Ref. Ares(2024)770826 - 01/02/2024

D4.9 a – Water Efficiency Optimization Services (Intermediate Report)

WP4 - Digital Twin with Smart Analytics and Cognitive Services for Water Efficiency

November 2023

Authors: Yiannis MOURTOS, Pavlos EIRINAKIS, Stavros LOUNIS, George ZOIS, Stathis PLITSOS, Stavros VATIKIOTIS (AUEB)

The AquaSPICE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958396.

Document Information

GRANT AGREEMENT NUMBER	958396	**ACRONYM**	AquaSPICE
FULL TITLE	Advancing Sustainability of Process Industries through Digital and Circular Water Use Innovations		
START DATE	1st December 2020	**DURATION**	51 months
PROJECT URL	www.AquaSPICE.eu		
DELIVERABLE	D4.9 a – Water Efficiency Optimization Services		
WORK PACKAGE	WP4 – Digital Twin with Smart Analytics and Cognitive Services for Water Efficiency		
DATE OF DELIVERY	**CONTRACTUAL** 11/2023	**ACTUAL**	11/2023
NATURE	Report	**DISSEMINATION LEVEL**	Public
LEAD BENEFICIARY	AUEB		
RESPONSIBLE AUTHOR	Yiannis MOURTOS, Pavlos EIRINAKIS, Stavros LOUNIS, George ZOIS, Stathis PLITSOS, Stavros VATIKIOTIS (Athens University of Economics and Business - AUEB)		
CONTRIBUTIONS FROM	-		
ABSTRACT	Deliverable D4.6 reports the outcomes of Task 4.6 “Water Efficiency Optimization Services” of Work Package (WP) 4 of the AquaSPICE project. It presents an extensive literature review on the optimization methods employed in (waste)water industries. This report, also, demonstrates the methodology, which will be applied on the Case Studies, as designed within the project and emphasizes on its novel aspects.		

Document History

VERSION	ISSUE DATE	STAGE	DESCRIPTION	CONTRIBUTOR
0.1	1st October 2023	ToC	Initial Version with ToC	Yiannis MOURTOS, Pavlos EIRINAKIS, Stavros LOUNIS, George ZOIS, Stathis PLITSOS, Stavros Vatikiotis (AUEB)
0.8	20 November 2023	Pre-final for internal review	The completed document for internal review	Stavros Vatikiotis (AUEB)
0.9	28 November 2023	Review with comments	Internal Review	Dimitris Apostolou (ICCS)
1.0	30th November 2023	Final	Final version of the Interim Deliverable	Stavros VATIKIOTIS (AUEB)

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS/ACRONYMS

CS	Case Study
CT	Cooling Tower
DMOALO	Dynamic Multi-Objective Ant Lion Optimization
DSS	Decision Support System
LBBD	Logic-Based Benders Decomposition
LP	Linear Program
MILP	Mixed Integer Linear Program
MINLP	Mixed Integer Non-Linear Program
NPV	Net Present Value
PSO	Particle Swarm Optimization
KPI	Key Performance Indicator
WaterCPS	Water Cyber-Physical System
WP	Work Package
WWRP	Waste Water Reuse Plant
WWTP	Waste Water Treatment Plant

1. Executive summary

The AquaSPICE "Advancing Sustainability of Process Industries through Digital and Circular Water Use Innovations" research project aspires to incorporate circular water-related practices in the process industry and to demonstrate contemporary innovative solutions under three pillars: process, circular, and digital. A Cyber-Physical System (WaterCPS) will be developed, focused on water-related processes, to facilitate the transition from "smart" to "cognitive" process industries. Cognition implies the ability to devise optimal decisions or to re-optimise current decisions, e.g., regarding the volume of water that flows over time within a process. Therefore, optimization models and methods are a pre-requisite for an effective WaterCPS approach and hence an indispensable service within this WaterCPS platform.

Optimization problems have been defined on the cases of CS1B: Dow Böhlen, CS2: Rosignano Solvay and CS6: Tüpraş refinery, since multiple and complex decisions on these specific cases need to take place on both dynamic and simulation context. The remaining CSs of AquaSPICE have not revealed the need for optimising decisions using elaborate mathematical models or algorithms, mostly because the associated decisions can be made in a rather straightforward manner.

The core aim of Deliverable 4.6 is the development of the components of the final optimization service, as designated within Task 4.6. This service has to stand upon an efficient and sound methodological framework in order to offer provably optimal or near-optimal decisions and, in addition, remain replicable well beyond the process settings and specifications defined within AquaSPICE. For this reason, we detail the optimization methods and models, in parallel with an extensive literature review. This will on one hand offer a thorough description for the non-expert in mathematical optimization and on the other hand show that our work moves beyond literature by creating a generic optimization tool, being able to solve problems with networks of different structures and objectives.

The validation of our methodology comes from our initial experimentation on instances comprised of synthetic – yet close to real- data. This procedure validates our approach via the results returned on the pilots.

In conclusion, the fundamental objective of this report is the demonstration of optimization techniques developed within the AquaSPICE project, and also, the proof of their applicability via our initial experimental process.

Due to the non-existence of the SynDi plant in CS#4, this CS is considered as void case and no work related to this CS is reported in this deliverable.

2. Introduction

2.1. Scope and Objectives

The fundamental objective of the AquaSPICE project is to promote water circularity, awareness of resource efficiency, and compact industrial solutions achieving these aims.

A continuously growing body of literature focuses on improving production processes, and subsequently, increasing potential profits by either handling inbound resources more efficiently or limiting related costs. These efforts are being employed on industrial plants via optimization and scheduling algorithms, simulation modules, machine learning tools and/or real-time monitoring platforms under the Digital Twin modeling paradigm [1]. In practice, however, little attention is paid to the outcomes of these heavy industrial processes, and especially, on the low quality of wastewater produced. These residual quantities are typically discharged into the environment e.g., rivers, lakes or the sea. A severe challenge for plant owners/managers is keeping the respective water quality parameters within the range of specified limits predefined from national and international legislation. We indicatively mention that the industrial sector is a major water polluter, as only up to 60% of industrial wastewater [2] receives treatment before being disposed into the environment.

On the other hand, an equally important aspect of the above-described situation originates from the vast amounts of fresh water inserted and used inside the industrial setting. It is also worth mentioning that an estimated 20% of freshwater is used by the industrial sector globally [2], with this, already huge, percentage rising to 50% in industrialized countries [2]. These statistics, when combined with severe water scarcity problems on multiple areas of the world, encourage initiatives to efficiently deal with both situations, aiming at reducing the fresh water quantities drawn. As a result, the problem we are facing is two-fold; i) to efficiently manage fresh water resources without hindering the production processes inside the plant and ii) to treat wastewater flows with the goal of reusing them and thus, indirectly contribute to the first goal of fresh water intake minimization. The complexity of the problem proposed calls for the design of an optimization module tasked to decide upon optimal (waste)water flows within a network on operational and/or design time. The aforementioned optimization method should be, then, encapsuled within a decision support system (DSS) in order to facilitate the work and choices to be made by the appropriate decision makers on the plant.

2.2. Structure of the Deliverable

The structure of this deliverable is as follows.

Chapter 2 presents an introduction regarding the current situation on (waste)water management and clarifies some of its problematic aspects, thus presenting a research gap.

Chapter 3 presents an extensive literature review on optimization related works on (waste)water management.

Chapter 4 defines the problem at hand from the optimization perspective and how this addresses the gap identified in Chapter 2.

Chapter 5 presents our novel methodology designed to tackle the (waste)water management problem, and also, demonstrates the mathematical programs formulated for this task.

Chapter 6 demonstrates the applicability of our techniques on each of the Case Studies, on which an optimization problem has been defined.

Last, Chapter 7 comprises conclusive remarks on this report and Chapter 8 presents the next steps to be followed.

3. Literature Review

Multiple research works have been focusing on the modeling of the processes included inside industrial networks. However, since the problem presented in our work considers the network on a systematic level and not on the interior activities of each process, we focus only on related research, where the problems dealt with involve optimization from the network perspective.

Let us begin our literature review by examining works, where the inlet streams consist of solely wastewater. [3] deal with the problem on design time aiming at deciding upon the configuration of the network in terms of pipe diameters and slopes. The method employed is a broadly used metaheuristic (Particle Swarm Optimization - PSO), where the objective is the total cost minimization. From another economic perspective, i.e. the Net Present Value (NPV), [4] model the wastewater treatment network as a Mixed Integer Non-Linear Program (MINLP), aiming at examining the economical sustainability of wastewater reuse as an energy resource. Their approach is employed on a Case Study with six possible treatment processes on the design time. A similar approach is, also, adopted in [5], where a mathematical programming (MINLP) framework, aiming at designing early-stage treatment networks is proposed. The final solution is obtained by examining paths derived from the optimization formulation under a cost minimization objective.

Interestingly, [6] focus on the reuse possibility of excessive sludge to produce energy; a multi-objective approach, aiming at optimizing operating costs and nitrogen discharge is proposed. [7] examine different structures of the wastewater treatment network depending on whether the treatment processes are in series, in parallel or with in a mixed manner, minimizing the overall cost. [8] take into account possible uncertainties ranging from the cost of each treatment processes or their respective capacities up to the available quantities of wastewater to be treated in the system. Thus, they propose a robust flexible chance-constrained model (RFCCM), which is applied on a Case Study provided by the authors. [9] focus on the wastewater treatment part of the problem and use a dynamic multi-objective ant lion optimization (DMOALO) approach coupled with a deep learning algorithm, aiming at optimizing both the energy consumption and the effluent quality.

On the other hand, related literature, also, includes works, where the inlet streams consist of fresh water. [10] focused on a problem where multiple freshwater resources have to be transferred on different applications (e.g. industrial sector, agriculture) under multiple objectives of either cost (i.e. cost of water consumption) or resource quantities (indicatively, the reuse of water or minimization of ground water extraction). A similar problem is also tackled in [11]. [12] examine a problem where fresh water resources need to satisfy different demands on nearby industrial plants aiming at cost minimization. The MINLP proposed is linearized in order to increase the efficiency of the method.

Additionally, researchers, also, focus on problems, where different types of units coexist (indicatively process and regeneration units), under different objectives such as the total fresh water intake and the number of possible interconnections between components ([13], [14]). [15] designed a nonlinear mathematical model under the objective of minimizing the construction cost of a water distribution network.

Finally, an extension not to be overlooked refers to the multiple-type inlet, i.e., both water and wastewater streams. [16] aim at mixing different inlet streams consisting of both fresh water and wastewater, which are, then, directed to a set of processes. The outcomes of these processes are treated in order to either be discharged within strict quality limits or be recovered. [17] and [18] use a MINLP, which coupled with a decomposition formulation, minimizes the total cost of treatment within the network. A similar problem is, also, tackled in [19] where a two-stage NLP-based heuristic is used to minimize the cost of fresh water intake one the first step and the cost on the second one. [20] designed a simple mathematical programming formulation in a petrochemical plant of three different types of components (input streams, tanks, applications) in order to efficiently treat wastewater, while, also, satisfying the demanded quantities. The objective set is the minimization of fresh water intake in order to complement the already treated wastewater.

