



MARCLAIMED

Predictive Models

Deliverable 1.1

WP1: Operational efficiency and economic sustainability in short, medium, and long-term perspective

Author: Cetaqua Andalucía

Date: April 2025



**Funded by
the European Union**

GRANT AGREEMENT NUMBER	101136799	
ACRONYM / FULL TITLE	MARCLAIMED: integrated Decision Support Tool for Reliable and Affordable Application of Manage Aquifer Recharge with Alternative Water Resources in River Basin and Drought Management Plans	
START DATE	01/02/2024	DURATION
END DATE	31/01/2027	
		36 months
PROJECT URL	https://marclaimed.eu/	
DELIVERABLE	D1.1 Predictive Models	
WORK PACKAGE	WP1	
CONTRACTUAL DATE OF DELIVERY	30/04/2025	
ACTUAL DATE OF DELIVERY	25/04/2025	
NATURE		DISSEMINATION LEVEL
		PUBLIC
LEAD BENEFICIARY	CETAND	
RESPONSIBLE AUTHOR	Sara Espinosa Martínez (CETAND) and Lupicinio García Ortiz (CETAND)	
CONTRIBUTIONS FROM	Joan Coines Lopez-Nieto (CETBCN)	
ABSTRACT	Report on the predictive models of quality monitoring and operational control of MAR systems with AWR (T1.1 MARVIS, T1.2 PREDMAR, T1.3 MARAISER)	

Disclaimer

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

© MARCLAIMED Consortium, 2024

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

Copyright message

TABLE OF CONTENTS

EXECUTIVE SUMMARY	4
1. Introduction	6
1.1 Context and background	6
1.2 Deliverable Objectives	6
1.3 Scope of work	7
2. THEORETICAL FRAMEWORK	7
2.1 Operational efficiency in MAR with AWR ‘key concepts’	7
2.2 State of the Art of MARVIS, PREDMAR and MARAISER Tools	10
3. Methodology: General approach	18
3.1. MARVIS - Virtual sensing tool for near real-time water quality monitoring	19
3.2. PREDMAR - AI-based water availability and demand forecasting tool	22
3.3. MARAISER - Short and long-term MAR performance optimization tool	26
4. Description of the Digital Tool	31
4.1. Conceptual scheme and main functionalities	31
4.2. Data Collection and Processing	44
4.3. Digital Tool Development: AI components, algorithms and models used	61
4.4. Technical and Implementation Requirements for the integration in the MARCLAIMED IDST	65
5. Application and preliminary results in Demo Sites	65
5.1 MARVIS - Virtual sensing tool for near real-time water quality monitoring	65
5.2 PREDMAR - AI-based water availability and demand forecasting tool	72
5.3 MARAISER - Module 1 (Short-term) MAR performance optimization tool	96
6. Discussion, Conclusions and Next steps (follow up, implementation, and validation)	121
7. Bibliographic References	131
ANNEX 1: MARAISER - SHORT AND LONG-TERM MAR PERFORMANCE OPTIMIZATION TOOL	135

EXECUTIVE SUMMARY

This deliverable presents an **integrated approach to enhance operational efficiency in Managed Aquifer Recharge (MAR) with Alternative Water Resources (AWR) systems** through three interconnected digital tools, addressing key operational challenges in water quality monitoring, resource forecasting, and risk management.

The first pillar of operational improvement comes through **MARVIS Tool**, a water quality monitoring tool by implementing cost-effective virtual sensing technology. This virtual sensing technology **replaces expensive traditional nitrate sensors - a parameter chosen due to its crucial role as a key indicator of water quality and its significant impact on aquatic ecosystems - while providing continuous monitoring without manual sampling requirements**. The tool's ability to detect water quality variations in real-time enables immediate operational adjustments, **while its adaptive machine learning models ensure sustained monitoring accuracy, significantly reducing operational costs and improving system responsiveness**.

PREDMAR Tool enhances resource optimization by **combining short-term forecasting for proactive operational planning with long-term projections for infrastructure capacity optimization**. The integration of climate scenarios enables adaptive management strategies, while sophisticated forecasting models support enhanced decision-making for water allocation and storage.

MARAISSER Tool strengthens risk management through **comprehensive monitoring of health, environmental, and operational risks**. Its early warning system for operational issues, combined with automated risk assessment and mitigation recommendations, **enables proactive maintenance and optimal MAR infrastructure performance**. The system's ability to **streamline regulatory compliance while optimizing maintenance schedules represents a significant advancement in operational efficiency**.

The integration of these tools within the **MARCLAIMED Integrated Decision Support System (IDST) platform creates a unified operational management system that optimizes MAR performance while reducing operational costs and risks**. Implementation across demonstration sites in Llobregat (Spain), Comporta (Portugal), and Texel (The Netherlands) will validate the Tool's adaptability to various operational contexts and requirements. The platform's data-driven approach enables informed decision-making, while its automated monitoring and control capabilities minimize human error and reduce operational overhead. This comprehensive approach to operational efficiency represents a significant advancement in MAR system management. By providing operators with sophisticated tools for system optimization, cost-effective operation, and sustainable resource management, **MARCLAIMED establishes a new standard for operational excellence in water management systems**. The platform's ability to integrate real-time monitoring, predictive analytics, and risk management **creates a robust framework for achieving and maintaining optimal operational efficiency in MAR systems**.

