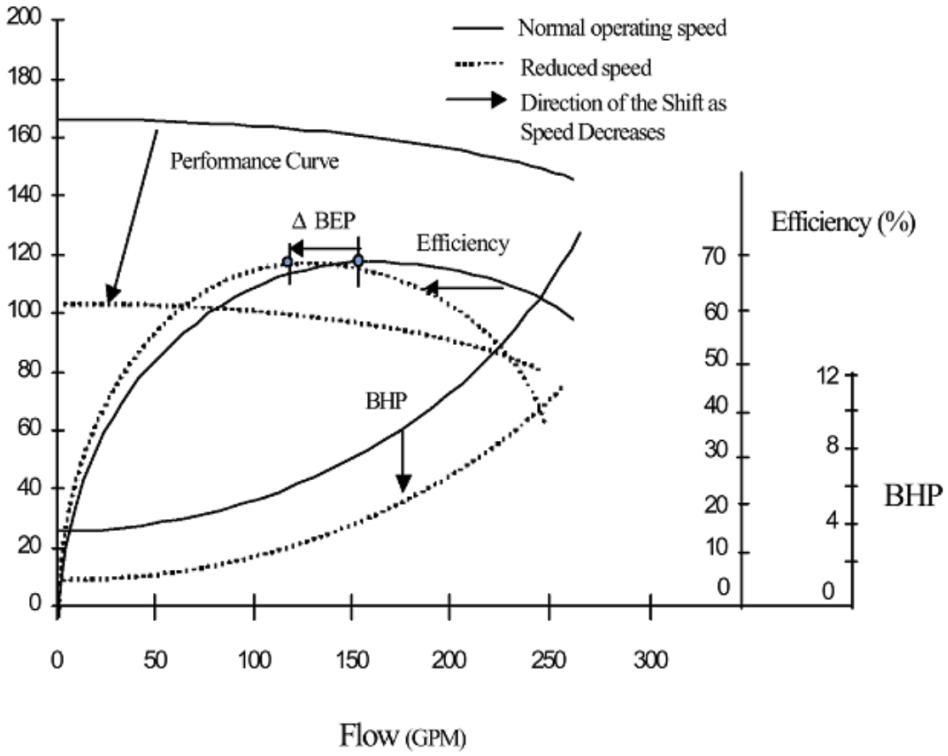


Figure 1.7: The effect of speed variation on the pump efficiency η.

- The impeller velocity.
- The size of solid particles in the pump casing.
- Mechanical losses.
- Viscosity of the liquid.

Another crucial parameter in centrifugal pump operation is the net positive suction head (NPSH). The NPSH value of a pump specifies the minimum pressure at the suction nozzle required for the pump to function properly, often set at a specific condition like a 1% reduction in head. Failure to meet NPSH requirements, which means insufficient liquid supply to the pump's suction side, can result in undesirable and costly consequences, such as cavitation, leading to increased maintenance expenses.

2 | The pump scheduling optimization problem

The water pump scheduling optimization problem involves finding the most efficient and cost-effective way to operate multiple water pumps in a WDS to meet demand while minimizing energy consumption, operational costs, and other constraints. Currently, in most Chinese cities, water distribution scheduling is managed by dispatcher teams who activate valves and pumps based on predefined shifts. While this ensures continuous system operation, it has drawbacks related to human intervention, team management, and long-term system stability. Dispatchers rely on past experiences to estimate water demand, leading to biased decisions. This approach can result in fluctuations in water supply pressure, requiring adjustments that decrease the quality of the urban water supply. Therefore, it is crucial to introduce optimization methods, such as the cross-entropy method, along with an accurate mathematical model of the problem. These tools enable more precise and efficient decision-making, addressing the challenges associated with human intervention and leading to improved water supply stability and quality. This benefits both the urban water systems and the consumers they serve.

2.1. A combinatorial optimization problem

The water pump scheduling is a well-known combinatorial optimization problem in the WDS domain. It is a type of mathematical problem in which the goal is to find the best possible solution from a finite set of discrete choices.

Considering a model $P = (S, \Omega, f)$, a combinatorial optimization problem is characterized by the following properties:

- The search space, denoted as S, encompasses a finite collection of components or discrete decision variables $C = \{c_1, c_2, ..., c_n\}$.
- A finite set L is established by selecting values from the elements within $\tilde{C}$, where $\tilde{C}$ is a subset of the Cartesian product $C \times C$. This set is represented as $L =$

$\{l_{c_i,c_j} | (c_i, c_j) \in \tilde{C}\}$, $|L| \leq N_C^2$, where N_C denotes the cardinality of C.

- A finite set of constraints, denoted as Ω, is assigned among the elements.
- The model comprises an objective function f that maps from the search space S to non-negative real numbers ($\mathbb{R}_0^+$) and is designed to be minimized, serving as the central focus of the model.
- A reasonable solution $s \in S$ is a solution that complies with all the constraints defined within the model.
- A global optimal solution s^* is found if a solution s is unique and satisfies the expression $s^* \in S \Leftrightarrow f(s^*) \leq f(s), \forall s \in S$.

A crucial aspect of combinatorial problems revolves around the management and handling of constraints within the problem-solving process. The manner in which constraints are mathematically modeled significantly impacts the performance of optimization algorithms. It's crucial to differentiate between direct and indirect constraints. Indirect constraints can be assimilated into the objective function, where optimizing the objective function inherently meets these constraints. In contrast, direct constraints stand apart from the objective function and are incorporated to ensure optimal outcomes.

Many optimization models integrate both direct and indirect constraints. Implementing indirect constraints in optimization involves reshaping the problem and employing the optimization function to penalize any breaches of constraints. These penalties represent the expense of narrowing the discrepancy between achieved and desired outcomes. They can be adjusted using two distinct penalty factors: one for constraint violations and the other for inaccurately initialized variables. We can express mathematically such conditions as in expressions (2.1).

Considering a set of constraints c_i $(i = 1, ..., m)$ applied to a set of variables v_j $(j = 1, ..., n)$, assuming C^i a set of constraints directly related to v_j, and w_i the parameters that determine the magnitude of the penalty imposed, the constraint conditions can be represented by the following penalty functions.

$$
\begin{aligned}
f_1(s) &= \sum_{i=1}^{m} w_i \chi(s, c_i) \quad \text{with} \quad \chi(s, c_i) = \begin{cases} 1, & \text{if } s \text{ violates } c_i \\ 0, & \text{otherwise} \end{cases} \\
f_2(s) &= \sum_{i=1}^{m} w_i \chi(s, c_i) \quad \text{with} \quad \chi(s, C^i) = \begin{cases} 1, & \text{if } s \text{ violates at least one } c \in C^i \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
\tag{2.1}
$$

2.2. The objective function

In combinatorial optimization, the objective function plays a central role in defining the goal of the problem and guiding the search for an optimal solution. This numerical value represents the quality or cost of the solution with respect to the problem's optimization goals. The goal of our combinatorial optimization problem is to find the best solution, which is the one that minimizes the value of the objective function, while respecting the constraints.

The objective function utilized to optimize the water pump scheduling problem considers several factors, with the primary focus being the power consumption of each pump. The total power usage is computed by summing the individual contributions from all operational pumps.

$$J = min\left\{\sum_{i=1}^{m} \omega_i(s_i^3 p_1 + s_i^2 p_2 Q_i + s_i p_3 Q_i^2) + \sum_{i=m+1}^{m+n} \omega_i(p_1 + p_2 Q_i + p_3 Q_i^2)\right\} \tag{2.2}$$

By utilizing equation (2.2), we may calculate the total power consumption of all pumps. The coefficients ω_i represent the on-off status of each pump and are pivotal in determining the optimal combination of pumps within the scheduling solution. The values p_1, p_2, p_3 represent the parameters of the power function associated with each particular type of pump.

2.3. The constraints

Constraints in optimization problems serve to capture the limitations and requirements of the problem at hand. Their role is to narrow down the solution space, ensuring that the final solution optimizes the objective function. Constraint modeling involves translating these real-world limitations into mathematical equations, inequalities, or logical expressions that can be integrated into the optimization problem. This process is crucial due to the fact that an accurate model is able to affect the optimization algorithm positively, for example, by improving its performance and accuracy.

In the water pump scheduling optimization problem, the main constraints are four, the speed ratio constraint, the flow balance constraint, the parallel operation of pumps, and the limitation related to the high-efficiency area.

2.3.1. Speed ratio constraint

Variable speed pumps can be controlled to change their speed ratio within the interval $s \in [0, 1]$. However, to increase efficiency and prevent issues like cavitation and shortened life cycles, the speed ratio of each pump has been constrained to the interval shown in (2.3).

$$s_{min,i} \leq s_i \leq 1, \quad \text{for } i = 1, 2, ..., m + n \tag{2.3}$$

Where m, n are, respectively, the number of variable speed pumps and fixed speed pumps. In the present project the interval has been set to $s \in [0.8, 1]$.

2.3.2. Flow balance constraint

The flow balance equation is a fundamental principle that describes how a fluid flows through a system while maintaining mass conservation. It states that the output flow of the system Q_s must be equal to the sum of the input flow generated by the pump station, which is the sum of the output flows from both the variable and fixed speed pumps.

$$Q_s = \sum_{i=1}^{m} Q_i + \sum_{j=m+1}^{m+n} Q_j \tag{2.4}$$

2.3.3. Parallel operation of pumps

A pump station typically employs multiple pumps operating in parallel. When these pumps run in parallel, it's essential that they operate at the same head. The total output flow of the system is calculated as the sum of the discharges from all the active pumps. In other terms, the discharge is obtained by superimposing the characteristic curves of all active pumps on the same horizontal axis. Fig. 2.1 illustrates this concept with two pumps running in parallel.

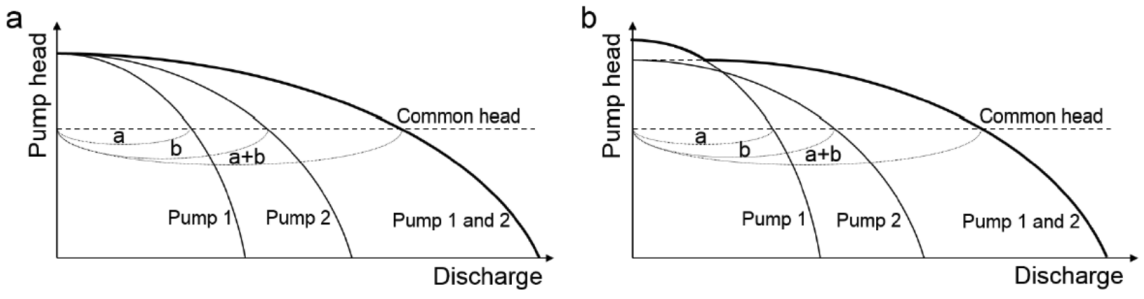


Figure 2.1: Pumps in parallel scheme.

Pumps operated in parallel will increase the flow but not the head. When this requirement is not met, there is the risk of water flowing backward toward the pump with lower head, potentially causing significant damage. As a result, condition (2.5) is imposed as a constraint.

$$H_s = H_1 = H_2 = ... = H_{m+n} \tag{2.5}$$

Where H_s is the system's operating head.

2.3.4. High-efficiency area

It is crucial to operate the pumps such that the control variables maximize the machine's efficiency. For fixed frequency pumps, this corresponds to an interval defined by a maximum value, $Q_{max,i}$, and a minimum value, $Q_{min,i}$, of pump flow.

$$Q_{min,i} \leq Q_i \leq Q_{max,i}, \quad \text{for } i = 1, 2, ..., n \tag{2.6}$$

In the case of variable frequency pumps, the speed parameter extends the interval to encompass a 'best efficiency area' (BEA). The BEA is delimited by flow and speed constraints, which are indicated by the curves connecting the four points A, B, E, and F.

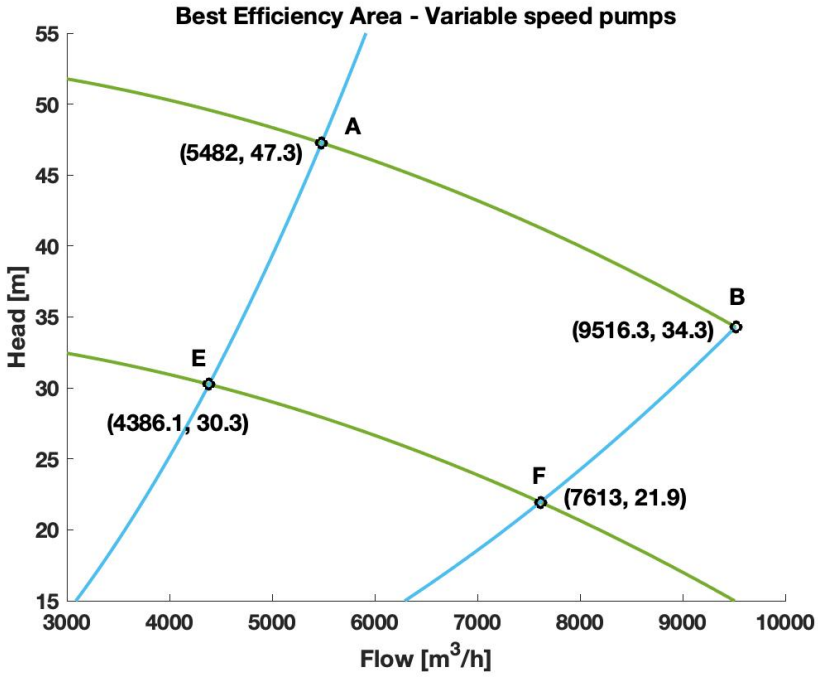


Figure 2.2: BEA delimited by A, B, E, and F.

This region determines the area where the pump efficiency, η, is equal to or greater than 85%. The curves AB and EF are the head-flow curves of the pump, respectively, at the maximum speed ratio and the minimum speed ratio. In this particular setting, the minimum speed has been limited to $S_{min} = 0.8$, in relation to equation (2.3) and to decrease maintenance costs associated to shortened life cycles. The parabolic curve EA corresponds to an efficiency of $\eta = 85\%$, while the FB curve to an efficiency of $\eta = 89.2\%$. It can be observed that as the pump's speed decreases, the parabolic curves EA and FB also exhibit a narrowing effect on the BEA.

The two boundary curves OA and OB, where O is the origin, and which respectively contain the curves EA and FB, are described by the equations:

$$\begin{aligned} H_{OA} &= k_1 Q^2 \\ H_{OB} &= k_2 Q^2 \end{aligned} \tag{2.7}$$

By reformulating the head expression in equation (1.9), we get the flow value Q.

$$Q = \sqrt{\frac{h_1 s^2 - H_s}{h_3}} \tag{2.8}$$

In the case of the AB curve, given that the maximum operating frequency sets $s = 1$, the flow value assumes the form (2.9).

$$Q = \sqrt{\frac{h_1 - H_s}{h_3}} \tag{2.9}$$

Whereas the EF curve, corresponding to the minimum operating frequency of $s_{min} = 0.8$, is described by the expression:

$$Q = \sqrt{\frac{h_1 s_{min}^2 - H_s}{h_3}} = \sqrt{\frac{h_1 \times 0.64 - H_s}{h_3}} \tag{2.10}$$

Therefore, it is possible to pinpoint the points A, B, E and F of the BEA specified in Fig. 2.2 by using equations (2.7), (2.8), (2.9) and (2.10). Finally, the maximum and minimum flow values in equation (2.6) are found by combining the parabolic curves in equations

(2.7) and the pump affinity laws described in equations (1.1).

$$
\begin{aligned}
Q_{max,i} &= \begin{cases} \sqrt{\frac{h_1 - H_s}{h_3}} Q_A & H_s \geq H_B \\ \sqrt{\frac{H_s}{H_B}} Q_B & H_s < H_B \end{cases} \\
Q_{min,i} &= \begin{cases} \sqrt{\frac{H_s}{H_A}} Q_A & H_s \geq H_E \\ \sqrt{\frac{h_1 s_{min}^2 - H_s}{h_3}} & H_s < H_E \end{cases} \\
&\text{for } i = 1, 2, ..., m+n
\end{aligned}
\tag{2.11}
$$

Where H_B, H_F are the head values at point B and point F, respectively.