However, it should be noted that the majority of the aforementioned literature works rely on nonlinear solvers, which hinders the scalability of the application of the models proposed. Furthermore, the designed mathematical formulations are tailored on the Case Studies proposed and thus, lacking the capability of general transferability on different cases or applications. As a result, we are aiming for a generic optimization tool, which would be indifferent to the structure and the objective set, and, avoiding the non-linearities faced on previous works. This is already a considerable step, whose value of course remains mostly technical and scientific. Its practical value is to be realized within the actual pilot implementation of AquaSPICE WaterCPS. Nevertheless, our lab-based experimentation might provide valuable indications and insights in that direction.

4. Problem Definition

The optimization module aims at deciding the optimal flow rates inside the system network. Before moving on to the optimization model, we should first expand on the basic components of any network, as it seems that a generic characterization of all the nodes inside any network is possible. This will help not only our mathematical formalization but also its data-driven interaction within the WaterCPS platform.

Specifically, typical node categories may be identified as follows:

1. Flow provider nodes, referring to a set of nodes, which provide flows to the network, yet they do not receive any flows by any other node (also called 'source nodes' in the literature).

2. Intermediate nodes, referring to a set of nodes, which both receive and provide flows within the network.

3. Flow receiver nodes, referring to a set of nodes, which only receive flows, yet they do not provide any flow to other nodes (also known as 'sinks' in the literature)

Additionally, the network we define consists of numerous further components, to sustain a generic perspective. Each such component has specific attributes, which should be provided as input to the optimization module. These attributes are inferred by the properties of each component. More precisely:

• Input streams, referring to a set of streams, which insert flow rates inside the system. In terms of attributes:

> *1. Material Flow*: A specific amount of flow, which is already determined and inserted in the system
>
> *2. Quality parameters*: The value of each quality parameter at the time the flow is inserted inside the system

• Treatment processes, referring to a set of processes, where the quality of materials is altered. In terms of attributes:

> *1. Flow rate reduction rate (S)*: The percentage of the (waste)water flow rate that exits each treatment process
>
> *2. Contamination reduction rate (R)*: The reduction percentage of the values of quality parameters on each treatment process
>
> *3. Capacity*: The maximum amount of flow rate which can traverse via the treatment process.
>
> Note that reduction rates (R and S) receive values between 0 and 1 – if the flow rate or the quality parameters are decreased – and greater than 1 if the aforementioned parameters are increased. As an example, if a flow rate of value

x is inserted in a component *i* with a flow rate reduction rate S_i, then the output is: $y = S_i \cdot x$

• Tanks, referring to a set of components, where (waste)water may be -temporarily- stored. In terms of attributes:

1. Capacity: The maximum amount of flow which can traverse via the tank. Note that for Tanks, we may assume that reduction rates *S*, *R* are equal to 1.

• Discharge points, referring to a set of points, where (waste)water may be discharged from the system. In terms of attributes:

1. Quality Requirements: The minimum and maximum quality values between which, water is eligible to be discharged

• Applications, referring to a set of final points, where water (or treated wastewater) may be used. In terms of attributes:

1. Quality Requirements: The minimum and maximum quality values between which, water is eligible to enter each application

2. Demand: The minimum flow which has to be directed to each application

• Edges, referring to the set of possible connections between different components inside the network. In terms of attributes:

1. Capacity: The maximum flow which can traverse each edge.

The next step needed regarding the formulation design is to associate each network component to an appropriate node category. (Waste)water streams are placed in the flow provider category. Similarly, discharge points and applications are considered as flow receiver nodes, while treatment processes and tanks are intermediate nodes. However, it should be clarified that attributes may be extended to be associated with more components than listed above. For example, if a treatment process is extremely sensitive in terms of the inserted water quality, then, the quality requirements attribute may be, also, assigned to this process. The following table sums up the relations between node categories and components.

Node Category	Components	Attributes
Flow provider nodes	(Waste)water Streams	Flow
		Quality Parameters
Intermediate Nodes	Tanks	Capacity
	Treatment Processes	Flow Rate Reduction
		Contamination reduction rate
		Capacity
Flow receiver Nodes	Applications	Quality Requirements
		Flow demanded
	Discharge	Quality Requirements

Table 1 - Node Categories, Components and Attributes of the network

5. Methodology

Given the data input of Section 4, we can present a mathematical formulation in as compact terms as possible. In this section, we provide a network-agnostic formulation for the problem presented. Due to the non-linearities occurring by the expressions related to weighted averaging of flows (Subsection 5.1), we show an efficient linearization approximation, previously discussed in [21] and finally, we demonstrate how networks of a specific structure can be dealt with by using a decomposition scheme.

5.1. Non-linear Model

When defining a mathematical formulation, it is necessary to include i) an objective function i.e., the target of the minimization or maximization procedure, ii) the variables, i.e., formulation components which change their values within the solution process and, iii) the constraints, i.e., the limitations of the solution space. Due to the different goals of each partner within the AquaSPICE project, it is not possible to define a horizontal objective for all Case Studies. However, it is reasonable to assume that the variables may be the flowrates on each edge of a given network. On the other hand, in terms of the constraints, we present the following list:

i) Flow conservation: The sum of flow rates entering a node is equal to the sum of flow rates exiting this node multiplied by the flow rate reduction rate (S).

ii) Demand satisfaction: The flow rate entering a flow receiver node should be equal or greater than the demand needed on this node (if demand is an attribute of the node).

iii) Quality Requirements: The quality parameter (e.g. pH, COD) value must be within the boundaries of quality requirements of each node

iv) Blending constraints: When two or more flows are mixed, the resulting stream is their sum in terms of flow rate and weighted average in terms of quality parameters

As a result, these verbal constraints have to be “translated” into a compact mathematical form to be incorporated to the final optimization model.

Let I be the set of components, E be the set of edges, and P be the set of pollutants. Each edge $e \in E$ is featured with a capacity C_e, while each $i \in I$ is featured with the following ones:

Annotation	Description
d_i	Flowrate demand (m^3/h)
q_i	Available flowrate to be provided (m^3/h)

δ^{i+}	Subset of *E*; The edges in which *i* is the origin node
δ^{i-}	Subset of *E*; The edges in which *i* is the destination node
S_i	Flowrate reduction rate
R_{ip}	Quality reduction rate for pollutant $p \in P$
l_{ip}	Lower limit for pollutant $p \in P$
u_{ip}	Upper limit for pollutant $p \in P$
$\bar{c}_{ip}$	Concentration of pollutant $p \in P$ for the available quantity of water

Table 2 - Annotations for each component $i \in I$

The input data do not provide labels for each component. The pre-processing stage determines the labeling of *I* as it follows:

- If $\delta^{i-} = \emptyset$, then component *i* is a flow provider. A maximum available flowrate q_i is optional.
- If $\delta^{i-} \neq \emptyset$ and $\delta^{i+} \neq \emptyset$, component *i* is an intermediate node.
- If $\delta^{i+} = \emptyset$, then *i* is a flow receiver. If $d_i > 0$, then *i* is an application; otherwise, it is a Discharge Body.
- If $|\delta^{i-}| = 1$, then *i* receives flow from exactly one node.
- If $|\delta^{i-}| > 1$, then *i* is a blending point (currently, only blending points of $|\delta^{i-}| = 2$ are applicable. For blending points of $|\delta^{i-}| > 2$, a construction of imaginary nodes should be preceded; see 5.2).

Given these, we consider the following variables:

- x_e: Flowrate of edge $e \in E$; $0 \leq x_e \leq C_e$
- c_{ip}: Quality of $i \in I$ for pollutant $p \in P$; $l_{ip} \leq c_{ip} \leq u_{ip}$

Flow Conservation:

$$\sum_{e \in \delta^{i+}} x_e \leq q_i \qquad \forall i : \delta^{i-} = \emptyset \qquad (1)$$

$$\sum_{e \in \delta^{i-}} S_i \cdot x_e = \sum_{f \in \delta^{i+}} x_f \qquad \forall i : \delta^{i-} \neq \emptyset, \delta^{i+} \neq \emptyset \qquad (2)$$

$$\sum_{e \in \delta^{i-}} x_e = d_i \qquad \forall i : \delta^{i+} = \emptyset, d_i > 0 \qquad (3)$$

Quality adjustment:

$$c_{ip} = \bar{c}_{ip} \qquad \forall i : \delta^{i-} = \emptyset, p \in P \qquad (4)$$

$$R_{jp} \cdot c_{jp} = c_{ip} \qquad \forall i : \delta^{i-} = \{j\}, p \in P \qquad (5)$$

$$c_{jp} = \frac{c_{ip} \cdot x_{e1} + c_{rp} \cdot x_{e2}}{x_{e1} + x_{e2}} \qquad \delta^{j-} = (i,j), (r,j), p \in P \qquad (6)$$

Variables:

$$0 \leq x_e \leq C_e \qquad \forall e \in E$$

$$l_{ip} \leq c_{ip} \leq u_{ip} \qquad \forall i \in I, p \in P$$

By (1), if i is a flow provider, then the maximum flowrate of all edges δ^{i+} do not exceed the maximum available quantity q_i. By (2), if i is an intermediate node, flows are conserved, given the reduction rate S_i. By (3), if i is a flow receiver with a demand value, then the received quantity is equal with this demand value d_i. By (4), if i is a flow provider, then the respective quality values are predefined. By (5), if i receives flow from exactly one node, then the quality values are conserved, given the respective reduction rates R_{ip}. Finally, constraints (6) calculate the value of each quality parameter p, as it occurs by the mixing of two (or more) edges on a component j.

The above mathematical formulation is non-linear due to the multiplication of variables c_{ip} and x_e on the weighted average calculations. This nonlinearity hinders the solution of the model from commercial solvers, as well as, it restricts its wide applicability due to apparent scalability issues. As a result, we are forced to employ a different approach in order to create a more efficient formulation, which would deal with a wider range of problems.

5.2. Linear Approximation

Let us proceed to explaining the linearisation of the non-linear formulation proposed. Let K be the number of parts, where each mixture may be divided into. Additionally, let k be the number of parts, which originate from edge e_1. That is, if the mixture consists of only two flows, then the number of parts from edge e_2 is equal to $|K\text{-}k|$. Furthermore, even though the capability of this procedure is restricted to up to two nodes, we may deal with this obstacle by creating dummy nodes (or duplicating nodes with the same characteristics of the original ones). As a result, this approximation scheme is applicable to the whole extent of the network and on networks of any structure. Finally, this idea needs to be fed to the model and thus, translated into a mathematical form.

Let us define a maximum number of discretization K (i.e., usually set to 100).

Given these, we add the following variables into the model:

- z_{ijk}: Binary variable; 1 if the flowrate from i to j occupies $\frac{k}{K}$ parts of the mixture in j (e.g., if $z_{ij50} = 1$ and $K = 100$, then the flow from i to j is equal with 50% of the mixture in blending point j).