ACRONYM GLOSSARY

ACA - Agència Catalana de l'Aigua	MAR-SP - MAR Safety Plans
AI - Artificial Intelligence	MCDA - Multi-Criteria Decision Analysis
APA - Agência Portuguesa do Ambiente	MDA - Mean Directional Accuracy
API - Application Programming Interface	ML - Machine Learning
ASR - Aquifer Storage and Recover	MSE - Mean Squared Error
ASTR - Aquifer Storage Transfer and Recover	NDSI - Normalized Difference Snow Index
AWR - Alternative Water Resources	NSZ - Not Saturated Zone
BEWSs - Biological Early Warning Systems	ORP - Operational Requirement Point
BOD - Biochemical Oxygen Demand	PES - Special Drought Plan
CDS - Climate Data Store	PGDCFC - Catalan River Basin Risk Management Plan
CEDEX - Center for Studies and Experimentation of Public Works	PIML - Physics-Informed Machine Learning
CECs - Contaminants of Emerging Concern	PRP - Pre-Requirement Point
CMIP6 - Coupled Model Intercomparison Project Phase 6	RCPs - Representative Concentration Pathways
COD - Chemical Oxygen Demand	RES - Exploitable Results
CRA - Comparative Risk Assessment	RMP - Risk Management Point
DS - Demo Site	SHAP - SHapley Additive exPlanations
EC - Electro Conductivity	sMAPE - Symmetric Mean Absolute Percentage Error
ECMWF - European Centre for Medium-Range Weather Forecasts	SO - Specific Objectives
EDA - Exploratory Data Analysis	SSF - Slow Sand Filtration
EPA - Environmental Protection Agency	SSPs - Shared Socioeconomic Pathways
ESGF - Earth System Grid Federation	SVM - Support Vector Machines
GAC - Granular Activated Carbon	SVH - Sant Vicenç dels Horts
GBM - Gradient Boosting Machines	TDW - Tail Drainage Water
GCMs - Global Climate Models	THMs - Trihalomethanes
GV - Guidance Value	TOC - Total Organic Carbon
IDST - Integrated Decision Support System	TV - Threshold Values
IPCC - Intergovernmental Panel on Climate Change	WHO - World Health Organization
IQR - Interquartile Range	WP - Work Package
JRC - Joint Research Centre	WWTP - Waste Water Treatment Plant
MAE - Mean Absolute Error	XAI - Explainable AI
MAR - Managed Aquifer Recharge	XGBoost - Extreme Gradient Boosting

1. Introduction

1.1 Context and background

Water scarcity is a critical 21st-century challenge, affecting 20% of European territory and 30% of its population [1]. This issue, driven by climate change, tourism, agriculture, and population density, impacts both water quantity and quality, leading to ecosystem degradation and aquifer overexploitation. To address this challenge, there's a growing shift towards circular water systems, with Managed Aquifer Recharge (MAR) using Alternative Water Resources (AWR) emerging as a promising solution, as proposed by MARCLAIMED.

MAR with AWR offers multiple benefits, including aquifer replenishment, saltwater intrusion prevention, and ecosystem preservation. However, the operational efficiency of MAR systems using AWR is a critical factor in their success and widespread adoption. Efficiency in this context encompasses various aspects, including:

- Water quantity: Maximizing the volume of water that can be recharged and recovered.
- Water quality: Ensuring that the recharged water meets or exceeds regulatory standards and does not degrade the aquifer.
- Operational reliability: Ensuring consistent performance over time, despite variations in input water quality and quantity.

Other relevant factors influencing the operational efficiency are: environmental impact, energy consumption and cost-effectiveness, which are considered within the economic sustainability pillar of MARCLAIMED and will be tackled in the D1.2 (MAR with AWR impact at municipal scales) and D1.3 (Economic and financial aspects) framework.

Achieving high operational efficiency in MAR with AWR presents unique challenges. These include: i) Variable quality of alternative water resources, necessitating adaptive treatment strategies, ii) Potential for clogging in recharge systems due to physical, chemical, or biological processes, iii) Long-term uncertainties regarding the behavior of contaminants in the subsurface environment.

The focus on operational efficiency in MAR with AWR is not merely a technical pursuit but a necessity driven by economic, environmental, and social factors. As water stress continues to increase globally, the ability to efficiently and safely integrate MAR with AWR plays a crucial role in sustainable water management strategies.

1.2 Deliverable Objectives

The general objectives for the developments presented in this deliverable are linked with the MARCLAIMED Project Specific Objective 1 (SO1): *To achieve high level operational efficiency of MAR systems with AWR in short, medium and long term perspective*. These general objectives are:

1. Better operational management of MAR with AWR improving water quality, operational control, and providing operational risk assessment (SO1).
2. Determination of current and future status of water bodies forecasting the water resources availability and demands in short-, medium- and long-term perspective (SO1).

To achieve the objectives, three tools have been developed: i) MARVIS, a virtual sensor for nitrates, to enhance our ability to monitor and manage water quality in real-time, ii) PREDMAR, which forecasts water resources availability in the short and long term, enables proactive management and optimization of MAR operations; and iii) MARAISER, a risk management tool for health, environmental, and operational risk control, ensures the safety and reliability of MAR systems.

Implementing advanced tools to improve operational efficiency of MAR systems with AWR will have significant implications for water security, ecosystem health, and sustainable development in water-stressed regions worldwide.

These innovative solutions collectively contribute to ensuring high quality of groundwater resources and optimal performance of MAR systems, thereby strengthening water resource management strategies and fostering resilience in the face of increasing water scarcity and climate change challenges.

1.3 Scope of work

This deliverable addresses the following items to offer a comprehensive overview on the MARVIS, PREDMAR and MARAISER Tools development:

- **Theoretical framework** including both key concepts and state of the art.
- **Methodology**, including the general approach, data collection and digital development.
- **Detailed description of each of the digital tools** with their main functionalities, conceptual scheme and technical and implementation requirements.
- **Application on demo sites** considering different scenarios and the preliminary results.
- **Discussion and conclusions** for further steps.

2. Theoretical Framework

2.1 Operational efficiency in MAR with AWR 'key concepts'

This deliverable focuses on three key concepts: *Operational efficiency in MAR with AWR*, *Artificial Intelligence (AI)* and *Risk Assessment and Management*.

Operational efficiency in MAR with AWR

This concept represents the system's capacity to effectively recharge, store, and recover (or transport and recover) water while maintaining optimal performance across technical,

environmental, and economic parameters. This efficiency is based on a thorough understanding of available water resources and demand patterns. This encompasses the management of water quality, infiltration rates, infrastructure performance, and energy consumption, alongside the prevention and control of operational challenges such as clogging.

The system's efficiency is measured through its ability to maintain consistent performance while adapting to varying conditions, including changes in source water quality, climate variations, and demand patterns, all while ensuring compliance with regulatory requirements and environmental protection standards. The precise knowledge of water resource availability - including seasonal variations, climate change impacts, and alternative water sources potential - combined with accurate demand forecasting, enables optimal system operation and infrastructure utilization.

