water mining
value for society

Deliverable 7.2

WM Dashboard deployments in the
3 demo cases- Part 1

Date: 29 August 2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 869474.

Deliverable 7.2	Project management guidelines
Related Work Package	WP7- Development of ICT tools supporting process monitoring, control & optimization (DATA-MINING), immersive stakeholder engagement (AR-applications) & market creation
Deliverable lead	KWR
Author(s)	Panagiotis Kossieris (ICCS), George Moraitis (ICCS) Nuria Nievas (EURECAT), Hamed Heravi (BRUNEL), Mohamad Zeidan (KWR), Maria Lousada Ferreira (KWR)
Contact	christos.makropoulos@kwrwater.nl
Reviewer	Daniel Dias (BRUNEL); Christos Makropoulos (KWR)
Grant Agreement Number	869474
Instrument	Horizon 2020 Framework Program
Start date	1st September 2020
Duration	48 months
Type of Delivery (R, DEM, DEC, Other)[1]	DEM
Dissemination Level (PU, CO, CI)[2]	**PU**
Date last update	
Website	www.watermining.eu
Name of researcher(s) with roles	Panagiotis Kossieris (Researcher, Nessie); George Moraitis (Researcher, Nessie); Mohamad Zeidan (Researcher, Process modelling); Nuria Nievas (Researcher, Reinforcement learning); Hamed Heravi (Researcher, Alerts and thresholds); Maria Lousada Ferreira (all other sections)

Revision no	Date	Description	Author(s)
D7.2-v01	4-8-2023	Deliverable sent to reviewer	
D7.2-v02	09-08-2023	Review	Daniel Dias
D7.2-v03		Implemented requested revisions	

[1] **R**=Document, report; **DEM**=Demonstrator, pilot, prototype; **DEC**=website, patent fillings, videos, etc.; **OTHER**=other

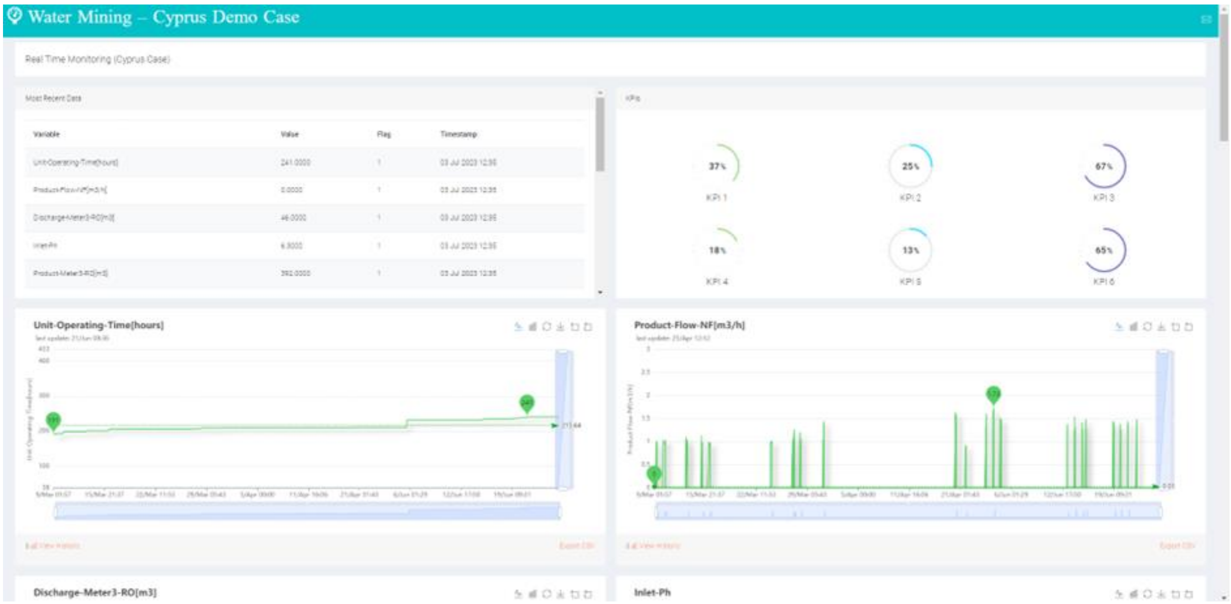
[2] **PU**=Public, **CO**=Confidential, only for members of the consortium (including the Commission Services), **CI**=Classified

Executive summary

Deliverable 7.2 demonstrates the deployment of dashboards to three Case-Studies (CS) of the Water-Mining (WM) project. The development and customization of dashboards requires a close collaboration between dashboard developers and CS owners. Knowledge, challenges, data and developed tools need to be exchanged between the stakeholders in an environment of trust and commitment to innovation. Additionally, to achieve the full potential of a dashboard, combined with advanced analytics, the pilot and demonstration units need to go beyond start-up and reach the optimization stage. The aforementioned conditions are not easily met. WP7 needed to make the strategic decision to change all the CSs initially attributed, which took place at different moments of task 7.2 development. Deliverable 7.1 describes the architecture of the dashboard, without the use of WM CS data, and also mentions the efforts made to reach out to the initially attributed CSs. The present deliverable describes the deployment of dashboards for CS2, CS4 and CS5, as a currently on-going task. Regarding the following amendment to the WM project, Task 7.2 will continue to work on the deployment of the dashboards to the 3 WM CSs, and will provide an additional deliverable, entitled “D7.7- WM dashboard deployment-final version”, due in M44, with the complete work. Lessons learned on way are described in section 6.

Dashboards were designed and deployed for CS2, CS4 and CS5. ICT services and the required infrastructure were defined and implemented on case-to-case basis, and even on a technology basis, when required. Solutions were found based upon on discussions and agreements between Nessie developers, CS owners, and when required, technology developers. The Nessie engine, as the core of the dashboard, consists of a Web Server and a Geolocation-aware Database. Nessie offers a suite of Application Programming Interfaces (APIs) functionalities that enables integration with data sources and analytics, hosted by the project partners. There is a system's token-based authentication to ensure secure data exchange. To facilitate the use and implementation of Nessie’s API, a user-manual has been developed and distributed to the involved partners (Annex II- User manual for Nessie’s API). Figure 11 shows the Nessie dashboard implementation in CS4, for the NTUA technologies, after customization by the CS owners. The dashboards of each CSs are presently at different stages of customization, namely depending on moment the new CSs started collaborating with WP7. The next steps for each CS dashboard’s are clarified in section 7.1.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement 869474.



(Reproduction of) Figure 11-Nessie dashboard implementation in CS4 for the NTUA technologies, displaying a portion the customized dashboard view as defined by the stakeholders.

The dashboards of 3 CSs of WM will be provided with advanced analytics tools. Each tool needs to be designed, developed, validated and discussed with the CS owners, since they will be applied for the use and optimization of the pilot plants. Three types of tools are being developed, namely: reinforcement learning; model-base control with Artificial Intelligence (AI) models; and alerts and thresholds. One example of each advanced analytics has been thoroughly described, per CS, in this deliverable, namely: reinforcement learning for CS2; alerts and thresholds for the Biophree technology (initially at CS4 and currently being deployed at CS5); and model-based AI models for the Partial Nitritation Reactor (PNR) of CS5. The development of the analytical tools will be extended to CSsCS2, CS4 and CS5, upon agreement with the CS owners, and integrated in the final versions of the dashboards.

As a reply to the EU review questions, concerning the novelty of the dashboard provided by the WM project, and the knowledge providing from previous EU projects, an on-going fact-sheet work about the dashboard is provided in Annex III- Fact-sheet on WP7 ICT technologies: the dashboard (@ TRL 4/5).

Table of Contents

List of Figures

List of Tables

Abbreviations

AI- Artificial Intelligence

API- application Programming Interface

CS- Case Study

CSP-Concentrated Solar Plants

DQN- Deep Q-Network

IP- Intellectual property

HMI- Human Machine Interface

HTTP- Hypertext Transfer Protocol

ICT- Information and Communications Technology

KPI- Key Performance Indicator

MED- Multi-Effect Distillation

MDP- Markov Decision Processes

NDA- Non disclosure agreement

NF- Nano-Filtration

OPEX-Operating Expenses

PLC- Programmable Logic Controller

PNR- Partial Nitritation Reactor

PPO- Proximal Policy Optimization

REST- Representational State Transfer

RL- Reinforcement Learning

RO- Reverse Osmosis

URL- Uniform Resource Locator

URN- Uniform Resource Name

WM- WM

WP- Work Package

WWTP- Wastewater Treatment Plant

1. Introduction

Background

WP7 of the Water-Mining (WM) project proposes to develop ICT tools to support process monitoring, control and optimization of data; Augmented Reality applications for stakeholder's engagement and tools for market creation. In particular, T7.2 focuses on the development of customized dashboards and deployment to selected case-studies (CS). A customized dashboard is a real-time process control tool, showing the development of operation over time of the process parameters, with additional advanced analytics providing information and advice on optimization routes to achieve specific goals, namely the overall KPIs defined for each pilot. The integration and use of a dashboard for process control and optimization potentially enables the end-users to save on operational costs, by allowing control of the system at a distance, and achieving quicker optimal performance. Additionally, by using the digital twins of the process system, trials of unexpected events or additional treatment goals can be simulated in advance, either as preparation for stringent discharge requirements or materials characteristics being produced at the plant, or as simulations regarding unexpected climate events. The architecture of the dashboard was described in Deliverable 7.1 (Kossieris et al. 2022). The present deliverable describes and demonstrates the deployment of the dashboards in 3 WM CSs.

The development and customization of a dashboard is a collaboration between developers and CS owners. The task can only be successfully achieved, when CS owners see the benefits of such a tool and actively collaborate with the developers, by investing their own time. Moreover, willingness to share data and knowledge is essential. The pilot or demo-installation, should also be operated and run for a considerable amount of time to reach an optimal stage, so that the benefits of a dashboard can be demonstrated in real-practice. Unfortunately, the aforementioned conditions were not gathered for the initially designated CSs that were to collaborate with WP7, namely CS1, CS3 and CS6. The trials, steps undertaken and lessons learned by WP7 to achieve a collaborative environment were described in Section 4 of Deliverable 7.1- *Stakeholder engagement and user-story collection/preparatory work* (Kossieris et al. 2022). Consequently, WP7 made the strategic decision to reach out to other WM CSs, willing to collaborate and develop customized dashboards. The change to CS4 occurred in an early phase of the project – substituting CS3 (the first CS to demonstrate little to no intention of collaborating with WP7). In September 2022, CS2 and CS5 followed. Joint letters of collaboration agreement were signed between the WP7 leader and CS2 and CS5 owners (Annex I – Jointly-signed collaboration letters between WP7, CS2 and CS5). Currently, Task 7.2 is successfully collaborating with CS2, CS4 and CS5.

Nevertheless, the change of CSs, brought an inevitable delay to complete the task within the stipulated timeframe (T7.2 was planned to last until M36). CS2 and CS5 were integrated into the workplan with a 1.5-year delay. The work was intensified and the usual methodology relying on user-stories collection, (as described and shown in Kossieris et al. (2022)), was substituted by one-to-one interactions with the CS owners and/or stakeholders. The T7.2. partners had multiple one-to-one meetings, either dealing with communication (with the Nessie system and with partners involved), or with advanced analytics, depending on the expertise of each partner. Despite joint efforts, the delay has not been overcome. T7.2 timeline will be extended from M36 to M44. Regarding the upcoming WM amendment, in M44, T7.2 will deliver an updated version of the current deliverable, therefore an additional deliverable "D7.7-WM dashboard deployment-final version", where the complete deployment of the dashboard, with advanced analytics, will be demonstrated.

Document's structure

Section 2 describes the dashboard, as a summary of D7.1 (Kossieris et al. 2022). Sections 3, 4 and 5 describe the implementation of the dashboards in CS2, CS4 and CS5, respectively. The pilot and main processes of each CS are summarized; ICT services and infrastructures are described; the deployment of Nessie is demonstrated; and the on-going advanced analytics are explained. Section 6 focusses on the lessons learned so far, and possible consequences for following projects and/or legal steps. Section 7 provides conclusions and the next steps regarding on-going work. In Annex III, as replied to previous EU review comments, we describe the innovation being exclusively developed in the WM project over previous EU projects. All these sections will be updated in "D7.7-WM dashboard deployment-final version", due in M44.

2. The WM on-line Dashboard

2.1. The Nessie system

The Nessie engine serves as the core of the WM dashboard architecture, designed as a visualization and analytics digital platform able to acquire, process, store, and manage high-resolution data from various sources (3rd party clients) like web applications, sensors, smart meters, etc., and provides an interface for third-party client applications to interact with the system through simple HTTP requests. The Nessie system is enhanced and implemented as a "cloud context broker" that enables seamless integration of information between multiple data sources (such as sensors at demo plants) and external analytics developed by third-party entities. This integration allows for a comprehensive and centralized view of data from different sources, eliminating the need for multiple individual dashboards. Moreover, Nessie offers a range of visualization capabilities through its Data Visualization Layer (see Figure 1), as the engine provides support for most of the common widgets (e.g., line charts, bar charts, pie charts, box plots, tables, gauges, etc.) that can be combined and adjusted into case-tailored dashboards, providing useful information to the users. This is particularly valuable in the real world, which often faces fragmentation in digital solutions. Having a centralized platform for visualization simplifies the user experience, eliminates the need to switch between multiple applications, and promotes collaboration among stakeholders.

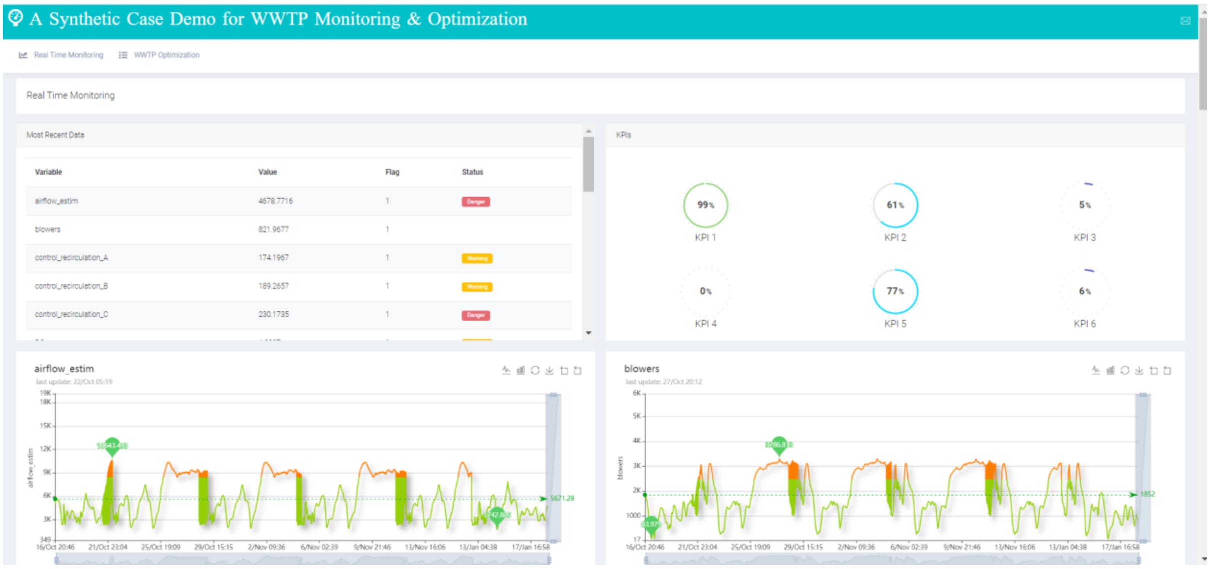


Figure 1- Indicative screenshot of Nessie's dashboard homepage

Within the WM project, an Application Programming Interface (API) was developed for Nessie, using HTTP requests to support and facilitate data exchange between Nessie and other components of the system (demo cases and scientific partners), as presented in Figure 2. In practice, the Nessie Engine consists of a Web Server and a Geolocation-aware Database. Utilizing the Web Server, Nessie offers a suite of APIs functionalities that enable seamless integration with data sources and analytics hosted by project partners.

The functionalities include authentication procedures, data retrieval from Nessie's database, insertion of data in Nessie's database as well as storing/retrieving/updating of outcomes (simulation scenarios) from third-party developers. The Nessie engine is responsible for receiving the data (either temporal, such as time-series data, or non-temporal, like parameter sets generated by AI algorithms) processing and storing them, while also providing a way to link data together (e.g., as in the case of an entity-relationship structure) and retrieve the data in a standard fashion using URNs. Nessie's API supports two models for storing time series data: a hierarchical model based on ontologies (*Entity*, *Device*, *Sensor*, *Variable*, and *SensorData*) and a plain model that assigns unique identifiers (*ObservedData*) to the data. Time series data can be accessed through dedicated API endpoints using HTTP GET requests, with the option to define custom time windows using URL parameters. The API also facilitates the exchange of simulation scenarios data, serving as a broker between clients and target servers. A detailed description of Nessie's API has been provided in the deliverable D7.1 (Kossieris et al. 2022).

Nessie's API functionalities allows developers of analytics to access raw data from demo sites in a straightforward manner, facilitating the retrieval of data without the need for complex procedures, and thus, enabling developers to focus on analytics development and efficiently communicate model outcomes to end-users.

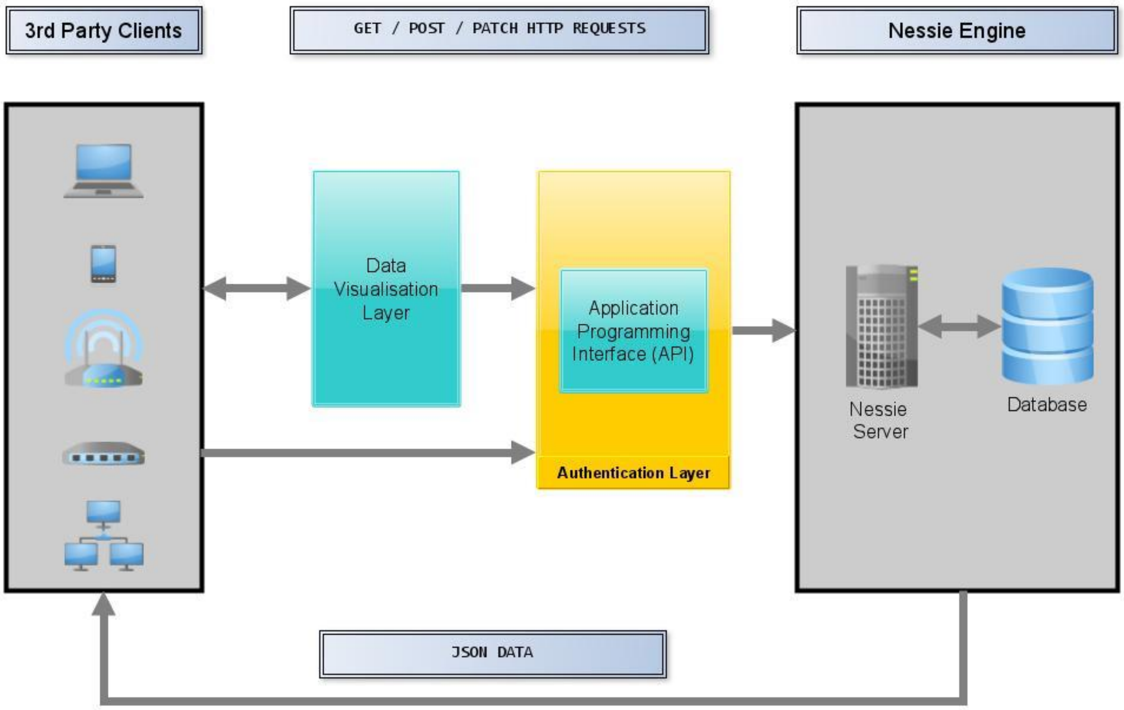


Figure 2- Components of Nessie platform and data exchange with third party clients

To ensure the security of sensitive data from demo cases, Nessie's API uses a customized token-based authentication procedure (see Figure 3). This enhanced security by using dedicated API keys, which is more reliable and secure than the simple password authentication, since it adds an extra layer of security, also known as two-factor authentication (instead of relying solely on usernames and passwords). To put it in simple terms, authentication tokens act like stamped tickets, allowing users to retain access as long as the token remains valid, granting access to applications, services, websites, and APIs, without having to enter their login credentials each time they visit. Instead, the user logs in once, and a unique token is generated and shared with the connected applications or websites to verify their identity. It should be noted that tokens expire after 1 hour of the last access to Nessie's API, adding an extra layer of security.

```
User log in – Token retrieval
Endpoint: 157.90.23.20:8000/service/api-auth/
Request method: POST
Request body: json { "username: "<the actual username>", "api_key": "<the provided api-key string" }
Response body: json { "user": <"user name">, "token":" <token's long string>", "expires_in": <seconds remaing> }
```

Figure 3- Nessie's authentication procedure

Overall, the Nessie system offers valuable capabilities in respect to real-time monitoring, control, and optimization of processes by integrating different metering systems and analytics into a unified digital solution. By providing a scalable and extendable architecture, Nessie enables seamless data integration, enhances collaboration among partners, and offers end-users access to a wide range of analytics through a single, case-tailored dashboard. The system's token-based authentication ensures secure data exchange, making Nessie a valuable tool for efficient water management decision-making[3].