2.3.5. Alternated pump usage

In the context of a water pumping station's crucial task to continuously meet the water demands of its customers for the entire day, prolonged operation of a water pump carries the risk of damage. Such damage could disrupt the station's normal operation and lead to undesired faults. Water consumption varies across seasons and weather conditions, prompting water plants to maintain surplus pump capacity to cover daily water supply demands.

To maximize the lifespan of each pump and reduce energy consumption, station dispatchers may periodically rotate which pumps are active. This rotation is not executed with every new dispatching instruction but is determined based on a chosen time frame. In our current project, we opt to select pumps based on the specific demand and date. For example, in a pumping station equipped with four pumps, if only two pumps are required to meet the daily demand, we run pumps 1 and 2 on even-numbered dates and pumps 3 and 4 on odd-numbered dates.

2.4. Mathematical model summary

Once the constraints and the objective function related to the characteristic curves and the Affinity Laws have been defined, we obtain the final problem formulation. The general model is characterized as a non-linear, multi-objective, multi-variable combinatorial problem, encompassing both equality and inequality non-linear constraints.

$$
J = min\left\{ \sum_{i=1}^{m} \omega_i (s_i^3 p_1 + s_i^2 p_2 Q_i + s_i p_3 Q_i^2) + \sum_{i=m+1}^{m+n} \omega_i (p_1 + p_2 Q_i + p_3 Q_i^2) \right\} \tag{2.12}
$$

such that

$$\begin{aligned} & Q_s = \sum_{i=1}^{m} Q_i + \sum_{j=m+1}^{m+n} Q_j \\ & H_s = H_1 = H_2 = ... = H_{m+n} \\ & s_{min,i} \leq s_i \leq 1, \qquad\qquad i = 1, 2, ..., m+n \\ & Q_{min,i} \leq Q_i \leq Q_{max,i}, \qquad\qquad i = 1, 2, ..., m+n \end{aligned} \tag{2.13}$$

Therefore, in addressing such a problem, the CE method proves to be suitable, whereas classical methods frequently falter when faced with intricate models.

3 | The Improved Cross-Entropy Method

The water pump scheduling optimization problem poses difficult challenges due to its high non-linearity and its complex system constraints. In the recent decades, it has been tackled by researchers and engineers with a variety of algorithms. Traditional methods such as linear programming and dynamic programming often show to be ineffective or unable to deal with complicated mathematical models, while more recent metaheuristics, such as the genetic algorithm or the ant colony optimization method, despite many improvements and adaptations, still face hardships while dealing with complex constraints. In this section, the cross-entropy method is presented, together with its improved version, with the aim of solving the pump scheduling optimization problem effectively. This work has also the goal of expanding the literature and background on the CE method and to provide guidelines on its application, eventually, for other scholars and future researches.

3.1. The classical CE method

The cross-entropy method is a relatively recent algorithm derived from the Monte Carlo method. The CE method initially proposed by Rubinstein was meant to be applied for rare event simulation, but it soon became clear that with slight adaptations it could be applied to combinatorial and multi-extremal optimization. The name derives from the process of minimization of the Kullback–Leibler divergence (or cross-entropy), defined as the distance or closeness between two sampling distributions. The key idea behind the cross-entropy method is that starting from an initial distribution of parameters, it iteratively refines the probability distribution of solutions to focus on the most promising regions of the solution space, ultimately converging towards an optimal or near-optimal solution. Over time, this process can converge to an optimal or near-optimal solution for the problem, making it suitable in domains where traditional optimization methods may struggle due to high dimensionality or non-convexity.

The CE method can be employed to solve two types of problems:

1. Estimation problems: it involves determining $l = E[H(X)]$, where X represents a random object with values in a set $\mathcal{X}$ and H is a function defined on $\mathcal{X}$. A significant case is the estimation of a probability, $l = P(S(X) \geq \gamma)$, where S is another function defined on $\mathcal{X}$, and γ is a threshold value.

2. Optimization problems: the goal is to maximize or minimize a specified objective function $S(x)$ across the entire set $x \in \mathcal{X}$. This function S might be a deterministic known function or subject to noise, introducing uncertainty into its evaluation.

To understand the logic behind the algorithm it is helpful to understand first how it is used to estimate the probability of rare events. Afterwards, it will be modified in order to be applied in combinatorial optimization problems.

3.1.1. Principles and mathematical formulation

Importance sampling

The foundation of the CE method is the concept of importance sampling. Consider the goal of wanting to compute the expectation of a function $z(x)$ when $x \sim P$ with the density function $p(x)$. The expectation is found according to

$$\mathbb{E}_p[z(x)] = \int z(x)p(x)dx. \tag{3.1}$$

Now, let's estimate this expectation through a Monte Carlo estimator, in other words, a sampling technique. By drawing samples x_i from P it is possible to estimate the expectation in the following way:

$$\begin{aligned} r &= \mathbb{E}_p[z(x)] \\ \hat{r} &= \frac{1}{n}\sum_{i=1}^{n} z(x_i) \end{aligned} \tag{3.2}$$

Consider a scenario where we find it impractical to utilize distribution P, perhaps due to difficulties in sampling from it due to an unknown exact form. In response, we introduce an alternative distribution, Q, accompanied by a straightforward density function $q(x)$. For instance, we may opt for a simple parametric distribution such as a Gaussian distribution. Employing this approach allows us to simplify our expectation computation, making it more manageable.

$$\begin{aligned}\mathbb{E}_p[z(x)] &= \int z(x)p(x)dx, \\ &= \int z(x)\frac{p(x)}{q(x)}q(x)dx, \\ &= \mathbb{E}_q[z(x)\frac{p(x)}{q(x)}]\end{aligned} \tag{3.3}$$

As a result, the expectation under distribution P has undergone a transformation to become an expectation under distribution Q, denoted as $x \sim q(x)$. To calculate this expectation through sampling, we generate samples x_i from Q and employ a correction factor known as the likelihood ratio. This correction factor is applied to compute the expectation.

$$r = \mathbb{E}_p[z(x)] = \mathbb{E}_q[z(x)\frac{p(x)}{q(x)}] \tag{3.4}$$

$$\hat{r} = \frac{1}{n}\sum_{i=1}^{n} z(x_i)\frac{p(x_i)}{q(x_i)}. \tag{3.5}$$

Now, it's important to select the optimal choice of q. By recalling the fact that an estimator for r is:

$$\hat{r} = \frac{1}{n}\sum_{i=1}^{n} z(x_i)\frac{p(x_i)}{q(x_i)} \tag{3.6}$$

We may choose the value $q^*(x)$, shown below, apply the importance sampling methodology, and substitute it into equation (3.6).

$$\begin{aligned}q^*(x) &= \frac{z(x)p(x)}{r} \\ \hat{r} &= \frac{1}{n}\sum_{i=1}^{n} z(x_i)\frac{p(x_i)}{q^*(x_i)} \\ \hat{r} &= r\end{aligned} \tag{3.7}$$

In this analysis, it becomes evident that the estimator $\hat{r}$ perfectly recovers r, resulting in zero variance. Consequently, in general the importance density $q(x)$ from which we draw samples should closely approximate $q^*(x)$. This is where the concept of cross-entropy becomes crucial, as it serves as a metric quantifying the dissimilarity between distributions.

Rare event probability estimation

Now, in the context of rare event probability estimation, suppose to estimate the probability l.

$$l = P_u(S(x) \geq \gamma) \tag{3.8}$$

where $x \sim f(\cdot; u)$.

In this context, S represents a function of x, which will subsequently evolve into the objective function of the optimization problem. The variable x adheres to the distribution $f(\cdot; u)$, where u serves as the parameter determining the probability distribution f.

To find this probability, we consider the equivalent expectation finding problem, which is derived from the observation that

$$l = \mathbb{E}_u[\mathbb{I}_{(S(x)\geq\gamma)}] \tag{3.9}$$

The Monte Carlo estimate of the above expression is

$$\hat{l} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(S(x_i)\geq\gamma)} \tag{3.10}$$

with $x_i \sim f(\cdot; u)$ and $\mathbb{I}$ representing the indicator function.

Give the assumption that the problem faced is a rare event estimation problem, the instances in which $S(x_i) \geq \gamma$ are very few. Therefore, it is useful to introduce another distribution $g(x)$ and substitute it into equation (3.10) to obtain the new estimator $\hat{l}$.

$$\hat{l} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(s(x_i)\geq\gamma)} \frac{f(x_i; u)}{g(x_i)} \tag{3.11}$$

where $x_i \sim g(x)$.

The problem of the optimal choice of the density function g(x) remains and the solution to it comes from the importance sampling technique, which suggests the optimal function:

$$g^*(x) = \frac{\mathbb{I}_{(s(x_i)\geq\gamma)} f(x; u)}{l} \tag{3.12}$$

The objective is to discover an importance sampling density that closely aligns with $g^*(x)$. Recognizing that the space of density functions is infinite, we simplify our task by confining ourselves to the same family as the original problem, f, as we have a known sampling

method for it. Consider our importance sampling density function to be parameterized by v, such that $g(x) = f(x; v)$. Therefore, our focus shifts to finding the optimal v that minimizes the cross entropy between $g^*(x)$ and $f(x; v)$, denoted as $D(g^*(x), f(x, v))$.

Therefore, the problem becomes the following minimization problem:

$$\begin{aligned} v^* &= \arg\min_v D(g^*(x), f(x, v)) \\ &= \arg\min_v \mathbb{E}_v[ln\frac{g^*(x)}{f(x; v)}] \\ &= \arg\min_v \int g^*(x) ln\, g^*(x) dx - \int g^*(x) ln\, f(x; v) dx \\ &= \arg\min_v - \int g^*(x) ln\, f(x; v) dx \end{aligned} \tag{3.13}$$

Now reformulating the expression into a maximization problem, multiplying by a minus sign, and substituting the expression of $g^*(x)$, the expectation l becomes independent from v.

$$v^* = \arg\max_v \int \frac{\mathbb{I}_{(s(x_i)\geq\gamma)} f(x; u)}{l} ln\, f(x; v) dx \tag{3.14a}$$

$$= \arg\max_v \int \mathbb{I}_{(s(x_i)\geq\gamma)} f(x; u) ln\, f(x; v) dx \tag{3.14b}$$

$$= \arg\max_v \mathbb{E}_u[\mathbb{I}_{(s(x_i)\geq\gamma)} ln\, f(x; v)] \tag{3.14c}$$

Hence, we are now faced with a maximization problem, the solution of which provides us with a density function for sampling x. This density function is aimed at closely approximating the optimal importance sampling density function. Equation (3.14c) is solvable analytically, in particular for distributions falling within the natural exponential family. The best way to find its solution is in an iterative way and through the refinement steps we may find compute the value v_{t+1} according to:

$$v_{t+1} = \arg\max_{v_t} \mathbb{E}_{v_t}[\mathbb{I}_{(s(x_i)\geq\gamma_t)} ln\, f(x; v_t)] \tag{3.15}$$

The process involves selecting elite samples based on their performance in $S(x)$, a crucial step where these elites establish the threshold level γ_t. This threshold is used to transform the original problem into a less rare scenario through the modified problem $S(x) \geq \gamma_t$, where x follows the distribution $f(x; v_t)$. Remarkably, the solution to equation (3.15) constitutes the maximum likelihood estimate derived from the elite samples. Consequently, a new threshold γ_t and parameter v_t is found at each iteration. This iterative approach incrementally improves the threshold γ_t and the distribution $f(x, v_t)$ gradually concentrating it more around the points x where $S(x) \geq \gamma$.

CE method for combinatorial optimization

The CE method methodically reshapes the distribution over the independent variable x, gradually focusing its entire mass on the realization of the targeted event. This approach can be adapted for combinatorial optimization problems, leveraging a modification of its scheme. Such adaptability is feasible when optimization problems can be adjusted and reformulated in a manner akin to a rare event estimation problem.

Considering the objective function

$$\max_x S(x) \tag{3.16}$$

suppose that its maximum value γ^* is found at the maximum x^*. Hence, we have:

$$\begin{aligned} x^* &= \arg\max S(x) \\ \gamma^* &= S(x^*) \end{aligned} \tag{3.17}$$

As shown in the previous sections, consider

$$l(\gamma) = P_u(S(x) \geq \gamma) \tag{3.18}$$

in which γ acts as the variable and its precise value is not known, which is the first difference from the rare event simulation scheme.

Our objective in this context is to determine γ^*. At very few points in the search space over x, $S(x)$ equals γ^*. Consequently, $S(x) \geq \gamma^*$ qualifies as a rare event under the uniform distribution, given that only a limited number of points x^* will achieve the optimum.

By doing so, we have transformed our optimization problem from equation (3.16) into a rare event simulation problem. The goal is to estimate $l(\gamma^*)$ under $x \sim unif$. It has to be noted that this approach makes sense also when γ^* is unknown, because we gradually increase the threshold value γ. The anticipation is that this incremental adjustment of γ will eventually lead to γ^*. Simultaneously, our distribution over x becomes increasingly focused, concentrating all its probability mass on x^*.

3.1.2. The CE method algorithm

Here the pseudo-algorithm of the cross-entropy method is presented. After defining the maximum number of iteration, the size of the population, the percentage of the elite elements to be considered and the boundaries of the variables, the algorithm proceeds in the following scheme.

- Initialization: the initial mean and covariance matrices are defined. This ensures that the search space is covered properly. Afterwards, a random generation of the initial population is obtained from the multivariate normal distribution.
- Iterative phase: until the convergence or stopping criteria is met, the algorithm loops over the same steps in order to achieve better solutions at every iteration. Given the initial population, their fitness is evaluated and are sorted by performance. In case the stopping criteria is met, the loop stops and the best value among the found elite population is returned.
- Selection phase: an elitist selection method is applied. Various iterations of the algorithm suggested that the elite percentage of 10% would be convenient.
- Update phase: the level counter t and other parameters are prepared for the next iteration. The mean and covariance parameters are updated according to the elite population set and according to $\hat{v}_t = \alpha\hat{w}_t$. Here, α is the update speed coefficient, which is bounded between $\alpha \in [0.4, 0.9]$ and $\hat{w}_t$ represent the new mean and covariance parameters related to the elite population set.
- new population sampling: A new population is sampled from the updated multivariate normal distribution. This population will be concentrated more around the best solutions found on the search space in the iteration.

Note that except for the initial parameters of mean, covariance, and other required to run the algorithm, the rest are "self-tuning" as they modify and improve at every iteration.

Algorithm 3.1 Cross-Entropy Method

Require: $max\_iterations, population\_size, elite\_percentage, variable\_boundaries$

Initialize:
Define initial mean and covariance matrices: $\hat{v}_0 = u$. Set level counter $t = 1$.
Randomly sample an initial population from the density $f(\cdot; v_{t-1})$.
while stopping criteria not met **do**
    Evaluate the current population
    Sort the population by performance
    Select the elite subset:
    elite_count = elite_percentage × population_size
    elite_set = top elite_count individuals
    Estimate the distribution of the elite set:
    Update the mean and covariance parameters based on elite_set.
    Sample a new population from the estimated distribution:
    Generate new candidates using the estimated distribution
    Replace the current population with the new population
end while
Return Best individual/solution found

3.2. Improvements on the CE method

In the CE method, the update step is responsible for retaining a record of the best solutions found and for the refinement of the probability density function (pdf) used in the subsequent iteration. The objective function is also important as it impacts directly the efficacy and performance of the optimization algorithm. An improved version of the CE method is proposed that enhances both aspects in order to yield superior results.