The updated mathematical formulation is as follows:

Flow Conservation:

$$\sum_{e\in\delta^{i+}} x_e \le q_i \qquad \forall i : \delta^{i-} = \emptyset \qquad (7)$$

$$\sum_{e\in\delta^{i-}} S_i \cdot x_e = \sum_{f\in\delta^{i+}} x_f \qquad \forall i : \delta^{i-} \neq \emptyset, \delta^{i+} \neq \emptyset \qquad (8)$$

$$\sum_{e\in\delta^{i-}} x_e = d_i \qquad \forall i : \delta^{i+} = \emptyset, d_i > 0 \qquad (9)$$

Quality adjustment:

$$c_{ip} = \bar{c}_{ip} \qquad \forall i : \delta^{i-} = \emptyset, p \in P \qquad (10)$$

$$R_{jp} \cdot c_{jp} = c_{ip} \qquad \forall i : \delta^{i-} = \{j\}, p \in P \qquad (11)$$

$$\sum_{k=0}^{K} z_{ijk} = 1 \qquad \forall (i,j) = e^1, \delta^{j-} = \{e^1, e^2\} \qquad (12)$$

$$(K-k)\cdot x_{e1} - k \cdot x_{e2} \le M \cdot (1 - z_{ijk}) \qquad \forall (i,j) = e^1, \delta^{j-} = \{e^1, e^2\}, k = 0, \ldots, K \qquad (13)$$

$$k \cdot x_{e2} - (K-k)\cdot x_{e1} \le M \cdot (1 - z_{ijk}) \qquad \forall (i,j) = e^1, \delta^{j-} = \{e^1, e^2\}, k = 0, \ldots, K \qquad (14)$$

$$\frac{k \cdot c_{ip} + (K-k)\cdot c_{rp}}{K} - c_{jp} \le M \cdot (1 - z_{ijk}) \qquad \forall \delta^{j-} = \{(i,j),(r,j)\}, k = 0, \ldots, K, p \in P \qquad (15)$$

$$c_{jp} - \frac{k \cdot c_{ip} + (K-k)\cdot c_{rp}}{K} \le M \cdot (1 - z_{ijk}) \qquad \forall \delta^{j-} = \{(i,j),(r,j)\}, k = 0, \ldots, K, p \in P \qquad (16)$$

Variables:

$$0 \le x_e \le C_e \qquad \forall e \in E$$

$$l_{ip} \le c_{ip} \le u_{ip} \qquad \forall i \in I, p \in P$$

$$z_{ijk} \in \{0,1\} \qquad \forall (i,j) = e^1, \delta^{j-} = \{e^1, e^2\}, k = 0, \ldots, K$$

Similar to the non-linear model, by (7), if *i* is a flow provider, then the maximum flowrate of all edges δ^{i+} do not exceed the maximum available quantity q_i. By (8), if *i* is an intermediate node, flows are conserved, given the reduction rate S_i. By (9), if *i* is a flow receiver with a demand value, then the received quantity is equal with this demand value d_i. By (10), if *i* is a flow provider, then the respective quality values are predefined. By (11), if *i* receives flow from exactly one node, then the quality values are conserved, given the respective reduction rates Rip. Now, if *j* is a blending point (i.e., it receives flows from two edges e_1, e_2), then edge e_1 covers k parts out of K in the mixture (12). If e_1 covers k parts, then e_2 should cover the remaining K – k, by constraints (13)-(14). Last, if e_1 covers k parts, then c_{jp} is determined by the mean average formula, by constraints (15)-(16).

5.3. Partitioning networks without discharge nodes

For simpler cases of the problem, it is possible that an exact method could replace the piecewise approximation approach, hence providing the optimal solution. We assume that the graph contains flow receiver nodes which are provided by fixed amounts of water flow (i.e., all flow receiver nodes are applications). To demonstrate the following approach more clearly, we consider an indicative graph of two applications, as shown in Figure 1, which is easily extendable to larger networks.

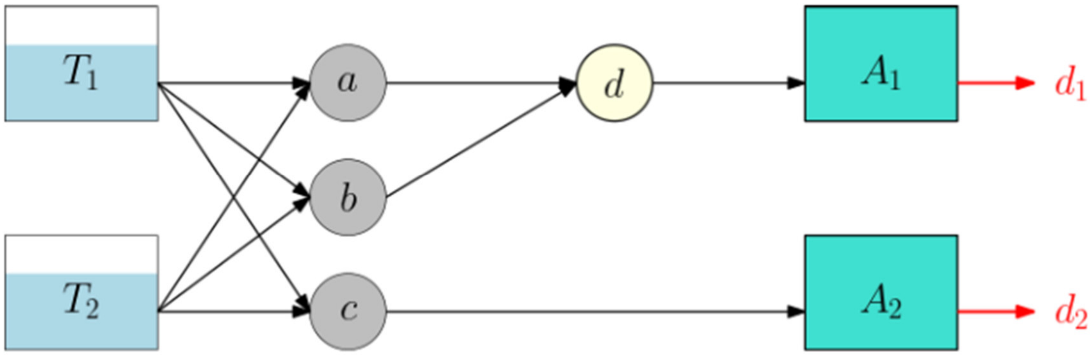

Figure 1 - An indicative graph without discharge nodes

In blending points *a, b, c* and *d*, the nonlinear constraints - reference when the NLP model is applied - should be replaced with equivalent linear expressions. Nonetheless, a straightforward partitioning of the graph is noticed in Figure 2. In this example, we considered a counterpart of exclusively fixed quality values (i.e., path to A_2) and a counterpart of variable quality values (i.e., path to A_1).

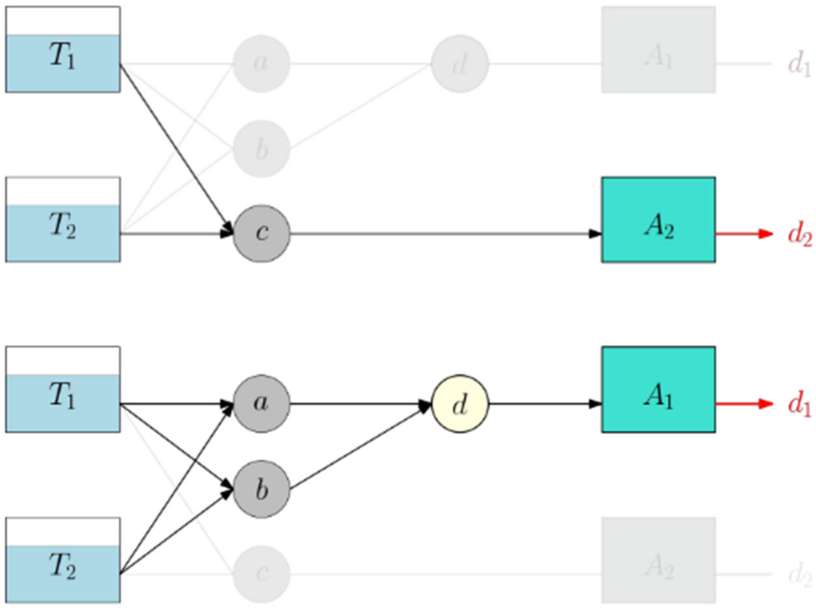

Figure 2 - Partitioning the graph into |A| counterparts

5.3.1. Fixed quality values

Starting with the former one, the problem is subject to the following flow conservation constraints:

$$x_{cA_2} = d_2$$

$$S_c \cdot \left(x_{T_1c} + x_{T_2c}\right) = x_{cA_2} = d_2$$

Given these, the blending constraint is easily linearized:

$$c_{cp} = R_{cp} \cdot \frac{\overline{c_{T_1p}} \cdot x_{T_1c} + \overline{c_{T_2p}} \cdot x_{T_2c}}{x_{T_1c} + x_{T_2c}} = R_{cp} \cdot \frac{\overline{c_{T_1p}} \cdot x_{T_1c} + \overline{c_{T_2p}} \cdot x_{T_2c}}{\frac{d_2}{S_c}} \quad \forall p \in P$$

If no additional restrictions which require binary variables are applied (e.g., candidate by-pass of nodes), the model contains strictly continuous variables, hence it can be solved quickly, even for instances of large scale.

5.3.2. Variable quality values

For the path to application A_1, we notice that a second blending point (i.e., d) is considered, therefore there will be variable quality values. Nevertheless, similar flow conservation constraints are used:

$$x_{dA_1} = d_1$$

$$S_d \cdot (x_{ad} + x_{bd}) = x_{dA_1} = d_1$$

$$S_a \cdot \left(x_{T_1a} + x_{T_2a}\right) = x_{ad}$$

$$S_b \cdot \left(x_{T_1b} + x_{T_2b}\right) = x_{bd}$$

For the blending points, the quality values adhere to the following ones:

$$c_{dp} = R_{dp} \cdot \frac{c_{ap} \cdot x_{ad} + c_{bp} \cdot x_{bd}}{x_{ad} + x_{bd}} = R_{dp} \cdot \frac{c_{ap} \cdot x_{ad} + c_{bp} \cdot x_{bd}}{\frac{d_1}{S_d}}, \forall p \in P$$

$$c_{ap} = R_{ap} \cdot \frac{\overline{c_{T_1p}} \cdot x_{T_1a} + \overline{c_{T_2a}} \cdot x_{T_2a}}{x_{T_1a} + x_{T_2a}} = R_{ap} \cdot \frac{\overline{c_{T_1p}} \cdot x_{T_1a} + \overline{c_{T_2a}} \cdot x_{T_2a}}{\frac{x_{ad}}{S_a}}, \forall p \in P$$

$$c_{bp} = R_{bp} \cdot \frac{\overline{c_{T_1p}} \cdot x_{T_1b} + \overline{c_{T_2b}} \cdot x_{T_2b}}{x_{T_1b} + x_{T_2b}} = R_{bp} \cdot \frac{\overline{c_{T_1p}} \cdot x_{T_1b} + \overline{c_{T_2b}} \cdot x_{T_2b}}{\frac{x_{bd}}{S_b}}, \forall p \in P$$

In these three equations, we observe that four variables (i.e., c_{ap}, c_{bp}, x_{ad} and x_{bd}) imply nonlinear expressions. However, the flow variables x_{ad} and x_{bd} have a summation of d_1/S_d, allowing us to implement a binary search approach. Let k be a value between 0 and 1. It is easily noticed that if the water flow of x_{ad} occupies a percentage of k × 100% of total flow d_1/S_d, then the water flow of x_{bd} will occupy the remaining (1 − k) × 100%. Hence, for each value of k, the respective values of x_{ad} and x_{bd} are obtained, thus all nonlinear expressions are omitted. After enumerating all candidate values of k - given a predefined level of decimal precision - the one which implies the optimal objective value is chosen.

Although the accuracy of the method is greater than the piecewise linear approximation of the aforementioned MILP, still some restrictions should be taken into consideration. Although each LP is expected to be solved in no more than a minimal number of seconds, the number of required iterations is significantly enlarged if multiple blending points are included in the network. Also, we should consider that this iterative procedure should be executed for all pollutants of interest simultaneously, meaning that the LP will be solved for vector of k values of cardinality $|P|$ (i.e., $L = \{k_1, k_2, ..., k_p\}, p \in P$).