To facilitate the use and implementation of Nessie's API, a user-manual has been developed and distributed to the involved partners. The manual comprises all the necessary steps and POST-type requests to perform the different tasks at hand. The manual is given in "Annex II- User manual for Nessie's API". Giving special attention to the protection of sensitive data, authentication credentials have been provided only to scientific partners involved, after direct communication with Nessie's developers (ICCS).

2.2. Optimization and Reinforcement Learning

Water management services with the extraction of renewable resources involve decision-making, parameter setting, and process control. Understanding their temporal evolution and process functioning is crucial for informed decision-making, especially considering potential stochasticity. There are various approaches in the literature for water treatment control, including conventional control such as proportional-integral-derivative control or model predictive control, and more intelligent control such as fuzzy and machine learning models.

Reinforcement Learning (RL), in the machine learning field, is utilized to dynamically adjust control techniques and provide optimal solutions in real-time. RL involves the interaction between a learning model and the environment, progressively improving decision-making through trial and error. The learning model selects actions, receives feedback in the form of reward signals, and continuously enhances its decision-making strategy, known as the policy. The primary objective of RL is to maximize a numerical reward signal, which can represent process Key Performance Indicators (KPIs). To achieve this, the model learns appropriate actions based on the given state, which involves acquiring a function that maps states to corresponding actions.

A Markov Decision Process (MDP) provides a mathematically defined representation of sequential decision-making scenarios, where actions influence both immediate and future rewards. MDP models

[3] More detailed information regarding the Nessie system, the API developments and functionalities and its progress within the WM project can be found in D7.1 "The WM Dashboard" (Kossieris, P., N. Nievas, M. Korevaar, A. Mousavi, M. Lousada-Ferreira and C. Makropoulos (2022). Deliverable 7.1-The Water-MIning Dashboard. Water-Mining-Next Generation Smart Water Management Systems- H2020, Grant agreemet n. 869474.).

discrete-time processes, where state changes result from both taken actions and uncontrolled environmental changes. In simpler terms, an MDP comprises five components (Sutton et al. 2018): a set of states and actions (S, A), a reward function R, a transition probability distribution determining state transitions, and a discount factor γ (between 0 and 1). The discount factor balances the significance of immediate versus future rewards, with $\gamma = 0$ considering only immediate rewards and $\gamma = 1$ assigning equal importance to future rewards.

The adaptability and immediacy of RL hold great promise for intelligent, dynamic, and optimal process control as a decision support system. An illustration of this iterative learning process is provided below in Figure 4.

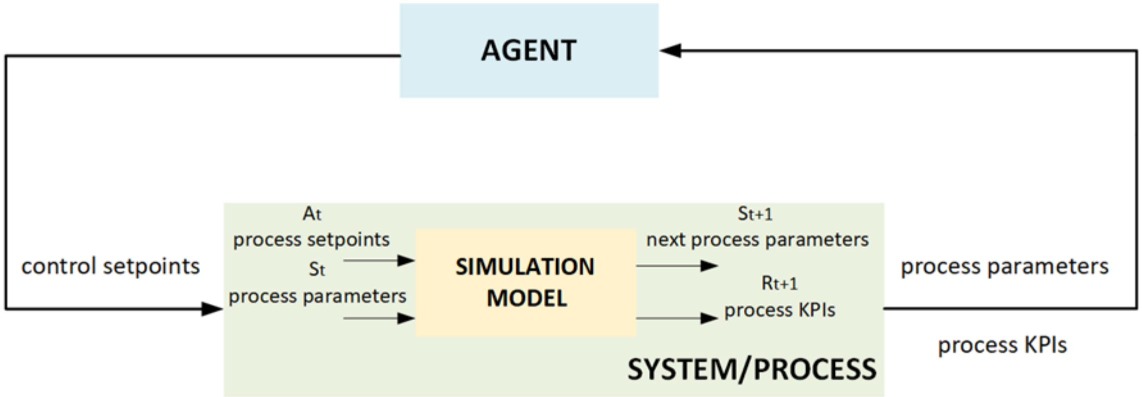


Figure 4- Illustration of the Iterative Reinforcement Learning Process.

Model-free RL algorithms directly learn to estimate the value function or policy by utilizing observed state-action pairs (Sutton et al. 2018). These models can be classified into three primary groups: value-based, policy-based, and actor-critic algorithms. Value-based approaches focus on estimating the value function, which predicts the cumulative reward associated with taking actions in specific conditions and following the optimal course of action. Policy-based approaches, on the other hand, learn a parameterized policy that directly maps states to actions. Actor-critic algorithms combine both value-based and policy-based approaches, leveraging two interacting processes to enhance learning speed, stability, and performance.

The selection of an algorithm for a given task is influenced by various variables, such as the characteristics of the environment, dimensionality of the problem, and control objectives.

In discrete control task, the conventional Q-Learning control algorithm (Watkins et al. 1992) learns a Q-value for each state-action pair. The Q-value represents the projected future cumulative reward that can be achieved by taking actions from a specific state. The Q-table serves as a repository for storing all the Q-values. However, like other tabular approaches, this method suffers from the curse of dimensionality (Mnih et al. 2015). To address these problems in real-world scenarios, Deep Q-Learning (DQN) approaches introduce neural networks (NN) for approximating nonlinear value functions. NNs offer scalable solutions for handling high-dimensional problems.

In continuous control tasks, actor-critic algorithms have proven to be the most effective and reliable model-free RL methods. More recently, the Proximal Policy Optimization (PPO) algorithm has gained popularity in literature. It consistently achieves state-of-the-art performance on benchmark tasks and finds practical applications in real-world scenarios.

2.3. Digital twin and model-based control with AI models

The pilots from the case-studies from the WM project are composed by a sequence of processes and/or operations. A process can be represented by a model, which can be mechanistic, (semi-)empirical or an Artificial Intelligence (AI) model.

Digital twins can simulate single properties, single processes and combinations of processes. Potentially, digital twins can represent complete treatment trains, i.e. complete pilots with the several processes and/or operations. A digital twin of a complete pilot system is particularly useful when a part of the pilot operates in a continuous mode and the remaining in batch mode. Output of the continuous mode processes, can be feed into the discontinuous mode processes, also designated as batch mode, providing guidance on the expected performance results, independently from operational modes (continuous vs discontinuous/batch) or possible physical distance between different processes steps. Nevertheless, digital twins of several processes in a treatment train, even when all share the same operational mode, will always enable a combined control optimization, aiming for a certain goal.

A process model can be combined with sensor data, achieving what is considered the digital representation of the physical process, also designated as digital twin. As the model-based control, the digital twin has the advantage to go beyond real practice. A digital twin can simulate a property, which is difficult to measure, but possible to calculate through a series of equations combining data from different sensors. A digital twin can also be used to simulate data ahead of time, aiming for a certain performance, therefore providing info about the most suitable operational settings to achieve a certain goal.

The aim of the model-based control is to produce an optimal set point. Optimal operational set-points vary according to the water/wastewater quality and quantity to be treated, external environmental conditions and the operational goal to be achieved. Usually, in a pilot, such as the pilots of the WM CSs, optimal set-points are obtained through practical experiences. With different operational conditions and operational goals, multiple settings are tested and the optimal settings selected. However, the experimental approach is dependent on the variety of operational conditions imposed by real practice. For instance, if a pilot is operated only during dry weather conditions, it won't be able to obtain the settings to operate during wet weather conditions. Therefore, optimal set-points based only on experimental data are limited to the conditions imposed by real practice. The advantage of a model-based control is to go beyond real practice. If a process, or a part of a process is represented by an equation, simulations based on the model can provide guidance for extraordinary events, not necessarily experienced in real practice. Furthermore, if the operational goal of the pilot change, model-based control simulations can provide indications on how to achieve a certain performance. For instance, if the goal of a pilot is to consume less energy instead of producing more water, a model-based simulation can provide guidance on the optimal settings to achieve the new goal.

When the process model is complex, i.e. either composed by several differential equations or requiring calculations of virtual sensors/parameters to feed the process equations, it becomes difficult to implement in the data collection system of a pilot. The data collection system, combined with the local PLC, ensures the normal operation of the pilot. In these cases, platforms such as Nessie, assure that the pilot data can be feed into more complex models, physically placed somewhere else.

2.4. Operational optimization regarding KPI (water consumption)

In the water treatment sector, optimizing the operation with respect to Key Performance Indicators (KPIs) plays a crucial role in ensuring efficient and sustainable water management. KPIs are essential metrics that provide valuable insights into the performance of water treatment processes and the overall system. These indicators may include parameters such as water quality, energy consumption, chemical usage, and equipment reliability.

By systematically analyzing and interpreting KPI data, water treatment plant operators and managers can identify potential areas for improvement and make informed decisions to enhance operational efficiency. Utilizing advanced monitoring and control systems, real-time data can be collected and processed to facilitate predictive maintenance and early detection of anomalies, leading to reduced downtime and operational costs.

Moreover, by aligning operational strategies with KPI objectives, water treatment facilities can achieve better compliance with regulatory standards and environmental guidelines. This approach also fosters a proactive approach to resource management, reducing wastage and supporting sustainable practices.

3. The WM Dashboard for “Plataforma Solar de Almería” (PSA) demo case

3.1. Demo case description

The demonstration pilot at the Solar Platform of Almería, Tabernas, Spain, is a desalination plant from sea water, powered by thermal energy, and operated by CIEMAT. The demo-plant at Almeria aims for the following goals: zero-liquid-discharge; record-breaking energy consumption in thermal desalination, i.e. below 25 kWhth/m^3; low OPEX costs; production of water for irrigation and production of high purity NaCl salts. The schematics of the demo-plant is presented in Figure 5.

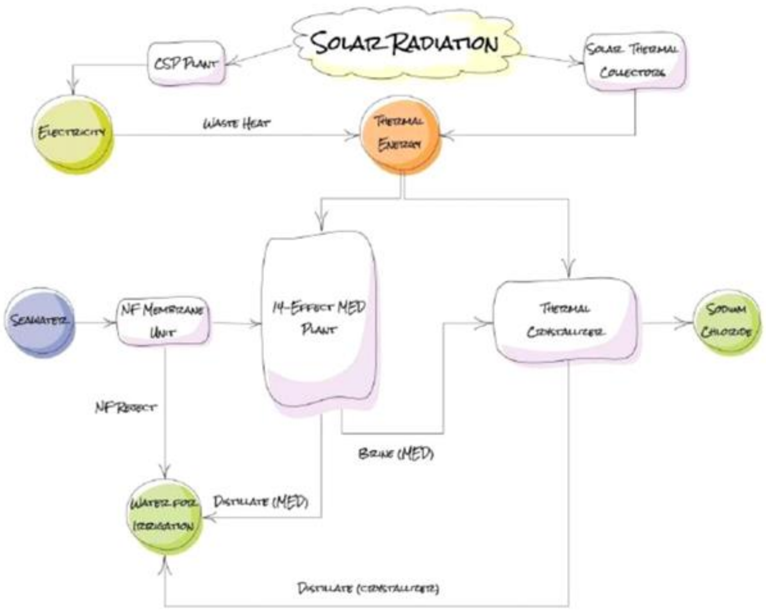


Figure 5- Schematics of the demo-plant at the solar platform of Alméria, Spain (CS2). (Source: Xevgenos (2022))

The main technology of the demo-plant is a Multi-Effect Distillation (MED) system, with 14 effects. The demo plant produces 3 m^3/h of demineralized water, with a maximum operating temperature of 70 °C, using a thermal energy of 190 kWth. The MED plant integrates a co-generation system, where thermal energy can be provided either by a solar field with 60 collectors or by waste heat of Concentrated Solar Plants (CSP). To avoid interference on the efficiency of the power cycle, the co-generation system integrates CSP configurations with high temperature power cycles, producing enough waste heat to power the demo- plant.

The sea water being feed into the system is pre-treated with a Nanofiltration (NF) step to remove divalent ions, such as Mg^{2+}, Ca^{2+} and SO_4^{2-}. Therefore, the resulting permeate rich in sodium chloride (NaCl), increases the recovery of the MED step and the operating temperature, usually limited to 70°C to avoid scaling. The concentrated brine produced in the MED, free of divalent ions, is further treated in a solar-powered crystallization step, aiming for high-purity NaCl salts. The concentrated stream (brine) produced by the NF step, is used to re-mineralize the distilled water, produced both by the MED and thermal crystallizer.

3.2. ICT services and infrastructures

The whole procedure to establish communication between the demo case infrastructure and Nessie comprises 4 steps:

Step 1: Getting authenticated.

Given a username and password Nessie receives the necessary access token to be used for validating the next HTTP requests.

Step 2: Receiving variables.

Using the previously retrieved access token, Nessie acquires a dictionary of the variables exported by the system. The values of the dictionary include the name of the variable, the description, the unit of the variable and a unique id for the variable. Nessie stores the variable data as necessary in its database, creating new variables or updating existing ones.

Step 3: Retrieving data.

For each variable retrieved in the previous step, Nessie requests and retrieves the average of operation data for the last ten minutes of plant operation. In case the plan is off-line an HTTP response error (400) is sent back to Nessie, which indicates that the treatment unit is not operating. If data is available, Nessie parses the dataset, and inserts it in its database, accordingly.

Step 4: Parsing and inserting the data.

Nessie creates a dictionary of all the variables and their corresponding data, it then inserts data in the database in a very fast and efficient way (since it is optimized for real-time data processing by design). In case timestamped data arrive in the database with overlapping timestamps, Nessie by default updates overlapping data with the newest available values.

3.3. Deployment of Nessie

The real-time monitoring of Plataforma Solar de Almería (PSA) demo site is available, with the use of the proper credentials. This instance monitors a total of 13 variables associated with the CS processes and defined by the stakeholders. The sampling frequency of received measurements varies from 10 to 30 minutes, depending on the customisation of sensors at the plant. Those variables and their measuring units can be seen in Table 1.

The deployment of the Nessie system commenced with an initial version of the dashboard, created on the basis of past historical data provided by CS2 stakeholders. The data were used to identify relevant variables and establish a baseline for data visualization. Following the deployment of the first version, a series of technical discussions were conducted to refine the dashboard's design and functionality in accordance with the CS needs. Additionally, discussions centered around the graphical arrangement of the variables to provide a clear and intuitive display for users. To ensure consistency and accurate data representation, the stakeholders and the Nessie team deliberated on the time zone used within the dashboard. This consideration aimed to align the displayed timestamps with the stakeholders' operational context and facilitate easy interpretation of the data. Lastly, and based on the specific requirements of the case study stakeholders, it was determined that a tabular display of the latest data would be more practical, while secondary information such as "flags" and "status" information for each device were deemed of low practicality for the Nessie use-case intended. As a result, the Nessie team customized the tabular data display to present the most recent data in a concise and streamlined format, enabling stakeholders to access relevant information quickly, as seen in Figure 6.

Table 1-List of variables monitored in real-time through NESSIE system for CS2- Plataforma Solar de Almería (PSA) demo case.

Nr.	Variable	Units
1	Ms (hot water flow rate)	L/s
2	Msw (cooling water flow rate)	m^3/h
3	Mf (feedwater flow rate)	m^3/h
4	Mprod (Distillate flow rate)	m^3/h
5	Ts_out (Outlet hot water temperature)	ºC
6	Pv_1 (Vapour pressure at first effect)	mbar
7	Tvc (Condenser vapour temperature)	mbar
8	Tcwout (Cooling water outlet temperature at condenser)	ºC
9	Tf (Preheated feedwater temperature)	ºC
10	Tcwin (Cooling water inlet temperature at condenser)	ºC
11	Tprod (Condensate temperature)	ºC
12	Tbrine (Brine temperature at outlet of last effect)	ºC
13	Ts_in (Inlet hot water temperature)	ºC

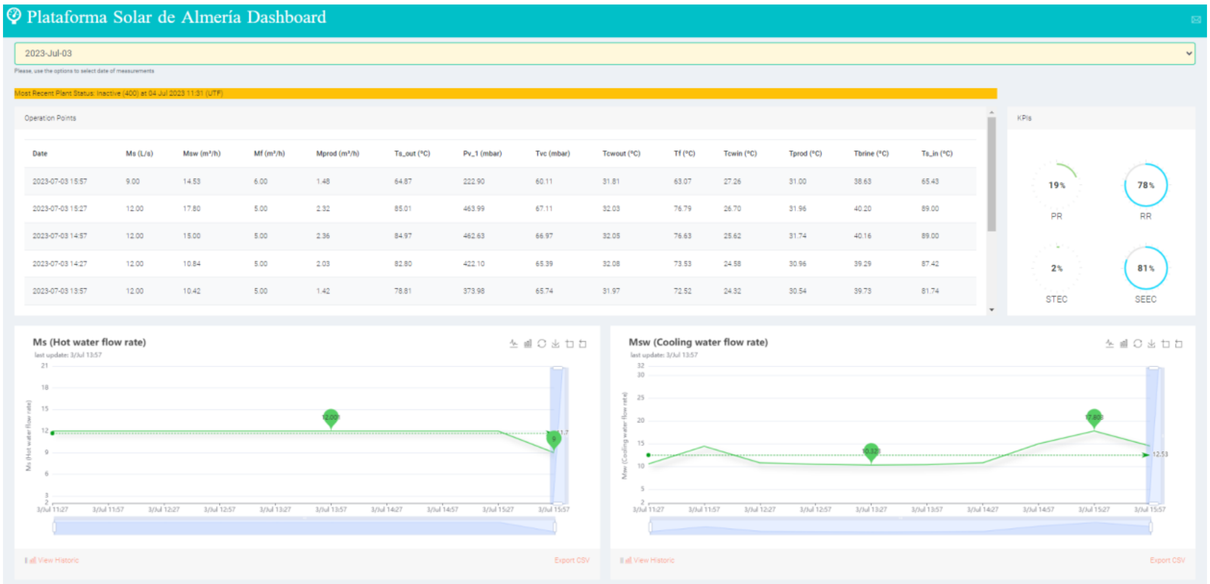


Figure 6- Nessie system implementation in CS2, displaying a portion the customized dashboard view as defined by the stakeholders.