3.2.1. Application on the water pump station

One primary goal of the CE method is to minimize the power consumption related to water pumps in pump stations. Therefore, it's fundamental to apply it correctly taking in consideration all the characteristics of the model, such as the search space and the constraints. The initial step in the CE method is to define the pdf that characterizes the set of solutions and from which the initial population is sampled. To accomplish this, a multivariate normal distribution is employed for sampling the initial population, which is set to contain 1000 elements, taking into account the trade-off between computational cost and execution time. As the CE method progresses through its iterative

updates, the multivariate normal distribution dynamically adjusts itself relative to the best solutions discovered through elitist selection. With each update, a new population is drawn from this adapted distribution. The primary objective here is to guide the set of solutions towards converging to the optimal point. This adaptive approach helps the CE method refine and improve its solution space, ultimately working to achieve more favorable outcomes.

We can define each population member with a two dimensional random vector $\boldsymbol{X} = (x_1, x_2)^T$ and sample the initial population $\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_N}$ from the multivariate normal distribution:

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.19}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \bar{s} \\ \bar{h} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \left(\frac{s_{max} - \bar{s}}{3}\right)^2 & 0 \\ 0 & \left(\frac{head_{max} - \bar{h}}{3}\right)^2 \end{bmatrix} \tag{3.20}$$

where $\bar{s}, \bar{h}$ are the mean values between the upper and lower bounds of the speed and head limitations, which are given by the constraints. This ensures the proper coverage of the search region.

$$\bar{s} = \frac{s_{max} - s_{min}}{2}, \qquad \bar{h} = \frac{h_{max} - h_{min}}{2} \tag{3.21}$$

Afterwards, the fitness of each individual is evaluated according to the cost function described in equation (2.12) and the individuals with the highest fitness are included in the elite population set. To account for the scheduling problem, which is to understand which pump combination results in the lowest power consumption while satisfying the requirements, the values of the status parameters ω_i are sampled from a binary uniform distribution.

The elite samples, which are composed by one tenth of the total population, are then used to compute the mean and covariance parameters implemented in the update step. In the classical CE method, the new multivariate normal distribution is equal to the one that best represents the elite population. The update step is computed by the expression:

$$\hat{v}_t = \alpha \hat{w}_t, \quad (0.4 \leq \alpha \leq 0.9) \tag{3.22}$$

Where w_t represents the mean and covariance matrices of the elite set and α is the learning rate coefficient. The updated multivariate normal distribution will be used in the next iteration to sample the new population.

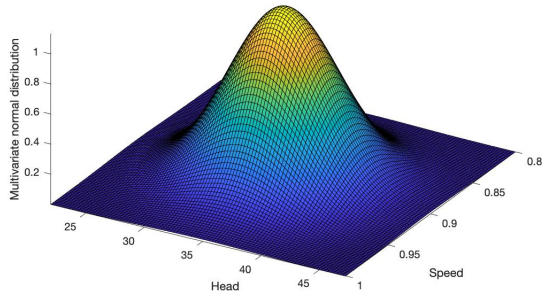
3.2.2. The asymmetric smoothed update

In the classic version of the CE method this step is straightforward. The parameter vector $\hat{v}_t$, which comprises the mean and covariance matrices, is updated at each iteration according to $\hat{v}_{t-1} = \hat{w}_t$, where t is the iteration and $\hat{w}_t$ is the new parameter vector that describes the mean and standard deviation of the elite samples. This update method shows to be inefficient and renders the algorithm vulnerable to local minima.

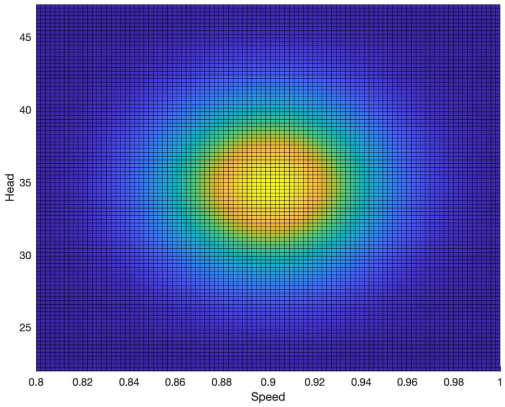
Decomposing the parameters vector into its its mean and covariance components and introducing a second parameter β, the previous issue is addressed by providing an asymmetric smoothed update.

$$\begin{aligned} \hat{v}_{\mu,t} &= \alpha\hat{w}_{\mu,t} + (1-\alpha)\hat{v}_{\mu,t-1} \quad (0.4 \leq \alpha \leq 0.9) \\ \hat{v}_{\sigma,t} &= \beta\hat{w}_{\sigma,t} + (1-\beta)\hat{v}_{\sigma,t-1} \quad (0.3 \leq \beta \leq 0.7) \end{aligned} \tag{3.23}$$

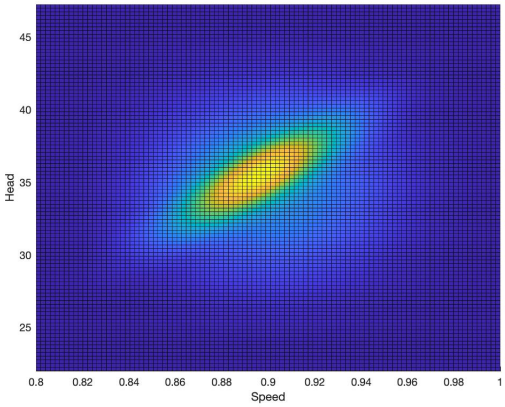
In optimization problems, the factor $(1-\alpha)$ is commonly employed in smoothed update steps. In this context, our focus extends beyond the introduction of a second parameter β, which enables an asymmetric smoothed update. The CE method samples from a Gaussian pdf, which, in 2 dimensions, forms a bell-shaped curve (Fig. 3.1 (a)). When using a single parameter α, in each update step, the bell-shaped curve 'moves towards the best solution' as the mean matrix updates and 'reduces in size' when the covariance matrix updates. Unfortunately, this simultaneous movement and scaling can lead the algorithm to converge into local minima, compromising accuracy. By applying two parameters instead, we enable the Gaussian pdf to 'move' and 'scale' in an asymmetric manner, as shown in Fig. 3.1 (b)-(d). This results in faster convergence, reduced susceptibility to local minima, and an improved ability to explore better solutions, particularly when specific directions offer a higher probability of yielding superior results.



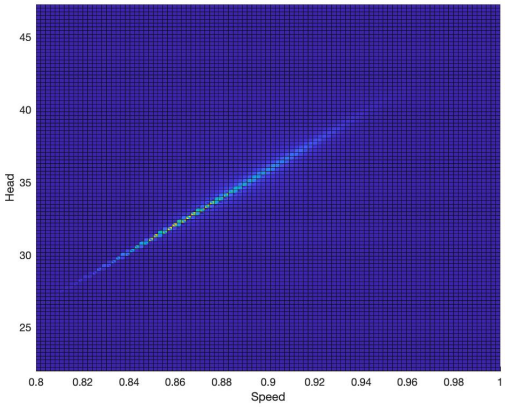
(a) Frontal representation.



(b) Iteration 1.



(c) Iteration 5.



(d) Iteration 15.

Figure 3.1: 'Moving' and 'resizing' in the update step.

3.2.3. The adaptive cost function

To achieve a faster and more accurate optimization a different parametrization has been applied to the expression (2.12). The constraints in the equations (2.13) are reformulated by writing flow values Q_i in terms of H_s and s_i.

$$Q_i = \sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} \tag{3.24}$$

In this way the head constraint, equation (2.5), is implemented directly in the model by setting for all pumps the system head value of H_s. In result, we obtain:

$$J = min\Bigg\{\sum_{i=1}^{m}\omega_i(s_i^3 p_1 + s_i^2 p_2\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} + s_i p_3\frac{h_1 s_i^2 - H_s}{h_3}) + \sum_{i=m+1}^{m+n}\omega_i(p_1 + p_2\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} + p_3\frac{h_1 s_i^2 - H_s}{h_3})\Bigg\} \quad (3.25)$$

The constraints and the choice of their mathematical representation can significantly impact the feasibility and quality of the solution. Therefore, two penalty functions, Pe_1 and Pe_2, that take into account the constraints are introduced in the model.

The penalty function Pe_1 represents the BEA constraint.

$$\begin{aligned} Pe_1 &= \sum_{i=1}^{n+m}(\Delta Q_i)^2 = \\ &= \begin{cases} \sum_{i=1}^{n+m}(\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} - Q_{min,i})^2, & \sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} < Q_{min,i} \\ 0, & Q_{min,i} < \sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} < Q_{max,i} \\ \sum_{i=1}^{n+m}(\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} - Q_{max,i})^2, & \sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} > Q_{max,i} \end{cases} \end{aligned} \quad (3.26)$$

The penalty function Pe_2 represents the flow balance constraint.

$$Pe_2 = (\sum_{i=1}^{n+m}\omega_i\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} - Q_s)^2 \quad (3.27)$$

Therefore, the problem's mathematical formulation is summarized by:

$$F = min\Bigg\{\Bigg[\sum_{i=1}^{m}\omega_i(s_i^3 p_1 + s_i^2 p_2\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} + s_i p_3\frac{h_1 s_i^2 - H_s}{h_3}) + \sum_{i=m+1}^{m+n}\omega_i(p_1 + p_2\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} + p_3\frac{h_1 s_i^2 - H_s}{h_3})\Bigg] + \sigma\Bigg[Pe_1 + (\sum_{i=1}^{n+m}\omega_i\sqrt{\frac{h_1 s_i^2 - H_s}{h_3}} - Q_s)^2\Bigg]\Bigg\} \quad (3.28)$$

Where σ is a weight coefficient that decreases with the iterations, giving less importance to the penalty functions as the solutions converge to the optima. It is characterized by

an initial value σ_0 and a cooling parameter γ.

$$\sigma = \sigma_0 \frac{1}{\gamma T}, \quad \gamma \in [0, 1] \tag{3.29}$$

3.3. Comparison with other metaheuristics

Comparative analysis in optimizing methods is crucial. It helps identify the best method for a specific problem, considering variations in structure, size, and constraints. By evaluating efficiency, solution quality, and reliability, it guides decision-making. In in this section two other metaheuristic algorithms, namely, the ant colony optimization and the genetic algorithm are presented with the aim of comparing thoroughly the improved CE method's performance.

3.3.1. The Ant Colony Optimization (ACO)

Ant System (AS) algorithms, classified as metaheuristic methods, draw their inspiration from the study "Self-organized shortcuts in the Argentine Ant" conducted by Goss et al. in 1989 [32]. This research delved into the foraging behavior of the Argentine ant, and its remarkable ability to optimize pathfinding through the use of trail pheromones. The ants' proficiency in identifying the shortest route between their nest and a food source was thoroughly examined and was used as an idea for the development of optimization algorithms.

Among AS algorithms, the Ant Colony Optimization (ACO) method received particular attention. Developed in the early 1990s by Marco Dorigo, ACO is an optimization algorithm that draws its inspiration from the foraging behavior of ants. It is a powerful and versatile method used to solve a wide range of combinatorial optimization problems. Its potential was analyzed in the double bridge experiments, which provided clear evidence of the inherent optimization capabilities of ant colonies. Given an arbitrary setting, ants were able to determine the shortest path between two points within their environment using probabilistic rules that rely on locally available information [33]. Notably, the study involved an environment featuring a bridge structure comprising two identical modules, as illustrated in Fig. 3.2.

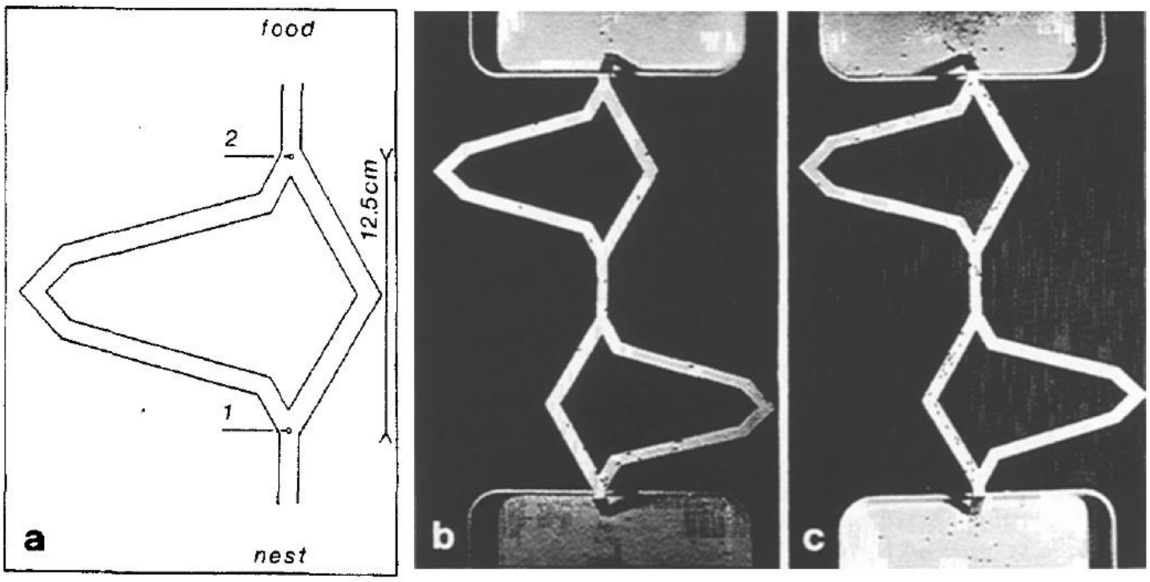


Figure 3.2: The double bridge experiment used to study the ants' behaviour.

Fig. 3.2 (a) shows a draft of the connection between the ant colony's nest and their food source. Image (b) is a snapshot taken four minutes after the bridge was placed. During this period, the ants' movement appears evenly distributed across both modules of the bridge. Image (c) is a photograph at the 8-minute mark following the bridge's placement. It reveals a change in behavior as the ants begin to preferentially use the shortest paths on both modules of the bridge.

The probability of ants choosing a shorter path increases as the difference in path lengths becomes more pronounced. This behavior is achieved through indirect communication within the ant colony. When argentine ants traverse the bridge, each ant deposits a chemical substance on the ground, which serves as a pheromone, triggering a social response in their fellow colony members.

Upon reaching a bifurcation point, which can be denotes with $(j = 1, j = 2)$, each ant makes a choice between the long and short paths based on a probability, p_j. This probability is directly linked to the amount of pheromone present on each module of the bridge. Since ants taking the shorter path reach the nest first, the short bridge accumulates a higher concentration of pheromone compared to the longer one. Consequently, the shorter path gains a higher probability of being selected by subsequent ants.

This natural behavior of the ants results in an autocatalytic process, where, as long as the food source remains available, an increasing number of future ants will opt for the paths with a higher probability, thus reinforcing the colony's efficient pathfinding behavior.

It is possible to quantify the probability p_j of choosing path $j = 1$ or $j = 2$ by setting a parameter, χ_s,that accounts for the amount of pheromone deposited on the short path and another parameter, χ_l for the longer path. Therefore, the probability that each ant chooses a certain path is:

$$p_1 = \frac{(k + \chi_s)^h}{(k + \chi_s)^h + (k + \chi_l)^h}$$
$$p_1 + p_2 = 1$$

Where the coefficients k, h are related to the particular system settings. In the case depicted in Fig. 3.2 the optimal value of $k = 20$ and $h = 2$ were found.

It is possible now to summarize the structure of the ACO algorithm into three key phases:

- Solution construction, where artificial ants iteratively construct solutions to the given problem, moving through adjacent states according to transition rules, which determine how ants make decisions about the next step to take.
- Pheromone updating, where pheromone trail reinforcement and pheromone trail evaporation occur.
- Deamon action, which is an optional step that involves applying global updates to the pheromone trail from a broader perspective, such as pheromone promotion and pheromone re-initiation.

The combination of local construction heuristics, pheromone communication, and optional global actions makes ACO a robust and adaptable optimization approach, applicable to a wide range of combinatorial problems.