The achievement of operational efficiency requires integrated management of multiple components, including pre-treatment facilities, infiltration structures, monitoring systems, and recovery infrastructure. This integration must optimize water quality transformation during aquifer passage, maintain appropriate hydraulic gradients, and ensure cost-effective operation through efficient use of energy, chemicals, and maintenance resources. The system's capacity to balance available water resources with varying demand scenarios, both in short and long term, becomes crucial for strategic planning and operational optimization, allowing for adaptive management strategies that maximize the efficiency of the MAR system while ensuring sustainable water resource management.

Artificial Intelligence (AI)

The integration of AI into MAR and AWR management is enhancing the operational and strategic capabilities required for sustainable water system governance. AI tools are contributing to better operational management of MAR systems by improving water quality monitoring, strengthening operational control, and enabling advanced risk assessment mechanisms. Through real-time data from sensor networks, remote sensing, and hydrological models, AI-driven systems facilitate the continuous supervision of recharge operations, detect anomalies in water quality, and support early warning systems for operational disruptions. This allows for prompt corrective actions, optimized maintenance, and compliance with environmental and health regulations, directly addressing General Objective 1.

At the same time, AI technologies are enabling comprehensive forecasting of water resources availability and demand, supporting strategic planning at short-, medium-, and long-term horizons (General Objective 2). In the short term, predictive analytics support real-time forecasting of streamflow, reservoir storage, and groundwater recharge, thereby informing dynamic water allocation for agriculture, urban, and industrial uses. For the long term, AI models incorporate climate change scenarios, socio-economic

projections, land-use changes, and aquifer behavior to simulate the evolution of water availability. These models assist in evaluating the resilience of MAR strategies, optimizing recharge schemes, and informing infrastructure investment decisions. AI also supports scenario-based analysis and risk-informed planning to anticipate extreme weather events and long-term water scarcity, enhancing the preparedness and sustainability of water resource systems. This integrated approach aligns with European objectives on digital transformation, climate adaptation, and sustainable water management.

Risk Assessment and Management

A third essential pillar in this framework is risk management. The integration of MAR with AWR requires comprehensive risk assessments supported by comprehension and mitigation strategies. Risk, in this context, pertains to the probability and severity of detrimental outcomes that may influence groundwater reserves, public health, ecological frameworks, and operational facilities. **Risk management** in MAR systems has to follow a systematic preventive approach that emphasizes the implementation of multiple protective barriers throughout the recharge process. This includes careful control of source water quality, robust treatment systems, continuous monitoring protocols, and well-defined emergency response procedures. The management framework must ensure reliable operation while protecting both the aquifer and end-users of the water resource.

The **risk assessment** process involves several interconnected components. Technical assessment begins with detailed hydrogeological evaluation to understand aquifer characteristics and behavior. This includes analyzing water quality parameters, evaluating infrastructure capabilities, and establishing comprehensive performance monitoring systems. The assessment considers three primary risk categories:

- Public health risks focus on potential threats to human wellbeing, including the presence of pathogens, chemical contaminants, and emerging pollutants that could affect water users. These risks require particular attention when the recharged water is intended for human consumption or comes into contact with drinking water supplies.
- Environmental risks encompass potential impacts on the aquifer system and surrounding ecosystems. This includes possible contamination of groundwater, changes in aquifer chemistry, alterations to groundwater levels, and effects on dependent ecosystems. Long-term environmental sustainability must be carefully evaluated and protected.
- Operational risks relate to the technical and infrastructural aspects of the MAR system. These include potential infrastructure failures, clogging of infiltration surfaces, and overall system efficiency. Operational risks can affect both the short-term performance and long-term viability of the recharge system.

The aim of **operational efficiency** combining **AI techniques** and **risk management & assessment in MAR systems with AWR** represents a cutting-edge approach to sustainable groundwater management. Furthermore, this integrated approach enhances decision-making processes by providing data-driven recommendations for both immediate operational adjustments and long-term strategic planning, ultimately resulting in more resilient and adaptive MAR systems that can effectively respond to changing environmental conditions, regulatory requirements, and water demand scenarios.

2.2 State of the Art of MARVIS, PREDMAR and MARAISER Tools

2.2.1 MARVIS - Virtual sensing tool for near real-time water quality monitoring

Current SoA and challenges: Virtual sensing has emerged as a cutting-edge application of AI among several fields, including water quality control [2]. The methodology entails the utilization of Machine Learning algorithms to develop virtual sensors. These sensors are designed to estimate parameters that are challenging to measure directly, through the establishment of correlations with parameters that are readily measurable. In the case of water quality monitoring, such readily measurable parameters may include pH, electrical conductivity, dissolved oxygen, temperature or turbidity, among others. Neural networks, random forest and multiple linear regression are currently the predominant ML techniques used for developing these virtual sensors. While virtual sensing exhibits potential for enabling real-time monitoring of water quality parameters that are traditionally analyzed in the laboratories, certain limitations and challenges persist. These include data quality assurance, insufficient labeled datasets, and the need for optimal feature selection. Recent advances in deep learning architectures such as auto-encoders or convolutional neural networks are showing promising results in addressing these challenges. However, the transition from research to practical implementation remains a critical challenge, with limited large-scale deployments across real-world settings. MARVIS aims to develop a virtual sensor capable of measuring nitrates concentration. Nitrate monitoring is crucial for water quality assessment as excessive levels can cause severe environmental and health issues, yet current monitoring methods are expensive and time-consuming, requiring laboratory analysis that delays response times. In that regard, recent developments demonstrate that ML techniques can predict nitrates concentration [3, 4, 5], including the application in the field of virtual sensing [6]. These advancements suggest a promising pathway for MARVIS to overcome current implementation barriers and significantly improve nitrate concentration predictions in diverse environmental contexts.

Previous projects: CETAQUA has been at the forefront of integrating virtual sensing technologies to enhance water quality monitoring and management. Experiences in previous projects are mainly focused on the development of reliable real-time virtual sensors for detecting trihalomethanes (THMs) in drinking water facilities. THMs are

chemical compounds that form as byproducts when chlorine or other disinfectants react with natural organic matter in water. These compounds are regulated due to their potential health risks, including an increased likelihood of cancer and other adverse health effects from long-term exposure. The virtual sensors employ ML algorithms to predict THM concentrations based on data from existing, less costly sensors. This approach was successfully validated at Barcelona's Sant Joan Despí Water Treatment Plant, Spain [7] and within the distribution network of Marbella, Spain [8], achieving high prediction accuracy and demonstrating the scalability of the solution. This research was performed in the context of **TRON**, **LAGAR THM** and **Lab Digital Twins** private projects (founded by AIGUES and Hidralia), whose results reached TRL 7 and have served as a foundation for developing the MARVIS virtual sensor to predict nitrate concentrations.