One of the case-specific requirements and challenges in the CS2 implementation is its intermittent operational schedule and batch mode data communication. This intermittent operation pattern presented unique challenges in terms of data collection, analysis, and visualization, necessitating tailored solutions to address these requirements. To facilitate effective data monitoring and analysis, the Nessie system was customized to automatically classify and assort relevant data received during each "experiment day". This feature simplifies the process of identifying and retrieving data specific to a particular day, streamlining the analysis workflow.

The automatic data classification feature provided by the Nessie system offers several benefits to the end-users. Firstly, it allows to perform focused analysis of data at specific days where experiments took place, enabling a detailed understanding of system performance during those periods. This capability is useful for identifying patterns, trends, and anomalies associated with different operational conditions and activities. Additionally, this feature enhances data traceability and reproducibility, as users can effortlessly access and compare data sets from multiple experiment days, aiding in research, troubleshooting, and performance optimization. By automatically categorizing data based on the experiments conducted on those days, users can easily access and analyze the data sets they require, minimizing the time and effort involved in manual data sorting.

The visual layer of the Nessie system was accordingly customized to enable users to visualize and monitor data corresponding to specific days where experiments took place ("experiment days") within the operational cycle, selected through the relevant menu located on the top of the screen, as seen in Figure 7.

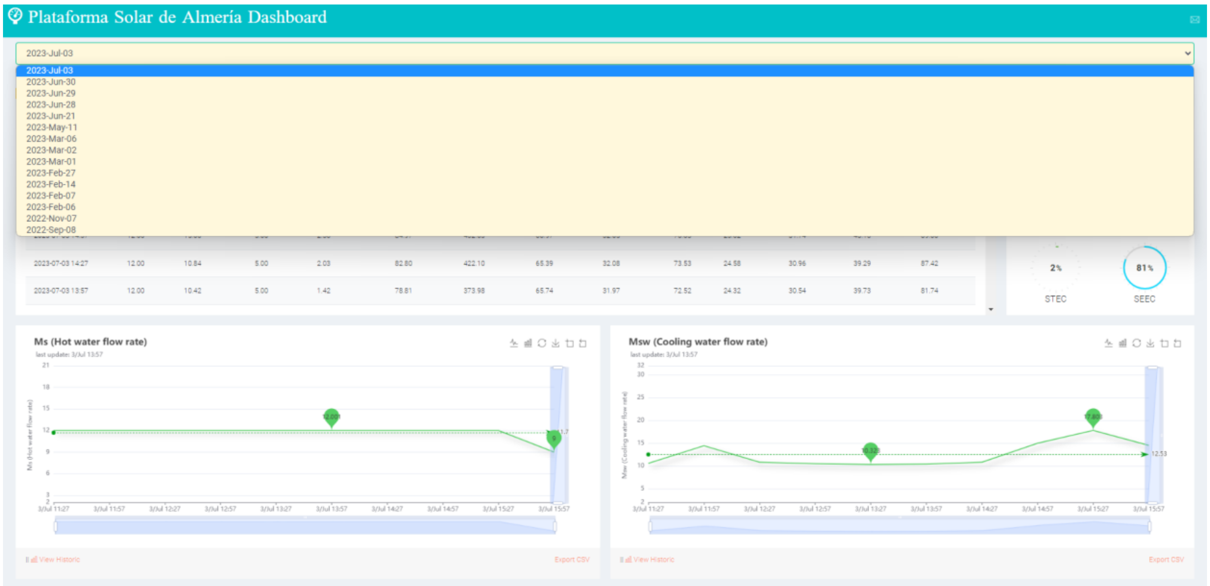


Figure 7-Example of selecting a specific experimental day of CS2 demo site through Nessie.

For the authentication step, the team provided a unique set of credentials, authenticated through a POST request. The generated token is included in every request, to ensure only authenticated clients access, retrieve or store data on the dashboard. A quick guide with step-by-step instructions for the a) authentication, b) sensor/device data retrieval and c) simulation scenario data storing, and retrieval was developed and shared among partners to facilitate the use of Nessie capabilities.

3.4. Advanced Analytics

Regarding the advanced analytics being developed for CS2, the most advanced is the reinforcement learning analytics, which are being developed by EURECAT. KWR is currently exploring with CS2 the possibility to apply AI data validation, as an additional tool to the current data validation techniques.

3.4.1. Reinforcement learning

The system described in CS2 for intelligent control in the renewable energy desalination process is a complex system comprised of multiple subsystems (Figure 8). Specifically, the Multi-Effect Distillation (MED) plant is supplied with low-temperature solar heat obtained from stationary flat plate collectors, utilizing solar thermal energy.

Typically, these plants are designed to operate under optimal conditions, ensuring maximum performance. However, when operating outside of these design parameters, heat exchangers may not function optimally due to variations in the operating conditions (temperature difference, flow rates, etc.).

Furthermore, the system optimization involves incorporating the solar field, primarily functioning as an electrical-to-thermal energy converter, and thermal storage. The thermal storage capacity varies depending on the temperature availability of thermal energy, which, in turn, relies on external factors such as solar radiation and ambient temperature. In this process, thermal energy is utilized to heat the circulating water within the MED. A heat exchanger transfers the heat from the water flow originating from the solar field to the water flow within the tanks (thermal storage).

The system employs a 2-tank configuration to ensure stratification. A tank operates at a higher temperature (red tank in Figure 8), receiving heat from the energy source and transferring it for consumption. The heat energy is consumed in the sense that it's utilized to power the distillation process in the MED system. The discharged water from the MED returns to the bottom of the cold tank (blue tank in Figure 8) before flowing into the hot tank, absorbing heat from the source. Both the water leaving the thermal storage and the water returning from the MED pass through a three-way valve, ensuring a controlled input temperature to the MED.

The objective is to efficiently manage the available energy within the system, starting with a fluid at a given temperature, in order to maximize the overall distillation output. This scenario presents a complex situation where deviating from the nominal conditions may prove to be a more suitable strategy to adapt to the variability of the combined system.

This complex process is illustrated in Figure 8.

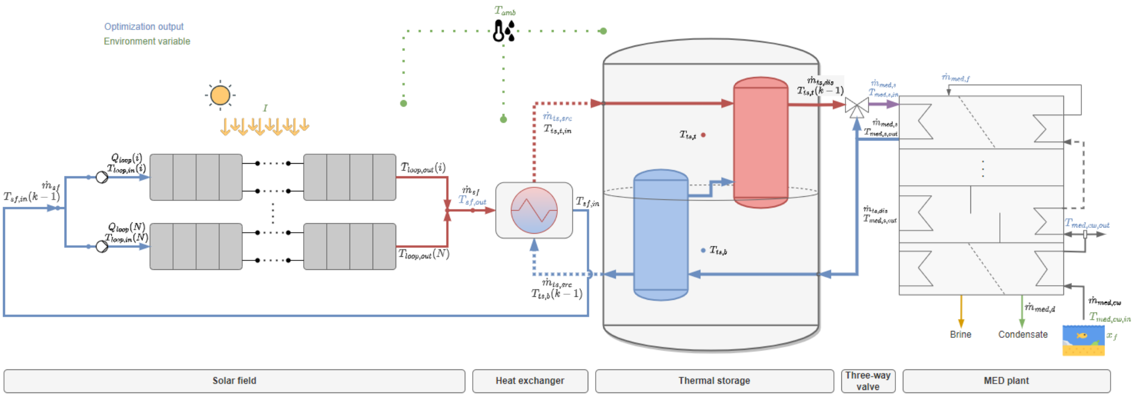


Figure 8- Solar-MED optimization general diagram

RL is proposed to control this system, aiming to achieve dynamic and real-time adaptation to all non-controllable variables, including solar radiation, ambient temperature, seawater temperature, and salinity. These data are acquired from both historical records and sensors, facilitating the training of the model to closely align with the operational reality of the plant. While water temperature and salinity may not vary significantly, effectively adapting to available energy through efficient control poses a challenge in the system.

3.4.1.1. Control problem modelling

To effectively model the system, several components need to be defined: the state space, the action space, and the reward function of the MDP. Additionally, a simulation environment encompassing the entire system is required to facilitate the training of the control models. This simulation environment should provide the system's state at each time step and predict the subsequent state based on the executed action.

The state of the system encompasses all relevant variables that contribute to decision-making including external factors to the system, such as the uncontrollable variables. It reflects the current situation of the system. On the other hand, the action is determined by selecting appropriate values for the control variables in response to the provided state. Lastly, the simulation environment is responsible for predicting the next state and evaluating the quality of the action taken in that particular situation, providing a measure of reward. This reward guides the learning process of the control model, aligning it with the defined objective in the case study.

The description of the different elements of the MDP is provided below.

Simulation environment

The system consists of five distinct models, which are also depicted in Figure 8. These models are interconnected in a way that conforms a new model of the combined system. Each individual subsystem model is described below.

Solar Field model: this model utilizes the inlet and outlet temperatures of the water flow, solar radiation, and ambient temperature to predict the flow rate and the conversion factor from electrical to thermal energy. The water flow from the outlet temperature circulates back to the solar field in a closed loop.

Heat Exchanger model (solar field – thermal storage): this model handles the flow of water from both the solar field and thermal storage and transfers heat between these flows. It predicts the temperatures of the water coming out of the heat exchanger for both the solar field and thermal storage.

Thermal Storage model: this model considers the temperature and inlet flow to the hot tank, the temperature and inlet flow to the cold tank, and the ambient temperature. The model outputs the temperature profile in the thermal storage, which is then used to evaluate the three-way valve and heat exchanger models, respectively.

Three-Way Valve model: the three-way valve model allows to determine the corresponding flow rate leaving the thermal storage, considering the temperature calculated by the thermal storage model, the desired water temperature entering the MED system and the temperature of the flow leaving the MED system. A portion of this flow, which has a lower temperature, is mixed with the flow coming from the storage until the desired temperature for re-entering the MED system is achieved. The remaining water from the MED system returns to the thermal storage, completing the flow heating cycle.

MED model: the MED model takes in various inputs, such as the temperature and flow rate of hot water, the flow rate and salinity of the feed water, and the temperatures of the cooling water going into and out of the condenser. Using this information, it predicts the amount of distilled water produced, the temperature of the hot water coming out, and the flow rate of the cooling water.

<u>**Action space**</u>

The control variables for the system are as follows:

- Outlet temperature of the flow from the solar field;
- Flow rate from the thermal storage to the secondary circuit of the solar field heat exchanger;
- Feed water flow rate to the MED;
- Hot water flow rate and inlet temperature in the MED;
- Cooling water temperature at the outlet of the MED condenser.

Considering the finite energy available in the system, these variables will have certain limitations or restrictions.

<u>**State space**</u>

The state variables for the decision-making system are as follows:

Independent variables

- Solar radiation;
- Ambient temperature;
- Cooling seawater temperature and salinity at the inlet of the MED condenser.

Dependent variables

- Inlet temperature of the flow from the solar field;
- Flow rate circulating through the solar field;
- Temperatures of the thermal storage, both at the inlet and outlet points as well as in the upper and bottom halves of the tanks;
- Flow rate of heat source from the thermal storage to the valve;
- Hot water outlet temperature from the MED.

<u>**Reward function**</u>

The reward function for training the control models has been developed by considering the economic benefits associated with operating the system. Specifically, the average benefit of distillate production flow rate, measured in euros per cubic meter per hour, is calculated by factoring in the price of water, operational costs, and fixed costs.

Operational costs (C_{op}) are determined based on the cost of electrical energy consumed for both electrical and thermal energy needs (Equation 1), taking into account the conversion factor of the solar field. Ideally, a conversion factor closer to zero signifies lower costs associated with the heat required by the MED process. Conversely, a higher conversion factor indicates increased expenses for thermal consumption in the MED, prompting the optimizer to prioritize enhancing thermal performance and minimizing consumption.

$$C_{op}\left(\frac{€}{m^3 \cdot h}\right) = C_e\left[\frac{€}{kWh_e}\right] \cdot \left(SEEC_{MED}\left[\frac{kW_e}{m^3}\right] + STEC_{MED}\left[\frac{kW_{th}}{m^3}\right] \cdot SEC_{SF}\left[\frac{kW_e}{kW_{th}}\right]\right)$$

Equation 1- Operational costs for the Reward function in the control problem modelling.

The operational costs are further explained below, with the variable *Ce* representing the cost of the electrical energy, which may vary based on the time, $SEEC_{MED}$ representing electrical energy consumption, $STEC_{MED}$ representing thermal energy consumption, and SEC_{SF} denoting the conversion factor from electrical to thermal energy.

If benefits fall below a specific threshold, it may indicate the necessity to halt operations. In such circumstances, it becomes crucial to consider the process' fixed costs, as they may have a substantial impact on the analysis of the process' overall profitability.

3.4.1.2. Control algorithms

The algorithms chosen for implementation in CS2, which are among the most used RL algorithms in the literature, are presented as follows. These two algorithms will be tested to compare their performance, ultimately determining the more appropriate choice for the specific use case.

Deep Q-Network (DQN)

DQN (Mnih et al. 2015) is a value-based RL method that approximates the optimal action-value function by employing a NN, known as the Q-function. This approach addresses the dimensionality limitations often encountered in traditional Q-Learning, commonly known as the 'curse of dimensionality.' It achieves this by obviating the requirement for a tabular format that encapsulates all value information within a NN. This NN takes the state as input and generates a set of Q-values that correspond to the potential actions available. To learn the Q-function, DQN iteratively adjusts the parameters of the NN, minimizing the mean squared error between the predicted and target Q-values.

DQN utilizes a separate target network, which is a duplicate of the NN used to calculate the target Q-values. This target network aids in stabilizing the learning process by decoupling the Q-value updates from the value estimations. By employing a different target network, the correlation between the Q-values and the target values is reduced, leading to improved learning stability.

Additionally, to mitigate the potential instability associated with using NNs for RL, DQN incorporates several strategies. One such strategy is experience replay and prioritized experience replay (Schaul et al. 2016), which involves storing the collected experiences in a replay buffer. This buffer is then used to sample experiences for learning, thereby reducing the correlation between consecutive samples over time.

Proximal Policy Optimization (PPO)

PPO (Schulman et al. 2017) is an actor-critic algorithm designed to maximize the learned policy while minimizing disruptive changes that hinder learning. It is similar to Trust Region Policy Optimization (TRPO) in that it uses a trust region to prevent large policy updates, but it simplifies the objective function by using a clipped surrogate instead of a KL divergence penalty. This makes PPO easier to implement and generally more efficient than TRPO. Unlike other actor-critic algorithms, TRPO considers a regularization term when updating the policy function, which introduces a constraint to minimize the changes from the previous policy. This constraint is important for preventing the algorithm from making drastic changes to the policy that could lead to divergences or overly negative outcomes.

PPO incorporates multiple epochs of stochastic gradient ascent in each policy update, leading to substantial enhancements in the quality and stability of the learned policies. Through iterative updates of policy parameters and the application of stochastic gradient ascent, PPO progressively refines the policy, ensuring that the modifications are proximal and avoiding substantial changes that could disrupt the learning process. By striking a balance between exploration and exploitation, PPO generates more robust and reliable policy optimization.

The choice of algorithm for a specific task depends on various factors, such as the environment's properties, available resources, and agent objectives. Actor-critic algorithms, including TRPO and PPO, are the most efficient and robust model-free RL algorithms for continuous control tasks. Recently, PPO has gained popularity due to its efficiency and simplicity of implementation, achieving state-of-the-art performance on benchmark tasks and applied in real-world scenarios.

3.4.1.3. Control benchmark

The primary objectives of the control benchmark task are as follows:

- **Evaluation**: Assess whether the trained learning models have acquired the ability to make decisions in line with the process objectives and relevant KPIs by analyzing their policy improvements during the training phases;
- **Baseline comparison**: Compare the behavior of trained learning models to the business-as-usual behavior. This comparison will determine if the RL agents can enhance the current outcomes;
- **Selection**: Compare DQN and PPO solutions to identify the most effective one(s).

To accomplish these goals, we propose automating the generation of a report. The report, generated as an HTML document with tables and graphics, will include:

- **Description of control model training**: A concise overview of the benchmark models, including information on training parameters and plots showcasing the training progress;
- **Evaluation episodes**: Selection of diverse and noteworthy periods for evaluating the models such as days abundant in solar radiation as well as cloudy instances. This approach aims to observe how the control agent adjusts to different scenarios;
- **Plots**: Graphical representation of the reward variables' evolution per evaluation episode;
- **Model comparison**: The comparison involves different RL models and their performance in relation to the business-as-usual behavior. The comparison is based on the reward function variables. To assess the performance of each model, the information from the different evaluation episodes will be aggregated into final metrics. These metrics will summarize the models' performance and serve as the basis for ranking their quality.

3.4.1.4. Real-time Adaptive Control

Once the control model has been selected for the specified system, it will be integrated with Nessie. Nessie will provide real-time sensor data, serving to define the state variables of the MDP. By leveraging its interactions with the simulation environment and employing the learned policy derived from trial and error, the control model will determine the values of the control variables. Subsequently, Nessie will present the user with real-time feedback, illustrating the response of the control models. While the dynamic and adaptive control mechanism is effective at adapting to uncontrollable environmental conditions that the system was trained for, it's important to consider its applicability to new and extreme events. In the context of these events and the potential scalability to larger systems, the control mechanism might indeed require retraining to ensure optimal performance and adaptation. This could also entail adjustments to the definitions of the state and action space, particularly if new variables are introduced into the system.

3.4.1.5. Reinforcement learning- next steps

We have already defined the specific use case that requires control, achieved through the identification of the most relevant state and action variables within the system. Additionally, we have successfully developed the training code for both DQN and PPO control models and generated benchmark reports to facilitate the comparison and selection of the optimal solution. Our current priority is focused on defining the configuration files for the trainings, while we eagerly await the full environment model from CS2. This model will allow us to accurately simulate the real system, enabling us to establish an effective training environment. Moving forward in this case study, we will execute the trainings, and select the best hyperparameters to control the process. Subsequently, we will evaluate and compare the solutions proposed (DQN and PPO algorithms) to determine the most suitable one for the studied system. The final step will involve integrating the selected model with the Nessie platform, allowing the control system to provide recommendations to end-users when needed. Table 2 provides a summary of the current status of the main tasks to be developed for the control system.

Table 2- Summary of task accomplishment for control system development

Task description	Status
Definition of the problem and description of control variables	Completed
Development of training code for DQN and PPO control models	Completed
Generation of benchmark reports for solution comparison	Completed
Full environment model for accurate system simulation	In progress
Definition of configuration files for trainings	In progress
Execution of trainings	Not started
Selection of the best hyperparameters	Not started
Evaluation and comparison of solutions	Not started
Integration of selected model with the Nessie platform	Not started

4. The WM Dashboard for Larnaca demo case

4.1. Demo case description

The pilot at Larnaca, Cyprus, is an advanced treatment facility for complementary removal of nutrients and salts, aiming to produce high quality irrigation water from treated municipal wastewater. The pilot-plant is located at the Wastewater Treatment Plant of Larnaca. The capacity of the pilot plant is 1 m^3/h production of irrigation water. Additionally, salt recovery occurs from the concentrated stream, while water from the concentrated stream can be used in an industrial context. The schematics of the pilot plant is shown in Figure 9.