The principles and mathematical formulation

The ACO algortihm searches iteratively for optimal or near-optimal solutions by simulating the foraging behavior of ants. The primary step of the ants before they deposit pheromone on the ground is to identify the shortest path connecting a pair of nodes within a graph. In this context, we define a construction graph as $G = (V, E)$, where V represents the set of vertices, and E denotes the set of edges. This graph, G, encompasses $n = |N|$ nodes. The optimization problem's solution is determined by the shortest path within this graph that can connect an initial node, i, to a final node, f. The artificial ants navigate from one vertex to another along the edges of the graph, incrementally constructing a partial solution.

For each arc (i, j) in the graph, a specific amount of pheromone is deposited and saved

by the variable $\tau_{i,j}$. The quantity of pheromone on a given arc is directly linked to the quality of the solutions found by the ants. Consequently, the presence and concentration of pheromone guides the ants towards regions in the search space where better solutions are likely to be discovered.

The solutions are constructed by the ants in a stepwise manner during each iteration of the algorithm. The decision regarding which node to move to is made probabilistically, adhering to a specific rule. For instance, when an ant k is located at node i and needs to determine the probability of choosing node $j = N_i$ as the next node to traverse, the ant considers the set of neighboring N_i associated with node i and takes into account the pheromone level $\tau_{i,j}$ found on the trail to calculate the probability.

$$p_{ij}^k = \begin{cases} \tau_{ij}, & \text{if } j \in N \\ 0, & \text{if } j \notin N \end{cases} \tag{3.30}$$

After several iterations, the pheromone levels deposited on the arc (i, j) get updated according to the following expression.

$$\tau_{ij}(t) = \tau_{ij}(t) + \Delta\tau \tag{3.31}$$

From equations (3.30), (3.31), it becomes evident that as the number of ants utilizing the connection between node i and node j increases, so does the likelihood that subsequent ants will opt for the same arc.

Nonetheless, updating pheromone levels by merely adding new amounts to arcs explored by new ants can lead to premature convergence. In such cases, local optimal solutions may mistakenly be regarded as the global optimal solution, which is contrary to the desired outcome of the model. Moreover, the behavior of real ants' pheromone does not mirror this simplistic approach. In reality, the pheromone deposited on a path naturally begins to evaporate over time. To counteract premature convergence and encourage exploration, in alignment with the natural behavior of pheromone and real ants, ACO incorporates a pheromone evaporation mechanism. This intentional reduction of pheromone intensity serves to promote the exploration of new arcs in the process. As a result, equation (3.31) is transformed into equation (3.32), facilitating a more dynamic and realistic representation of the pheromone update process in the ACO algorithm.

$$\tau_{ij}(t) = (1 - \rho) \times \tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}^k \tag{3.32}$$

In which ρ is the evaporation rate coefficient, with $\rho \in (0, 1]$, and m is the number of ants. $\Delta\tau_{ij}^{k}$ is the amount of pheromone deposited on the trail (i,j) by the ant k, according to:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k}, & \text{if ant k used arc (i,j)} \\ 0, & \text{otherwise} \end{cases} \tag{3.33}$$

Where Q is a constant and L_k is the length of the trail covered by the ant (k).

The probability of equation (3.30) can also be reformulated. Assuming that ants construct solutions and select nodes to visit through a stochastic mechanism, when ant k is situated at node i and has constructed the partial solution S^P, the probability of transitioning to node j is expressed by the equation below.

$$p_{ij}^{k} = \begin{cases} \frac{\tau_{ij}^{\alpha} \times \eta_{ij}^{\beta}}{\sum_{c_{il} \in N(S^P)} \tau_{il}^{\alpha} \times \eta_{il}^{\beta}}, & \text{if } c_{ij} \in N(S^P) \\ 0, & \text{otherwise} \end{cases} \tag{3.34}$$

Here, $N(S^P)$ represents the set of neighbors or feasible components of node i in the context of the partial solution S^P. When considering arcs (i,j), the node l is one that has not been visited by ant k. The parameters α and β play a significant role in controlling the balance between the influence of pheromone, represented by τ_{ij}, and heuristic information, represented by η_{ij} in (3.35), during the decision-making process.

$$\eta_{ij} = \frac{1}{d_{ij}} \tag{3.35}$$

where d_{ij} is the distance between i and j.

The ACO algorithm

This section presents the basic structure of the ACO and its pseudocode. At the start of the algorithm, initial pheromone levels are assigned to all edges in the problem graph. These pheromone levels represent the information that ants use to make decisions about their paths. Afterwards, the ants are positioned randomly in the search space and each ant will the problem graph to find a solution.

The basic ACO parameters to be defined are:

- α: pheromone influence, controls the impact of pheromones on ant decision-making.
- β: heuristic influence, determines the influence of heuristic information, such as the

distance, on ant decisions.

- Q: total pheromone deposit, represents the total amount of pheromone deposited by ants on their paths.
- ρ: evaporation rate, controls the rate at which pheromone evaporates over time.

The main loop of the algorithm contains the three key steps of ACO. The construction of solutions, in which the ants probabilistically select paths based on the combination of pheromone information and heuristic values. The probability of choosing a path is influenced by both the amount of pheromone on the edge and the heuristic information. The pheromone update step, in which the pheromone deposition is based on the quality of the solutions found by the ants. Better solutions result in higher pheromone deposition. The pheromone evaporation step, in which the global pheromone levels based on the solutions found by the entire ant colony are adjusted according to an evaporation coefficient. Finally, when the stopping or convergence criteria is met, the algorithm stops and return the optimal solution.

Algorithm 3.2 Ant Colony Optimization

Require: $population\_size, variable\_boundaries, termination\_criteria$

Initialize:
Initialize pheromone levels on all edges in the graph
Initialize ant population, best solution and ACO core parameters:
Set parameters: α, β, ρ, Q.
while termination criteria not met **do**
  for each ant in the population **do**
    Initialize ant's current position
    Initialize an empty solution for the ant
    while solution is not complete **do**
      Compute probabilities for unvisited neighbors using α and β
      Select the next node to visit based on the probabilities
      Add the selected node to the ant's solution
      Update pheromone level on the chosen edge
    end while
    Evaluate the fitness of the ant's solution
    if ant's solution is better than the best solution **then**
      Update the best solution
    end if
  end for
  Evaporate pheromone on all edges
  Deposit pheromone on edges of the best solution found
end while
Return Best individual/solution found

3.3.2. The Genetic Algorithm (GA)

Evolutionary Computation (EC) is an reknown branch in the field of metaheuristic optimization. It started to gain popularity in recent decades due to its versatility and its heuristic-driven problem-solving strategies, which showed to be effective in tackling complex optimization problems. Within the EC landscape, the genetic algorithm (GA) stood out for its effectiveness.

Developed by John Holland in the 1960s, the GA is a type of optimization and search algorithm inspired by the process of natural selection and genetics. It is used to find approximate solutions to optimization and search problems by mimicking the process

of evolution within a population of candidate solutions. It utilizes mechanisms such as selection, crossover, and mutation to iteratively refine solutions with the hope that over time, the population converges toward better results. GA is particularly useful in optimization problems where the search space is large, complex, or when gradient-based optimization techniques are not applicable. Furthermore, due to the fact that it allows for a family of solutions, it can obtain various optimal solutions or solutions close to the optimal case. GA only employs the information given by the objective function and thus avoids complications associated with the determination of derivatives and other auxiliary functions.

The algorithm generates a population composed of candidate solutions for the specified optimization problem by emulating the characteristics of gentic selection. Each individual in the population is marked by its own features, which results in being more or less prone to survive in the next generation. The individuals with the fittest or best chromosomes have an edge compared to other weaker individuals and so, the algorithm increases their chances of reproducing. As the chromosomes of the best individuals get passed to future generations, the population gets fitter. This procedure, observed from nature, is one of the key characteristics of GA, together with the reproduction, crossover and mutation steps. This algorithm offers the GA method a remarkable degree of versatility, making it adaptable to a diverse spectrum of optimization challenges.

The principles and mathematical formulation

The first step of the GA method is the generation phase, in this step the population is defined. The initial population in GA can be chosen using various techniques, but the most straightforward one is the through random generation, which is implemented in this work. It is known that characteristics related to the initial population affect the search capabilities of optimization algorithms. Fore example, if a large population is considered, a higher amounts of computations are needed and the algorithm might have a slow execution time. In case a small population is taken, it might not be enough to cover the search space effectively, with the risk of not finding the optima or falling into a local minima. Therefore, as summarized in Fig. 3.3, it's important to pay attention to the aspects that might influence the population initialization.

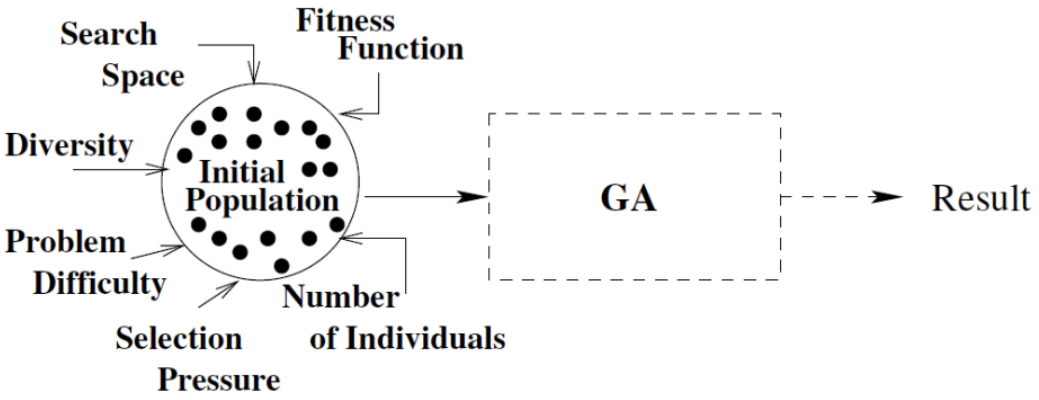


Figure 3.3: Influence of the problem's characteristics on of the initial population.

In accordance with the mathematical model of the problem, the pump station incorporates m variable speed pumps and n fixed speed pumps, hence, we can adapt the GA for this specific circumstance. During the generation phase, the information of each individual is represented by a binary string, containing the pumps' configuration, speed and system head value.

$$\omega_1\omega_2...\omega_{m+n}S_1S_2...S_{m+n}Hs \tag{3.36}$$

Where ω_i is the binary status of the pump i, which is 1 for 'On' and 0 for 'Off', S_i is the speed ratio of the pump i, and H_s is the system head value. In this case, the system head value constraint is implemented directly in the formula (3.36), instead of defining all the values $H_{S,1}, H_{S,2}, ..., H_{s,m+n}$ and then imposing the condition of equation (2.5). In order to cover effectively the search space, traced by the speed and head boundaries given by the system constraints, each value of the speed ratio is represented with 8 bits and the head value with 12 bits. Since the algorithm uses a binary encoding, an effective decoding method is important in order to smoothly proceed to the selection phase. For each of the variables S and H_s, the following decoding formula is applied.

$$\begin{aligned} S &= S_{min} + (\sum_{i=1}^{8} b_i \times 2^i) \times \frac{S_{max} - S_{min}}{2^8 - 1} \\ H_s &= H_{min} + (\sum_{k=1}^{12} b_k \times 2^k) \times \frac{H_{max} - H_{min}}{2^{12} - 1} \end{aligned} \tag{3.37}$$

In which the variable boundaries, $S \in [S_{min}, S_{max}]$ and $H_s \in [H_{min}, H_{max}]$ are indicated by the system constraints.

The next step is the selection phase, which chooses the fittest individuals and dictates a

string's eligibility for involvement in the reproductive phase. Occasionally referred to as the reproduction operator, this step significantly influences GA's convergence rate, which hinges on what's termed as selection pressure. Various established methods of selection include the roulette wheel, rank-based selection, tournament selection, Boltzmann selection, and stochastic universal sampling. In the roulette wheel selection, each potential string is represented on a wheel proportional to its fitness value. A random rotation of this wheel determines the selection of particular solutions contributing to the formation of the subsequent generation. The main drawbacks of this method are related to its stochastic nature, which may introduce errors and making the algorithm more sensible to local minima. These issues are avoided with the use of the tournament selection technique, which dates back to Brindle's proposal in 1983. This method involves the selection of individuals based on their fitness values using a stochastic roulette wheel mechanism applied to a pair of population members. Subsequently, individuals with superior fitness values are incorporated into the succeeding generation's pool. Within this selection approach, each individual is evaluated against all but one other individuals, should it advance to the ultimate population of solutions [34]. The advantages of the implementation of the tournament selection method is that is preserves the the genome diversity and does not require a preventive sorting step, which slows down the execution time.

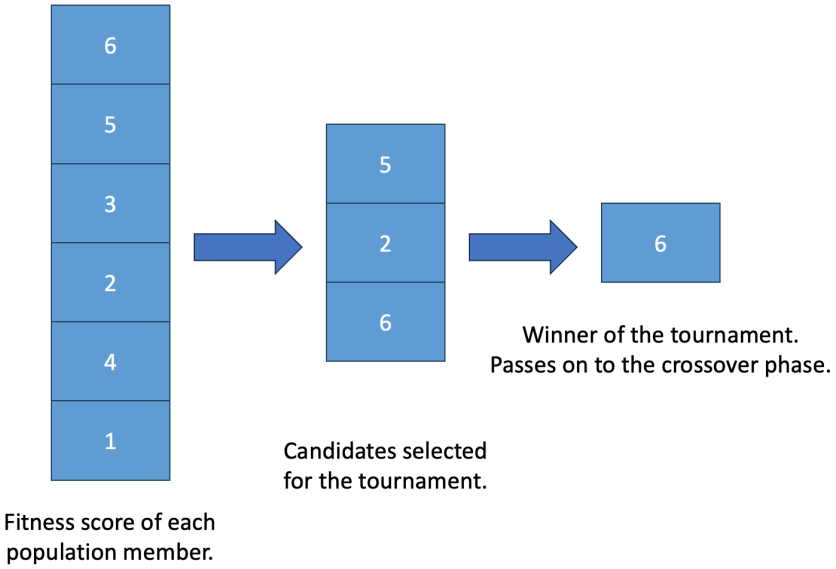


Figure 3.4: A 3-way tournament selection.

The crossover operation is a key step in the GA. It is used to generate the population of offspring by merging the genetic data of a pair or multiple parents. In other words, it mixes the genetic information of the best found solutions in the hope of providing a fitter population for the next iteration [35]. Among a variety of crossover methods, common

ones include the single-point, k-point, or the shuffle method. For a clearer understanding of the crossover operator, we may consider an example applied to the single water pump case, which utilizes the k-point crossover method. Given to candidate solutions for the operation of a water pump, the crossover operator swaps genetic information in

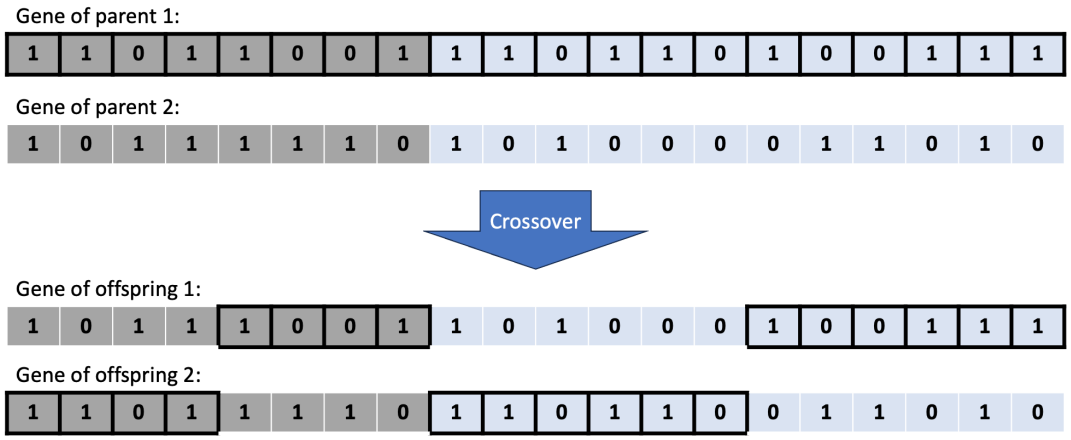


Figure 3.5: Crossover of the genome containing the operational parameters of a single water pump.