5.3.3. Decomposing conflicting subgraphs

For the example above, the partitioning was easy, as each branch was independent to each other. The linearization of the problem is more complicated if the water flows of the tanks are blended before the distinction to the branches occurs, as shown in the following example:

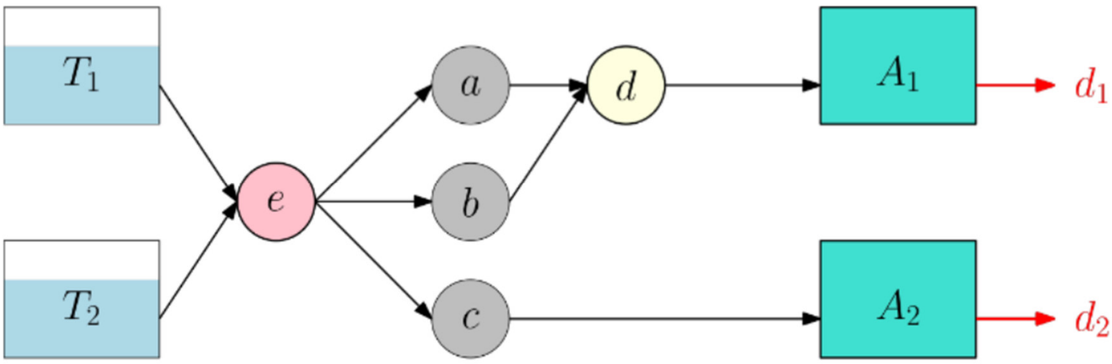

Figure 3 - A graph with conflicting water flows

As the blending point precedes of the branching of water flows, we should opt for a different approach to linearize the blending equation for point *e*. That is, we decompose the LP into two different optimization problems.

We consider a pair of counterparts for each flow x_{T1e} and x_{T2e}:

$$x_{T_1e} = x^1_{T_1e} + x^2_{T_1e}$$

$$x_{T_2e} = x^1_{T_2e} + x^2_{T_2e}$$

Each counterpart addresses the partial flow which is provided to exclusively cover the demand of the indicated application (e.g., $x^1_{T_1e}$ is the water flow which is provided by tank T_1 and it will be used for the branch of A_1 only). Then, we partition the graph into two different Linear Programs, similarly as in Figure 2, each one addressing one branch. The implication of this will be a scheme which can be solved using a Logic-Based Benders Decomposition framework: the master problem M solves the LP of branch to A_2, using the linearization techniques which have been mentioned earlier. Given the solution of M, the quality value of e is obtained as a parameter, since the following relation between the flows of the tanks holds:

$$\frac{x^2_{T_1e}}{x^2_{T_2e}} = \frac{x^1_{T_1e}}{x^1_{T_2e}}$$

For the subproblem S, we compute the optimal flows to application A_2, given the restrictions that the solution of M implies (i.e., the quality values in point e). If $\frac{x^2_{T_1e}}{x^2_{T_2e}} = \alpha$, then the following constraint is used:

$$x^1_{T_1e} = \alpha \cdot x^1_{T_2e}$$

and if $c_{ep} = \overline{c_{ep}}$ for each pollutant $p \in P$, then:

$$\overline{c_{ep}} = \frac{\overline{c_{T_1p}} \cdot \alpha \cdot x^1_{T_2e} + \overline{c_{T_2p}} \cdot x^1_{T_2e}}{(1+\alpha) \cdot x^1_{T_2e}}, \quad \forall p \in P$$

which is a set of linear expressions. Thereinafter, the remaining formulation of the subproblem is identical with the constraints of the branch to A_1 of the aforementioned LP, integrating the binary search approach for all pollutants of interest. Sequentially solving M and S will either imply a feasible solution of the flow network or an infeasible subproblem (i.e., it is possible that the quality values of some pollutants in e cannot ensure that the requirements of the branch to A_1 will be respected). For the former case, there is no proof that the obtained solution is the globally optimal one; for the latter one, a new solution should be explored. In both cases, a Benders cut should be added to M, so that resolving it will imply a different solution for branch to A_1. To ensure the validity of the method, the proposed cuts should ensure that either $x^2_{T_1e}$ or $x^2_{T_2e}$ should receive different values in the upcoming iterations. Since no binary variables are involved, we consider the construction of auxiliary variables, mentioned as literals in literature. Let l be the value of variable $x^2_{T_1e}$ at some iteration t. We consider a pair of binary variables $b^{\geq l+1}$ and $b^{\leq l-1}$ being equal to 1 if $x^2_{T_1e} \geq l+1$ or $x^2_{T_1e} \leq l-1$ respectively:

$$b^{\geq l+1} + b^{\leq l-1} = 1 \qquad (11)$$

$$l + 1 - x^2_{T_1e} \leq M \cdot (1 - b^{l+1}) \qquad (12)$$

$$x_{T_1e}^2 - (l - 1) \leq M \cdot (1 - b^{l-1}) \tag{13}$$

Constraint (11) ensures that one literal will be set to 1. Constraint (12) ensures that if $b^{\geq l+1}$ is set to 1, then the flow of $x_{T_1e}^2$ will be greater than $l + 1$. On the contrary, if $b^{\leq l-1}$ is set to 1, then $x_{T_1e}^2$ will be less than $l - 1$, by constraint (13). In any case, adding these three cuts to M will ensure that, for all upcoming iterations, $x_{T_1e}^2 \neq l$.

6. Case Studies

Within the AquaSPICE project, optimization problems have been identified on three different Case Studies, namely CS#1B: Dow Böhlen, CS#2: Solvay, Aretusa and CS#6: Tüpraş refinery. For convenience, we follow the structure of “D4.3 – Cognitive Model and Simulation for Water-Related Processes”, where simulation results and network structures have been described. As a result, the optimization module assumes two different states for each Case Study: i) the situation before the AquaSPICE (AS-IS) and ii) the situation after the AquaSPICE project (TO-BE). On each different situation, the methodology will be tested, in order to be validated in terms of feasibility. Since the Decomposition method depends on specific objectives and network structures, we will test only the piece-wise linear approximation method, which is horizontal and applicable to any network.

There are two main differences between these three cases - apart from the ones expected on the network structure. The first one is the optimization objective, where we aim at

- maximizing the wastewater quantities to be treated for Tüpraş,
- minimizing the total cost in the case of Rosignano Solvay, and,
- minimizing the freshwater intake from the available water sources in the Dow Böhlen case.

The second difference relates to the purpose of use regarding the optimization module. In the cases of Tüpraş and Dow Böhlen, the optimizer will be used on an operational scale, whereas the Rosignano Solvay case opts for using it during the design phase of the flow network.

These differences show the versatility and broad applicability of our mathematical formulation, which is effective regardless of the purpose of use and the optimization objective. In addition, these differences yield the three cases we focus on as quite complementary and therefore representative of how our approach could be replicable beyond AquaSPICE pilots.

6.1. CS#1B: Dow Böhlen, Germany

6.1.1. Case Study Overview

The Dow Böhlen industry operates in a region south of Leipzig, where the sources of freshwater are limited. That causes significant water stress due to the high water requirements of the industrial plant. So, to diminish the freshwater intake, it is necessary to enhance the internal recycling of process water streams by incorporating sustainable solutions. Within the AquaSPICE project, the allocation of fresh water sources can be optimized for the production of higher quality water, utilizing it for fit-for-purpose applications.

The AS-IS network is shown in Figure 4. The objective is the minimization of the fresh water intake from lake Witznitz, while, also, satisfying predefined demands on the applications (Cooling Towers and Demin Water).

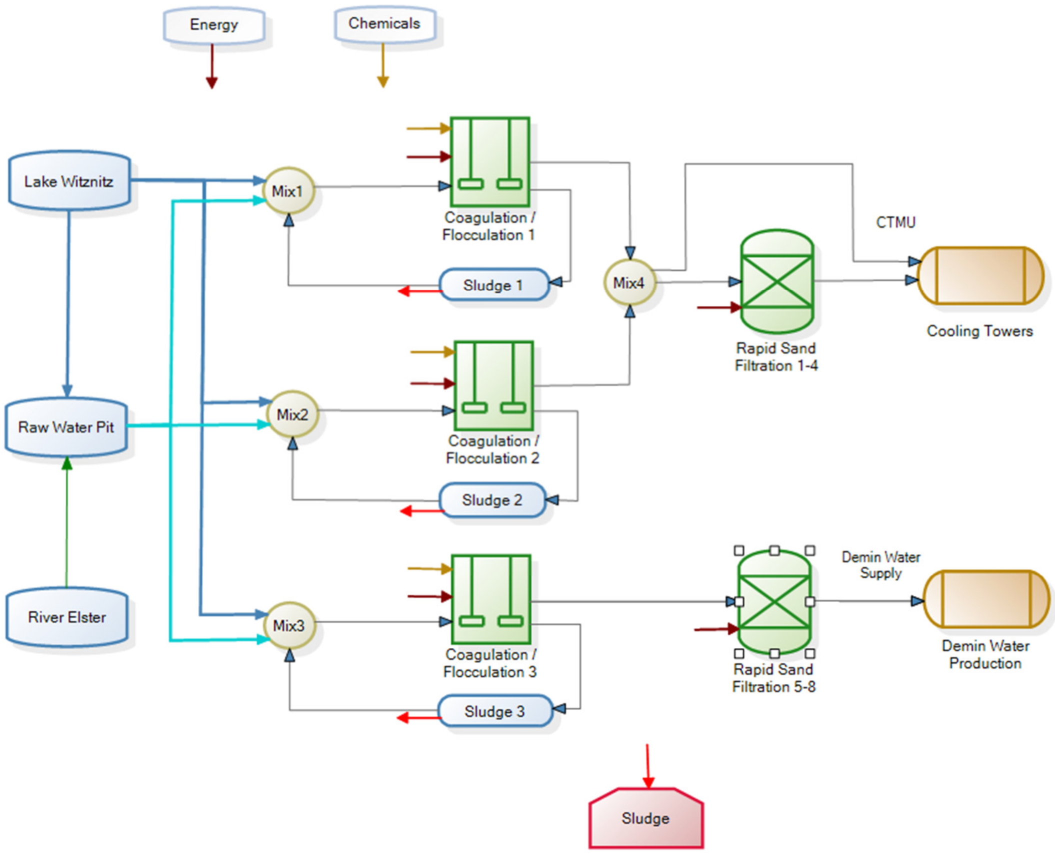

Figure 4 – Network for the As-Is Scenario in CS#1B

As shown above, the network consists of two freshwater streams (namely Lake Witznitz and River Elster), three wastewater streams (Sludge 1-3), the raw water pit, which is considered as a Tank, Coagulation/Flocculation, and Rapid Sand Filtration, which are considered as treatment processes and two applications (Demin water and Cooling Towers). The objective which has been agreed upon is the minimization of the fresh water intake from River Elster with respect to the aforementioned constraints. Another important clarification to be made is the difference between the data provision mechanism. We distinct two different cases; i) online data, referring to data which should be provided each time the optimization module is triggered and ii) offline data, referring to data which remain unchanged. The following Table sums up the data generation mechanisms used during the testing of the mathematical formulation application on Case Study 1B. Data used have been obtained from the historical data provided by Dow

Böhlen, and 'min', 'max' values are the corresponding minimum and maximum values observed.