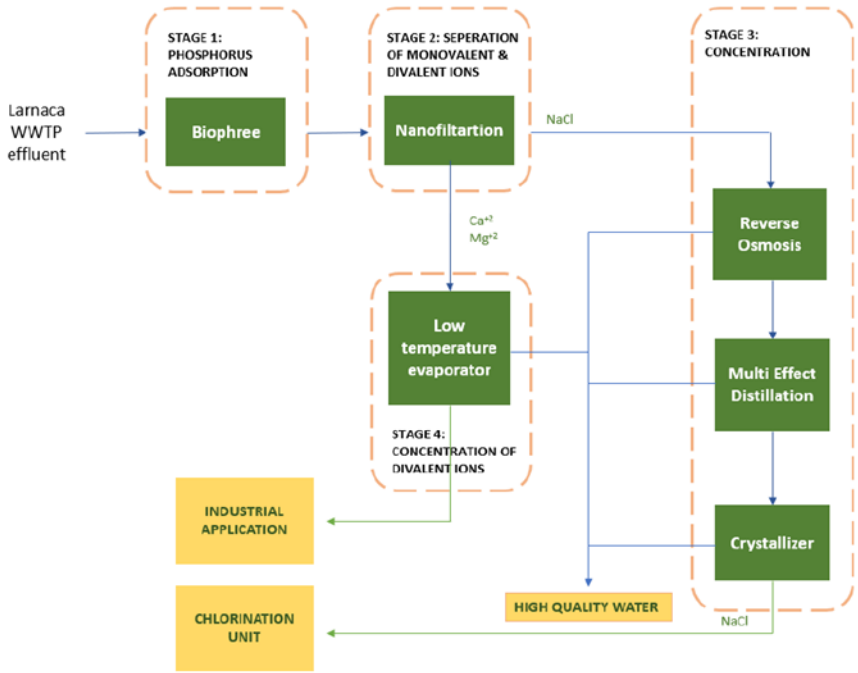


Figure 9- Schematics of the pilot-plant at the WWTP of Larnaca, Cyprus (CS4) (Source: Ramos et al. (2022))

CS4 pilot receives the effluent from the Larnaca WWTP. WETSUS developed the Biophree technology, while NTUA developed the remaining technologies. The first unit to receive the effluent is the Biophree technology, to remove phosphorus. Phosphorus is a limiting factor on biological growth, therefore additional removal of phosphorus improves the water quality, and limits bio-fouling on the following membrane filtration steps of the pilot. The Biophree effluent is subsequently treated by a Nano-membrane Filtration (NF), followed by Reserve Osmosis (RO) membrane filtration. The NF step will remove divalent ions, such as Ca^{2+} and Mg^{2+}, boosting the RO efficiency by the prevention of scaling. The RO step will further reduce the conductivity of the water, to values below 100 µS/cm. The concentrated streams resulting from the NF and RO units are submitted to further treatment. The NF concentrate, rich in divalent ions removed from the water line, is further treated with a low temperature evaporator. The condensed water from the evaporator (after previous distillation) is added to the RO permeate water, due to its high quality standards, while the produced brine might be used for industrial applications. The RO concentrate is further treated with a multi-effect distillation (MED) evaporator unit. The condensed water of the MED is added to the RO permeate, as high-quality water production, and the brine further treated with a crystallizer. The recovered distillate water from the crystallizer, also designated as condensed water, is gathered to the high quality water produced by the RO, low temperature evaporator and MED. The brine of the crystallizer, a saturated solution of sodium chloride (NaCl), can be recovered and used in the disinfection step (chlorination) of the WWTP. In the Larnaca pilot, the Biophree, NF and RO installations are designed to work continuously. The remaining technologies, applied to treat the concentrated streams, operate in batch mode.

4.2. ICT services and infrastructures

To access the online data from the CS4 system regarding NTUA's technology, access rights over the plant's Virtual Private Network (VPN) were required. The plant's VPN holds additional systems, connections, and data from the entire plant, making it restricted to external personnel. Initially, the data export process relied on a manual procedure where data from the Programmable Logic Controller (PLC) were manually fetched and transferred to an Excel file with a predefined format. This process required personnel to access the plant's VPN. Although there was a need to automate this data export process, there were security concerns about granting access to the plant's VPN to the technology providers. As a result, the Larnaca partners sought the assistance of a trusted external programmer who had access rights to the plant's VPN to handle this task and develop an automated procedure, allowing the CS4-NESSIE connection. This procedure involved accessing the plant's PLC, exporting the latest measured data, formatting the data appropriately, and pushing it through the NESSIE API. To facilitate the establishment of a steady communication, the NESSIE team provided a server located in its premises, operating in an isolated room. This server serves as a secure platform for exchanging data between CS4 and NESSIE, ensuring the successful integration of the two systems.

To achieve the connection, a special software application was developed based on Microsoft Visual Studio platform and using .Net technology. The application continuously runs on the dedicated PC/server and transfers all measurements in real-time from the PLC into a predefined file format. Each file has a ".csv" structure and includes all the measurements of the corresponding date in rows. Each row is marked with a unique time-stamp and includes the set of measurements for the given time. In addition to the measurement files, the application uses 2 more files, (a) a configuration-file, which keeps the values of the various parameters mentioned before as assigned by the operator and (b) a history-file (log), which tracks historical data regarding application usage applying relevant time-stamps (e.g. when each transfer was performed, the communication and transfer status, etc.). This functionality allows traceability of the application and helps to detect possible malfunctions in the systems' connection, adding an extra security layer.

To access the Biophree system dataset, the Nessie team collaborated closely with WETSUS partners. A proxy application called "SiteManager Client" was installed on ICCS's servers that also hosts Nessie. The proxy application acted like a tunnel to receive the files using the standard File Transfer Protocol (FTP) and was chosen to increase security between data transmitting. SiteManager is an Industrial IoT gateway typically installed with a PLC/HMI or other industrial equipment (including servers and virtual machines) to enable remote access. The remote access granted to Nessie involves connecting to a File Server which stores data files in binary and proprietary format. For this reason, the PLC required certain intervention routines in order to produce .doc files (Microsoft's Excel format for Open Documents is also a format that Nessie understands).

After installing the SiteManager, a safe connection had to be successfully established. To accomplish this, the Data Provider was contacted in order to white-list the defined domain, and configure the GateManager. Finally, provided credentials were used in order to tunnel the FTP connection and retrieve the listing of data files. Figure 10 depicts the basic successions of proxy elements needed to securely retrieve data from the PLCs.

water mining

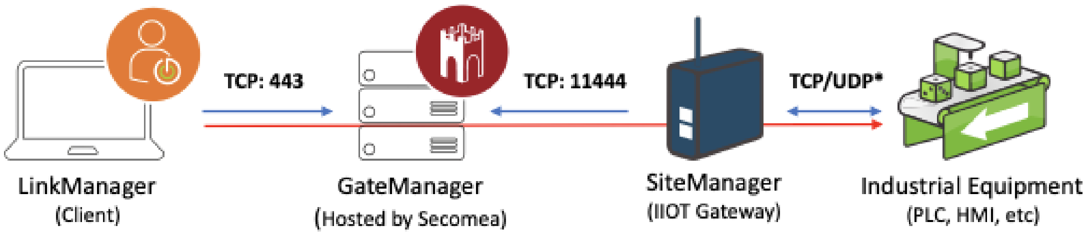


Figure 10- Data retrieval from PLCs in Biophree system

After everything was in place and running, Nessie could connect to the industrial FTP server, retrieve files on a timely basis and store data in its database.

4.3. Deployment of Nessie

4.3.1. NTUA technology

Two dashboards were developed and deployed for the CS4 demo case, connecting to a) the WM technologies of NTUA and b) the WETSUS Biophree system. For each system connection, different connection requirements and end-user needs were defined.

Regarding real-time monitoring of the CS4 NTUA technology, this is only accessible with approved user credentials. It displays real-time data of 17 variables associated with the CS processes, as defined by the stakeholders/end-users (see Deliverable D7.1). The variables, along with their units, can be seen in Table 3Table 3. The sampling frequency of all variables measured is 1 min.

Table 3-List of variables monitored in real-time through NESSIE system for CS4- Larnaca Cyprus (NTUA technology).

Nr.	Variable	Units
1	Unit-Operating-Time	hours
2	Product-Flow-NF	m^3/h
3	Discharge-Meter3-RO	m^3
4	Inlet-Ph	Ph
5	Product-Meter3-RO	m^3
6	Inlet-Pressure-NF-PT1	bar
7	Discharge-Flow-RO	m^3/h
8	Discharge-Meter3-NF	m^3
9	Discharge-Flow-NF	m^3/h
10	Discharge-Pressure-NF-PT2	bar
11	Inlet-REDOX	-
12	Product-Meter3-NF	m^3
13	Conductivity-RO	mS
14	Inlet-Pressure-RO-PT3	bar
15	Discharge-Pressure-RO-PT4	bar
16	Product-Flow-RO	m^3/h
17	Conductivity-NF	mS

The deployment of Nessie instance began by identifying the key data and sensors that stakeholders were interested in, based on the first list of specified requirements. Historical data corresponding to these parameters were provided and used to create an initial version of the CS4 dashboard. Following the demonstration of the first version, the Nessie team collaborated closely with the CS4 stakeholders to define their specific needs and requirements to proceed with the case-specific tailoring of the dashboard – both for the proper establishment of connection and the displayed content. In respect to establishing a connection, the team collaborated closely with NTUA partners in order to provide guidance for the development of the data fetching software, through dedicated technical meetings. Following the software development, the Nessie team established the connection with the CS4 system through a dedicated FTP server and using the Nessie API. For the authentication step, the team provided a unique set of credentials, authenticated through a POST request. The generated token is included in every request, to ensure only authenticated clients access, retrieve or store data to the dashboard. A quick guide with step-by-step instructions for the a) authentication, b) sensor/device data retrieval and c) simulation scenario data storing, and retrieval was developed and shared among partners to facilitate the use of Nessie's capabilities.

By navigating through Nessie's Data Visualization Layer for CS4 demo site, seen in Figure 11, end users are able to explore the most recent data from each device/sensor of the demo site, along with a "flag" indicating the quality of measurement, and the relevant timestamp (left side of screen), a concise list of user-selected KPIs that quickly summarizes the system or process performance (right side of screen) and the system's variables timeseries charts (bottom of the screen), according to user specifications. The latter also provide indications for the maximum, minimum and mean value for the selected time window (x-axis). The user may wish to explore only a portion of the given timeseries presented in the chart, by sliding the axis' upper and lower limits, for which the system will update the statistical characteristics. The Nessie system also allows for the export of timeseries values to .csv format, by selecting "Export CSV" located on the bottom of each chart.

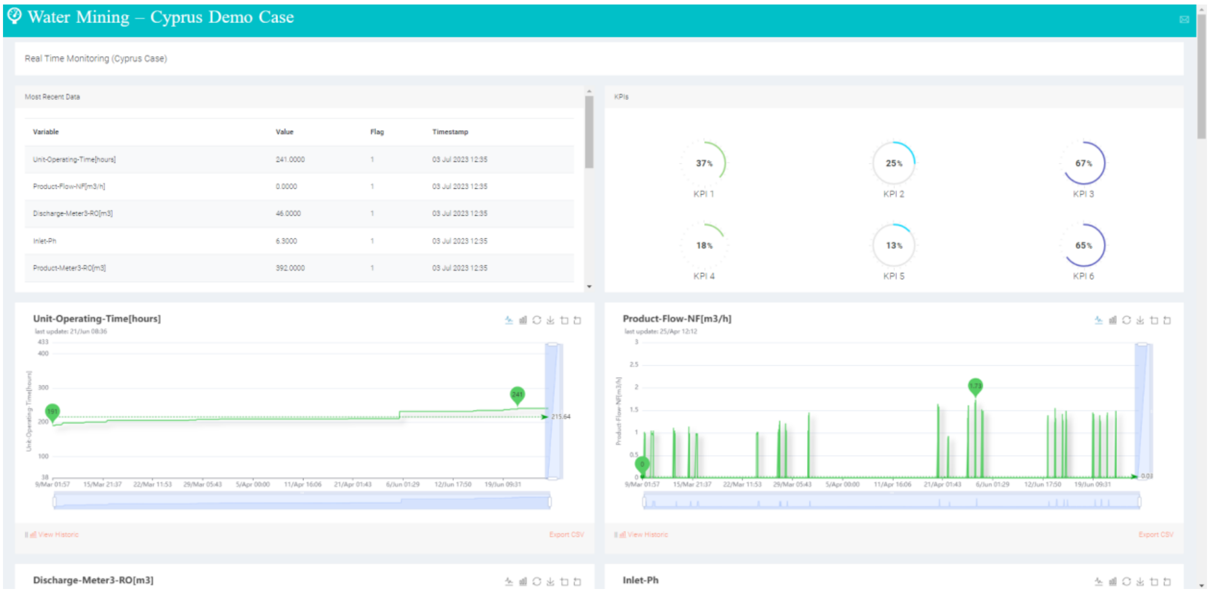


Figure 11-Nessie dashboard implementation in CS4 for the NTUA technologies, displaying a portion the customized dashboard view as defined by the stakeholders.

By selecting "View Historic", the Nessie system provides additional filtering capabilities, and allows users to filter timeseries from a list of qualitative date ranges (i.e., today, yesterday, this month etc.), and aggregating fine-timescale, raw data to predefined intervals (i.e., hourly, daily, weekly etc.). An example of this use case can be seen in Figure 12.

Figure 12 -View of historical data through the Nessie dashboard implementation in CS4 for the NTUA technologies, displaying the latest measurements for variable "Product-Flow-RO", aggregated in an hourly step. Y-axis m3/h.

4.3.2. Biophree technology

The second dashboard deployed for CS4 is that of the Biophree system operations, which is accessible with the use of the granted credentials. This Nessie instance monitors a total of 107 unique variables associated with the Biophree system processes, as defined by the stakeholders. Following the specifications of the system operators, the Nessie dashboard was modified to accommodate a total of 26 unique dashboard views – each associated with a specific portion of the system and its complex processes. The list of custom dashboard views that the users can navigate through Nessie and associated monitored variables can be seen in Table 4. The sampling frequency is 1 min.

Table 4-List of customized dashboard views and associated variables monitored in real-time through NESSIE system for CS4- Larnaca Cyprus (Biophree system)

Nr.	Dashboard view (alias)	Variables name per dashboard view
1	Main	Flow R01
		Flow R02
		Flow R03
		P04 current measurement
		Effluent flow
		Total effluent flow
2	AirPressure	Pressure air compressor
3	Compressor	Compressor state
4	EffluentAnalog	Effluent flow
		P02 Effluent setpoint
		Tanklevel T01 effluent
		Total effluent flow
5	EffluentValves	V401 State
		V402 State
6	InfluentAnalog	Setpoint P01
7	InfluentValves	V101 State
		V102 State

8	PF01Analog	Pressure after PF01 Pressure before PF01
9	PF01Valves	V110 State V111 State V112 State V113 State
10	PF02Analog	Pressure after PF02 Pressure before PF02
11	PF02Valves	V120 State V121 State V122 State V123 State
12	PIDEffluent	Effluent CV Effluent Kd Effluent Ki Effluent Kp Effluent PV Effluent SP
13	PIDInfluent	Influent CV Influent Kd Influent Ki Influent Kp Influent PV Influent SP
14	PIDV211	V211 CV V211 Kd V211 Ki V211 Kp V211 PV V211 SP
15	PIDV221	V221 CV V221 Kd V221 Ki V221 Kp V221 PV V221 SP
16	PIDV231	V231 CV V231 Kd V231 Ki V231 Kp V231 PV V231 SP
17	PO4MeasurementAnalog	PO4 current measurement PO4 influent measurement PO4 reactor 1 measurement PO4 reactor 2 measurement PO4 reactor 3 measurement
18	PO4MeasurementValves	V501 State V502 State V503 State V504 State
19	R01 Analog	Actual position V211 Flow R01 Pressure after R01 Pressure before R01 Setpoint V211
20	R01 Valves	V210 State

		V212 State
		V214 State
		V215 State
		V216 State
		V217 State
21	R02 Analog	Actual position V221
		Flow R02
		Pressure after R02
		Pressure before R02
		Setpoint V221
22	R02 Valves	V220 State
		V222 State
		V224 State
		V225 State
		V226 State
		V227 State
23	R03 Analog	Actual position V231
		Flow R03
		Pressure after R03
		Pressure before R03
		Setpoint V231
24	R03 Valves	V230 State
		V232 State
		V234 State
		V235 State
		V236 State
		V237 State
25	TanksAnalog	Tanklevel T02 effluent
		Tanklevel T03 effluent
26	TanksValves	P03 State
		P04 State
		P05 State
		P06 State
		P07 State
		V410 State
		V411 State
		V420 State
		V421 State

The landing view of this Nessie instance is the “Main” dashboard view, which displays data of 6 variables identified as most important by the stakeholders. Specifically, it includes data related to the phosphate (PO4) and effluent measurements. To navigate through the other 25 customized dashboard views, Nessie offers a “Dashboard Selector” menu, located on the top of the UI, as shown in Figure 13. The Nessie’s Data Visualization Layer in this implementation also offers a concise view of user-defined KPIs and the “Export CSV” and “View Historic” features, as described in the NTUA technology dashboard of CS4.

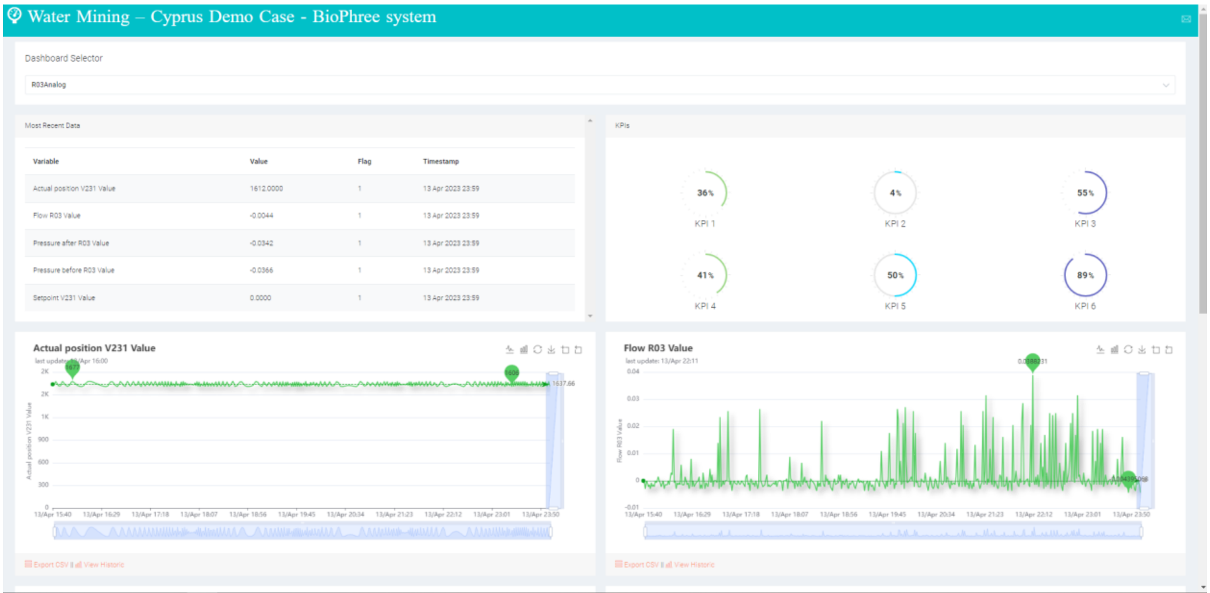


Figure 13-Example of the Nessie implementation in CS4 for the Biophree system, displaying the customised “R03 Analog” view, selected through the “Dashboard Selector” drop-down menu

The deployment of the Nessie instance for the Biophree system has been evolved in a similar manner as for the NTUA technology described above. The first version of the Nessie system was customized and deployed based on an ensemble of offline datasets provided by WETSUS partners, from the initial implementation of the Biophree system in CS4. Following the first demonstration of the Biophree-Nessie system, a new, more elaborate set of devices/sensors were defined to integrate in the next version. In close collaboration with the partners, and following dedicated technical meetings and communications, the Biophree-Nessie connection was established and the dashboard was further customized to accommodate the case-defined devices/sensors and the various dashboard views that can better help organize and monitor the performance of the system and the individual operations/processes. As per every other Nessie instance, a quick guide with step-by-step instructions for the a) authentication, b) sensor/device data retrieval and c) simulation scenario data storing, and retrieval was developed and shared among partners to facilitate the use of Nessie capabilities.