Mutation involves the random alteration of components within a solution, enhancing the population's diversity and offering a means to break free from a local optimal solution. Without it, the GA might converge prematurely to suboptimal solutions, as it would lack the exploration needed to discover potentially better solutions. The mutation rate, p_m, is a parameter that determines the probability of mutation occurring in each gene of an individual's chromosome. It's usually kept low to balance exploration and exploitation, since too high a mutation rate may lead to excessive randomness, hindering convergence towards good solutions, while too low a rate might limit exploration. In Fig. 3.6 is shown the effect of the simple inversion mutation method applied to a chromosome representing the operational parameters in the case of a single water pump.

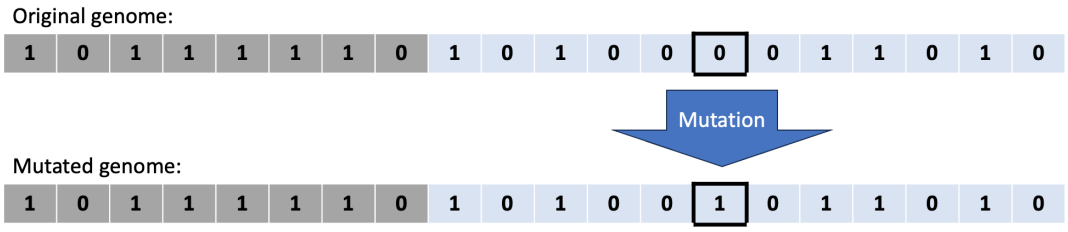


Figure 3.6: Mutation of the genome containing the operational parameters of a single water pump.

The GA algorithm

This section presents the basic structure of the GA and its pseudocode. The first step is the generation of the initial group of potential solutions, the starting population. Each solution is represented as a set of parameters or characteristics, often called chromosomes or genes, which contain the values of the operational parameters of the pumps in a binary format. The evaluation step assesses how well each solution performs regarding the problem's cost function and computes a fitness score or evaluation metric for every individual in the population based on how close its solution is to the desired outcome. After a termination criterion check, the algorithm proceeds to the three key operators of the GA, the selection step, the crossover step and the mutation step. In the first, individuals are selected to act as parents for the next generation based on their fitness scores. Higher fitness individuals are more likely to be selected, according to a tournament selection method. In the second, new solutions are created by combining genetic information from the selected parents. The crossover technique takes parts of two parents' genome and swaps or combines them to create the offspring, hoping to combine advantageous traits from both parents. In the third, small, random changes in the genetic material of the offspring are made. This random alteration is applied with a low probability to maintain genetic diversity within the population and explore new potential solutions. Finally, the performance of the newly created offspring solutions is assessed and compared with the parents. Based on the fitness scores, the most performing among the parents and offspring will populate the next generation of population. When the termination criteria is met, the loop breaks and the best found solution is returned.

Algorithm 3.3 Genetic Algorithm

Require: $population\_size, variable\_boundaries, termination\_criteria$

1: Initialize population
2: Evaluate fitness of each individual in the population
3: **while** termination criterion not met **do**
4: Tournament selection of parents based on fitness
5: Perform crossover to create offspring
6: Apply mutation to offspring with a certain probability, p_m
7: Evaluate fitness of new offspring
8: Select individuals for the next generation
9: **end while**
10: **Return** Best individual/solution found

3.4. Benchmarking and validation

Benchmark functions provide a common ground for comparing the performance of different optimization algorithms. Since these problems have known solutions, they are used to assess and validate the performance of optimization methods. The validation step is important in order to ensure generalizability and to allow the comparison of newly developed optimization techniques against established standards, ensuring the reliability and credibility of their results. In the present work, six benchmark functions are applied to analyze the improved CE method. Each of them presents specific characteristics that provide an unbiased evaluation of the algorithm.

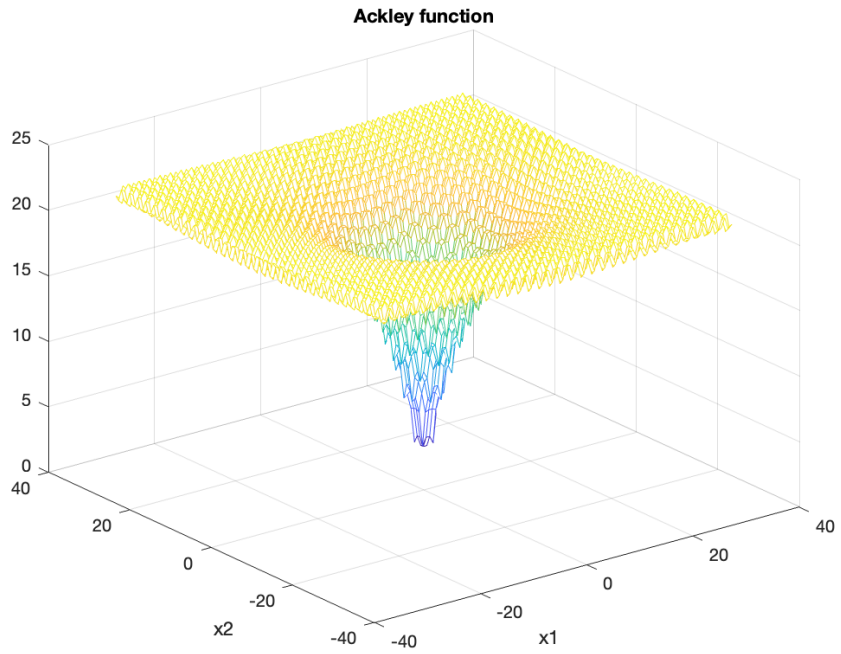


Figure 3.7: The Ackley function plot.

The Ackley function is a multimodal function with a global minimum at $f(0,0) = 0$, characterized by a nearly flat outer region and a large hole at the center.

$$f(\boldsymbol{x}) = -a \times \exp\left(-b \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(c \times x_i)\right) + a + exp(1) \quad (3.38)$$

Where $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ represents the n input variables, a, b, and c are constants typically set to $a = 20, b = 0.2, c = 2\pi$ in many implementations.

Global minimum	Optimal solution	Search space
$f(\boldsymbol{x}^*) = 0$	$\boldsymbol{x}^* = (0, ..., 0)$	$x_i \in [-32.768, 32.768]$

Table 3.1: The Ackley function characteristics.

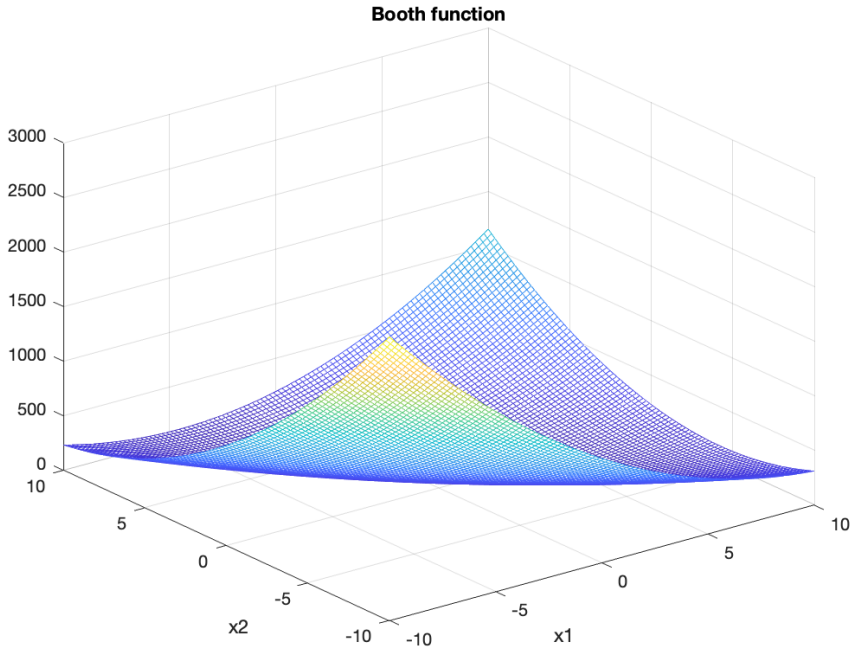


Figure 3.8: The Booth function plot.

The Booth function is a two-dimensional function with a single global minimum $f(1, 3) = 0$, characterized by a well-defined, valley-shaped search space.

$$f(x, y) = (x + 2y - 7)^2 + (2x + y - 5)^2 \tag{3.39}$$

Global minimum	Optimal solution	Search space
$f(x^*) = 0$	$x^* = (1, 3)$	$x_i \in [-10, 10]$

Table 3.2: The Booth function characteristics.

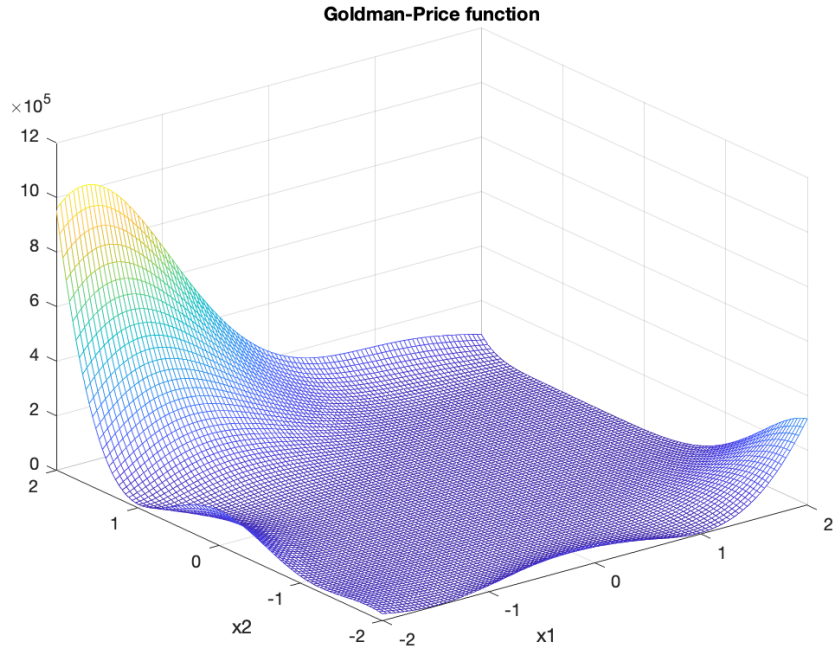


Figure 3.9: The Goldstein-Price function plot.

The Goldstein-Price function is another two-dimensional function with a single global minimum $f(0,-1)=3$.

$$f(x,y)=\left[1+(x+y+1)^2\times\left(19-14x+3x^2-14y+6xy+3y^2\right)\right]\times \left[30+(2x-3y)^2\times\left(18-32x+12x^2+48y-36xy+27y^2\right)\right] \quad (3.40)$$

Global minimum	Optimal solution	Search space
$f(x^*)=3$	$x^*=(0,-1)$	$x_i\in[-2,2]$

Table 3.3: The Goldstein-Price function characteristics.

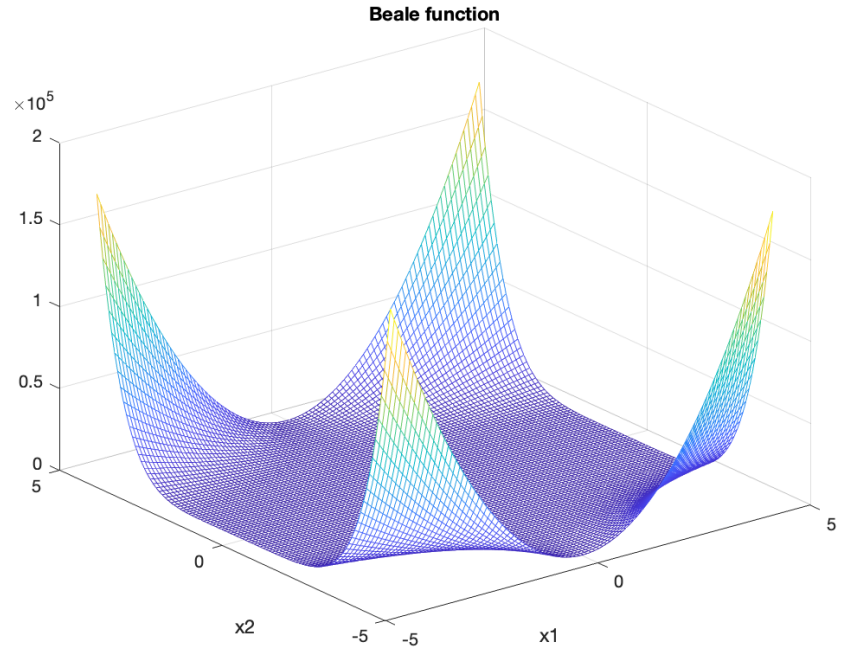


Figure 3.10: The Beale function plot.

The Beale function is a two-dimensional function with a single global minimum $f(3, 0.5) = 0$, characterized by a well-defined, valley-shaped search space.

$$f(x, y) = (1.5 - x + xy)^2 + (2.25 - x + xy^2)^2 + (2.625 - x + xy^3)^2 \tag{3.41}$$

Global minimum	Optimal solution	Search space
$f(x^*) = 0$	$x^* = (3, 0.5)$	$x_i \in [-4.5, 4.5]$

Table 3.4: The Beale function characteristics.

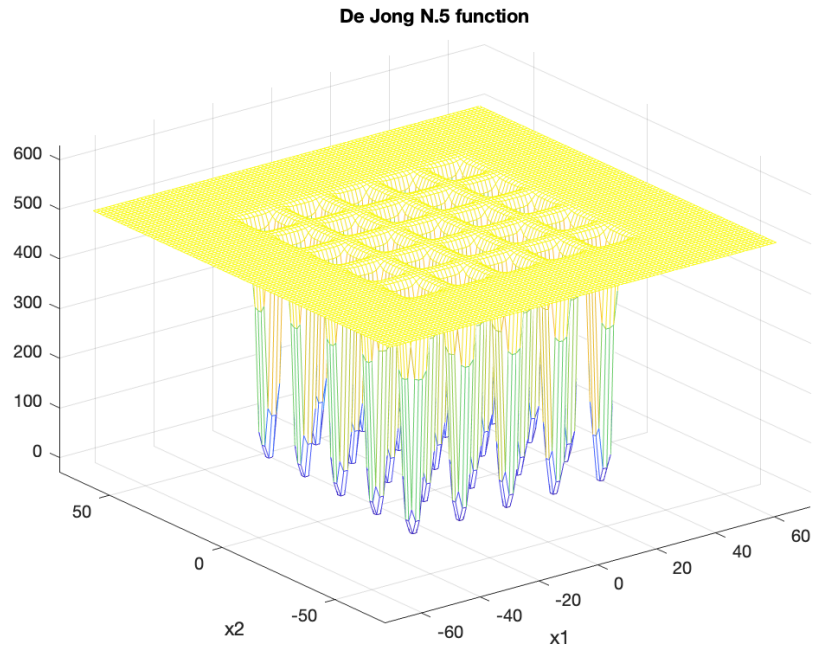


Figure 3.11: The De Jong function N.5 plot.

The De Jong function N.5 is a multimodal function also known as a particular version of the Shekel's Foxhole function. The function has multiple local minima located in a narrow, deep structure, making it challenging for optimization algorithms to find the global minimum.

$$f(\boldsymbol{x}) = \left(0.002 + \sum_{i=1}^{25} \frac{1}{i + (x_1 - a_{1i})^6 + (x_2 - a_{2i})^6}\right)^{-1} \tag{3.42}$$

where,

$$a = \begin{pmatrix} -32 & -16 & 0 & 16 & 32 & -32 & -16 & \dots & 0 & 16 & 32 \\ -32 & -32 & -32 & -32 & -32 & -16 & \dots & 32 & 32 & 32 \end{pmatrix} \tag{3.43}$$

Global minimum	Optimal solution	Search space
25 local minima	25 optim. sol.	$x_i \in [-65.536, 65.536]$

Table 3.5: The De Jong function N.5 characteristics.