Component Type	Component	Attribute Category	Attribute	Type	min	max	Online/Offline	Data provider
Fresh water Input Streams	Lake Witznitz	Quality Parameters	pH	u.a.r.	7.39	8.12	Online	Dow Böhlen
			EC	u.a.r.	466	690	Online	Dow Böhlen
			Fe_tot (Iron)	u.a.r.	0.01	1.7	Online	Dow Böhlen
			Sulphate	u.a.r.	1	139	Online	Dow Böhlen
			Zn_dissolved (Zinc)	u.a.r.	0.01	0.03	Online	Dow Böhlen
			Cl (Chloride)	u.a.r.	50	60	Online	Dow Böhlen
			PO4-P_tot (Phosphate)	u.a.r.	0.1	0.54	Online	Dow Böhlen
			Ca (Ca Hardness)	u.a.r.	0.8	1.8	Online	Dow Böhlen
			Ca/Mg (Hardness)	u.a.r.	1.1	2.3	Online	Dow Böhlen
			KS4.3	u.a.r.	1.3	2.56	Online	Dow Böhlen
			TOC	u.a.r.	5	9	Online	Dow Böhlen
	Lake Witznitz	Quality Parameters	pH	u.a.r.	7.74	8.2	Online	Dow Böhlen
			EC	u.a.r.	571	1415.9	Online	Dow Böhlen
			Fe_tot (Iron)	u.a.r.	37	400	Online	Dow Böhlen
			Sulphate	u.a.r.	0.01	0.1	Online	Dow Böhlen
			Zn_dissolved (Zinc)	u.a.r.	16	34	Online	Dow Böhlen
			Cl (Chloride)	u.a.r.	1.4	5.9	Online	Dow Böhlen
			PO4-P_tot (Phosphate)	u.a.r.	0.01	0.11	Online	Dow Böhlen
			Ca (Ca Hardness)	u.a.r.	1.69	9.41	Online	Dow Böhlen
			Ca/Mg (Hardness)	u.a.r.	0.01	0.8	Online	Dow Böhlen
			KS4.3	u.a.r.	1.4	3.75	Online	Dow Böhlen
			TOC	u.a.r.	0.01	0.05	Online	Dow Böhlen
Wastewater Stream	Sludge 1-3	Fixed Flowrate	Fixed Flowrate	u.a.r.	10	30	Online	Dow Böhlen
		Quality Parameters	Similar to Lake Witznitz (randomly) *	u.a.r.	-	-	Online	Dow Böhlen
Tank	Raw Water Pit	Flow Rate reduction rate	Flow Rate reduction rate	constant	1	1	Offline	Dow Böhlen
		Contamination Reduction rate	All Quality parameters	constant	1	1	Offline	Dow Böhlen
Applications	Cooling Tower	Demand	Demanded Flow rate	u.a.r.	496	1014	Online	Dow Böhlen

		Quality Requirements	pH	constant	7.9	8.5	Offline	Dow Böhlen
			EC	constant	-	3000	Offline	Dow Böhlen
			Fe_tot (Iron)	constant	-	0.5	Offline	Dow Böhlen
			Sulphate	constant	-	1000	Offline	Dow Böhlen
			Zn_dissolved (Zinc)	constant	-	3	Offline	Dow Böhlen
			Cl (Chloride)	constant	-	500	Offline	Dow Böhlen
			PO4-P_tot (Phosphate)	constant	-	1	Offline	Dow Böhlen
			Ca (Ca Hardness)	constant	-	45	Offline	Dow Böhlen
			Ca/Mg (Hardness)	constant	-	70	Offline	Dow Böhlen
			KS4.3 (m-value)	constant	-	4	Offline	Dow Böhlen
	Demin Water	Demand	Demanded Flow rate	u.a.r.	169	285	Online	Dow Böhlen
		Quality Requirements	pH	constant	6.5	7.5	Offline	Dow Böhlen
			EC	constant	-	0.1	Offline	Dow Böhlen
			TOC	constant	-	0.2	Offline	Dow Böhlen
Treatment Processes	Rapid Sand Filtration	Flow Rate reduction rate	Flow Rate reduction rate	u.a.r.	0.85	0.95	Offline	TUC
		Contamination Reduction rate	pH	constant	1	1	Offline	TUC
			EC	constant	0.31	0.31	Offline	TUC
			Fe_tot (Iron)	constant	0.05	0.05	Offline	TUC
			Sulphate	constant	0.92	0.92	Offline	TUC
			Zn_dissolved (Zinc)	constant	0.06	0.06	Offline	TUC
			Cl (Chloride)	constant	1	1	Offline	TUC
			PO4-P_tot (Phosphate)	constant	1	1	Offline	TUC
			Ca (Ca Hardness)	constant	1	1	Offline	TUC
			Ca/Mg (Hardness)	constant	1	1	Offline	TUC
			KS4.3 (m-value)	constant	1	1	Offline	TUC
			TOC	constant	0.75	0.75	Offline	TUC
	Coagulation / Flocculation	Flow Rate reduction rate	Flow Rate reduction rate	constant	1	1	Offline	TUC
		Contamination Reduction rate	pH	u.a.r.	1	1.13	Offline	TUC
			EC	constant	1	1	Offline	TUC
			Fe_tot (Iron)	constant	1	1	Offline	TUC
			Sulphate	constant	1	1	Offline	TUC
			Zn_dissolved (Zinc)	constant	1	1	Offline	TUC
			Cl (Chloride)	constant	1	1	Offline	TUC
			PO4-P_tot (Phosphate)	constant	1	1	Offline	TUC
			Ca (Ca Hardness)	constant	1	1	Offline	TUC
			Ca/Mg (Hardness)	constant	1	1	Offline	TUC
			KS4.3 (m-value)	constant	1	1	Offline	TUC
			TOC	u.a.r.	0.75	0.99	Offline	TUC

Edges	Edges	Capacity	Edge Capacity	constant	10000 **	10000	Offline	Dow Böhlen

*: No data have been provided on the sludge streams and as a result, they are currently randomly generated.

**: edges are considered incapacitated – a big number is used.

Table 3 - Data generation mechanism for optimization experimentation on CS#1B

Preliminary testing results have been performed, using the bounds of the parameters from Table 3, and 10 instances of synthetic data have been created. For our experimentation we use only the quality parameters of pH – attributes which are available on all components either as requirement or as an input quality parameter. The algorithm runs optimally, in under 2 minutes – however, optimality under that time may depend on the range of data values provided.

The next round of experimentation will focus on the comparison of the AS-IS situation and the TO-BE situation, in laboratory terms rather than in terms of everyday use, in order to provide numerical improvements on KPIs caused within the actions of the AquaSPICE project.

6.2. CS#2: Solvay, Aretusa, Italy

6.2.1. Case Study Overview

The Rosignano Solvay plant is located in Italy and produces numerous products, such as sodium carbonate, calcium chloride, chlorine, hydrochloric acid, plastic materials, peracetic acid, and hydrogen peroxide. The Solvay chemical plant collaborates with Consorzio Aretusa, attempting to allocate the water more efficiently by implementing a utility-industry symbiosis system.

6.2.2. Scope

The objective of the project is to treat and reuse the industrial wastewater of Solvay. In particular, the generated wastewater derives from the peroxide and peracetic acid production plant. In this regard, the target contaminants must be removed by deploying the appropriate water treatment technologies. For this reason, the wastewater will be treated within the WAPEREUSE system. Water is aimed to be reused within the plant for cooling purposes in order to cover the freshwater demand.

However, in this Case Study, the synergies that can be developed are particularly interesting. The chemical plant of Solvay, the Aretusa Wastewater Reuse Plant (WWRP), and the Rosignano Waste Water Treatment Plant (WWTP) collaborate by exchanging, among others, water and wastewater. There are three alternative closed-loop scenarios to follow regarding this synergy. So, the allocation of water among them should be simulated and optimized in order to achieve a water-efficient system and promote water circularity.

6.2.3. Alternative Scenarios

The alternative scenarios have been already defined in D4.1, and are also, presented on an abstract level on the following figure.

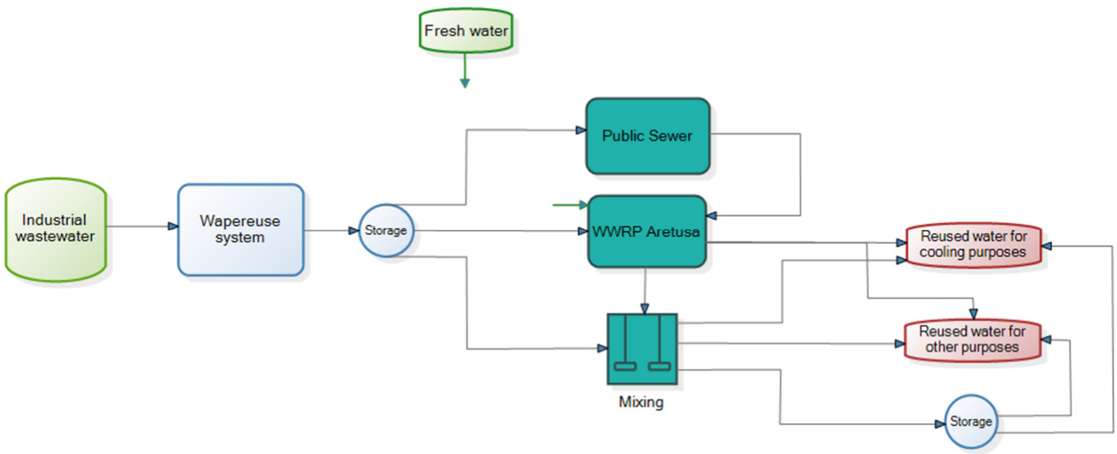

Figure 5 – Representation of alternative scenarios in CS#2

<u>Scenario 1:</u> The effluent from the WAPEREUSE system (10 m^3/h) is discharged to the public sewerage system, where it is further treated. The same water quantity is, then, sent to the WWRP Aretusa, and this actor supplies 440 m^3/h of water in total to Solvay,

15 m^3 water/h to be reused as CTMU water and 425 m^3 water/h for other cooling purposes.

Scenario 2: The effluent stream is provided to WWRP Aretusa, where it is treated. If the water quality parameters for certain contaminants are met, regarding to the required quality of the cooling towers, the stream is sent back to the Solvay plant.

Scenario 3A: The treated wastewater (10 m^3/h) is mixed with fresh water coming from WWRP Aretusa (430 m^3/h). The water quality and quantity of Aretusa are fixed. It is then used for cooling and other purposes of the Solvay chemical plant. It is worth mentioning that the cooling towers require water of specific quality. In order to use the mixed stream as CTMU water for the cooling towers of Solvay, the effluent from the WAPEREUSE has to be within an acceptable range. This way, after mixing with the water from Aretusa, the water quality will meet the quality requirements for the CTs.

Scenario 3B: The treated wastewater by WAPEREUSE is mixed with water from WWRP Aretusa to be sent to the Solvay chemical plant. However, in this scenario, the ratio of the two streams varies. WAPEREUSE stream is in the range of 1-10 m^3/h and the water from Aretusa of 5-14 m^3/h. The ratio depends on the water quality provided by WWRP Aretusa and the water quality required for the CTs. So, the total amount of water stream flow is 15 m^3/h. In addition, WWRP Aretusa will supply approximately 425 m^3/h for other purposes within the industry.

Similar to our approach in the CS#1B, we assign each component with its appropriate type, and attributes. Since the complete (finalized) data are yet at our disposal, we move on with an abstract representation.