4.4. Advanced Analytics

Regarding the advanced analytics applied to CS4, BRUNEL has developed the alerts and thresholds for the Biophree technology of Wetsus. The Biophree technology was recently relocated from CS4 to CS5.

4.4.1. Alerts and thresholds

Through research and analysis, we created an application that reports the real time state of the Biophree filters and predictor of future performance and remaining useful life, for maintenance planning.

The development of this KPI computing model required a working model of the Biophree filters. We created a KPI computing application for the Biophree filters using scientific principles and advanced math. This RUL model accurately simulates how the filters work and gives valuable insights regarding the status of the filters. Data is fetched from the Nessie Platform and fed to the developed RUL model to compute the KPIs. This helps operators and maintenance staff address issues promptly and keep the filtration process running smoothly.

Historical data from the Biophree filter CS4 was used to develop and customize the elements below: kinetic model of filter, KPIs widgets, creation of custom dashboards to insert the widgets, customize, customizing alert thresholds for both real and virtual sensors, customizing the specific analytics required in CS4 and deploying them.

Figure 14 shows a screenshot of the HMI of the PLC control box for CS4. This is where the Biophree (CS4) pilot is controlled.

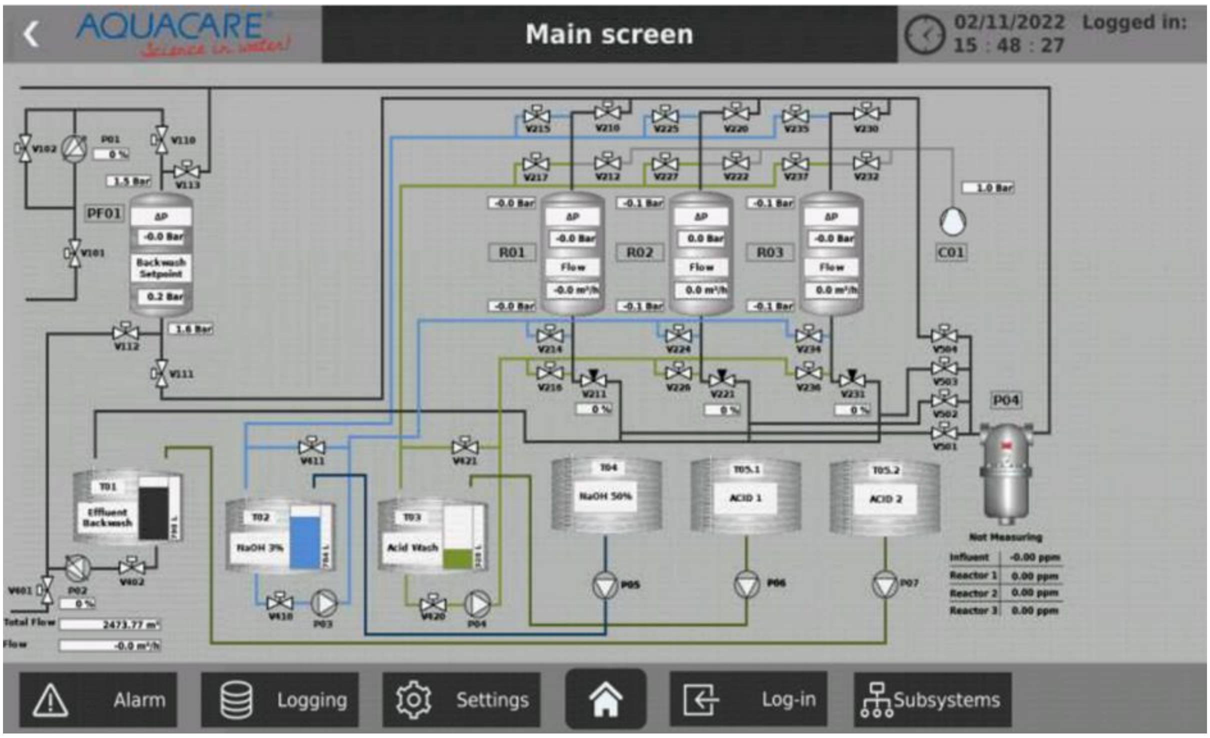


Figure 14- Screenshot of the HMI and PLC control box for CS4.

At BRUNEL, a data processing and analytical layer was designed, connected to data from the Nessie Platform through its Application Programming Interface (API). The API acts as a bridge, facilitating the seamless flow of information between BRUNEL and the Nessie Platform, ensuring a smooth exchange of data essential for scientific analyses and investigations.

To visualize this interconnection, Figure 15 provides a clear representation of how the BRUNEL server communicates and exchanges raw data and processed information with Nessie Platform.

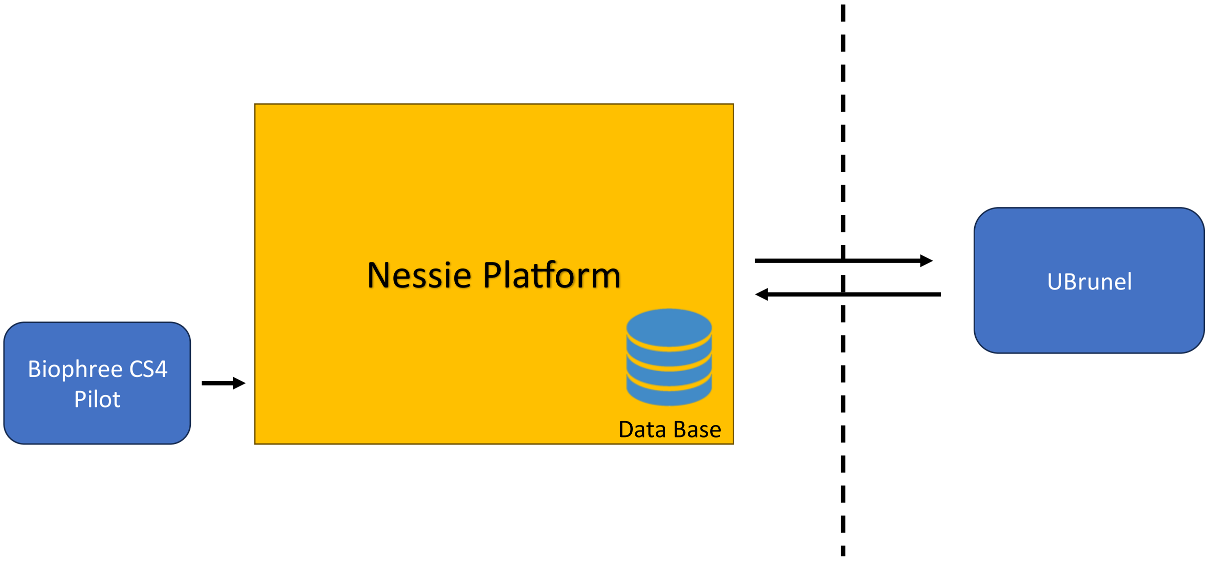


Figure 15- Connection between BRUNEL server and Nessie Platform.

4.4.1.1. Developed KPIs and Alerts

Based on the phosphorus filter specifications in CS4, we have designed and developed a specific set of Key Performance Indicators (KPIs) and an alert system. The list of KPIs is as follows:

1. **Phosphorus Removal Efficiency** (PRE): This KPI measures the percentage of phosphorus effectively removed by the filter from the influent.

2. **Filter Status**: This KPI tracks the operational status of the phosphorus filter during operation time.

3. **Backwash Frequency**: This KPI determines the frequency of backwashing required to maintain the filter's efficiency.

4. **Phosphorus Concentration in Effluent**: Measuring the concentration of phosphorus in the treated effluent provides insights into the filter's overall effectiveness.

5. **Breakthrough time**: the breakthrough time KPI measures the efficiency of the filter in retaining phosphorus before it starts to pass through the filtration media and appears in the treated water. The longer the breakthrough time, the more effective the filter is at removing phosphorus from the influent.

6. **Saturated time**: used to measure the duration for which a phosphorus filter remains saturated with captured phosphorus before reaching its capacity. In water treatment systems employing phosphorus filters, the filter media or material has a limited capacity to adsorb or retain phosphorus from the influent water.

7. **Filter efficiency**: the mass transfer zone KPI assesses how well the system facilitates the transfer of phosphorus from the influent (the water entering the system) to the filter media or adsorbent material where it is captured and removed from the water.

The Alert system is designed to promptly notify operators when any of the KPIs deviate from the predefined thresholds. This ensures timely interventions, maintaining the filter's efficiency and extending its operational life. The combination of these KPIs and the Alert system enhances the monitoring and performance optimization of the phosphorus filter in CS4, contributing to improved water treatment processes and environmental protection.

4.4.1.2. Kinetic model

The data gathered from the continuous mode studies' column was utilized to determine the highest concentration of MB (Methylene Blue) on the adsorbent's solid phase, along with the adsorption rate constant. These calculations were performed using the kinetic model established by Thomas (1944). The columns behavior was modelled based on shared historical data.

Madsorbent is the amount of adsorbent mass which is calculated following as Equation 2:

$$m_{DCF\_ads} = \frac{Q \times C_0}{1000} \times \int_0^{t_s} \left(1 - {C_t}/{C_0}\right) dt$$

Equation 2

removal efficiency (%R) is calculated based on the ratio of adsorbed phosphorus to total phosphorus passed through the filter as Equation 3:

$$\%R = \frac{m_{DCF\_ads}}{m_{DCF\_total}} \times 100$$

Equation 3

mP total shows Total mass of P introduced into the adsorbent until the saturation point and mP_ads indicates Total mass of P adsorbed until the saturation point (Thomas 1944, Han et al. 2006, Han et al. 2007). Table 5 shows the all-parameter definitions for each filter.

Table 5- Parameter definition for performance of a column.

List of relevant parameters
C0 – Initial concentration of P
Q – output flowrate
Z – height of the adsorbent
madsorbent – adsorbent mass
t_B – breakthrough time (Ct/C0 = 0.05)
t_E – saturation time (Ct/C0 = 0.95)
Vefl – Total volume at the end of the cycle
mP total – Total mass of P introduced into the adsorbent until saturation point
MTZ – Mass transfer zone
mP_total – Total mass of P adsorbed until saturation point
qtotal – Adsorbtion capacity at the saturation point or total capacity of the adsorbent
%R – Removal efficiency

The inputs and outcomes are outlined in Table 6, which displays the comprehensive specifications of a Biophree filter, along with the associated Key Performance Indicators (KPIs) recorded at the conclusion of a complete cycle. The utilization of historical data and the model's results presented in Table 6 shows the performance and efficacy of the Biophree filter throughout the full operational cycle.

Table 6- Specification of a Biophree filter and its KPIs.

C0 (mg/ml)	Qout (ml/min)	Z (cm)	madsorbent (g)	t_b (min)	t_s (min)	Vefl (ml)	mP_total (mg)	MTZ	mP_ads (mg)	qtotal (mg/g)	%R
1.44	8280.398773	75.00	90000	5100	11880	98372379	141278.84	42.803	124554.32	1.3839	88.16

Figure 16 below shows a typical plot of the ratio of outlet solute concentration to inlet solute concentration in the fluid as a function of time from the start of flow. The S-shaped curve is called the breakthrough curve.

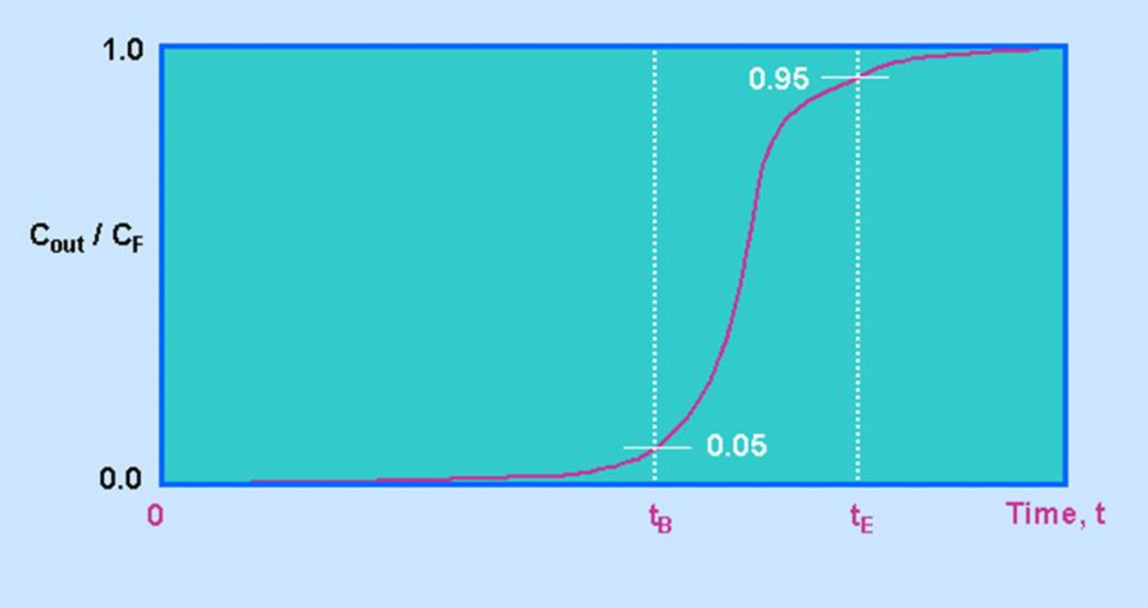


Figure 16- Ratio of outlet solute concentration to inlet solute concentration in the fluid as a function of time (Han et al. 2006)

Prior to t_B, the outlet solute concentration is less than the maximum allowable of 0.05. At t_B, this value is reached, and the adsorption step should be discontinued. If the adsorption step were to be continued for $t > t_B$, the outlet solute concentration will rise rapidly, eventually approaching the inlet concentration as entire bed become saturated. The time required to reach Cout/Cin = 0.95 is designated t_E.

The architecture of an analytical layer developed for a data processing system is show Figure 17 and comprises of two principal stages: model training and inference. During the model training stage, historical data is utilized to train the model, which subsequently transitions to the inference stage. The inference stage handles new data inputs from the Nessie Platform, enabling real-time data processing and analysis. In this process, Key Performance Indicators (KPIs) and Alert computations are performed, facilitating valuable insights and triggering necessary notifications. These computed KPIs and Alerts are then sent back to the Nessie Platform, ensuring a feedback loop to enhance the system's overall performance and accuracy. The integration of these stages and the feedback mechanism enables the analytical layer to continually improve and adapt to evolving data, optimizing its performance and provide insight for decision-making and monitoring purposes.

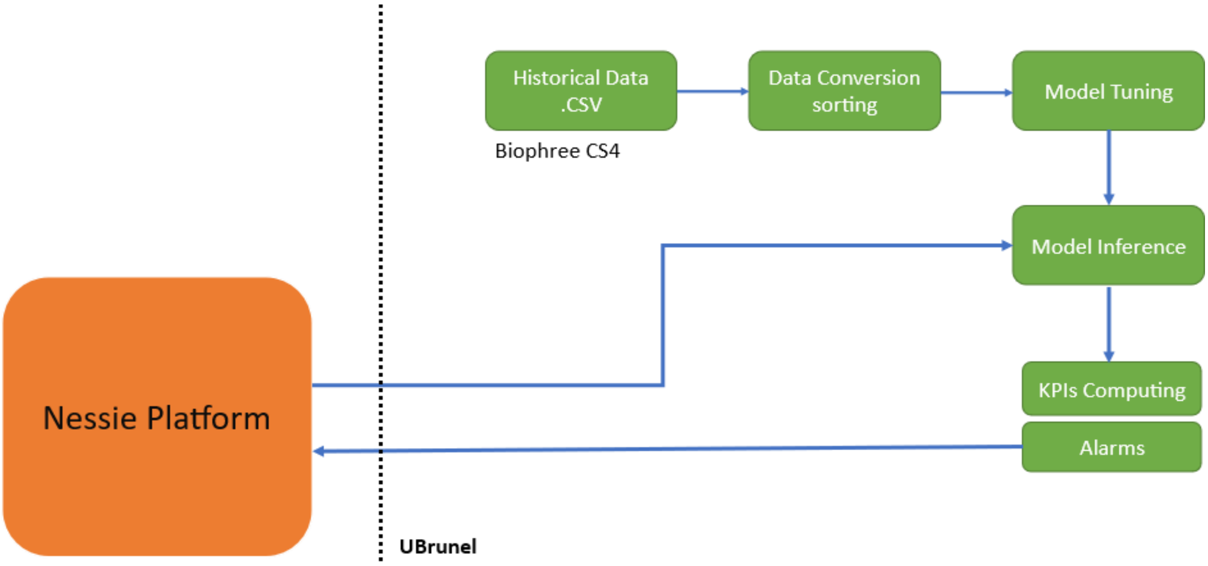


Figure 17- Architecture of an analytical layer developed at BRUNEL.

At BRUNEL, we developed our own test-dashboard to represent computed KPIs and Alert system for the Biophree filter. It provides valuable information about the filter's performance and helps operators understand when the filter is approaching its saturation point. By monitoring the breakthrough time, operators can schedule timely regeneration cycles or initiate appropriate maintenance procedures to prevent excessive phosphorus levels from reaching the effluent, ensuring the water quality remains within the desired regulatory standards or project requirements. Figure 18 shows the BRUNEL developed test-dashboard.