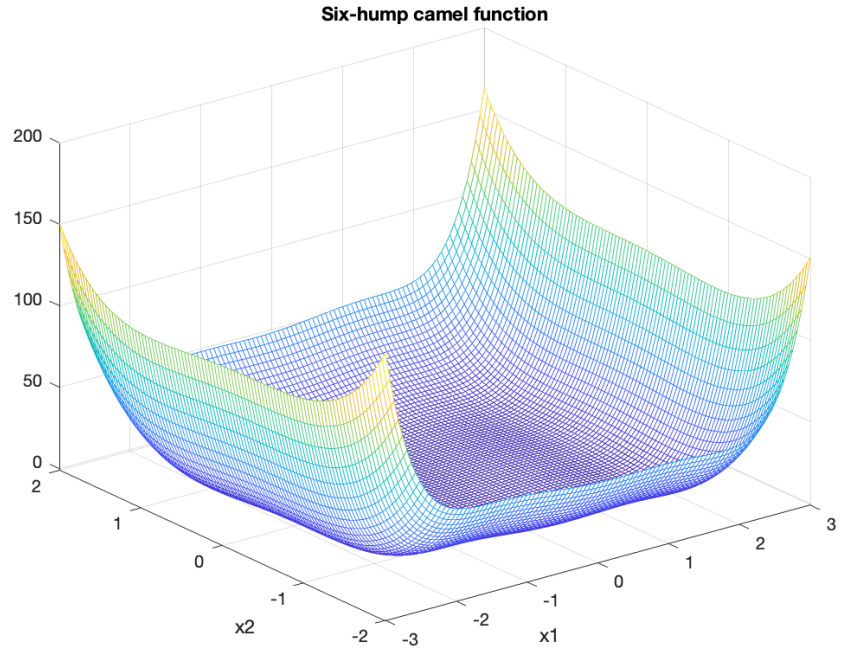


Figure 3.12: The six-hump camel function plot.

The Six-Hump Camelback function is a two-dimensional function with multiple global and local minima and maxima, making it a challenging function to optimize. The function has six prominent peaks and is typically defined in the domain $(-3 \leq x \leq 3), (-2 \leq y \leq 2)$.

$$f(x,y) = \left(4 - 2.1x^2 + \frac{x^4}{3}\right) x^2 + xy + \left(-4 + 4y^2\right) y^2 \tag{3.44}$$

Global minimum	Optimal solution	Search space
$f(x^*) = -1.0316$	$x^* = \pm(0.0898, -0.7126)$	$x_1 \in [-3,3], x_2 \in [-2,2]$

Table 3.6: The Six-Hump Camelback function characteristics.

The improved cross-entropy method has been applied to the six benchmark functions and it returned favorable outcomes, successfully converging to the global minimum and achieving the optimal solution for each function. The results are reported in the tables below. The results given by the ACO and GA are also provided for a comparative analysis.

Improved cross-entropy Method

Function	Global minimum	Optimal solution	Search space
Ackley	0	$(0, 0)$	$[-32.768, 32.768]$
Booth	0	$(0.999, 3)$	$[-10, 10]$
Goldstein-Price	3	$(0, -0.990)$	$[-2, 2]$
Beale	0	$(2.999, 0.499)$	$[-4.5, 4.5]$
De Jong 5th	1.990	$(-15.900, -31.600)$	$[-65.536, 65.536]$
Six-Hump Camelback	-1.0315	$(-0.0895, 0.7124)$	$x_1 \in [-3, 3], x_2 \in [-2, 2]$

Table 3.7: Benchmark performance of the improved CE method.

Ant Colony Optimization

Function	Global minimum	Optimal solution	Search space
Ackley	0.180	$(-0.032, 0.032)$	$[-32.768, 32.768]$
Booth	0	$(1.006, 3.001)$	$[-10, 10]$
Goldstein-Price	3	$(0, -0.990)$	$[-2, 2]$
Beale	0	$(3.020, 0.500)$	$[-4.5, 4.5]$
De Jong 5th	0.998	$(-31.960, -31.960)$	$[-65.536, 65.536]$
Six-Hump Camelback	-1.030	$(0.1026, -0.7059)$	$x_1 \in [-3, 3], x_2 \in [-2, 2]$

Table 3.8: Benchmark performance of the ACO method.

Genetic Algorithm

Function	Global minimum	Optimal solution	Search space
Ackley	0.180	$(0.032, 0.016)$	$[-32.768, 32.768]$
Booth	0	$(1.006, 3)$	$[-10, 10]$
Goldstein-Price	3	$(0, -1)$	$[-2, 2]$
Beale	0	$(2.986, 0.497)$	$[-4.5, 4.5]$
De Jong 5th	1	$(-31.900, -31.900)$	$[-65.536, 65.536]$
Six-Hump Camelback	-1.031	$(0.0967, -0.7135)$	$x_1 \in [-3, 3], x_2 \in [-2, 2]$

Table 3.9: Benchmark performance of the GA method.

4 | Experimental implementation on a pump station

The model has been adapted to optimize a real-life-inspired scenario featuring the water pump station of a megalopolis equipped with four pumps, including two variable frequency pumps, referred to as pump '1' and '2', and two fixed frequency pumps, referred to as pump 'A' and 'B'.

The real performance curves of each pump have been incorporated. It is a common issued in pump stations not to have the actual performance curves of each pump, in many cases, caused by the fact that once the pumps have been put in operation their charts are no longer taken in consideration.

A preliminary data collection step was needed to gather the values of head, power and efficiency at different flow values. Afterwards, through polynomial interpolation, the real performance curves were obtained reflecting the functions expressed in equations (1.9).

Two types of pumps were analyzed. The variable frequency pumps showed to be larger in size, capable of a higher flow output, a higher power consumption and could reach a higher efficiency with respect to the other. The other type represented the fixed frequency water pumps, which were smaller in size and their speed could not be modified once activated. The characteristic functions of the two types of pumps are reported below.

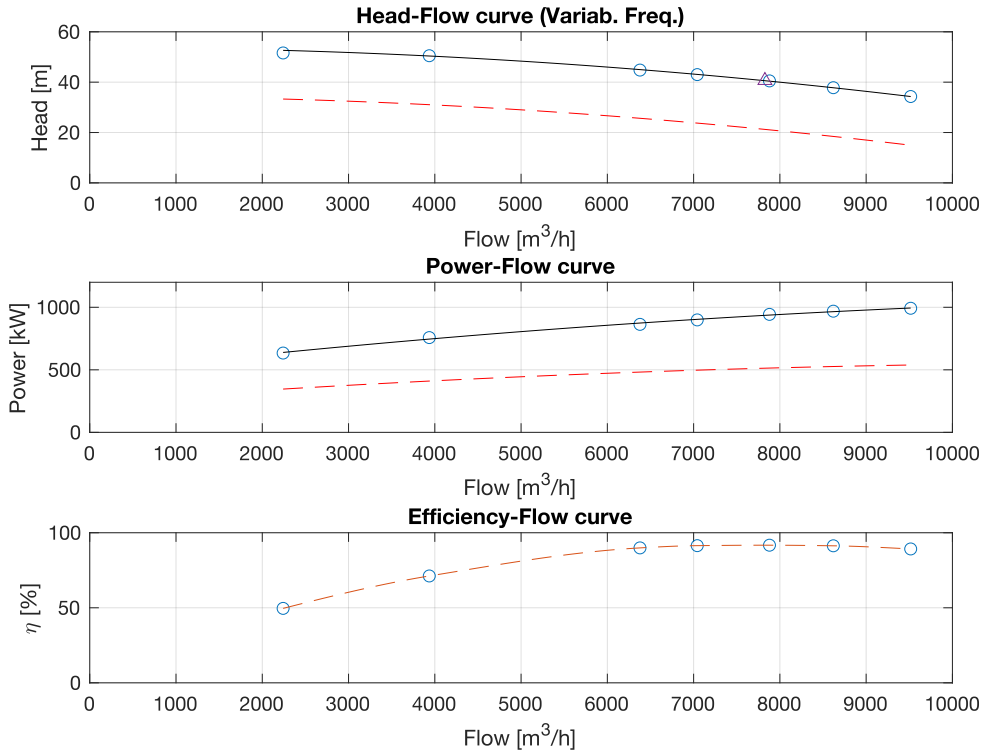


Figure 4.1: Performance curves of the variable frequency pump.

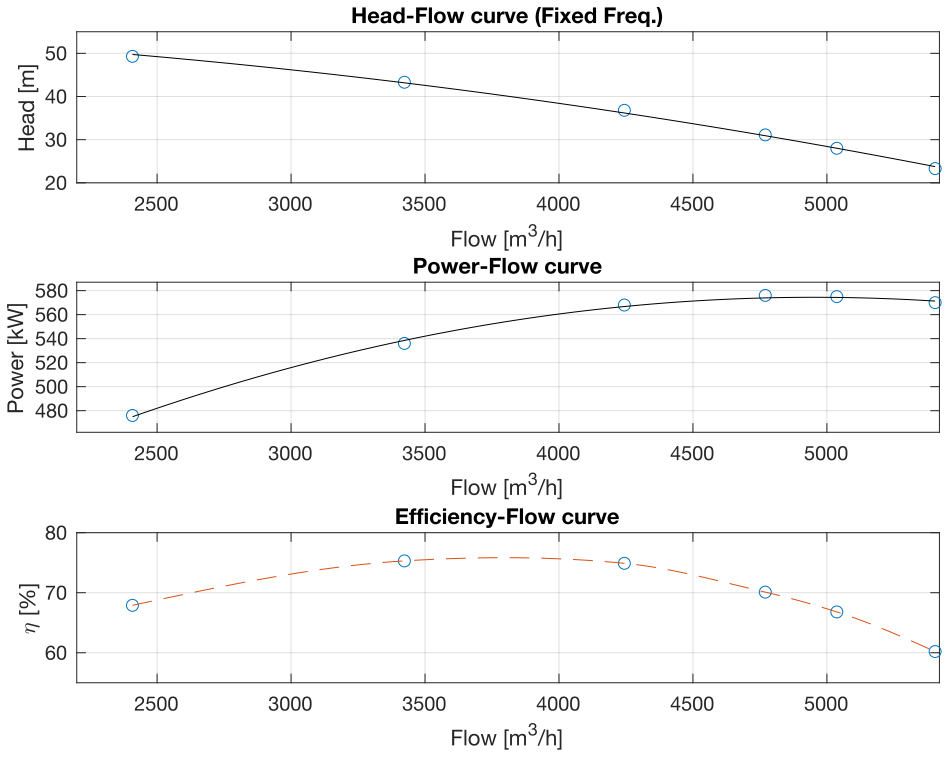


Figure 4.2: Performance curves of the fixed frequency pump.

Fig. 4.1 depicts the variable frequency pump curves showcasing two distinct speed ratios: one at $S = 1$ (shown in black) and another at $S = 0.8$ (highlighted in red). The plotted data points, represented by small circles, were effectively interpolated to derive the head, power, and efficiency curves. Notably, the fitted curves demonstrated a strong alignment with the data points. For instance, the head-flow curve exhibited excellent goodness of fit with an $RMSE$ (Root Mean Square Error) of 0.2432 and a $R - square$ value of 0.9988, indicating the high quality of the fit. As the speed varies, the curve smoothly transitions from the black to the red curve. This dynamic movement through speed control enables attainment of any value between these two curves. In contrast, Fig. 4.2 showcases the fixed frequency pump curves obtained through a similar methodology.

The water pump station's operation across multiple months forms the foundation of the database in use. This comprehensive database catalogs system flow demand values at five-minute intervals. Notably, the flow demand pattern displayed consistent daily cyclicality, repeating itself day after day. Given this repetitive nature, it proves both interesting and valuable for this project to analyze and consider the daily water flow demand chart.

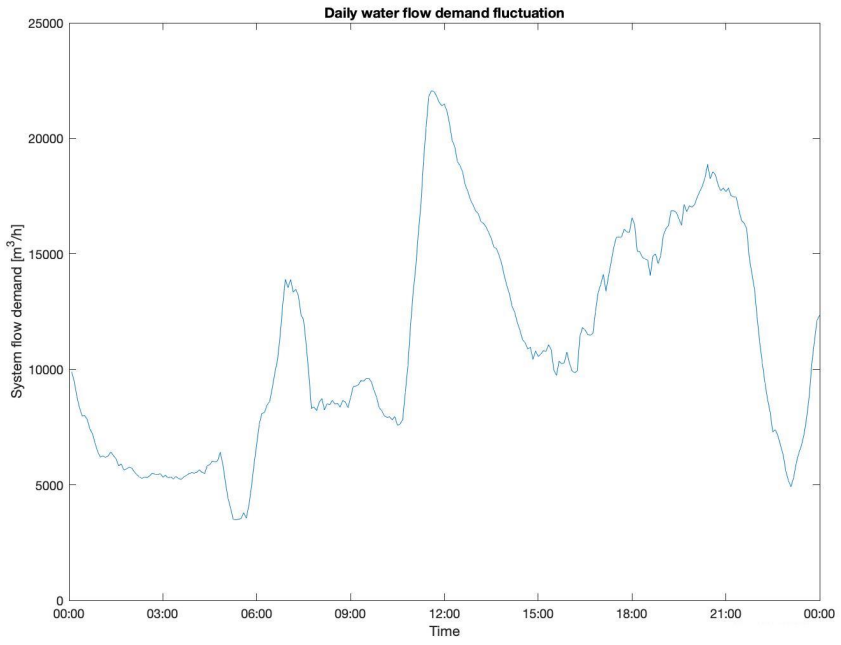


Figure 4.3: Water flow demand fluctuations in a day.

Fig. 4.3 reveals a discernible pattern recurring consistently each day. The peak demand for water aligns roughly around 13:00 and 18:00, corresponding to meal times when the population's water usage surges. Additionally, peaks at 8:00 and 21:00 coincide with

periods when people typically use water for personal hygiene and household appliances. In contrast, the valley points consistently occur during nighttime, likely corresponding to sleeping hours, or during typical working hours when water consumption tends to decrease.

Through the implementation of the improved CE method, optimal operational points for the pump set have been successfully established, prioritizing adherence to all constraints, notably ensuring operation within the BEA as illustrated in Fig. 2.2, while meeting the flow demand requirements. Additionally, the method has provided insights into the pumps' statuses (on-off), indicating the optimal combination of pumps that should be operated to minimize the costs of the pump station.

The subsequent sections detail the outcomes related to the improved CE method's performance, the resolution of the pump scheduling problem, and the comprehensive enhancements achieved within the water pump station.

4.1. Comparative analysis of performance

The improved CE method has introduced enhancements that improve best existing solutions achieved by other metaheuristic methods, including the GA and ACO. In this section, its performance is thoroughly analyzed from various perspectives to highlight its contributions and improvements.

The core part of the objective function is the quantity describing the power consumption of each pump. Therefore, it is meaningful to understand the effectiveness of the improved CE method on minimizing the power function of the single pump, since the overall optimization is directly related to how well the algorithm optimizes the single variable frequency pumps. Therefore, the improved CE method is applied to find the best operational values of speed and head that minimize power, while providing the required amount of output flow. Since in general the flow varies in time, two key flow demands are considered, the first at $6000m^3/h$ and the second at $8500m^3/h$, which show the effectiveness of the algorithm in different conditions. The improved CEM is compared with its classic version, with the ACO and with the GA. The resulting graphs are provided and the numerical values are organized in tables for a clear understanding. Due to the fact that the graphs for the two flow demand conditions are similar from a performance point of view, just the ones related to the $6000m^3/h$ are reported.