MARVIS breakthrough: MARVIS represents a significant breakthrough in water quality monitoring through the innovative application of virtual sensing technology for real-time nitrate concentration measurement. This development is specifically designed for MAR application, thus contributing to water scarcity mitigation strategies. Unlike previous virtual sensing applications that focused on other water quality parameters like THMs, MARVIS pioneers the use of virtual sensing for nitrate monitoring by establishing correlations between easily measurable parameters (such as pH, electrical conductivity, temperature, and turbidity) and nitrate concentrations.

Focus on nitrates as the target contaminant responds to several pressing needs:

- Regulatory compliance: Meeting the requirements of EU Directive 91/676/EEC necessitates close and continuous monitoring of nitrate concentrations
- Site-specific water quality limitations: Two of the three MARCLAIMED demonstration sites are currently affected by high levels of nitrate contamination, leading to economically unsustainable penalties.
- Limitations of current sensor technologies: Existing nitrate sensors fall into two categories:
 - Ion-selective sensors, which are affected by interference from other compounds and tend to lack robustness.
 - Optical sensors, which are accurate and reliable but often unaffordable for the majority of the potential users.

The advantages that sustain the selection on this contaminant are:

- Model generalizability: The virtual sensor will be generated with a model not dependent on site-specific training conditions. This ensures an increased replicability and scalability of the solution.
- Data availability and model robustness: All demonstration sites will be equipped with optical nitrate sensors, enabling the collection of large datasets, accounting for more variability and contributing to the development of a more robust and accurate model.

- **Cost-efficiency:** From an economic perspective, the proposed virtual sensing approach offers a cost-effective alternative for nitrate monitoring. In the short to medium term, this facilitates the broader adoption of MAR technologies beyond big users, making implementation more accessible to smaller stakeholders.

The development of this kind of sensor is particularly valuable for MAR operations, where continuous monitoring of water quality is crucial for ensuring the safety and efficiency of aquifer recharge using alternative water sources. The system initially employs XGBoost algorithm for its proven capability in handling complex water quality relationships, with plans to transition to a more sophisticated deep learning model as more data becomes available. MARVIS also introduces an innovative adaptive calibration methodology that allows the virtual sensor to maintain its accuracy over time by automatically adjusting to seasonal variations and changing water quality conditions, which is essential for the dynamic nature of MAR operations. This breakthrough enables real-time decision-making in MAR facilities, optimizing the use of alternative water resources while ensuring compliance with water quality standards. The implementation of virtual sensing for nitrates monitoring in MAR contexts represents a significant advancement in making water recharge operations more efficient and sustainable, particularly in regions facing water scarcity challenges.

2.2.2 PREDMAR - AI-based water availability and demand forecasting tool

Current SoA and challenges: The integration of **AI** into water availability and demand forecasting has seen significant advancements in recent years, leveraging **ML** and data-driven analytics to improve the accuracy, efficiency, and adaptability of hydrological models. These AI-driven tools provide enhanced capabilities in predicting water supply variability, optimizing water allocation, and supporting decision-making for sustainable water management.

Artificial Intelligence is increasingly employed to enhance the resolution of Global Climate Models (GCMs), including those from the CMIP6 framework, by downscaling their outputs for regional-scale applications. Techniques such as neural networks and ensemble learning are applied to improve bias correction, detect anomalies, and assimilate climate data in real time, thereby complementing established dynamical and statistical downscaling approaches.

Moreover, AI supports multi-model ensemble analysis to refine the interpretation of climate projections. By integrating data from various Representative Concentration Pathways (RCPs) and Shared Socioeconomic Pathways (SSPs), AI enables more robust assessments of long-term climate risks under different emission and socio-economic scenarios. It also contributes to improving the prediction of extreme weather events, thereby enhancing the performance of early warning systems in terms of both accuracy and timeliness.

The CMIP website [9] is a valuable resource for the climate research community, providing access to data, tools, and documentation. Researchers can use the CMIP data to investigate a wide range of climate phenomena, from regional changes in temperature and precipitation to global sea level rise and extreme weather events. The CMIP project has been instrumental in informing international climate policy, including the Intergovernmental Panel on Climate Change (IPCC) Assessment Reports. However, challenges remain in interpretable AI, uncertainty quantification, and computational scalability, necessitating further research on integrating explainable AI (XAI), probabilistic modeling, and hybrid AI-physics frameworks to improve climate resilience and adaptation strategies.

The development of an AI-based water availability and demand forecasting tool faces several key challenges, including data availability, quality, and heterogeneity, as water-related datasets originate from multiple sources (e.g., remote sensing, IoT sensors, hydrological models, climate projections) with varying spatial and temporal resolutions. Uncertainty quantification is another critical issue, as AI models must integrate stochastic hydrological processes, climate variability, and socio-economic factors while ensuring reliable predictions. The interpretability and transparency of AI models remain a challenge, particularly in regulatory and decision-making contexts where XAI is essential for stakeholder trust. Scalability and computational efficiency are also concerns, as large-scale water systems require high-performance computing for real-time forecasting. Additionally, integration with existing hydrological and climate models demands robust coupling frameworks that balance physics-based modeling with AI-driven insights.

Previous projects: **PRIMA GOTHAM** (led by CETAND [10]): development of a governance tool for sustainable water resource allocation, **MAR2PROTECT** (CETAND and AQUA are partners [11]): development of a DSS tool for MAR integrating technological and societal engagement information using an AI-based evaluation to improve groundwater quality. **PRIMA AGREEMAR** (LNEC is partner [12]): application of a daily sequential soil-water budget hydrological model based on adjusted projections of rainfall and potential evapotranspiration to better understand the magnitude and extent of the expected climate change impacts in the water availability, and assess the potential conventional water available for MAR. **ICARIA** (led by CETBCN and AQUA is partner [13]) aims to increase knowledge of the impacts of natural disasters on strategic infrastructures with a calibrated hydrological model as a main tool to drought impact assessments.