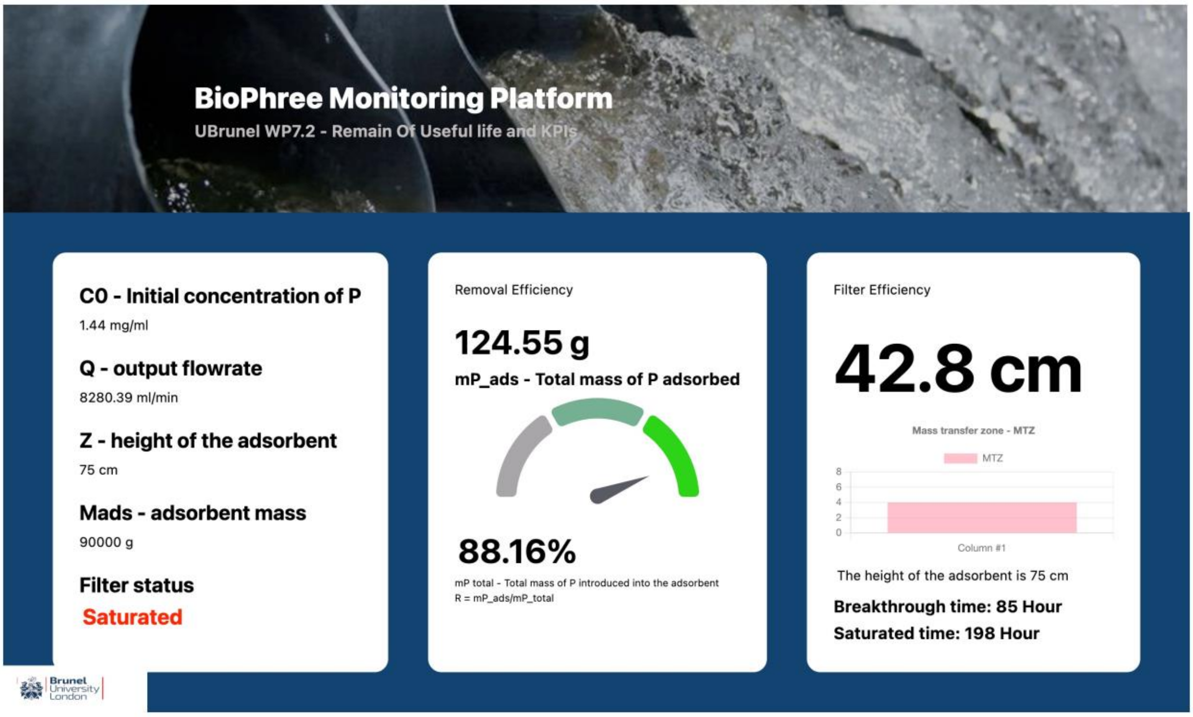


Figure 18- Development of a test dashboard to demonstrate the alerts and thresholds developed for Biophree.

In conclusion, this task represents the KPIs and Alert system which has been designed and developed for Biophree filtration systems. Our custom-made KPIs and Alert system is a valuable addition to the Biophree filters, as digitalization and intelligence is now added to the physical system. Therefore, building the grounds for sustainability, efficiency, predictability, and sustainability in WWTP as proof of concept and demonstration.

4.4.1.3. Alerts and thresholds- next steps

A test dashboard to demonstrate the KPIs system based on the graphical widgets has been developed. The KPIs and alert are connected to a pretrained model which model and represent the Filters status. This model was trained and calibrated based on the Biophree historical data.

Due to the physical relocation of CS4 to CS5, the live data of Biophree system that was available to us has been halted until the system is reinstalled in the new location. Provided the connections are re-established, we will be able to demo the results as before. Once the connection is established and ready, data from CS5 will be fetched and fed into the analytical layer, where Key Performance Indicators (KPIs) and Alerts will be computed and subsequently provided as feedback to the Nessie Platform. As the other systems become online, we expect to allocate our resources to conduct the same processes of data valorization to feed into the corresponding KPIs defined by each technology provider and site managers.

5. WM Dashboard for La Llagosta demo case

5.1. Demo case description

The demonstration pilot in the WWTP of La Llagosta, Barcelona, Spain (CS5) treats municipal wastewater through innovative technologies aiming to produce energy, reduce energy consumption and generate by-products for industrial and agricultural purposes. The leader of CS5 is EURECAT, who is working in close collaboration with Soringué. WETSUS is responsible for the phosphorus removal technologies. The schematics of the La Llagosta pilot is shown in Figure 19. The demo-plant has a capacity of 10 m3/d.

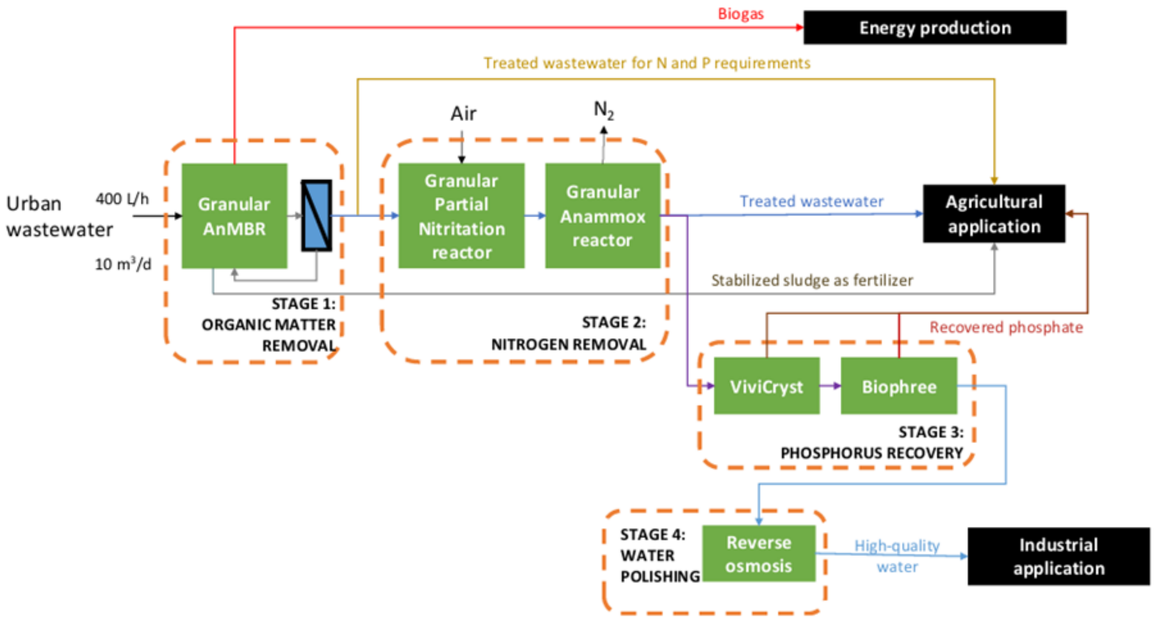


Figure 19- Schematics of the pilot-plant at the WWTP of La Llagosta, Spain (CS5) (Source: Ramos et al. (2022))

The vast majority of the municipal WWTPs applies activated sludge treatment, as the main biological treatment for removal of organic contaminants and eventually nutrients, namely nitrogen and phosphorus. The removal of organic contaminants and certain nutrients in activated sludge systems requires oxygen, therefore extended aeration is required. To lower the energy costs, the La Llagosta pilot applies anaerobic treatment, to remove the carbonaceous materials (organic matter) from the wastewater. The proposed granular Anaerobic Membrane Bioreactor (AnMBR) combines an anaerobic degradation of carbonaceous materials, therefore in the absence of oxygen, with an Ultra-Filtration (UF) membrane, which mainly retains the biomass in the biological reactor. Anaerobic technologies have the advantage of producing biogas, which is a combination of methane with other gases and in itself an energy source. The selected AnMBR is an Expanded Granular Sludge Bed (EGSB) configuration.

Because the AnMBR does not remove nutrients, other biologically based technologies follow, aiming to remove nitrogen and phosphorus from the water line. Granular Partial Nitritation combined with Annamox, is the transformation of ammonia, as nitrogen source, to nitrogen gas, without the need of organic matter.

During Granular Partial Nitritation, about half of the ammonium is oxidized to nitrite; while during Annamox anaerobic oxidation bacteria transform the ammonium nitrate into nitrogen gas. The resulting effluent from the Granular Partial Nitritation/Annamox will have low concentrations of carbonaceous materialsand nitrogen. To remove phosphorus from the water line, two subsequent technologies are included, ViviCryst and Biophree. Vivicryst is based on chemical precipitation of phosphorus, using iron, for the production of vivianite. Vivianite is a hydrated iron phosphate mineral (Fe_2+Fe22+(PO4)2.8H_2O), which potentially can be used as a mineral. The effluent from Vivicryst is further treated with Biophree, as described in section 4.1, obtaining very low phosphorus concentrations. The obtained effluent with low concentrations of carbonaceous material, nitrogen and phosphorus is further treated with a Reverse Osmosis (RO) membrane filtration step to obtain de-mineralized water for industrial applications. The water produced by the AnMBR, Granular Partial Nitritation/Annamox, ViviCryst and Biophree can be combined to obtain water with nutrients (nitrogen and phosphorus) in the correct ratio for agricultural applications.

5.2. ICT services and infrastructures

In the case of the CS5, the development of Nessie's instance has been based on data provided by the demo case, while currently the demo case is working to make the data available on real-time basis. Specifically, we built on .csv files that contained data from two PLCs: the Sorigué and EURECAT PLC. Data files are stored in a specific storage on ICCS server and Nessie accesses and parses them in order to insert into the database.

5.1. Deployment of Nessie

The real-time monitoring of La Llagosta demo site, is available with the use of the granted credentials. This instance integrates and monitors a total of 34 variables associated with the CS processes and defined by the stakeholders, composing a dual dashboard display. A set of 20 devices/sensors is associated with Sorigue's dashboard and 14 devices/sensors are associated to EURECAT's dashboard view.

In the initial phase of data connection exploration between CS5 and Nessie, the team was faced by a technological restrain. The available system in the pilot only allows data sharing via e-mail (in an Excel file) with a certain periodicity – for example, 3 times a day. Through dedicated technical meetings, it was proven that building on such procedures deviates from best practices in real-time data connections, leading to less practical and possibly more unstable connection between the systems due to inherent limitations and potential drawbacks. Firstly, email platforms are not designed to handle real-time data updates efficiently. The process of sending and receiving emails introduces delays, which can lead to outdated or delayed information, especially in dynamic scenarios requiring immediate data access. Additionally, email attachments have size limitations, making it challenging to share large datasets or continuous streams of real-time data effectively. Last but not least, managing data access and permissions becomes cumbersome when relying on email, as it lacks granular control and security features. In view of such limitations, the team collaborated with the CS owners, providing technical advices and guidance in order to initiate the development of a dedicated FTP server in the CS premises, that will enable the seamless integration of devices and sensors to the dashboard in a secure and stable manner. With the server deployment under progress, a first version of the Nessie system was set-up using again offline historical data from the system operations, seen in Figure 20.

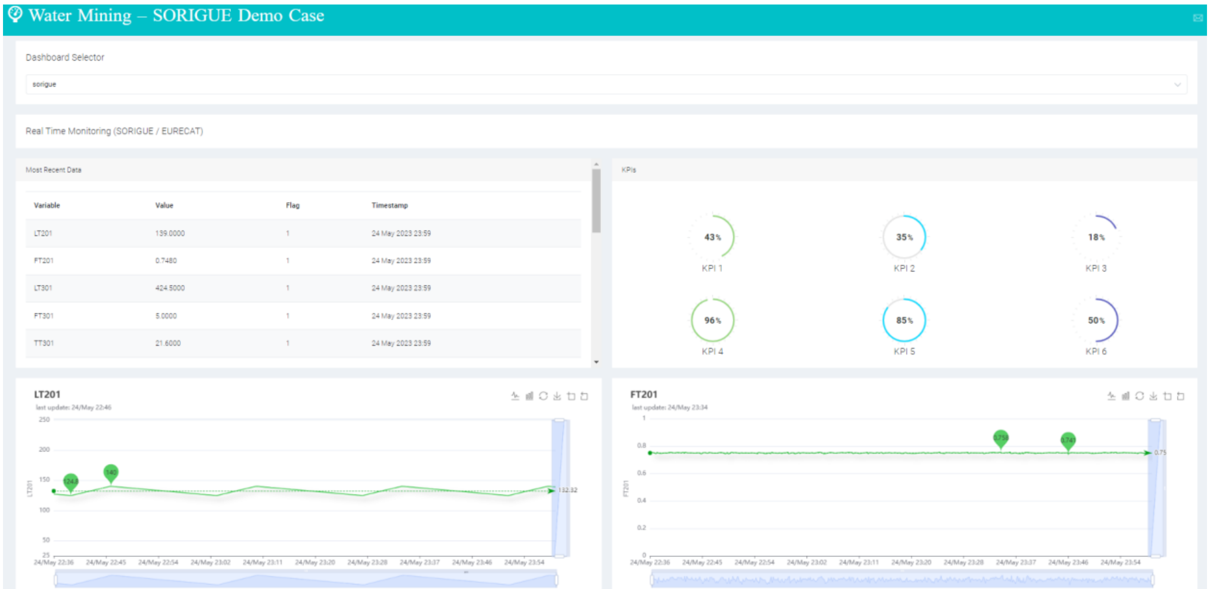


Figure 20- Portion of the Nessie dashboard set-up for CS5, using historical data.

It is worth mentioning that for this case study, a Nessie instance will be developed for Biophree System which is currently being deployed on the site. A Nessie instance was already developed for Biophree, while the technology was placed at CS4 (section 4.3.2).

5.2. Advanced Analytics

Currently, KWR is working with CS5 to develop an AI model for the Partial Nitration Process.

5.2.1. Model based control with AI models

In this report is described the development of an AI model named Fine Tree using the Regression Learner app in MATLAB. The AI model is designed to predict the performance of a Partial Nitritation Reactor (PNR) based on 12 input parameters measured through continuous sensors. The input parameters include various flow rates, pressures at the intake, flow rates, temperature, and oxygen levels at the reactor circulation pipe.

The primary objective of the AI model is to predict the success of the PNR based on the removal of Ammonia (NH_4). The success criterion is defined as achieving a measured Ammonia concentration at the circulation pipe that is 50% of the Ammonia concentration at the intake.

5.2.1.1. Problem description

The problem at hand revolves around predicting the success of the PNR which is a critical process in wastewater treatment. The reactor's primary function is to convert Ammonia (NH_4) to Nitrite (NO_2) Nitrate (NO_3) through a partial nitrification process. This conversion is essential for further stages of wastewater treatment, where Nitrate and Nitrite will be converted in Nitrogen gas. Success is defined as an Ammonia ratio of 50% between the intake and the circulation pipe. The PNR reactor operates as a complex biological process, challenging its representation through traditional mechanistic models. It involves multiple interconnected parameters, microbial biochemical dynamics, and chemical reactions, leading to a non-linear behavior. Additionally, the PNR's behavior is dynamic, influenced by changes in influent conditions, temperature, microbial activity, and other factors. Achieving accurate predictions under varying conditions necessitates a flexible and adaptable modeling approach. Understanding the intricate interactions between different parameters and their impact on the reactor's performance poses difficulties due to the dependency on microbial activity and environmental conditions. Furthermore, establishing a precise mechanistic model is challenging as the relationships between input parameters and reactor performance may not be immediately apparent and may exhibit non-intuitive behavior. Overcoming these challenges through the development of an AI model offers a promising solution to predict the PNR's NH_4 removal ratio accurately.

5.2.1.2. Data Quality

The success of an AI model heavily depends on the quality of the training data. In developing the Fine Tree model for predicting the performance of the PNR, significant emphasis was placed on ensuring the data quality to achieve reliable and accurate results. Several important considerations were taken into account during the data preprocessing stage:

Sensor Selection: The choice of which sensors to include in the dataset was a crucial step. We carefully selected sensors that provided essential measurements, including flow, pressure, temperature, and oxygen levels at various points within the reactor system. The selected sensors were deemed critical for understanding the dynamics and behavior of the reactor.

Data Filtering: The raw data received from the sensors contained observations spanning different time periods, some of which were characterized by instability and erratic behavior. To obtain reliable insights, we focused on extracting data from a stable period of the system. By excluding data from the initial stage of the reactor when it was undergoing significant changes and adaptations, we ensured that the selected data reflected a more stable and consistent operating state.

To assess the stability of the parameters used in the Fine Tree model, we analyzed their statistical properties, including the maximum, minimum, mean, and standard deviation. Stability is a crucial aspect when working with real-world data, especially in predicting the performance of the PNR as unstable or fluctuating parameters may introduce noise and hinder accurate predictions. Table 7 displays the statistical summary of the parameters used in the Fine Tree model for predicting the performance of the PNR. It includes maximum, minimum, mean, and standard deviation values, ensuring data quality and model stability.

To enhance coherence and clarity, Figure 21 shows the PNR layout diagram, illustrating the placement of the sensors.

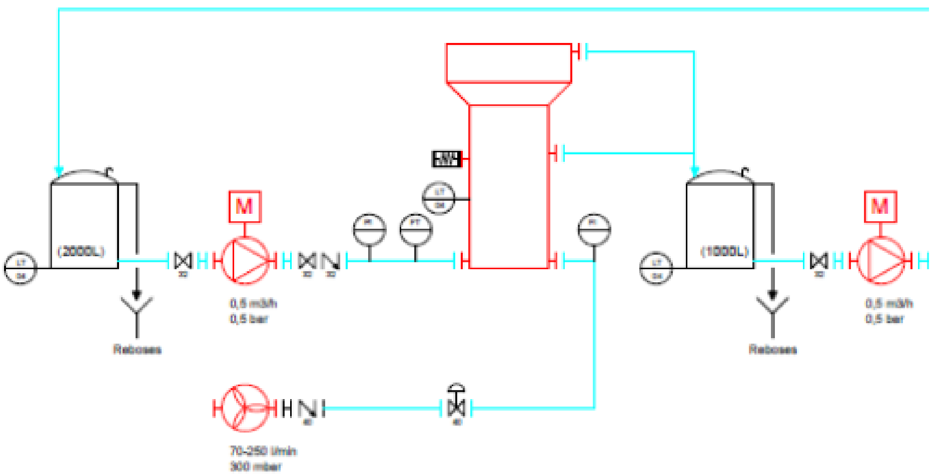


Figure 21- Partial Nitritation reactor schematics (source: (Ramos et al. 2022))

Table 7- Summary of Parameter Statistics (Partial Nitritation Reactor CS5)

Sensor	Max	Min	Mean	Stdv
LT1001	144.40	29.90	97.02	38.55
FT1001	1.50	0.00	0.09	0.36
NHT1001	198.90	44.70	94.05	28.41
LT2001	191.00	124.70	180.38	15.95
NHT2001	599.60	0.00	17.77	57.11
PHT2001	8.30	5.70	7.01	0.49
OXT2001	10.60	0.10	4.21	3.31
TT2001	48.50	25.00	29.54	1.60
LT3001	158.90	25.00	99.59	47.14
FT3001	0.52	0.00	0.11	0.14
LT4001	292.80	25.00	108.71	117.76
TT4001	61.70	27.30	33.74	4.26
LT4101	162.00	51.90	84.78	43.43
FT4101	0.38	0.00	0.08	0.11

By addressing these data quality considerations, we have laid a strong foundation for the Fine Tree model, enabling it to learn meaningful relationships between the input parameters and the reactor's performance. The use of high-quality data from stable periods ensures that the model can effectively predict the success of the PNR and serves as a reliable tool for guiding reactor optimization and decision-making processes.

5.2.1.3. Fine Tree Model

The Fine Tree model is a powerful decision tree-based regression algorithm widely employed in various domains for predictive modelling tasks. It belongs to the family of supervised learning algorithms and is particularly effective in handling both numerical and categorical data, making it suitable for diverse datasets. The Fine Tree model builds a tree-like structure, consisting of decision nodes and leaf nodes, to create a series of binary decisions based on the input features and arrive at a final prediction. The Fine Tree model offers several advantages that make it an excellent choice for predicting the performance of the PNR:

1. **Interpretability:** The Fine Tree model produces human-readable decision rules, making it easy to interpret and understand how the model arrives at its predictions. This interpretability is particularly valuable in critical applications like reactor performance prediction, as it allows domain experts to gain insights into the underlying relationships between input parameters and reactor success.
2. **Handling Non-Linearity:** Decision trees, including the Fine Tree model, are inherently capable of capturing complex non-linear relationships between input features and the target variable. The PNR process is known to exhibit intricate and non-linear behavior due to its biological and chemical complexities. The Fine Tree model's ability to handle non-linearity allows it to accurately represent these complex interactions, ensuring more accurate predictions.
3. **Feature Importance:** The Fine Tree model provides a measure of feature importance, indicating which input parameters have the most significant impact on the reactor's performance. This feature importance analysis aids in identifying critical factors that contribute to the success of the reactor and guides process optimization efforts.
4. **Robustness:** Fine Tree models are relatively robust to noisy data and can handle missing values without requiring extensive data preprocessing. This characteristic is essential when dealing with real-world sensor data, which can often be imperfect due to measurement errors or technical issues.
5. **Efficiency:** Fine Tree models are computationally efficient, making them well-suited for large-scale datasets and real-time prediction scenarios. This efficiency enables quick model training and prediction, ensuring timely responses to changes in operating conditions.