The central focus of the objective function lies in quantifying the power consumption of each pump. Consequently, it holds significance to evaluate the efficacy of the improved

CE method in minimizing the power function of individual pumps, as this directly influences the overall optimization process. This section employs the improved CE method to identify the optimal operational values of speed and head, aiming to minimize power while meeting the required output flow. Given the time-varying nature of the flow, we assess two key flow demands: $6000m^3/h$ and $8500m^3/h$, showcasing the algorithm's performance under diverse conditions. Comparisons between the improved CEM, its classic counterpart, ACO, and GA are illustrated through graphs, and accompanying numerical values are organized in tables for clarity and comprehension. As the performance graphs for both flow demand conditions exhibit similar trends, only those corresponding to $6000m^3/h$ are detailed herein.

4.1.1. Comparison with the classic CE method

The improved CE method (ICEM) performs better compared to the classic CE method (CEM), as reported in Fig. 4.4. The improved CE method is able to converge to the global minima in power consumption, achieving a remarkable power reduction of approximately 11% in the $6000m^3/h$ case, where the classical CE method converged to a local minima with a higher power value equal to $526.77kW$. Furthermore, it demonstrated superior effectiveness in refining its search for the optimal working point, resulting in a very accurate flow value. Even though both results are satisfactory, it's meaningful to note that the increased accuracy of the improved CE method in both cases of $6000m^3/h$ and $8500m^3/h$ to reach the required flow demand indicate that in real case scenarios the improved version might be more robust to noise or sudden variations.

The improved CE method performs better compared to the classic CE method, as evident in Fig. 4.4. Notably, the improved CE method exhibits the capability to converge to the global minimum in power consumption, achieving a substantial power reduction of approximately 11% in the $6000m^3/h$ scenario. In contrast, the classical CE method converged to a local minimum with a higher power value of $526.77kW$. Moreover, the improved method showcases superior efficacy in refining its search for the optimal operating point, resulting in a highly accurate flow value. While both methods yield satisfactory outcomes, it's worth highlighting that the improved CE method's increased accuracy in meeting the required flow demand for both $6000m^3/h$ and $8500m^3/h$ cases suggests its potential robustness in real-world scenarios, particularly when faced with noise or sudden variations.

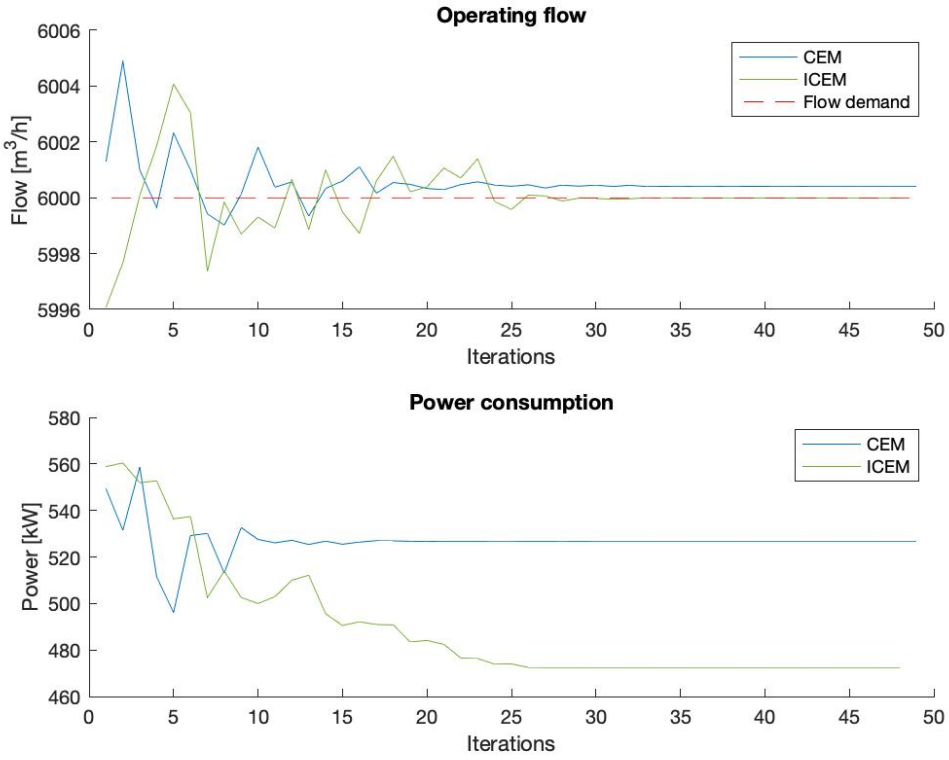


Figure 4.4: Comparison of the improved CE method in a single pump case.

Table 4.1: Comparative analysis with classic CEM.

	Classic CEM		Improved CEM	
Flow demand (m^3/h)	6000	8500	6000	8500
Computed flow (m^3/h)	6004.0	8499.87	6000	8499.88
Power (kW)	526.77	708.42	472.43	708.42
Iterations	30	12	32	22

The disparities in convergence are a direct consequence of the underlying logic of the improved CE method. By smoothing the update of the multivariate normal distribution, the algorithm doesn't converge directly toward the best solution but has a margin to explore possibilities in the vicinity of the current optimal point. Consequently, it exhibits improvements in its ability to search for global minima and accuracy.

4.1.2. Comparison with ACO

The comparison against ACO presents promising findings. As depicted in Fig. 4.5, the improved CE method exhibits faster convergence and greater accuracy in both flow and power metrics. Notably, its superior effectiveness in setting a more precise operational flow stems from the heightened efficiency of its search algorithm compared to ACO. Specifically, in the $8500m^3/h$ scenario, the improved CE method identifies a better minimum, showcasing the CE method's ability to explore potential solutions around the current best-found solution. This emphasizes its capacity for exhaustive evaluation and optimization.

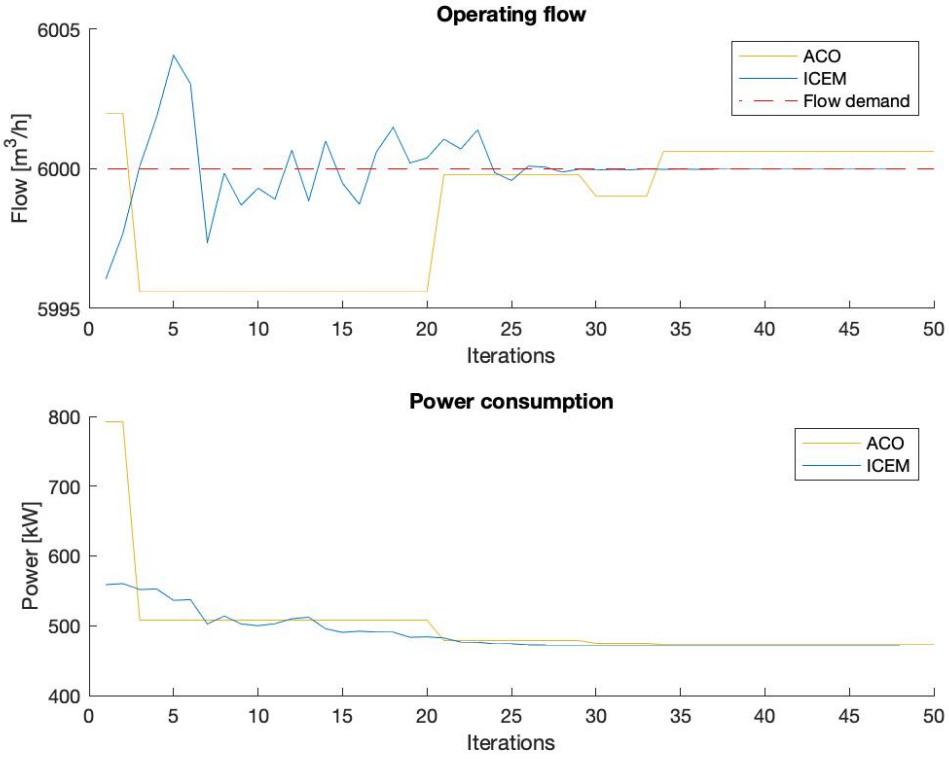


Figure 4.5: Comparison of the improved CE method with ACO in a single pump case.

Table 4.2: Comparative analysis with ACO.

	Improved CEM		ACO	
Flow demand (m^3/h)	6000	8500	6000	8500
Computed flow (m^3/h)	6000	8499.88	6000	8500.6
Power (kW)	472.43	708.42	472.43	727.78
Iterations	32	22	71	32

In comparison to ACO, the improved CE method showcased a slight but noticeable advantage in terms of convergence speed and accuracy. Specifically, the improved CE method achieved convergence in 32 iterations, whereas ACO required 71 iterations to reach the specified flow demand of $6000m^3/h$. Similarly, in the alternative scenario, it required 31% fewer iterations to achieve convergence. This efficiency translated into a substantial reduction of $480kW$ in power consumption during the transition period, a trend similarly observed in the $8500m^3/h$ case. This highlights the improved method's efficacy in achieving faster convergence and minimizing power consumption during transitional phases, illustrating its superior performance compared to ACO.

4.1.3. Comparison with GA

The graphs in Fig. 4.6 illustrate the analysis conducted against GA. A notable observation is the improved CE method's swifter convergence, particularly evident in the power consumption graph. Remarkably, it achieves convergence in roughly two-fifths fewer iterations compared to GA while reaching the necessary operating flow condition. This accelerated convergence results in a quicker transition period, ultimately reducing power costs by 2.2%. This highlights the improved CE method's efficiency in achieving faster convergence and consequent cost reduction compared to GA.

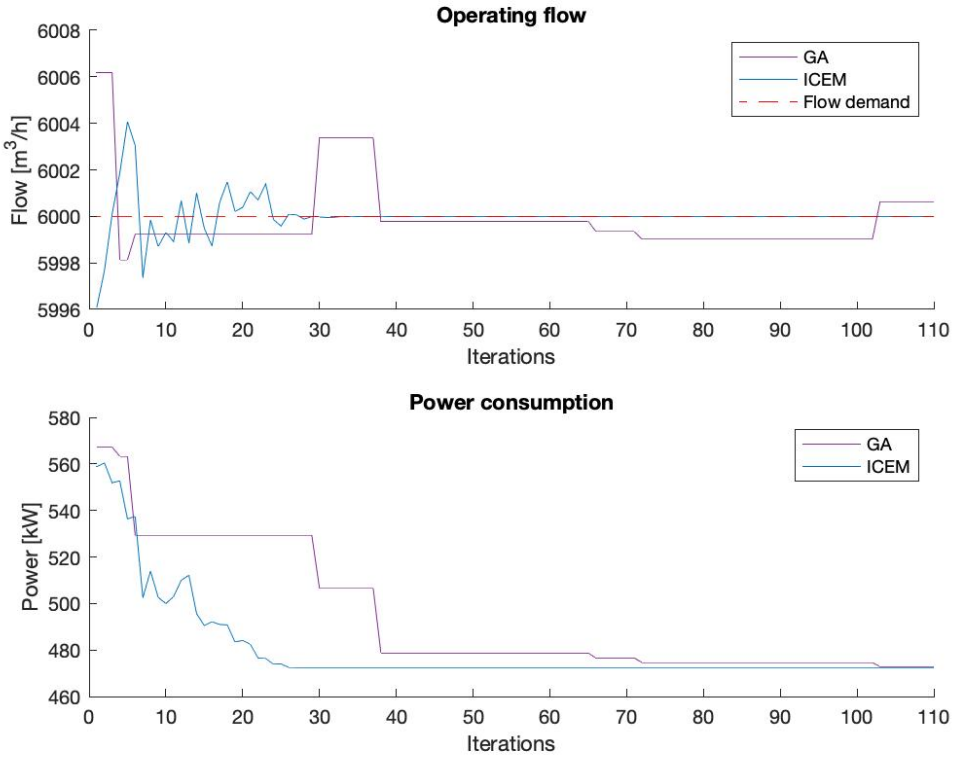


Figure 4.6: Comparison of the improved CE method with GA in a single pump case.

Table 4.3: Comparative analysis with GA

	Improved CEM		GA	
Flow demand (m^3/h)	6000	8500	6000	8500
Computed flow (m^3/h)	6000	8499.88	6000.63	8500.66
Power (kW)	472.43	708.42	472.84	709.16
Iterations	32	22	77	47

Indeed, it is interesting to observe the distinctive graphical representations between the GA and CE method. The GA tends to showcase an iterative refinement process resembling a ladder, whereas the CE method's graph appears notably smoother. This difference in graphical patterns could have practical implications. The smoother trajectory of the CE method potentially enhances its robustness in real-world scenarios by exhibiting resilience against biases or outliers in the data. This smoother nature might offer a more stable and reliable optimization process, proving advantageous when dealing with diverse or less predictable datasets.

4.2. Optimized water pump scheduling

The incorporation of the enhanced CE method leads to the effective operation of the pumps within their optimal efficiency range. This accomplishment is achieved while adhering to the water system's constraints and fulfilling the necessary head and flow requirements.

Since the system flow demand is constantly varying, three key flow demand values have been chosen to describe the workings of the pump scheduling system: $11000m^3/h$, $17000m^3/h$ and $25000m^3/h$. For each case its respective performance values are detailed in the associated tables, which report the actual values of the pumps' parameters present in the pump station. Among the various indexes, the key ones include the pump status, the rotational speed ratio, and the power consumption. The optimal operating conditions for the variable-speed pumps are shown in Figs. 4.7, 4.8, 4.9 and are pinned with a star symbol. The graphs are related only to the variable frequency pumps since the fixed speed pumps, once activated, will operate at their maximum efficiency, providing constant output flow and head values.

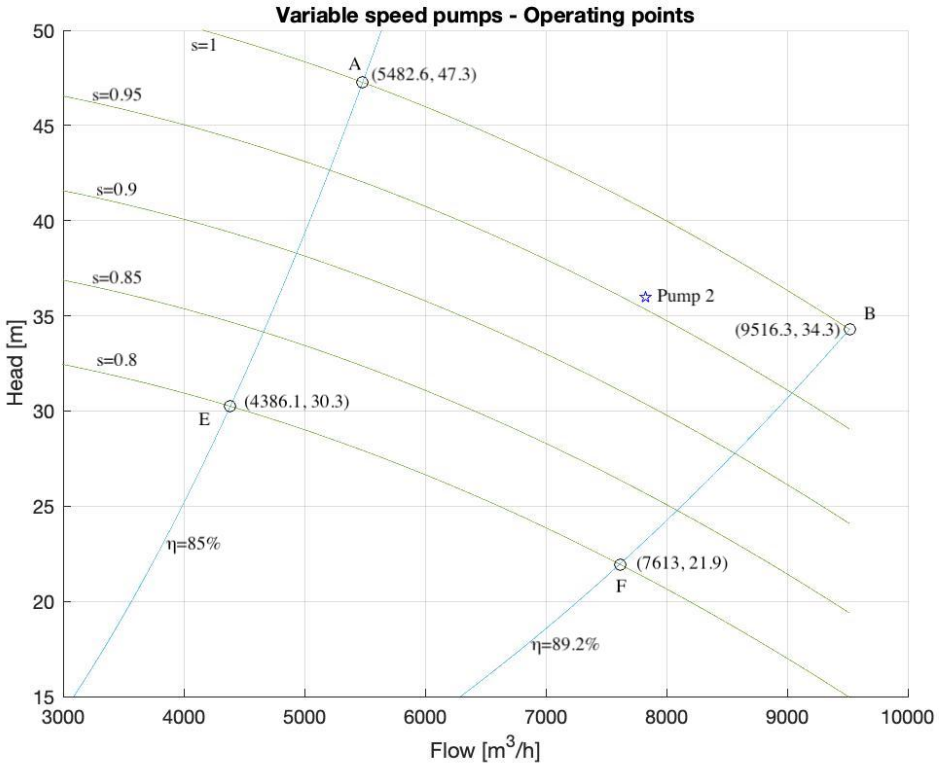


Figure 4.7: Working points when $Q_s = 11000m^3/h$.