PREDMAR Breakthrough: PREDMAR will enhance ML Models for forecasting water resource availability across both short-term (operational) and long-term (design and planning) horizons by integrating **Physics-Informed Machine Learning (PIML)** principles. This approach will lead to the development of a hybrid modeling framework, combining deterministic models with ML techniques, to improve predictive accuracy and decision-making capabilities. While PIML methodologies have been successfully applied

[14] in weather and climate sciences, their application to AWR management remains an emerging field.

PREDMAR will pioneer its use in the context of water resource forecasting by addressing the specific needs at each demonstration site:

- Llobregat, Spain (DS1), PIML will be applied to predict the basin's water scarcity indicator, as defined in Regional Drought Management Plans: "Plan de gestión del distrito de cuenca fluvial de Cataluña (PGDCFC)" (Catalan River Basin Risk Management Plan) 2022-2027 [15].
- Comporta, Portugal (DS2) will focus on identifying surface water surpluses at the basin and sub-basin scales, evaluating the potential for optimizing water resource management strategies in the region.
- Texel, Netherlands (DS3), the approach will be utilized to forecast groundwater extraction needs for the upcoming growing season. This represents a novel application of PIML, providing an innovative pathway to enhance the accuracy, resilience, and sustainability of water resource forecasting in the face of climate variability and evolving demand patterns.

2.2.3 MARAISER - Short and long-term MAR performance optimization tool

Current SoA and challenges: AWR, specifically reclaimed water, is subject to stringent quality standards, and boasts a multitude of applications. While irrigation and industrial uses currently represent the most prevalent applications, the role of reclaimed water to increase groundwater storage and ensuring a sustainable drinking water supply is progressively gaining relevance.

There are **key guidelines** that provide recommended water quality standards for water reuse such as the ones published by the World Health Organization's (**WHO**) health-based targets, focusing on microbial risk and log-removal requirements, and the US Environmental Protection Agency's (**EPA**) chemical and microbial quality criteria. In Europe, the Center for Studies and Experimentation of Public Works (**CEDEX**) from Spain, integrates WHO methodologies into risk plans using probabilistic risk scoring, and also the Joint Research Centre (**JRC**) of the European Commission's which published the "Technical Guidance Water Reuse Risk Management for Agricultural Irrigation Schemes in Europe".

Regarding **water normative**, one of the most important is the **Water Framework Directive 2000/60/EC** [16], (**amended by Directive 2008/32/EC** [17]) which sets **guidelines for aquifer recharge management and surface water conservation**. Even if this is considered the most significant directive related to water management, it's important to note that before its approval, in December 1991 Directive **91/676/EEC** [18] along with **Directive 91/271/EEC** [19] were approved. The primary objective of these two directives was to protect water quality by establishing parameters and guidelines for

wastewater treatment (91/271) and by controlling nitrate pollution from agricultural sources and urban wastewater (91/676).

Regarding wastewater treatment, the European Union recently approved a **new Directive (2024/3019)** [20] for urban wastewater. The previous one **Directive (91/271/EEC)** focused mainly on secondary treatment and pollution reduction, whereas **Directive (2024/3019)** strengthens environmental protection (by introducing quaternary treatment for micropollutants), digitalization, and circular economy goals.

The **EU Groundwater Directive 2006/118/EC** [21] plays a crucial role too: it primarily focuses on preventing pollution and ensuring good chemical status.

Regarding water reuse and risk management, **EU Regulation 2020/741** [22] establishes minimum quality and control standards for water usage and the **2024/1765** [23] establishes the technical specifications of the key elements of risk management.

Outside the European Union, policy frameworks are also evolving: China, starting in 2022, integrated the “Water Security plan” to address water challenges in the Nation. In the USA local governments implement programs under the **Safe Drinking Water Act** [24] and the **Clean Water Act** [25], as defined by the EPA. Similarly, Australia and the Singapore Water Agency (PUB) have developed authoritative guidelines to support indirect potable reuse applications for treated municipal wastewater and stormwater.

Diving deeper in the meaning of **risk**, the probabilistic approach refers to it as the probability that a hazard, whether biological, chemical, physical, or radiological, will cause harm within a given period, considering the severity and potential consequences. **Risk assessment frameworks** generally evaluate hazards based on their **severity** (insignificant to catastrophic) and **likelihood** (rare to almost certain). In water management, the three key guidelines (**WHO, EPA, and CEDEX**) provide methodologies to assess risks in **drinking and reclaimed water**. These organizations consider **multiple hazard types** affecting both human health and the environment, integrating assessments of biological, chemical, and physical risks to ensure compliance with safety standards.

Given the complexity and potential severity of risks associated with water management, it is crucial to transform these **risk factors into quantifiable indicators**. By doing so, it is possible to systematically monitor and assess the risks in real-time. Digitizing this monitoring process allows for the continuous collection and analysis of data, enabling timely interventions and adjustments to mitigate risks effectively. **This approach, considered in MARAISER Tool** ensures that water quality remains within the established regulatory standards, thereby protecting public health and the environment.

The considered methodologies primarily focus on risks to human health; however, for MAR systems, it is essential to also assess **environmental and operational risks**.

Environmental risk includes potential contamination of groundwater ecosystems, accumulation of persistent pollutants, and alterations to natural hydrological cycles and

habitats. **Operational risk** involves system performance issues, maintenance challenges, and unintended consequences of recharge activities, such as clogging or geochemical reactions. Effective **risk management in MAR systems** and water reuse hinges on reliable monitoring and analytical techniques to evaluate water quality, including physicochemical properties, biological composition, and contaminants.

Real-time monitoring systems, such as IoT-based platforms, (bio)sensors, and remote sensing tools, provide immediate water quality data. Additionally, systems like SAMOS, Biological Early Warning Systems (BEWSs), and passive samplers enhance pollution detection, while SCADA systems monitor filtration plants, pump stations, and the entire distribution network, improving **operational control**.

Beyond monitoring and operational control, ensuring effective risk mitigation in MAR systems requires **integrating decision-support tools** that aid in evaluating potential risks and optimizing management strategies.