5.2.1.4. Results

Table 8 presents the performance results of three tree-based models: Fine Tree, Medium Tree, and Coarse Tree, which were trained to predict the performance of the PNR. Generally, a more complex tree model, such as the "Fine Tree," tends to offer higher accuracy in predicting the reactor's success. This heightened accuracy stems from the model's ability to capture intricate patterns and relationships within the data. However, the greater complexity of the "Fine Tree" model makes it more susceptible to overfitting. Overfitting occurs when the model becomes overly tailored to the training data, potentially resulting in reduced performance on new, unseen data.

The table displays various evaluation metrics, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), R-square (R^2), Mean Absolute Error (MAE), and Root Squared Error for Normalized Data (NRMSE).

Table 8- Performance results of the three tree-based models

Model	RMSE	MSE	R-square	MAE	NRMSE
Fine Tree	0.176083	0.031005	0.968596	0.010728	0.012179
Medium Tree	0.215884	0.046606	0.952795	0.015535	0.014533
Coarse Tree	0.302159	0.0913	0.907527	0.027528	0.021385

The metrics provide valuable insights into the performance of each model in predicting the target variable (NH_4 ratio). Lower values of RMSE, MSE, and MAE indicate better predictive accuracy, while higher R-squared values indicate a better fit of the model to the data. From Table 8, it can be observed that the "Fine Tree" model outperforms the other two models, as it exhibits the lowest values for RMSE, MSE, and MAE, and the highest R-square value, as expected. However, it's worth noting that the "Coarse Tree" model also provides promising results, demonstrating its effectiveness in predicting the reactor's performance.

The subsequent section presents the outcomes of the Fine Tree model utilized to forecast PNR performance. The evaluation of the model's effectiveness entails a comprehensive analysis involving visual representations and statistical assessments.

Figure 22 displays the actual response (NH_4 ratio) and the corresponding predicted response generated by the AI model. The y-axis represents the response values, while the x-axis represents the record number from the dataset. The graph visually presents a comparison between the model's predictions (yellow) and the true values (blue Xs) for each record in the dataset.

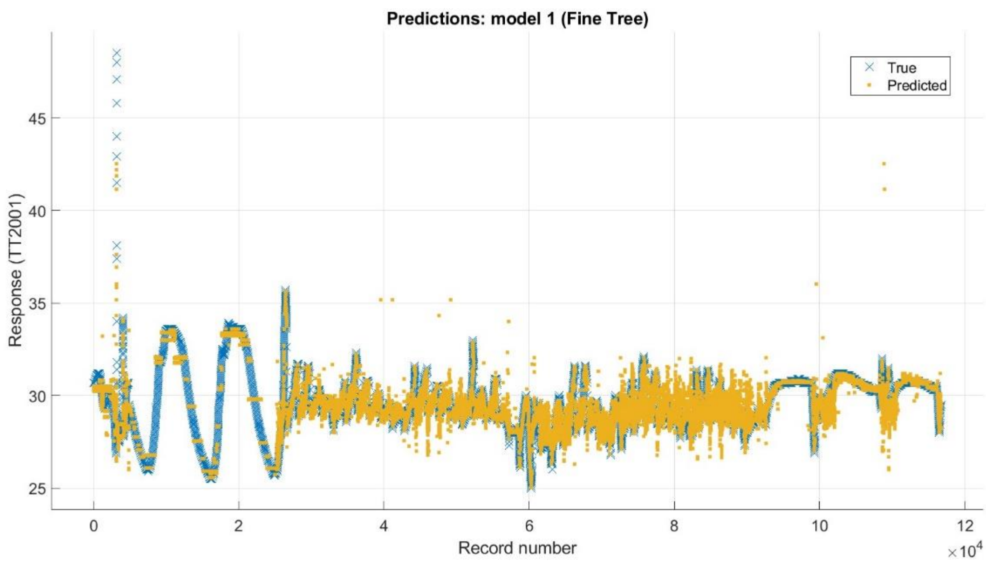


Figure 22- Response plot for the fine tree model.

A well-aligned prediction model graph indicates that the AI model effectively captures the underlying patterns and relationships in the data, resulting in predictions that closely match the actual responses. On the other hand, deviations between the real data points and the predicted line or curve may indicate areas where the model performs less accurately or requires further refinement.

Figure 23 provides a visual representation of the discrepancies between the model's predicted values and the actual true responses (NH4 ratio). Each data point on the graph corresponds to a specific record in the dataset. The horizontal axis represents the true response values, while the vertical axis shows the corresponding residuals. A well-behaved Residuals vs. True Response graph should exhibit random scatter around the zero line, indicating that the model's predictions are unbiased and closely aligned with the true responses. In this case, the residuals should be randomly distributed, without any discernible patterns or trends. Such a pattern signifies that the model accurately captures the underlying relationships between the input parameters and the reactor's performance. Deviation from the zero line, such as systematic patterns or clusters of points, might indicate areas where the model is underestimating or overestimating the reactor's success. Consistent overestimation or underestimation could point to areas where the model requires further refinement or additional data to improve accuracy.

Illustrated in the Prediction plot Figure 23, a notable pattern emerges: towards the upper range of the data, the residuals are predominantly positive. This pattern signifies that the model's predictions tend to underestimate the true values in this region. Conversely, in the lower range of the data, negative residuals indicate instances where predictions surpass the actual values. While certain prediction values exhibit deviation, a substantial portion of the residual points closely hug the zero line. From this analysis, it becomes apparent that enhancing the model's performance at the data extremes may necessitate a larger and more balanced dataset. Future refinement could involve data cleaning to identify and exclude outlier points that stem from sensor malfunctions, particularly those observed in relation to other sensors. This strategic approach could lead to improved model training and enhanced accuracy in capturing the edges of the data spectrum.

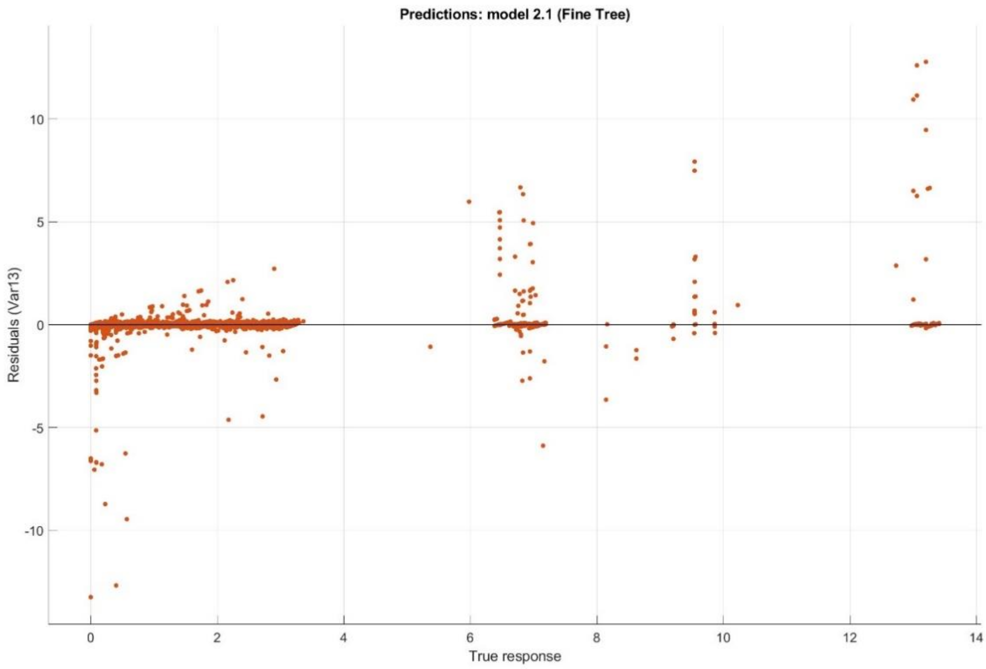


Figure 23- Residuals vs. True Response Graph.

Figure 24 showcases the true response (actual NH_4 ratio) against the model's predictions. This graph further emphasizes the model's predictive ability and highlights areas where the predictions deviate from the true values. A well-fitted graph indicates that the model closely follows the true responses, reinforcing its reliability in forecasting reactor performance. As shown in Figure 23, a majority of the data points form tight clusters around the 45-degree line, primarily in three distinct groups. However, notable deviations from this alignment are evident among specific points. In the effort to address these outliers, potential future investigations into sensor correlations could prove beneficial for their identification and removal. Importantly, similar conclusions can be drawn as those derived from the analysis of the preceding graph.

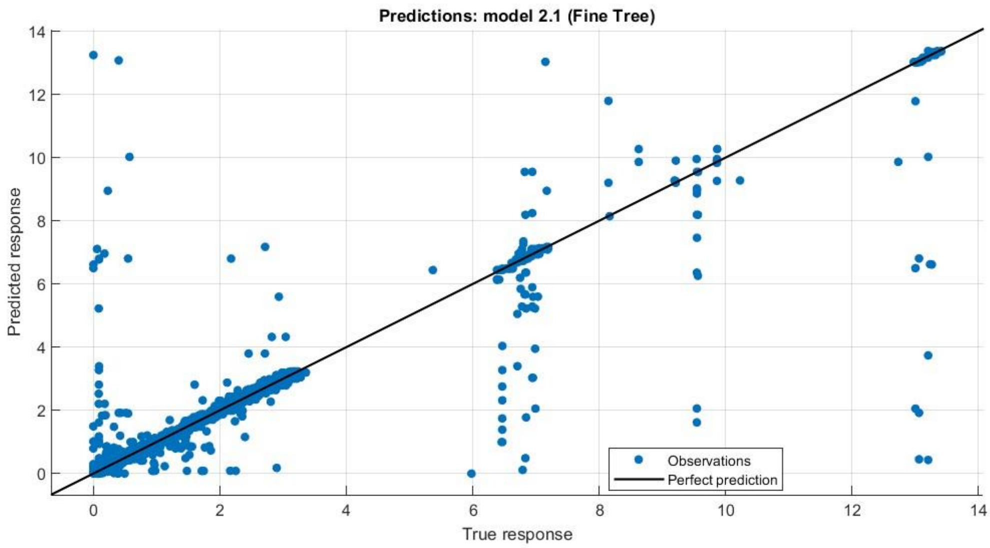


Figure 24- True Response vs. Prediction Graph

Figure 25 illustrates the results of the F-test conducted on the model's parameters. The F-test is a statistical test used to evaluate the significance of individual predictor variables in the model. A higher F-test value suggests that a parameter contributes significantly to the model's predictive power. By examining the F-test results, we can identify the most influential parameters and gain valuable insights into their impact on the reactor's success. Clearly observed in the F-test graph (Figure 24), the initial 8 sensors emerge as pivotal contributors to PNR performance prediction. However, the final 4 sensors exhibit varied significance in this predictive context. Notably, sensor FT4101 stands out with its persistent zero readings, prompting the recommendation of excluding it should similar patterns recur. Among the remaining trio—LT4001, TT4001, and LT4101—each demonstrates a discernible yet limited influence on the prediction. Their potential inclusion in future model training remains a viable avenue for enhancement.

Overall, the graphical representations and statistical tests of the Fine Tree model demonstrate its efficacy in predicting the PNR performance. The close alignment of data points in the Response vs. Prediction graph, unbiased residuals, and well-fitted True Response vs. Prediction graph affirm the model's accuracy and reliability. However, expanding the dataset to encompass the extremities of the spectrum holds the potential to enhance the prediction accuracy specifically within these regions. The significant F-test values also highlight the key parameters driving the model's predictions. The insights obtained from these results contribute to a deeper understanding of the reactor's behavior and facilitate data-driven decision-making for reactor optimization and performance enhancement.

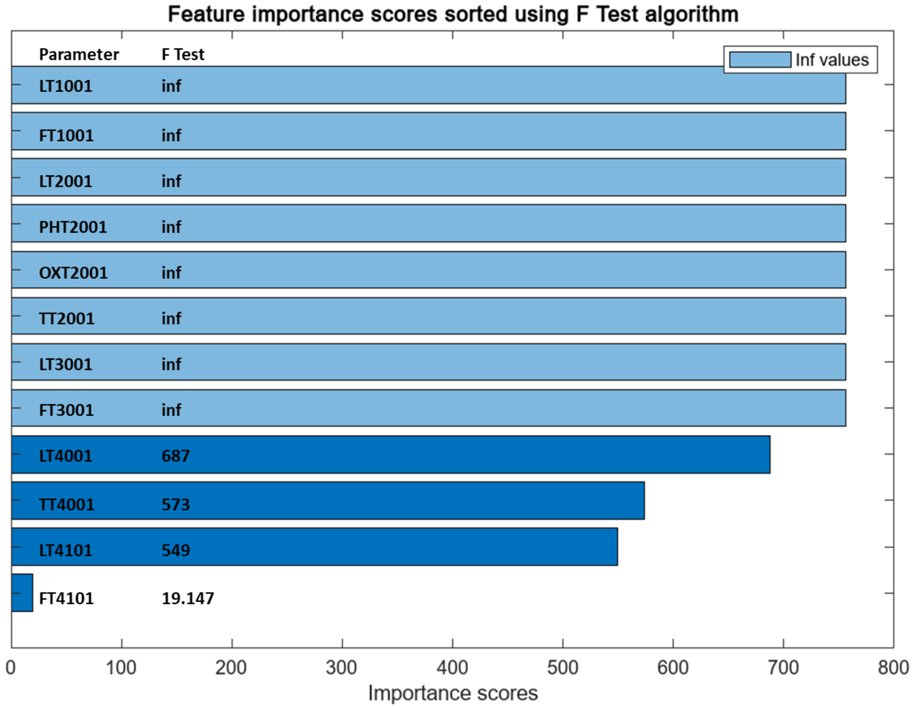


Figure 25- F-Test for Parameters Graph

5.2.1.5. Model based control with AI models - next steps

For CS5, we successfully developed an AI model, specifically the Fine Tree model, to predict the ammonia input/output ratio with remarkable accuracy, achieving an impressive 96%. The model was trained on a comprehensive dataset that included essential parameters measured through continuous sensors, such as flow, pressure, temperature, and oxygen levels in the Partial Nitritation Reactor.

Our analysis revealed that the selected parameters were sufficient to yield highly accurate predictions. Additionally, the F-test results provided valuable insights into the significance of each parameter, indicating that some parameters might be less important in the prediction process. As part of future model refinement, we can consider removing less relevant parameters to optimize the model's performance and streamline its computations. The tree models, including Fine Tree, Medium Tree, and Coarse Tree, have demonstrated their adequacy for predicting the Partial Nitritation Reactor performance. Among these models, the "Fine Tree" model exhibited superior performance, outperforming the other models with lower RMSE, MSE, and MAE values and higher R-squared value. While we initially intended to include the blower configuration/operation mode as the active parameter, we opted for the oxygen level in the recirculated pipes due to its significant impact on the reactor's performance. This substitution proved to be effective, resulting in promising results. In conclusion, the AI model developed in this study offers a reliable and accurate tool for predicting the performance of the Partial Nitritation Reactor. The model's high accuracy and robustness, along with the identification of influential parameters, provide valuable insights for reactor optimization and decision-making. As we continue to enhance the model and explore additional variables, we anticipate even greater accuracy and application potential in addressing challenges related to wastewater treatment systems.

Following steps include the discussion of the obtained results with CS5 stakeholders, to evaluate how the insights of the AI model presented can be translated into actions, supporting the operation of the PNR reactor leading to the optimization of the process. When the discussions with the CS5 stakeholders reach a satisfactory result for the CS, the analytical tool will be integrated in the pilot dashboard.

6. Early insights and lessons learned

This chapter summarizes the main lessons learned, within our experience of setting real-time dashboards for CS. Our goal is to possibly contribute to the definition of content for follow-up research projects, where ICT tools as dashboards are being developed; or to enforce policy changes, required in a world of accelerated digital change. Our insights will be both in terms of procedure and main requirements for implementation as well as follow-up scientific knowledge. More specifically, our insights will mention: commitment of partners, availability of data, previous agreements on data sharing, planning of the connections and ownership of tools.

Real-time monitoring through dashboards allows continuous optimization of operations; active action towards stringent environmental goals; increased resilience in time of climate and social change. The advantages seem undisputed; however, they are not necessarily shared goals within all stakeholders of a circular economy. In WM, WP7 changed all the CSs initially attributed. All industrial CS were changed to CS where universities and knowledge centers lead technologically. A technology or a service provider, with a product or service to sell is not necessarily interested in its optimization. Apparently, participation in an EU research founded project does not seem enough to ensure a commitment towards innovation, particularly the ones beyond the technology or service being provided. On the other hand, active engagement and collaboration are essential for successful implementation. Regular technical discussions and meetings played - and will continue to play- a vital role in defining requirements, clarifying expectations, and addressing challenges throughout the process.

A real-time dashboard requires sharing of real-time data. The WM project is an open source data project. However, the data sharing requirement was one of the most difficult to overcome. Once more, the signed EU agreements to participate in an open-source project were not enough to ensure smooth data-sharing. In particular, the industrial environment has severe restrictions on data sharing with external servers. To ensure confidentiality, the Nessie developers prepared authentication procedures and the remaining T7.2 partners remained available to sign Non-Disclosure-Agreements (NDA), whenever required. For example, additional one-to-one NDAs were signed, between T7.2 partners and new CS owners, when relevant (KWR and CIEMAT (CS2) signed a NDA). It was important to establish clear agreements on data sharing protocols and ensure that all parties were aligned and committed to the project goals. This collaborative effort demonstrated the importance of strong partnerships and the value of collective problem-solving. The partners' commitment fostered a shared understanding of the challenges at hand, the expectations and most importantly the need for practical solutions for real-world applications. The shared willingness to engage in technical discussions and contribute to the process played a vital role in addressing data connection issues.

A real-time dashboard, with advanced analytics, requires real-time data being produced during a considerable time-period. Only an operational period of several months, justifies the effort made on implementing a real-time dashboard and allows the testing of the advanced analytics in real-practice. Short-term pilot experiments are, in principle, not compatible with the development of a dashboard. Its full potential is only achieved, when a pilot or a demonstration site has an optimization phase beyond the start-up phase. Operation periods of at least 6 months are desirable.

Additionally, in WM project it became evident that not all processes of the pilot systems operated 24/7, due to maintenance, cleaning procedures, and adjustments, which affected the continuous creation and sharing of real-time data. In the case of intermittent operations for CS2, where the Nessie system is receiving data when the system operates on "experiment days," additional insights and added value emerged during the deployment process.