Component Type	Component	Attribute Category	Attribute	Online/Offline
Wastewater Input Streams	Exit Stream from WAPERUSE	Quality Parameters	Conductivity	Online
			COD	Online
			TOC	Online
			pH	Online
			Nitrates	Online
			Sulfates	Online
			Aluminium	Online
			E. coli	Online
			Iron	Online
			Hydrogen peroxide	Online
Applications	Cooling Towers	Quality Requirements	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
		Demand	Demanded Flowrate	Online
	Other purposes	Quality Requirements	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline

			Aluminium	Offline
			Iron	Offline
		Demand	Demanded Flowrate	Online
Treatment Process	Scenario 1	Flow Rate change	Flow Rate change	Offline
		Cost	Cost	Offline
		Contamination Reduction	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
		Quality Requirements	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
	Scenario 2	Flow Rate change	Flow Rate change	Offline
		Cost	Cost	Offline
		Contamination Reduction	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
		Quality Requirements	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
	Scenario 3A	Flow Rate change	Flow Rate change	Offline
		Cost	Cost	Offline
		Contamination Reduction	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
		Quality Requirements	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
	Scenario 3B	Flow Rate change	Flow Rate change	Offline
		Cost	Cost	Offline
		Contamination Reduction	pH	Offline
			Conductivity	Offline

			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
		Quality Requirements	pH	Offline
			Conductivity	Offline
			COD	Offline
			Nitrates	Offline
			Sulfates	Offline
			Hydrogen peroxide	Offline
			Aluminium	Offline
			Iron	Offline
Edges	Edges	Capacity	Edge Capacity	Offline

Table 4 - Data needs for the optimization model on CS#2

As mentioned above, the optimization role in this Case Study resides on the design time level. The candidate optimization objective is the cost minimization, while, also, satisfying the quality requirements set by the pilot on each application and scenario. Note that on this pilot, each scenario has entry quality requirements, in the sense that if these bounds are violated, then no flow may traverse this scenario. From the optimization perspective, we opt for validating the implementation decisions by simulating different settings of input data from the network perspective and then, examine the robustness of the choice made. More precisely, we are focusing on whether another choice would be more suitable given that the values of input data change. This would be possible via multiple executions of the optimization tool. Experimental work and results are noted on the next steps.

6.3. CS#6: Tüpraş refinery, Turkey

6.3.1. Case Study Overview

Tüpraş is the largest industrial enterprise in Turkey since it operates four oil refineries with a total processing capacity of 30 million tons of crude oil every year. The refinery consumes significant amounts of freshwater. On the other hand, the treated wastewater is discharged into the sea. However, there is the potential to either be further treated by using the biological treatment system, which Tüpraş has at its disposal, or be reused within the plant to cover the intensive freshwater requirements. AquaSPICE aims to promote water circularity in Tüpraş by treating adequately the industrial wastewater while, also, minimizing environmental pollution by producing water whose composition will comply with the legal discharge limits of Turkey.

In the refinery, wastewater treatment is considered necessary in order to reuse it within the plant. Treated water may substitute raw water, which is abstracted from a lake. In addition, by treating the generated industrial wastewater, in case of discharging it into the sea, the environmental pollution will be decreased due to lower contamination levels. Therefore, the system may comply with the legal discharge limits. Figure 6 shows the current state of the network, while edges and components in blue demonstrate the extra additions within the project.

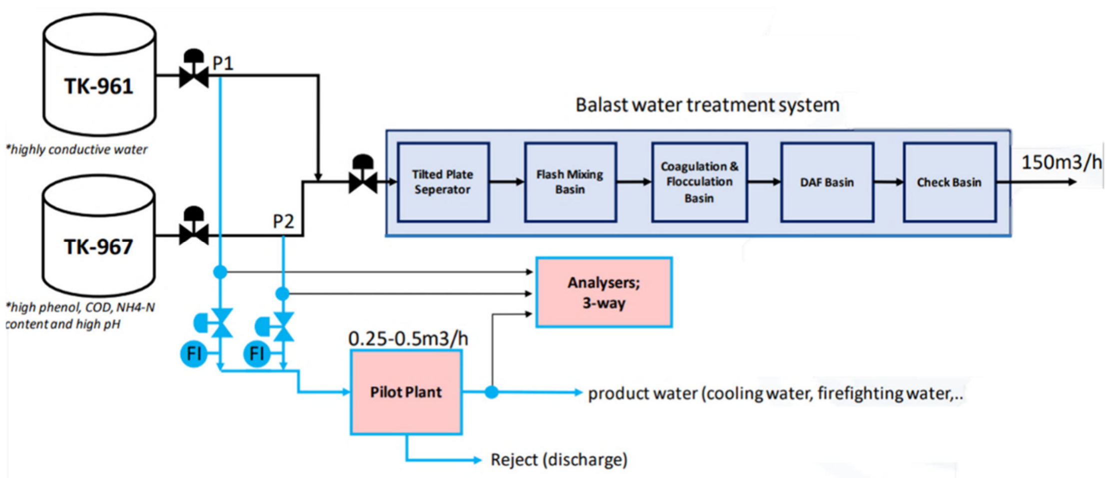

Figure 6 - To-Be Scenario for the treatment of industrial wastewater in CS#6

As shown above, the network consists of two wastewater streams (namely TK-961 and TK-967), the pilot plant (union of Activated granular sludge, Ultrafiltration and Reverse osmosis) and the biological treatment which are considered as treatment processes, the discharge body and two applications (firefighting water and cooling water). We, also, include the Ballast Wastewater treatment, which is located on another branch of the network and its role is to treat wastewater quantities in order to be discharged. The objective which has been agreed upon with the colleagues from TÜPRAŞ is the maximization of the wastewater flow rate from TK-961 with respect to the aforementioned constraints. Another important clarification to be made is the difference between the data provision mechanism. We distinct two different cases; i) online data, referring to data which should be provided each time the optimization module is triggered and ii) offline data, referring to data which remain unchanged. The following Table sums up the data generation mechanisms used during the testing of the mathematical formulation application on Case Study 6. Data used have been obtained from the historical data provided by TÜPRAŞ, and 'min', 'max' values are the corresponding minimum and maximum values observed.

Component Type	Component	Attribute Category	Attribute	Type	min	max	Online/Offline	Data provider
Wastewater Input Streams	TK-961	Quality Parameters	pH	u.a.r.	6	9	Online	TÜPRAŞ
			Conductivity	u.a.r.	1500	15000	Online	TÜPRAŞ
			COD	u.a.r.	100	250	Online	TÜPRAŞ
			S	u.a.r.	0.5	5	Online	TÜPRAŞ
			NH_4N	u.a.r.	2	10	Online	TÜPRAŞ
			Suspended solids	u.a.r.	10	60	Online	TÜPRAŞ
	TK-967	Quality Parameters	pH	u.a.r.	9	14	Online	TÜPRAŞ
			Conductivity	u.a.r.	1000	50000	Online	TÜPRAŞ
			Phenol	u.a.r.	10	3500	Online	TÜPRAŞ
			COD	u.a.r.	1000	10000	Online	TÜPRAŞ
			S	u.a.r.	20	1000	Online	TÜPRAŞ
			NH_4N	u.a.r.	20	250	Online	TÜPRAŞ
			Oil	u.a.r.	50	250	Online	TÜPRAŞ
Discharge Points	All Discharge Points	Quality Requirements	pH	constant	6	9	Offline	TÜPRAŞ
			Conductivity	constant	750	2500	Offline	TÜPRAŞ
			Phenol	constant	0	1	Offline	TÜPRAŞ
			COD	constant	0	200	Offline	TÜPRAŞ
			S	constant	0	1	Offline	TÜPRAŞ
			NH_4N	constant	0	20	Offline	TÜPRAŞ
			Suspended solids	constant	0	60	Offline	TÜPRAŞ
			Oil	constant	0	10	Offline	TÜPRAŞ
			Hydrocarbons	constant	0	10	Offline	TÜPRAŞ
Applications	Cooling Water	Quality Requirements	pH	constant	7	8.5	Offline	TÜPRAŞ
			Conductivity	constant	15	150	Offline	TÜPRAŞ
			Phenol	constant	0	0	Offline	TÜPRAŞ
			COD	constant	0	15	Offline	TÜPRAŞ
			Suspended solids	constant	0	1	Offline	TÜPRAŞ
			Oil	constant	0	1	Offline	TÜPRAŞ
			Hydrocarbons	constant	0	0	Offline	TÜPRAŞ
			Total Alkalinity	constant	0	90	Offline	TÜPRAŞ
	Firefighting Water	Quality Requirements	pH	constant	6	7.5	Offline	TÜPRAŞ
			Conductivity	constant	90	250	Offline	TÜPRAŞ
			Phenol	constant	0	0	Offline	TÜPRAŞ
			COD	constant	5	50	Offline	TÜPRAŞ
			NH_4N	constant	0	0	Offline	TÜPRAŞ
			Suspended solids	constant	0	3.5	Offline	TÜPRAŞ
			Oil	constant	0	1	Offline	TÜPRAŞ
			Hydrocarbons	constant	0	0	Offline	TÜPRAŞ

Treatment Process	Biological Treatment	Flow Rate reduction rate	Flow Rate reduction rate	u.a.r.	0.97	1	Offline	TUC
		Contamination Reduction rate	pH	constant	1	1	Offline	TUC
			Conductivity	constant	1	1	Offline	TUC
			Phenol	u.a.r.	0.1	0.5	Offline	TUC
			COD	u.a.r.	0.02	0.7	Offline	TUC
			S	constant	1	1	Offline	TUC
			NH_4N	u.a.r.	0.06	0.23	Offline	TUC
			Suspended solids	u.a.r.	0.04	0.5	Offline	TUC
			Oil	u.a.r.	0.04	0.42	Offline	TUC
			Hydrocarbons	u.a.r.	0.1	0.5	Offline	TUC
	Pilot Plant	Flow Rate reduction rate	Flow Rate reduction rate	u.a.r.	0.77	0.81	Offline	TUC
		Capacity	Pilot Plant Capacity	constant	0.5	0.5	Offline	TÜPRAŞ
		Contamination Reduction rate	pH	u.a.r.	0.82	1.11	Offline	TUC
			Conductivity	u.a.r.	0.01	0.03	Offline	TUC
			Phenol	u.a.r.	1.29	1.74	Offline	TUC
			COD	u.a.r.	0.01	0.03	Offline	TUC
			S	constant	0.01	0.01	Offline	TUC
			NH_4N	constant	0.01	0.01	Offline	TUC
			Suspended solids	constant	0.01	0.01	Offline	TUC
			Oil	constant	0.05	0.05	Offline	TUC
			Hydrocarbons	u.a.r.	0.3	0.8	Offline	TUC
	Ballast	Flow Rate reduction rate	Flow Rate reduction rate	u.a.r.	0.85	0.95	Offline	TUC
		Contamination Reduction rate	pH	constant	1.01	1.01	Offline	TUC
			Conductivity	constant	1.26	1.26	Offline	TUC
			Phenol	u.a.r.	0.1	0.44	Offline	TUC
			COD	u.a.r.	0.96	1.27	Offline	TUC
			S *	u.a.r.	0	1	Offline	TUC
			NH_4N	constant	1	1	Offline	TUC
			Suspended solids	u.a.r.	0.4	1.22	Offline	TUC
			Oil	u.a.r.	0.44	1.26	Offline	TUC
			Hydrocarbons	u.a.r.	0.28	0.82	Offline	TUC
Edges	Edges	Capacity	Edge Capacity	constant	200	200	Offline	AUEB**

*: insufficient data

**: A necessary assumption to create bounds on the flow rates.

Table 5 - Data generation mechanism for optimization model testing on CS#6

Preliminary testing results have been performed, using the bounds of the parameters from Table 5, and 10 instances of synthetic data have been created. For our experimentation we use only the quality parameters of pH and Conductivity – attributes which are available on all components either as requirement or as an input quality parameter. The algorithm runs optimally, in under 1 minute – however, optimality under that time may depend on the range of data values provided.