Supporting tools for decision-making in water recharge systems offer valuable insights for optimizing operations and mitigating risks. **Comparative Risk Assessment** (CRA) helps evaluate risks and prioritize environmental issues, while **Multi-Criteria Decision Analysis** (MCDA) offers a structured approach to compare project alternatives using decision matrices and stakeholder preferences. Within MCDA, various methods exist: Multi-Attribute Utility Theory assigns numerical scores to different options to enhance decision-making, and the Analytical Hierarchy Process uses pairwise comparisons to establish priority rankings. Building on the risk assessment frameworks established by WHO, EPA, and CEDEX, advanced decision-support and operational control tools further enhance risk management capabilities in MAR systems.

To further strengthen risk assessment and enhance decision-making, there are other types of tools that play a crucial role in continuously monitoring contaminants and optimizing system performance. **Automated risk assessment tools** (e.g. Global Water Tool [26], Water Risk Filter [27], etc...) and **digital tools** (sensors, early warning system, webGIS platform, etc...) continuously monitor contaminants and analyze data trends, improving response mechanisms and ensuring system reliability. SCADA system provides real-time insights, enabling immediate maintenance detection and remote adjustments. By integrating real-time monitoring, advanced analytics, and cloud-based solutions, MAR systems achieve greater efficiency, resilience, and sustainability, ensuring successful risk management.

A MAR operation tool has to be designed to support MAR systems management, given their tendency to lose performance over extended periods. Several tools for MAR have been developed in the framework of different EU and international projects. In particular, MARSOL (LNEC was a partner) developed a generic evaluation tool (MARRISKAPP) with probabilistic risk assessment based on the subjective opinion of an expert panel; RISMAR (from RECLAIM); FLOOD-MAR (regulation of flood water via aquifer storage); ASRRI (ASR

Risk Index); DINA-MAR Hydrogeoportal; or GRAT (Groundwater Recharge Assessment Tool). These tools have in common that the integration of the different key variables in MAR with AWR (water safety, quality, hydrogeological, soil, and environmental and human health risks) need to be enhanced with an interactive user interface, which is accomplished in LIFE MATRIX (led by CETAND). Moreover, these applications use tools not specifically developed for MAR such as MODFLOW, FEFLOW or SEAWAT, all of them models for groundwater simulation which require great effort to be built, calibrated, and validated with poor or inexistent risk assessment approaches.

The successful implementation of digital decision support tools for risk assessment and operational control in water management is exemplified by Spain's PERTE Plan [28], which integrates Digital Twins, IoT sensors, smart metering, and AI-driven predictive analytics. Similar initiatives in Tunisia and the Mediterranean region leverage SCADA systems and cloud-based analytics for real-time monitoring and predictive maintenance. Nevertheless, these approaches have not been applied to MAR systems, but in extractions wellfields and drinking water distribution networks.

From these case studies, **key lessons** were learned: a holistic approach is required, integrating IoT, AI, cloud computing, and predictive analytics rather than isolated solutions.

Previous projects: **MARSOL** (Grant agreement ID: 619120; LNEC was partner [29]): demonstration of MAR as a solution to water scarcity and drought where an assessment tool for risk evaluation and potential mitigation activities was developed; **QUEEN** (private funded project by CETBCN and AIGUES): development of a guide of critical parameters to prevent clogging in deep injection MAR systems; **LIFE MATRIX** (LIFE20 ENV/ES/000788, led by CETAND [30]): integration of an aquifer recharge operational optimizer (short-term); **MAR2PROTECT** (HORIZON-CL6-2022-ZEROPOLLUTION-01-01 101082048 MAR2PROTECT, CETAND and AQUA are partners [11]): development of a new generation of MAR technologies and tools to prevent impacts on groundwater pollution in climate change scenarios; **MARSoluT** (Marie Skłodowska-Curie Grant Agreement ID: 814066, LNEC was partner [31]) advance of MAR to an accepted key technology for the mitigation of bottlenecks in water resources management concepts.

MARAISER breakthrough: While conventional MAR operational tools are focused on short-term management, MARAISER will include modules with two times perspectives; 1) short-term: aiming at improving the "day by day" system operations providing a complete set of environmental, health, and performance risk indicators for both MAR schemes infiltration ponds and deep injection wells and 2) medium long-term: aiming at optimizing the volume recharged through ML algorithm and based on water resources availability and demand forecasting.

The combination of the DSS as early warning alert systems complement the traditional risk assessment methods by providing dynamic, data-driven insights that enhance

decision-making processes. With this aim, **MARAISER empowers decision-makers and operators to manage risks more effectively**, ensuring human health, and ecological integrity and optimal yield in water recharge systems.

3. Methodology: General approach

The concept of MARCLAIMED targets the effective integration of MAR with AWR in the strategic plans for water resources management at river basin scale, taking into consideration the relevant EU regulatory framework. This effective integration remains an important and complex challenge in the EU. Nevertheless, it is essential to ensure long-term sustainable use of the surface and groundwater resources. In this sense, there is a need for solutions that pave the way for increasing social acceptance, awareness, trust, and reliability on these AWR while improving their efficiency and cost-effectiveness. MARCLAIMED will integrate all the exploitable results (RES1- RES10) with a strong engagement of stakeholders, from decision makers and public/private water; and ii) promote a behavioural transition towards the awareness, reliability and trust of MAR with AWR. MARCLAIMED methodology is based on the following three pillars (Figure 1). Within the first two ones (Operational efficiency and Economic sustainability) 6 tools are developed. In Figure 1 the relationship between these 6 tools is shown.