One key insight was the need to develop a customized approach for handling intermittent data. The Nessie system was adapted to automatically classify and organize the relevant data received during each "experiment day." This allowed users to effectively monitor and analyze the data specific to each operational period. This insight highlighted the importance of flexibility in data handling and the ability to adapt the system to accommodate non-continuous data streams. This allowed Nessie to provide a centralized platform for collecting, organizing, and analyzing intermittent data, where stakeholders could easily access and review specific experimental days. This feature enables the gain of insights into system performance and evaluate the impact of operational variations, based on the given system operation and conditions. It can facilitate the identification of patterns, trends, and potential areas for improvement in discontinuous pilot-operations. However, it is important to note that discontinuous operations also present challenges, as the “irregular” data streaming may pose difficulties in establishing continuous trends and long-term analysis, requiring tailored approach to data interpretation and decision-making. Consequently, it highlights the importance of flexibility in data handling and ability to adapt to non-continuous data streams.

A valuable insight for the deployment of real-time monitoring systems in real-world applications, is that it often encounters the challenge of the lack of suitable data connection technologies. This bottleneck can hinder the seamless integration and transmission of data, impeding the effectiveness of the monitoring system. The availability and compatibility of data connection technologies play a critical role in ensuring timely and reliable data sharing. Identifying and addressing these limitations requires careful consideration of requirements and constraints, and close collaboration among practitioners and technology providers. Therefore, the partners commitment becomes once again a key feature for the implementation of a dashboard.

Regarding the development of advanced analytics, it became clear, upon the first exposure to the raw data, that there are adjustments, calibration and formatting issues that need to be addressed to ensure that the input data reflects the needs of the advanced analytics tools, such as for the KPI calculator. Therefore, a detailed signal processing, data filtering and modulation needs to be conducted to valorize the input data. The clean information is then feed into the internal model calculations that are conducted in an orderly manner (time series). The process needs to be repeated, as the various dashboards became available.

During the design, implementation and development of the dashboard and additional advanced analytics an underlying issue is the ownership of tools and the protection of intellectual property (IP). So far, ownership of tools is being assumed on the side of the developers. However, a dashboard, namely the development of advanced analytics, means that one partner will for instance develop a process model, that will be validated by another. So far, the protection of the IP is being openly addressed in joint meetings. Solutions, able to satisfy all partners, are still being discussed.

So far, the main lesson learned from this experience is the significance of effective communication, cooperation, and shared commitment among project partners. It highlights the need for proactive engagement and a shared sense of responsibility to overcome technical challenges. By leveraging the expertise and dedication of all stakeholders, a suitable solution can be identified and implemented more swiftly.

7. Conclusions

The present report presents the implementation of dashboards to three CS of the WM project, namely CS2, CS4 and CS5. At the WM project proposal phase, WP7 was meant to work with CS1, CS3 and CS6. Despite the efforts of WP7, described in Deliverable 7.1., developing dashboards for the initial CSs was not possible. Therefore, a strategic decision of changing CSs was made. The current stakeholders, namely those involved in CS2, CS4 and CS5, agreed on the potential of dashboards, as a support to operation and optimization tool, of water and wastewater demonstration pilots. Lessons learned during the process are reported in section 6, addressing commitment of partners, availability of data, previous agreements on data sharing, planning of the connections and ownership of tools.

The present deliverable presents and demonstrates the first versions of the dashboards for all the present CSs. The CSs processes are summarized, ICT services and communications described and the deployment of Nessie demonstrated. Examples of the advanced analytics in development, associated to each CS, are provided demonstrating the goals, development and applications of the advanced analytics tools. A dashboard of a pilot, as a digital twin of the full combination of processes and operations taking place at each CS pilot, supports, not only daily operation by visualizing real-time data, but also the optimization route towards previously defined KPIs. Furthermore, it increases system resilience by allowing simulations beyond real practice.

Nevertheless, due to the strategic change of CSs, the current deliverable does not present the complete customization of the dashboards to the CSs needs, and complete demonstration of the advanced analytics tools to all the CSs. Currently, customization of the dashboards and advanced analytics tools are an on-going task. To overcome the delay, one-to-one meetings were implemented, relying on the full collaboration with the new CSs, which in practice meant multiplying the interactions with the relevant stakeholders. Nevertheless, the delay has not yet been overcome, therefore Task 7.2 proposes to present the full implementation of the dashboards on an extra deliverable, designated "D7.7- WM dashboard deployment-final version", to be delivered in M44. Following work steps are described section 7.1.

7.1. Next steps

The next steps include tasks related with: connection, preliminary customization, advanced analytics development, calibration and validation of advanced analytics and final customization of the dashboards (including advanced analytics). The tasks are performed by different partners and the current status of each task varies according to each CS. The tasks of connection and customization (preliminary and final) are performed by ICCS, associated with the Nessie technology. Table 7 show the following steps in the hands of ICCs, related to the Nessie technology.

Table 9- Next steps, per CS, related to the Nessie technology, aiming for the final customization of the dashboards.

Technology	Case study	Next step
Nessie	CS2	Further customization to CS specific requirements
		Integrate additional devices/sensors according to CS needs – list to be provided by CS
		Integration of WP7 analytics
		Final customization
	CS4	Integration of WP7 analytics
		Final customization
	CS5	Establish data connection – awaiting CS technical developments (FTP server under development)
		Explore the need for further customization or additional devices/sensors for integration
		Potential integration of Biophree system devices/sensors after its successful implementation in the CS
		Integration of WP7 analytics
		Final customization

The development of advanced analytics is at the hands of EURECAT (reinforcement learning), BRUNEL (alerts and thresholds) and KWR (digital twins and model-based control with AI models). So far, each partner has focused in exemplifying one their analytical tools. Further steps, associated with specific CSs, are described per CS. In particular:

- Section 3.4.1.5 describes the following steps for reinforcement learning to CS2;
- Section 4.4.1.3 describes the following steps for alerts and thresholds associated with the Biophree technology, previously in CS4 and currently being placed at CS5;
- Section 5.2 describes the following steps regarding development of AI models in CS5.

The ultimate goal, is to extend the three types of advanced analytics to all CSs. The scope of each analytical tool is currently being agreed with the CS owners.

References

Han, R., Y. Wang, W. Zou, Y. Wang and J. Shi (2007). "Comparison of linear and nonlinear analysis in estimating the Thomas model parameters for methylene blue adsorption onto natural zeolite in fixed-bed column." Journal of Hazardous Materials **145**(1-2): 331-335.

Han, R., W. Zou, H. Li, Y. Li and J. Shi (2006). "Copper (II) and lead (II) removal from aqueous solution in fixed-bed columns by manganese oxide coated zeolite." Journal of hazardous materials **137**(2): 934-942.

Kossieris, P., N. Nievas, M. Korevaar, A. Mousavi, M. Lousada-Ferreira and C. Makropoulos (2022). Deliverable 7.1-The Water-MIning Dashboard. Water-Mining-Next Generation Smart Water Management Systems- H2020, Grant agreemet n. 869474.

Mnih, V. and *e. al.* (2015). "Human-level control through deep reinforcement learning." Nature **518**(7540): 529–533.

Ramos, C., C. Sielfeld, M. Kyriazi, M. Avramidi and W. Wijdeveld (2022). Deliverable 5.2: Design parameters for CE processes (focus on Phosphorus, water and energy recovery). Water-Mining-Next Generation Smart Water Management Systems- H2020, Grant agreemet n. 869474.

Schaul, T., J. Quan, I. Antonoglou and D. Silver (2016). Prioritized Experience Replay. Int. Conf. Learn. Represent.

Schulman, J., F. Wolski, P. Dhariwal, A. Radford and O. Klimov (2017). "Proximal Policy Optimization Algorithms."

Sutton, R. S. and A. G. Barto (2018). Reinforcement learning: An introduction., MIT press.

Thomas, H. C. (1944). "Heterogeneous ion exchange in a flowing system." Journal of the American chemical society **66**(10): 1664-1666.

Watkins, C. J. C. H. and P. Dayan (1992). "Q-learning." Mach. Learn **8**(3-4): 279–292.

Xevgenos, D. (2022). Deliverable 3.4: Demo systems used in CS1 & CS2. Water-Mining-Next Generation Smart Water Management Systems- H2020, Grant agreemet n. 869474.

Annex I – Jointly-signed collaboration letters between WP7, CS2 and CS5

KWR

From	Subject	Date
Christos Makropoulos	Collaboration between case study 2 and work package 7 in Water Mining	4 May 2023
Att.	**Copy**	**Page**
Project management team of Water Mining		1/1

Dear Water-Mining Coordinators,

For the past two years, Work Package seven (WP7) has been working together with Case Study one (CS1) to identify diverse ways in which modern, digital tools can benefit the pilot of the case study. Unfortunately, this work has not resulted in a tangible way forward, so we investigated suitable alternatives. Through extensive coordination with the project and work package coordinators, we have identified the possibility to work with CS2 instead of CS1.

We (WP7 and CS2) are happy to inform that we see several promising possibilities for a fruitful collaboration. One major benefit of working with CS2 instead of CS1 is the Living Lab that CS2 is developing which is planned to attract regular visitors. This makes it a strong use case for the Augmented Reality (AR) app which will improve the engagements of these visitors and, as such, allow for a better communication of the project's achievements. Another major benefit of this pilot is that its long-term perspective also benefits the development and implementation of the dashboard, Nessie system and advanced analytics. This is because there will be more time to demonstrate the use of the dashboard and assess the advanced analytics. As opposed, the MED installation in CS1 will only be installed for a brief period.

We (CS2 and WP7) have agreed on the willingness of data exchange of the pilot's design and process operation, as well as its sensor values; this ensures a quick start-up of this collaboration. Therefore, sharing of data is not expected to lead to issues, however, some technical difficulties need to be overcome before the data can be shared over the internet to the Nessie system. This is viewed as a challenge that a research project as Water Mining should investigate.

To conclude, with this communication, both CS2 and WP7 want to express their confidence that they can develop the dashboard and accompanying analytics and that both will give full support to achieve it. And as such we would like to ask for an amendment in which the demonstration of WP7 is changed from CS1 to CS2.

With kind regards,

ZARAGOZA DEL AGUILA GUILLERMO - 27524540A

Christos Makropoulos

Guillermo Zaragoza

water mining

KWR

From	Subject	Date
Christos Makropoulos	Collaboration between case study 5 and work package 7 in Water Mining	4 May 2023
Att.	**Copy**	**Page**
Project management team of Water Mining		1/1

Dear Water-Mining Coordinators,

For the past two years, Work Package seven (WP7) has been working together with Case Study 6 (CS6) to identify ways in which modern, digital tools can benefit the pilot of the case study. Unfortunately, this work has not resulted in a tangible way forward, so we investigated suitable alternatives. Through extensive contacts with the project and work package coordinators, the possibility to work with case study 5 (CS5) instead of CS6 presented itself.

We (WP7 and CS5) are happy to inform that we see several promising possibilities for a fruitful collaboration. One major benefit of working with CS5 is that EURECAT, the work-package leader of CS5, is already a member of WP7, facilitating communication, exchange of information and data. Moreover, since CS5 is currently kick-starting their pilot, WP7 can be involved in the design of the pilot at an early stage, assuring that communication of data with systems such as Nessie is an early request to the technology suppliers. Least but not least, the CS5 pilot can benefit from the Information and Communication Technologies (ICT) tools provided by WP7, since the pilot kick-off, namely the real-time dashboard.

We (CS5 and WP5) have agreed on the willingness of data exchange of the pilot's design and process operation, as well as its sensor values; this ensures a quick start-up of this collaboration. Therefore, sharing of data is not expected to be an issue, however, some technical challenges need to be overcome before the data can be shared over the internet to the Nessie system. This is viewed as a challenge that a research project as Water Mining should investigate.

To conclude, with this communication, both CS5 and WP7 want to express their confidence that they can develop the dashboard and accompanying analytics and that both will give full support to achieve it. And as such we would like to ask for an amendment in which the demonstration of WP7 is changed from CS6 to CS5.

With kind regards,

Christos Markopoulos
Collaboration

Teresa de la Torre
ACSA, Obras e Infraestructuras

sorigué

Annex II- User manual for Nessie's API

Nessie's API has been developed for accessing and managing sensor/device data as well as simulation scenarios through a secure and user-friendly interface. Built on the RESTful architecture and utilizing HTTPS protocol, it guarantees secure communication between clients and the Nessie platform. One of the key features of the API is its authentication layer, which ensures that only authorized clients can access, retrieve, or store data on the platform. To initiate the authentication process, a unique set of credentials is provided, and clients authenticate themselves through a POST request. Once authenticated, a token is generated and included in every subsequent request to validate the client's identity. To assist users in leveraging Nessie's capabilities, a comprehensive guide with step-by-step instructions has been developed and shared among partners. This guide covers the authentication process, as well as the retrieval and storage of sensor/device data and simulation scenario data. With Nessie's API and its user-friendly documentation, partners can seamlessly interact with the platform and harness its full potential.

Authentication

First of all, the user must be authenticated by sending a `POST` request at https://wm-ntua.uwmh.eu/service/api-auth/ with the following request JSON body:

```
{
"username": "<username",
"password": "<user_password>"
}
```

The respond body will be similar to the one below:

```
{
"user": "demo_user",
"expires_in": "3599.991223",
"token": "f40bc693da0e670f2dbd8660dbc2bd1d1265385d"
}
```

The token's long string, preceded by the keyword "Token", need to be included to every request headers section, under the key "Authorization", like the one below:

```
{
"Authorization": "Token f40bc693da0e670f2dbd8660dbc2bd1d1265385d"
}
```

Retrieve data

We can retrieve data by sending a `GET` request at:

https://wm-ntua.uwmh.eu/service/data/observed/<unique_id>/

Where `<unique_id>` is a unique string that describes the device or sensor that makes the measurements

Retrieve a list of all available device/sensor unique_ids

We can retrieve all available sensor/device "unique_id"s by sending a `GET` request at https://wm-ntua.uwmh.eu/service/data/available/

Store Simulation Scenario data (analysis results)

We can store simulation scenario data or any analysis data by sending a POST request at https://wm-ntua.uwmh.eu/service/sim/scenarios/. The request body must be in JSON format and must contain the following parameters:

- ***input_params***: A JSON object with all the input variables used in the process

- ***result***: A JSON object with all the variables that resulted after the process
- ***name*** : A user defined name for the process. The name MUST contains only alphanumerical charachters, hyphens or underscores.

Bellow we can see an example of such request body:


```
{
    "input_params": {"water_flow": 4.5, "pressure": 998 , "temperature": 31.4},
    "result": {"result": 18.2},
    "name": "data_set_13"
}
```


Retrieve Simulation Scenario data

We can retrieve Simulation Scenario data by sending a `GET` request at:

https://wm-ntua.uwmh.eu/service/sim/scenarios/<name>/

Where `<name>` is the name that we have previously given to our simulation scenario.

Retrieve a list with all available Simulation Scenario data

We can retrieve all available Simulation Scenario data by sending a `GET` request at https://wm-ntua.uwmh.eu/service/sim/scenarios/

Annex III- Fact-sheet on WP7 ICT technologies: the dashboard (@ TRL 4/5)

This Annex is a reply to the question made by the EU reviewer, concerning previous knowledge providing from other EU projects and current knowledge being development for the WM project, regarding the real-time dashboards of WP7. This annex is an on-going work, which will be concluded in the upcoming extra deliverable "D7.7- WM dashboard deployment-final version", due in M44.

The Nessie Platform

Description: A visualization and analytics digital platform able to acquire, process and store, and in general to manage, high-resolution data from IoT agents, such as sensors and smart meters, and transform the data into information to support end-users in decision making.

Significant challenges stem from the fact that the dashboard is being implemented in 3 Demo Cases with different treatment processes, different metering solutions and resolution of measurements, as well as of different requirements from the end-user regarding the analytics of interest. In parallel, the analytics of the dashboard are being developed by different partners involved in the project (KWR, BRUNEL, EURECAT). In this context, a key challenge is the integration of different metering systems and analytics into a unique digital solution.

To provide a remedy to this challenge and create a scalable and extendable digital solution, transferable across the demo cases, we developed a decentralized and scalable system architecture that supports seamless integration of different data sources. To do so, we put at the very center of this architecture, the Nessie system, a technology that has been developed throughout the years in the context of different EU-funded research projects, and more specifically:

- ***iWIDGET FP7* (2012 - 2015):** Web applications to support the householders and water utilities to monitor, analyse and get advice on water consumption, using real-time high-resolution data from smart water (and energy) meters. Key analytics: alerting for bursts, forecast water demand/bill, household water-related carbon footprint, breakdown to water uses, control of smart appliances, water-efficiency competitions among users.

- ***DESSIN* FP7 (2014 - 2017):** Remote monitoring and control of Sewer Mining Units, using real-time quality data (e.g., dissolved oxygen, conductivity, turbidity, nitrate, chloride, ammonium, temperature and suspended solids) from low-energy field sensors.

- ***SUBSOL* FP7 (2015 - 2018):** A transferable dashboard for the remote monitoring and management of Subsurface Water System. The widgets of the dashboard can be customised on the fly according to the needs of the end-user and the task of interest.

- **Fiware4Water H2020 (2019 - 2021):** A Fiware-compliant Nessie to support **r**eal-time monitoring, management and control of the external conveyance system of EYDAP (the largest water utility in Greece serving the city of Athens), by integrating data from flow and quality sensors, along with analytics, using FIWARE protocols. Key analytics: advice on optimal flow regulation, forecast of water supply volumes, forecast of turbidity across channel, warning for unusual conditions in the system.

In the WM era of Nessie, the system's role is twofold. Following the paradigm from past experience, we develop a customizable dashboard to allow the end-user to monitor on real-time the conditions of water treatment plants and receive support on decision-making through the outcomes of analytics. Moreover, in the framework of WM project, we are going one step further, transforming Nessie to a "cloud context broker" that allows the seamless integration of information between data sources (sensors at demo plants) and external analytics (developed by third-party entities). To implement this data flow, a new API for Nessie, based on HTTP requests, was developed to support time series retrieval, insertion and storage of time series, as well as store and retrieval of simulation scenarios data (inputs and outcomes) from third party analytics. Another key development around Nessie in WM project is the implementation of a token-based authentication procedure at data exchange procedures, instead of using a single username and password for logging in, to ensure the security of sensitive data from demo cases.

Advanced Analytics

RL solutions for water treatment control

Description: A real-time control model in a water treatment process, capable of receiving the current system's state and providing an immediate response to adapt one or several parameters, with the objective of optimizing quality indicators, emissions, energy consumption, and costs not only in the short term but also in the medium and long term. This control system will serve as a decision support system for the plant operator, aiding in the efficient management of plant operations.

In the domain of water treatment, a comparable control system was developed as part of the Fiware4Water H2020 (2019 - 2021) project, specifically for the case study of the WWTP in Amsterdam. The WWTP employed an Activated Sludge Process (ASP), where bacterial biomass played a vital role in eliminating pollutants like nitrogen, phosphorus, and organic carbon substances. Real-time oxygen setpoints were dynamically adjusted every 15 minutes, considering the system's state and external variables such as the influent flow, aiming to optimize energy consumption, reduce emissions, and maintain compliance with effluent quality standards. The intelligent control system developed in the project was compared with the plant's regular operation through simulations of various potential scenarios, thereby validating the RL agent's efficacy in controlling the oxygen setpoint and demonstrating its capacity for improvement.

In Water Mining, the intelligent control solution, based on RL, is customized for each case study, considering the specific system, operation, variables, and limitations of each pilot plant, along with its unique objectives. This involves comprehending the project case studies, identifying potential use cases for dynamic control, and adapting the knowledge to accommodate new systems. Moreover, in Fiware4Water, RL technology was initially applied to control a single system variable. Now, the proposal is to extend its application to control multiple variables, introducing greater complexity to the problem. This advancement showcases the application of intelligent technology to new water treatment systems and the consideration of multiple control variables, representing a significant improvement over the Fiware4Water project.