Let us provide a numerical example.

```
▼ WWS:
  ▼ TK-961:
      ID: 0
    ▼ QP:
        pH: 7.748
        Conductivity: 1679.584
  ▼ TK-967:
      ID: 3
    ▼ QP:
        pH: 10.786
        Conductivity: 39612.479
  ▼ Oily:
      ID: 8
      Water_flow: 860.568
    ▼ QP:
        pH: 8.296
        Conductivity: 1762.627
▼ Discharge:
  ▼ AB:
      ID: 2
    ▼ Req:
      ▼ Lower:
          pH: 6
          Conductivity: 0
      ▼ Upper:
          pH: 9
          Conductivity: 2500
  ▼ AP:
      ID: 9
    ▼ Req:
      ▼ Lower:
          pH: 6
          Conductivity: 0
      ▼ Upper:
          pH: 9
          Conductivity: 2500
▼ Treatment:
  ▼ Ballast:
      ID: 1
      S: 0.933
      Capacity: 250
    ▼ R:
        pH: 1.01
        Conductivity: 1.26
  ▼ Pilot_plant:
      ID: 4
      S: 0.775
      Capacity: 0.5
    ▼ R:
        pH: 1.003
        Conductivity: 0.029
  ▼ Biological_treatment:
      ID: 7
      S: 0.998
      Capacity: 1250
    ▼ R:
        pH: 1
        Conductivity: 1
    ▼ Req:
      ▼ Lower:
          pH: 0
          Conductivity: 0
      ▼ Upper:
          pH: 9
          Conductivity: 2500
  ▼ WW_Recovery:
      ID: 10
      S: 1
    ▼ R:
        pH: 1
        Conductivity: 1
▼ App:
  ▼ Fire:
      ID: 6
    ▼ Req:
      ▼ Lower:
          pH: 6
          Conductivity: 0
      ▼ Upper:
          pH: 8.5
          Conductivity: 1500
  ▼ CW:
      ID: 5
    ▼ Req:
      ▼ Lower:
          pH: 7.5
          Conductivity: 15
      ▼ Upper:
          pH: 8.5
          Conductivity: 250
▼ Edges:
  ▼ Capacities:
      0-1: 1000
      0-4: 1000
      3-1: 500
      1-2: "No"
      3-4: 500
      4-5: "No"
      4-6: "No"
      4-7: "No"
      4-9: "No"
      8-7: "No"
      7-10: "No"
      10-9: "No"
      10-5: "No"
      10-6: "No"
```


Figure 7 - Input instance for CS#6

On the Figure above, we present the different component types (Wastewater Streams, Discharge points, Treatment processes, Applications and Edges), which encapsule the components of the system. Each component is, also, associated with its respective attributes, containing a unique ID number to be properly identified.

After the execution of the algorithm, the output returned is shown on the following Figure.

```
Components:
  App:
    5:    "CW"
    6:    "Fire"
  Discharge:
    2:    "AB"
    9:    "AP"
  Treatment:
    1:    "Ballast"
    4:    "Pilot_plant"
    7:    "Biological_treatment"
    10:   "WW_Recovery"
  WWS:
    0:    "TK-961"
    3:    "TK-967"
    8:    "Oily"
Flow_rates:
  0-1:    250
  0-4:    0.41
  3-1:    0
  3-4:    0.09
  8-7:    860.57
  1-2:    233.25
  4-5:    0.39
  4-6:    0
  4-7:    0
  4-9:    0
  7-10:   858.85
  10-9:   858.85
  10-5:   0
  10-6:   0
```

```
Inlet:
  0:
    pH:             "No"
    Conductivity:   "No"
  1:
    pH:             7.748
    Conductivity:   1679.584
  2:
    pH:             7.825
    Conductivity:   2116.276
  3:
    pH:             "No"
    Conductivity:   "No"
  4:
    pH:             8.295
    Conductivity:   8507.505
  5:
    pH:             8.32
    Conductivity:   246.718
  6:
    pH:             0
    Conductivity:   0
  7:
    pH:             8.296
    Conductivity:   1762.627
  8:
    pH:             "No"
    Conductivity:   "No"
  9:
    pH:             8.296
    Conductivity:   1762.627
  10:
    pH:             8.296
    Conductivity:   1762.627
```

```
Outlet:
  0:
    pH:             7.748
    Conductivity:   1679.584
  1:
    pH:             7.825
    Conductivity:   2116.276
  2:
    pH:             "No"
    Conductivity:   "No"
  3:
    pH:             10.786
    Conductivity:   39612.479
  4:
    pH:             8.32
    Conductivity:   246.718
  5:
    pH:             "No"
    Conductivity:   "No"
  6:
    pH:             0
    Conductivity:   0
  7:
    pH:             8.296
    Conductivity:   1762.627
  8:
    pH:             8.296
    Conductivity:   1762.627
  9:
    pH:             "No"
    Conductivity:   "No"
  10:
    pH:             8.296
    Conductivity:   1762.627
```


Figure 8 - Output for instance on CS#6

On the first image, we provide the list of components, as well as, the optimal flowrates derived from the algorithm's solution. 'Inlet' demonstrates the values of the quality parameters when the optimal flowrate(s) enter(s) the component, while 'Outlet' shows the respective values on the exit.

For visualization purposes, we, also, provide an indicative diagram.

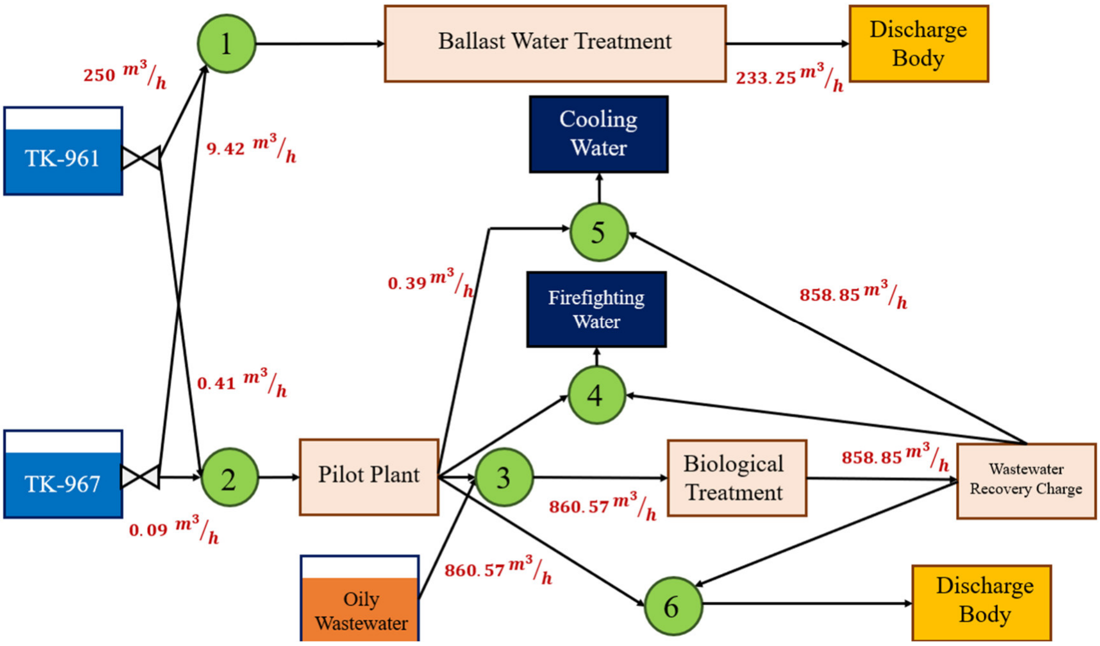

Figure 9 - Diagram representing the solutions on CS#6

The next round of experimentation will focus on the comparison of the AS-IS situation and the TO-BE situation, to provide numerical improvements on KPIs caused within the actions of the AquaSPICE project. These again will be simulated within a laboratory setting, resembling of course realistic settings but without embarking upon everyday actual use of the AquaSPICE by its pilots.

7. Conclusions

This deliverable presents the development of the optimization methodology in the Case Studies of the AquaSPICE project, where an optimization problem can be defined. Optimization needs are addressed from a two-fold perspective: i) the dynamic decision support in terms of providing the values of the flowrates on an online manner and ii) the simulation of different scenarios in order to facilitate long term decisions. Our extensive literature review shows that already published works are focusing only on specific Case Studies with multiple restrictions, without the ability to be transferred to other similar problems. This situation gives floor to the severe challenge of creating an optimization model to be applied on different cases.

To achieve the aforementioned task, it is necessary to create an optimization module, which may solve networks of different structure (horizontal applicability). On the other hand, this module needs to be both efficient (reach near-optimality) and fast (return solutions on a matter of minutes.) We provide the mathematical formulation, which, however appears to be non-linear. As a result, a linear approximation is formed for the horizontal approach, while, also, providing a high-level – in terms of expertise - exact decomposition method, which may be applied on a set of specific cases in terms of structure and objective.

Apart from our methodological contribution, we provide initial experimentation on the Case Studies, with instances comprised of synthetic – yet close to real- data. By this procedure, we are able to validate that our methodological framework is efficient and appropriate for the problem at hand.

8. Next steps

The next steps will start with the extension of our experimentation (this applies, also, to CS#2) and the inclusion of more quality parameters. To achieve this, bilateral discussions with the Case Studies is necessary in order to finalise the possible eligible sets of the input parameters. This experimentation procedure is of high importance, so as to estimate whether future data inputs will be feasible.

An additional step to be followed is the comparison between the two situations encountered in the project (AS-IS and TO-BE). This experimentation will be carried out as follows: firstly, a set of instances will be created – generated as before from synthetic data. Then, we will execute the algorithm twice: i) to solve the AS-IS problem and ii) to solve the TO-BE problem. The output solutions will be compared and provide numerical proof that the work done within the AquaSPICE project has helped the pilot partners to improve on the KPIs set.

Finally, we, also, have the integration of the optimization module developed into the WaterCPS platform. From the optimization perspective, this process demands the definition of the appropriate generic data models and .json structures. As a result, we aim at creating a single component, which may be called by any of the end users and due to its design, no customisation will be needed.

As a result, we briefly mention our tentative approach regarding the 'optEngine'. It is clarified that this approach will be revised and finalized on our final deliverable. The class diagram below depicts the data requirements of optEngine.