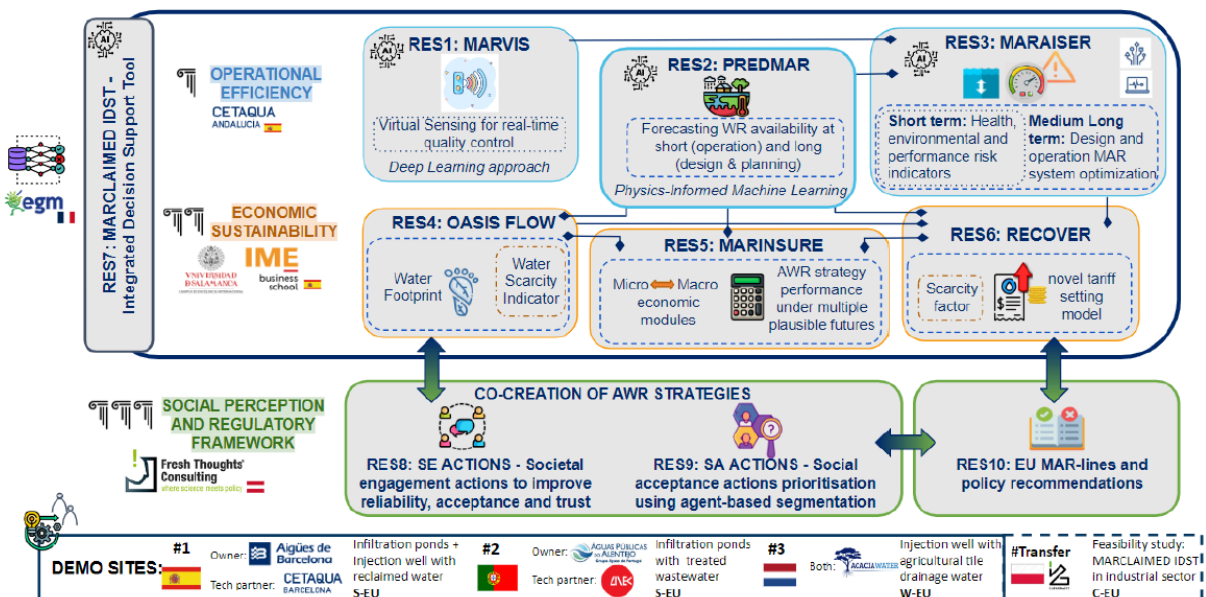


Figure 3. MARCLAIMED methodology scheme along with the participants involved, to be validated in the MARCLAIMED demo sites (Legend: light blue boxes: operational efficiency tools, orange boxes: economic sustainability tools, and green boxes: social perception assessment).

3.1. MARVIS - Virtual sensing tool for near real-time water quality monitoring

MARVIS, as a nitrates virtual sensor, requires data collection from physical sensors integrated with ML models capable of estimating the nitrate concentrations. Therefore, the first step within the general methodology consists in installing and deploying such physical sensors at strategic locations, which in this case are DS1 and DS3. These demo sites contain extensive historical records of nitrate levels, a primary groundwater contaminant affecting these region's water quality. These physical sensors require regular maintenance and cleaning to ensure accurate measurements. Parameters such as pH, turbidity, temperature, among others, provided by these sensors, are later used as data input to develop the ML algorithms.

For DS1 and DS3 a separate model will be developed to account for site-specific characteristics and variations in the corresponding environmental conditions. For DS2, a transfer learning approach using one of the models from DS1 or DS3 is planned.

Once the data of each site has been collected and validated, the following step when developing ML models are: i) feeding the collected sensor data into the corresponding ML algorithm, which learns the complex relationships between the input parameters and nitrate concentrations and ii) validating the models using a subset of data to ensure their reliability and accuracy in predicting nitrate concentrations. iii) Calibration is performed periodically by comparing the model's predictions with the actual nitrate measurements obtained from water samples in the laboratory, allowing for adjustments to improve precision.

Considering the availability of data, two different types of models will be developed during the execution of the project: an initial Extreme Gradient Boosting (**XGBoost**) model, which is less dependent on large datasets compared to the subsequent deep learning model, for which currently limited data is available. XGBoost is a powerful, optimized ML algorithm based on gradient boosting, designed for speed and performance. It improves predictive accuracy by combining multiple weak models (decision trees) while efficiently handling missing data, regularization, and parallel computation. A deep learning model is a neural network that can be also trained on labeled data (the nitrate concentrations in our case), learning patterns to map inputs to correct outputs. It adjusts weights through backpropagation and optimization (e.g., gradient descent) to minimize prediction errors. Since such models typically require large amounts of data to achieve high accuracy and generalization, they will be explored and developed in further steps when such amounts of data are available.

The **easy-to-measure parameters** employed to train the ML models are related with the concentration of nitrates in water, including aquifers. These include temperature, turbidity, conductivity, pH, total organic carbon (TOC) and UV254 absorbance, which can help indicate the sources, mobility, and transformation processes of nitrate in

groundwater systems. **Temperature** plays a key role in regulating biological and chemical reactions that affect nitrate levels. Higher temperatures can enhance microbial activity, leading to increased denitrification, a process in which nitrate is reduced to nitrogen gas under anaerobic conditions. Seasonal temperature variations also impact nitrate transport and leaching, with higher temperatures often associated with increased biological uptake and chemical transformations. **Turbidity**, which measures the presence of suspended particles in water, does not directly include nitrate since nitrates are dissolved, but it can serve as an indicator of contamination sources. Elevated turbidity often correlates with agricultural runoff or wastewater infiltration, both of which are common sources of nitrate pollution. Additionally, surface water influence in an aquifer can contribute to higher turbidity and increased nitrate levels, particularly in areas with intensive land use. **Conductivity**, which reflects the presence of dissolved ions in water, is often positively correlated with nitrate concentrations. Since nitrate exists in water as the anion NO_3^- , higher concentrations contribute to increased conductivity, especially in areas impacted by fertilizers or wastewater effluent. However, because other dissolved salts such as chloride and sulfate also affect conductivity, the relationship may not always be straightforward, requiring additional analysis to confirm nitrate sources. The **pH** of groundwater influences nitrogen transformations and nitrate stability. In neutral to slightly alkaline conditions (pH 6.5-8), nitrate remains stable in its dissolved form. In contrast, acidic conditions may inhibit microbial activity, reducing denitrification efficiency and allowing nitrate to accumulate. Conversely, highly alkaline conditions can promote ammonia volatilization, which may indirectly reduce nitrate formation by limiting nitrogen availability for nitrification. **UV254** absorbance, a measure of dissolved organic matter (related also to **TOC**), can be linked to nitrate concentrations. Organic matter and nitrate often originate from similar contamination sources, such as agricultural runoff or wastewater discharge, leading to a positive correlation between UV254 absorbance and nitrate levels. However, in anaerobic conditions where organic matter is abundant, microbial denitrification may occur, leading to a reduction in nitrate concentrations. This suggests that while UV254 can indicate the presence of contamination, its relationship with nitrate depends on the specific hydrogeochemical conditions of the aquifer. By leveraging these water quality parameters in a ML model, complex patterns and relationships governing nitrate concentrations can be uncovered, enhancing predictive accuracy and enabling more effective groundwater quality management.