Table 4.4: Pump station parameters

	System flow demand = 11000 (m^3/h)			
	Pump A	**Pump B**	**Pump 1**	**Pump 2**
Status	On	Off	Off	On
Tot. output flow (m^3/h)	11005.0			
Power (kW)	975.78			
Head (m)	40.1			
Pump flow (m^3/h)	3128.20	0	0	7876.86
Speed ratio	0.90	0	0	0.99
Efficiency (%)	75.83	-	-	91.70

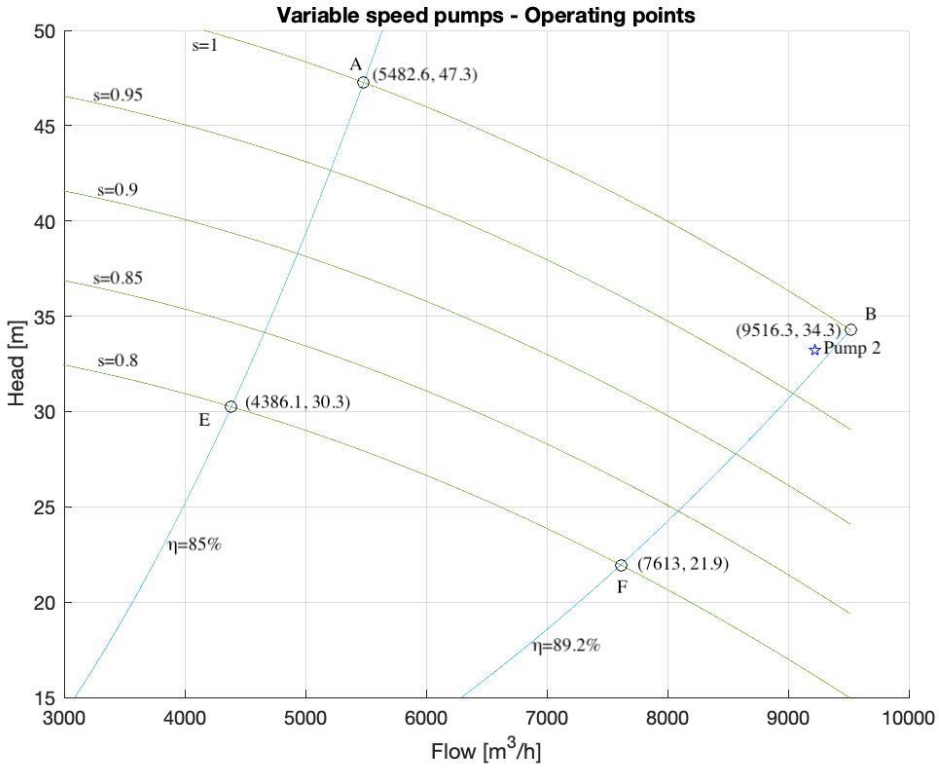


Figure 4.8: Working points when $Q_s = 17000m^3/h$.

Table 4.5: Pump station parameters

	System flow demand = 17000 (m^3/h)			
	Pump A	**Pump B**	**Pump 1**	**Pump 2**
Status	On	On	Off	On
Tot. output flow (m^3/h)	17052.6			
Power (kW)	1283.07			
Head (m)	33.2			
Pump flow (m^3/h)	4027.88	4100.57	0	8924.20
Speed ratio	0.91	0.92	0	0.93
Efficiency (%)	75.83	75.83	-	90.07

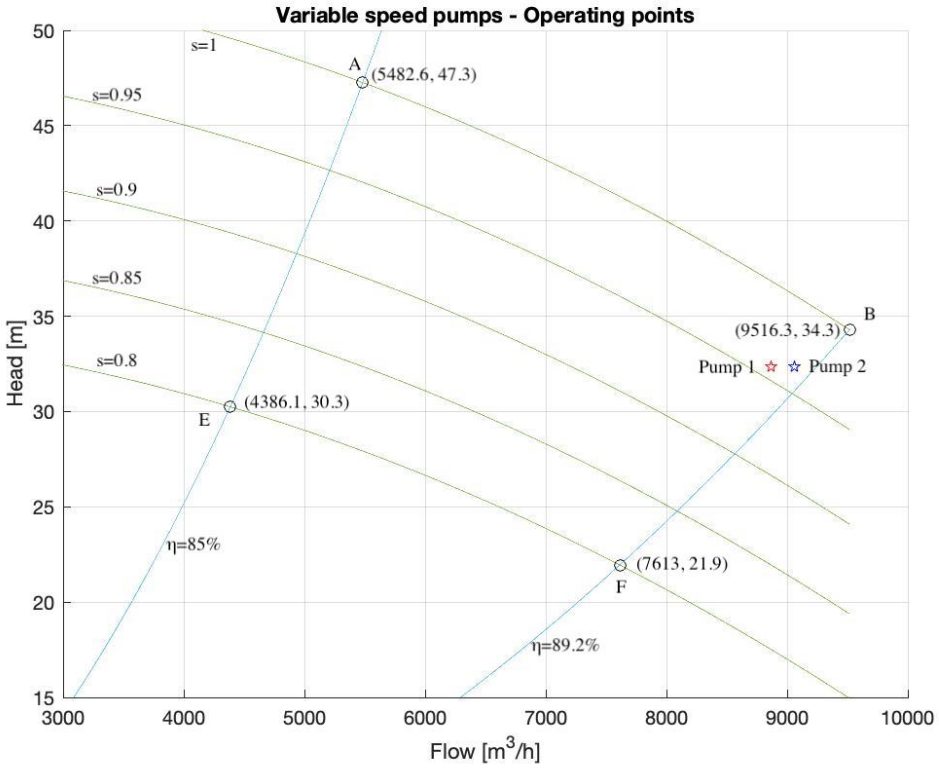


Figure 4.9: Working points when $Q_s = 25000m^3/h$.

Table 4.6: Pump station parameters

	System flow demand = 25000 (m^3/h)			
	Pump A	**Pump B**	**Pump 1**	**Pump 2**
Status	On	On	On	On
Tot. output flow (m^3/h)	24969.7			
Power (kW)	2394.05			
Head (m)	32.33			
Pump flow (m^3/h)	4196.71	3902.94	8299.98	8570.11
Speed ratio	0.92	0.87	0.87	0.89
Efficiency (%)	75.83	75.83	90.95	90.51

Figs. 4.7 to 4.9 delineate the BEA marked by points A, B, E, and F, signifying the intersections between pump performance curves at various speeds and boundary curves ensuring optimal operational efficiency. These results are notably satisfactory within this context. The variable-speed pumps consistently exhibit efficiency surpassing 85%, while adhering to the head constraint.

Fig. 4.9 illustrates how pump 1 and pump 2 operate simultaneously while maintaining the equal system head value constraint. The scheduling problem, tackled using the improved CE method, effectively recommends activating the minimum number of pumps necessary to meet flow demand and minimize power consumption.

The accompanying tables reveal that as the output flow demands increase, the power consumption of the pump station also rises. However, the results ensure the concurrent minimization of maintenance costs. Furthermore, the data suggests that, for a fixed flow value and pump sizes, activating a fixed frequency pump consumes less power than its variable-speed counterpart. In the particular case reported in Table 4.5, the fixed frequency pump B, gets activated before the variable speed pump 1, because the variable frequency ones are smaller, hence consume less power.

This implementation of the resultant scheduling system effectively slashes operational and maintenance costs for the water pump station, all while meticulously adhering to system constraints and demands.

5 | Conclusions

The optimization of the water pump station scheduling was explored using an improved CE method. Among the improvements made on the classical version of the cross-entropy method, a key one is the asymmetric smoothed update of the mean and covariance matrices parameters. By introducing a smoothed update mechanism for the multivariate normal distribution, the algorithm doesn't directly converge to the optimal solution. Instead, it allows for exploration around the current optimal point, enhancing its ability to search for global minima and improve accuracy.

To evaluate the effectiveness of this method, it was compared against the classic CE method, ant colony optimization, and the genetic algorithm. In all instances, the enhanced CE method consistently outperformed the others, successfully minimizing power consumption for the considered water pumps. Moreover, its generalizability was tested across six benchmark functions, demonstrating its credibility, reliability, and reproducibility.

Applied to a real-world scenario involving a major water pump station within a large city, the improved CE method was tailored to determine optimal operational values for each pump, including head, operating speed, and pump status. Leveraging this information, a pump scheduling system was developed. This system dictates the most efficient combination of active pumps to minimize power consumption while meeting required water flow for various demand scenarios. It also assists in determining pump activation during peak and off-peak hours, aiding human operators in decision-making processes and directly impacting maintenance and labor costs associated with the pump station.

The algorithm was developed and simulated in MATLAB, and its performance was tested in a real-life-inspired scenario, representing a megalopolis water pump station. The results of this study were highly satisfactory, with the system and its components operating at their most efficient points while adhering to all specified constraints.

Bibliography

[1] V.K. Shankar, S. Umashankar, S. Paramasivam, N. Hanigovszki. "A comprehensive review on energy efficiency enhancement initiatives in centrifugal pumping system". Applied Energy., Nov. 2016, pp. 495-513.

[2] European Commission. Study on improving the energy efficiency of pumps; 2001.

[3] F. Vieira, H.M. Ramos, "Optimization of operational planning for wind/hydro hybrid water supply systems", Renewable Energy, vol. 34, pp. 928-936, 2009.

[4] Jowitt P. W., Germanopoulos G. "Optimal Pump Scheduling in Water-Supply Networks." Journal of Water Resources Planning and Management, vol. 118, no. 4, 1992, pp. 406-422.

[5] V. Nitivattananon, E. Sadowski, R. Quimpo, "Optimization of water supply system operation". J. Water Resour. Plan. Manage, pp. 374– 84, 1996.

[6] Sampson, Jeffrey R. "Adaptation in natural and artificial systems (John H. Holland)." (1976): 529.

[7] Dorigo M., Maniezzo V., Colorni A., "The ant system: optimization by a colony of cooperating ants." IEEE Trans Syst Man Cybern 26:29–42, 1996.

[8] Dorigo M., Gambardella L.M., "Ant colony system: a cooperative learning approach to the traveling salesman problem." IEEE Trans Evol. Comput. 1:53–66, 1997.

[9] Goldberg D.E., Holland J.H. "Genetic Algorithms and Machine Learning." Machine Learning, vol. 3, 1988, pp. 95-99.

[10] Hajela P., Lee E., Lin CY. "Genetic Algorithms in Structural Topology Optimization." In: Bendsøe, M.P., Soares, C.A.M. (eds) Topology Design of Structures. NATO ASI Series, vol 227. Springer, Dordrecht, 1993.

[11] Fonseca C.M., Fleming P.J., "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization." Proceedings of the ICGA-93: Fifth

International Conference on Genetic Algorithms, 17-22 July 1993, San Mateo, 1993, pp. 416-423.

[12] Oliveira Turci L. d., Sun H., Bai M., Wang J., Hu P. "Water pump station scheduling optimization using an improved genetic algorithm approach." 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, Jun. 2019, pp. 944-951.

[13] Tang, Kit-Sang, Kim-Fung Man, Zhi-Feng Liu, and Sam Kwong. "Minimal fuzzy memberships and rules using hierarchical genetic algorithms." IEEE Transactions on Industrial Electronics, vol. 45, no. 1, 1998, pp. 162-169.

[14] Zyl, J.E, Savic, D.A., and Walters, G.A. "Operational optimization of water distribution systems using a hybrid genetic algorithm." J. Water Resour. Plann. Manage, vol. 130, no. 2, 2004, pp. 160-170.

[15] Torregrossa D., Capitanescu F. "Optimization models to save energy and enlarge the operational life of water pumping systems." Journal of Cleaner Production, 213, 89-98.

[16] Mambretti S., Villacampa Y., Brebbia C.A., et al. "Optimization of the pumping station of the Milano water supply network with Genetic Algorithms." Nagoya Mathematical Journal, vol. 102, no. 65, 2011, pp. 185-194.

[17] Jang, Se-Hwan, Jae Hyung Roh, Wook Kim, Tenzi Sherpa, Jin-ho Kim, Jong-Bae Park. "A Novel Binary Ant Colony Optimization: Application to the Unit Commitment Problem of Power Systems." 2011.

[18] S. S. Hashemi, M. Tabesh, B. Ataeekia. "Ant-colony optimization of pumping schedule to minimize the energy cost using variable-speed pumps in water distribution networks." Urban Water Journal, 11:5, 335-347, 2014.

[19] L. Manuel, T. D. Prasad, B. Paechter. "Ant Colony Optimization for Optimal Control of Pumps in Water Distribution Networks." J. Water Resour. Plann. Manage, pp-337-346, 2008.

[20] F. Zheng, A. C. Zecchin, J. P. Newman, H. R. Maier and G. C. Dandy. "An Adaptive Convergence-Trajectory Controlled Ant Colony Optimization Algorithm With Application to Water Distribution System Design Problems." IEEE Transactions on Evolutionary Computation, vol. 21, no. 5, pp. 773-791, Oct. 2017.

[21] López-Ibáñez, Manuel, T. Devi Prasad, and Ben Paechter. "Ant colony optimization for optimal control of pumps in water distribution networks." Journal of water resources planning and management 134.4, 2008: 337-346.

[22] Rubinstein, Reuven Y. "The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning." 1997.

[23] Liu, Zaifei, Arnaud Doucet, and Sumeetpal S. Singh. "The cross-entropy method for blind multiuser detection." International Symposium on Information Theory, 2004.

[24] Chepuri, Krishna, and Tito Homem-De-Mello. "Solving the vehicle routing problem with stochastic demands using the cross-entropy method." Annals of Operations Research 134, 2005: 153-181.

[25] Ernst, Damien, et al. "The cross-entropy method for power system combinatorial optimization problems." IEEE Lausanne Power Tech. IEEE, 2007.

[26] Szabó, Zoltán, Barnabás Póczos, and András Lőrincz. "Cross-entropy optimization for independent process analysis." International Conference on Independent Component Analysis and Signal Separation. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[27] Busoniu, Lucian, et al. "Cross-entropy optimization of control policies with adaptive basis functions." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 1, 2010, pp. 196-209.

[28] Kothari, Rishabh P., and Dirk P. Kroese. "Optimal generation expansion planning via the cross-entropy method." Proceedings of the 2009 Winter Simulation Conference (WSC), IEEE, 2009.

[29] Kroese, Dirk P., Sergey Porotsky, and Reuven Y. Rubinstein. "The cross-entropy method for continuous multi-extremal optimization." Methodology and Computing in Applied Probability, vol. 8, 2006, pp. 383-407.

[30] L. Bachus, A. Custodio. "Know and understand centrifugal pumps." pp. 1-12, 2003.

[31] SULZER PUMPS LTD. "Sulzer Centrifugal Pump Handbook," Elsevier, 1989, pp. 40-86.

[32] S. Goss, S. Aron, J. L. Deneubourg, J. M. Pasteels. "Self-organized shortcuts in the Argentine ant." Naturwissenshaften. 76:579-581, 1989.

[33] Dorigo M, Stutzle T, "Ant colony optimization." Massachusetts Institute of Technology, MA, 2004.

[34] Jebari K. "Selection methods for genetic algorithms." Abdelmalek Essaâdi University. International Journal of Emerging Sciences. 2013; 3(4):333–344.

[35] Soon GK, Guan TT, On CK, Alfred R, Anthony P. "A comparison on the performance of crossover techniques in video game," 2013 IEEE international conference on control system. Mindeb: Computing and Engineering; 2013. pp. 493–498.

List of Figures

List of Tables

Publications

[1] Zeng F., Wang JC, Xu JH. “An improved cross-entropy method for the water pump scheduling optimization problem.” Industrial Electronics Society Annual Online Conference, 2023 (IES ONCON 2023).

Acknowledgements

I want to thank God for being with me in both difficult and joyful moments, and for constantly reminding me that there is much more to life than achieving recognition and success. A special thanks to Prof. Wang Jingcheng and my laboratory friends Wu Shunyu, Gan Ziyi, Xu Jiahui, and Rao Jun from Shanghai Jiao Tong University for their support with my thesis.

I would also like to thank my family for encouraging me to continue even in hard times. A heartfelt thanks to my friends, Daniele, Francesco, Matteo, as well as to the 'patio friends' and my flatmates Adriano, Dario, and Thomas. With you, I experienced the best moments of my university years.