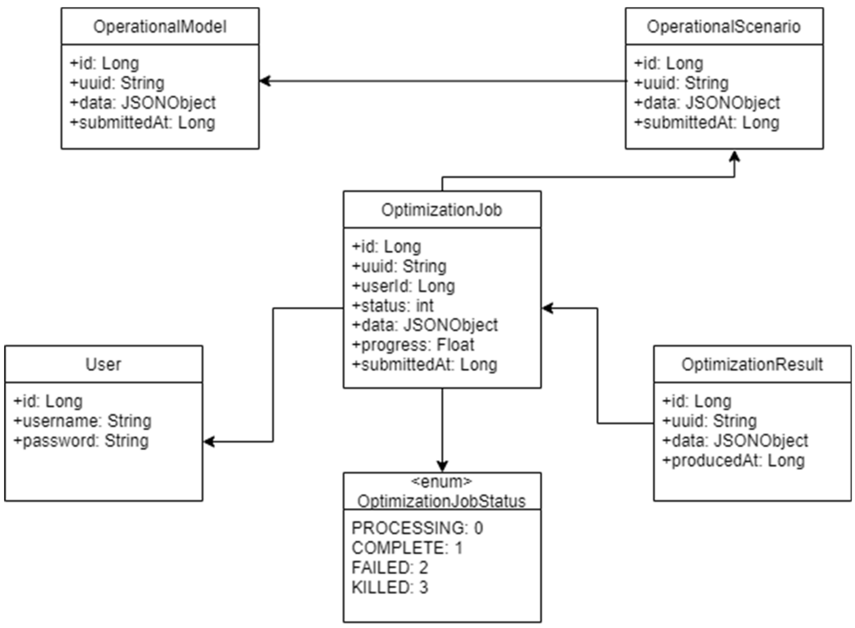

Figure 10 - optEngine Data Requirements

OptEngine works as a shell around the optimization. Its architecture follows an asynchronous approach and is agnostic to optimization-specific data requirements. That is, optEngine receives, stores, and forwards the optimization data to the optimization service requested by the end-user.

Optimization requests along with the respective data are received via a web API. This API allows the actions described in the previous section. The communication with the API requires authentication, is encrypted (https) and asynchronous, i.e., once an optimization job is submitted, the callee does not wait for its completion.

Data stores within optEngine work in a twofold manner:

- Permanent storage via a database (db): this is where optimization requests and the related data are permanently stored or retrieved and updated when necessary.
- Temporal storage via the use of queues: this is where optimization data is stored up to the point where they get consumed by the optimization services that read these queues.

Regarding the optimization data, both the db and the queues are data-agnostic following a general json schema. This allows the storage, permanent and temporal, of different data structures required from different optimization services.

The employed queues allow the asynchronous processing of an optimization job. Additionally, by being durable they ensure that when optEngine or an optimization service fails, the job along with data are available in the respective queue. This means that optEngine, upon reception of a new optimization job, forwards it to the requested optimization service via a queue. Each optimization service listens for a new optimization job to a specific queue and writes status/progress updates to another queue. Last, optEngine listens to (a) a queue for status/progress updates and (b) multiple queues for optimization results.

The flow of data and the architectural approach of optEngine are depicted below.

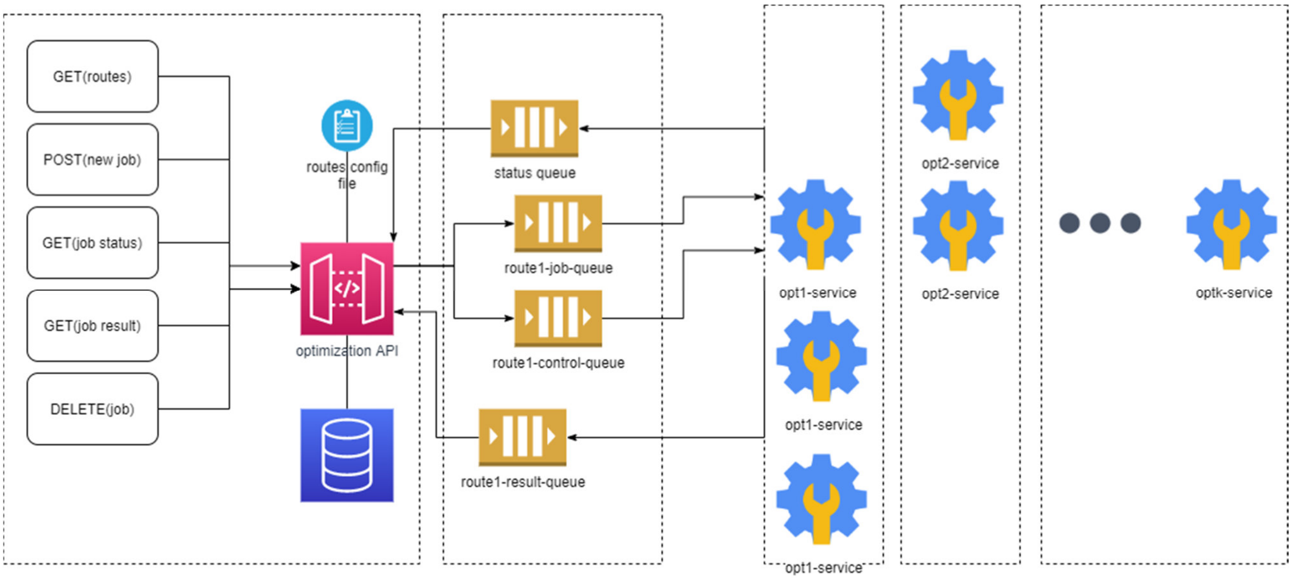

Figure 11 - optEngine data Flow

The technologies used to develop optEngine are the following:

- Java 8
- Spring-boot 2.4.5
- Springdoc openAPI 1.5.2
- Hibernate 1.0.0
- PostgreSQL 11
- RabbitMQ 3.8.16
- Docker 18.09.7

Figure 12 - optEngine data stack

9. References

[1] P. Eirinakis, K. Kalaboukas, S. Lounis, I. Mourtos και J. M. Rozanec, «Enhancing Cognition for Digital Twins,» σε *IEEE International Conference on Engineering, Technology and Innovation*, 2020.

[2] J. Forster, «Eurostat - Statistics Explained,» 2014. [Ηλεκτρονικό]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php/Archive:Water_use_in_industry. [Πρόσβαση 1 10 2023].

[3] J. Izquierdo, I. Montalvo, R. Pérez και V. S. Fuertes, «Design optimization of wastewater collection networks by PSO,» *Computers & Mathematics with Applications,* τόμ. 56, αρ. 3, pp. 777-784, 2008.

[4] C. Puchongkawarin, C. Gomez-Mont, D. Stuckey και C. B., «Optimization-based methodology for the development of wastewater facilities for energy and nutrient recovery,» *Chemosphere,* τόμ. 140, pp. 150-158, 2015.

[5] H. Bozkurt, A. Quaglia, K. V. Gernaey και G. Sin, «A mathematical programming framework for early stage design of wastewater treatment plants,» *Environmental Modelling & Software,* τόμ. 64, pp. 164-176, 2015.

[6] R. Hreiz, N. Roche, B. Benyahia και M. Latifi, «Multi-objective Optimization of Small-size Wastewater Treatment Plants Operation,» σε *12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*, Elsevier, 2015, pp. 2495-2500.

[7] M. Ang, J. Duyag, K. Tee και C. L. Sy, «A Multiple Input Type Optimization Model Integrating Reuse and Disposal Options for a Wastewater Treatment Facility,» *Chemical Engineering Transactions,* τόμ. 70, pp. 199-204, 2018.

[8] B. M. Tosarkani και H. A. Saman, «A robust optimization model for designing a wastewater treatment network under uncertainty: Multi-objective approach,» *Computers & Industrial Engineering,* τόμ. 146, p. 106611, 2020.

[9] G. Niu, X. Li, X. Wan, X. He, Y. Zhao, X. Yi, C. Chen, L. Xujun, G. Ying και M. Huang, «Dynamic optimization of wastewater treatment process based on novel multi-objective ant lion optimization and deep learning algorithm,» *Journal of Cleaner Production,* τόμ. 345, p. 131140, 2022.

[10] M. Al-Zahrani, M. Ammar και C. Shakhawat, «Multi-objective optimization model for water resource management: a case study for Riyadh, Saudi Arabia,» *Environ Dev Sustain,* τόμ. 18, p. 777–798, 2016.

[11] G. Chung, K. Lansey, P. Blowers, P. Brooks, W. Ela, S. Stewart και P. Wilson, «A general water supply planning model: Evaluation of decentralized treatment,» *Environmental Modelling & Software,* τόμ. 23, αρ. 7, pp. 893-905, 2008.

[12] D. Abdulbaki, M. Al-Hindi, A. Yassine και M. Abou Najm, «An optimization model for the allocation of water resources,» *Journal of Cleaner Production,* τόμ. 164, pp. 994-1006, 2017.

[13] M. Boix, L. Montastruc, L. Pibouleau, C. Azzaro-Pantel και S. Domenec, «A multiobjective optimization framework for multicontaminant industrial water network design,» *Journal of Environmental Management,* τόμ. 92, αρ. 7, pp. 1802-1808, 2011.

[14] M. A. Ramos, M. Boix, L. Montastruc και S. Domenech, «Multiobjective Optimization Using Goal Programming for Industrial Water Network Design,» *Industrial \& Engineering Chemistry Research,* τόμ. 53, αρ. 45, p. 17722–17735, 2014.

[15] G. Cassiolato, E. P. Carvalho, J. A. Caballero και M. A. S. S. Ravagnani, «Optimization of water distribution networks using a deterministic approach,» *Engineering Optimization,* τόμ. 53, αρ. 1, pp. 107-124, 2021.

[16] J. M. Ponce-Ortega, A. C. Hortua, M. El-Halwagi και A. Jiménez-Gutiérrez, «A property-based optimization of direct recycle networks and wastewater treatment processes,» *AIChE Journal,* τόμ. 55, αρ. 9, pp. 2329-2344, 2009.

[17] L. Yang, R. Salcedo-Diaz και I. E. Grossmann, «Water Network Optimization with Wastewater Regeneration Models,» *Industrial \& Engineering Chemistry Research,* τόμ. 53, αρ. 45, pp. 17680-17695, 2014.

[18] L. Yang, I. E. Grossmann, M. S. Mauter και R. M. Dilmore, «Investment optimization model for freshwater acquisition and wastewater handling in shale gas production,» *AIChE Journal,* τόμ. 61, αρ. 6, pp. 1770-1782, 2015.

[19] K. Pungthong και K. Siemanond, «MINLP Optimization Model for Water/wastewater Networks with Multiple Contaminants,» σε *12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*, Elsevier, 2015, pp. 1319-1324.

[20] E. Hansen, M. A. S. Rodrigues, M. E. Aragao και P. Monteiro de Aquim, «Water and wastewater minimization in a petrochemical industry,» *Journal of Cleaner Production,* τόμ. 172, pp. 1814-1822, 2018.

[21] C. D’Ambrosio, A. Lodi, S. Wiese και C. Bragalli, «Mathematical programming techniques in water network optimization,» *European Journal of Operational Research,* τόμ. 243, αρ. 3, pp. 774-788, 2015.