With the data already gathered, the applied methodology for training ML models within MARVIS includes:

- Data preprocessing.
- Model selection.
- Validation and deployment for predictions.

The **data preprocessing** may include data cleaning, handling the missing values and normalizing features. The dataset is then separated into training and test sets, in our case employing a standard ratio of 80/20, to facilitate model training and evaluation. Our first approximation with the XGBoost model allows us to investigate which variables are relevant for predicting nitrate concentrations since XAI methods can be applied. It is not only capable of measuring the contributions to the model's predictions of each variable, but also if they are positive or negative, indicating if it matches the physicochemical expected behaviour. This can serve for further feature engineering steps when considering the final model. Cross-validation is employed during the training phase to ensure the model's robustness and generalizability; this involves dividing the training set into multiple subsets and training the model iteratively on different combinations of these subsets while validating on the remaining data. Evaluation follows, using the test set to assess the model's performance with metrics such as the Mean Squared Error (MSE) and the coefficient of determination (R^2). Hyperparameter tuning is conducted to optimize the model's performance, often using techniques like grid search. Once validated, the model is deployed to make predictions on new data, and ongoing monitoring and maintenance ensure it remains accurate and effective over time.

After obtaining enough data, a **neural network model will be chosen** for its ability to capture complex patterns through its layered architecture. It is trained using backpropagation and optimization techniques like stochastic gradient descent. The model is evaluated with the test set, and hyperparameter tuning optimizes its architecture. Deep learning models may include multiple hidden layers for hierarchical data representation. **Once validated, the model is deployed for predictions**, with ongoing monitoring to ensure accuracy. After obtaining a neural network model for each demo site where sensors were installed to develop the virtual sensor (DS1 and DS3), **transfer learning methods** can be applied to DS2, leveraging the knowledge gained from these models. The general methodology starts by trying both pre-trained models from DS1 and DS3, assessing their performance on DS2 data to determine which model performs best. The selected model's architecture and learned weights serve as the starting point for the new model at DS2. It is important to consider that physical sensors will be installed at DS2 as well in order to provide the local data. The model is then fine-tuned using this data by freezing the initial layers of the pre-trained model to retain the general features learned, while allowing later layers to adapt to the specific characteristics of DS2. The model is trained on the new dataset, adjusting weights through backpropagation and optimization techniques. During the training, hyperparameter tuning is conducted to optimize the performance of the model, ensuring that it effectively captures the unique patterns of the DS2. After evaluation, the transfer learning model is deployed to make predictions at DS2, providing a cost-effective and efficient solution by reducing the need of extensive training from scratch.

3.2. PREDMAR - AI-based water availability and demand forecasting tool

A general methodology for an AI-based water resource availability tool that operates across both short-term and long-term forecasts involves the integration of ML models with meteorological, hydrological, and socio-economic data.

In the **short term**, the tool typically uses real-time data, such as precipitation, temperature, streamflow, and water demand, to forecast daily or monthly water availability. ML algorithms, including support vector machines (SVM), decision trees, and random forests, are trained on historical data to make real-time predictions on water availability. The short-term component of the model often incorporates weather forecasts and climate predictions for the upcoming months, along with IoT sensor data from reservoirs and monitoring stations to refine predictions and improve decision-making for immediate resource allocation [32]. These models can also integrate real-time observations from satellite-based remote sensing and ground-based sensors, making them robust tools for addressing fluctuations in water availability due to seasonal variability and sudden climate events.

For **long-term** water resource forecasting, AI models utilize climate change projections, such as those from **Coupled Model Intercomparison Project Phase 6 (CMIP6)** models, and socio-economic data to predict water availability over decades. These long-term models rely on neural networks and deep learning algorithms, particularly long short-term memory (LSTM) networks, which are effective in capturing temporal dependencies and forecasting trends over extended periods. The tool integrates multiple climate scenarios, such as **RCPs and SSPs**, to provide insights into how water resources might evolve under different future emission and development pathways. By combining these scenarios with historical data, the AI tool can generate projections of water availability, helping to inform long-term planning for water storage, agricultural practices, and infrastructure development [33]. Through adaptive learning techniques, the model continuously updates its predictions as new data becomes available, ensuring accuracy and relevance across changing climate conditions. This two-tiered approach allows for a comprehensive strategy that addresses immediate needs while planning for sustainable water resource management in the face of long-term climate change.

Specifically, within the structure of the MARCLAIMED project, the overarching methodology to be implemented will be tailored to the unique requirements and conditions of each individual demonstration site, while maintaining the following core objectives:

- **DS1 - Llobregat (Spain)**
 - The development of a predictive model ***to forecast the stored surface water volume in the four reservoirs of the Ter-Llobregat system over a period of up to 6 months*** is essential for accurately assessing the drought status of the system. This model will allow water resource managers to

anticipate potential water scarcity conditions and implement proactive measures to mitigate the impacts of drought. By incorporating real-time and historical data, along with advanced AI techniques, the model will provide early warnings of impending drought conditions, facilitating informed decision-making regarding water allocation, restrictions, and drought management strategies.

To enhance the model's predictive accuracy, autoregressive AI models will be employed, specifically focusing on ML-based autoregressive models. These models will leverage time series data, accounting for past water volume patterns and environmental factors, to predict future water availability. The autoregressive nature of these models allows for the incorporation of temporal dependencies, which is crucial for capturing the cyclical nature of water availability influenced by seasonal changes, weather patterns, and other external variables. The model will offer highly accurate, short-term forecasts of water storage and provide a reliable indicator for assessing the drought status of the system over the next 6 months. This approach will empower water resource managers to adopt adaptive and data-driven strategies, ensuring sustainable water management and minimizing the adverse impacts of drought conditions.

Specific objective: *Forecast of stored surface water volume (up to 6 months) in the four reservoirs of the Ter-Llobregat system → Drought status of the system.*

- ***A 7-day flow forecast for the Sant Vicenç dels Horts (SVH) gauging station.***
The forecast is designed to support decision-making within the MAR system, particularly concerning water resource allocation, flood risk management, and environmental protection. The flow forecast incorporates a variety of hydrological and meteorological data, including historical flow records, real-time rainfall measurements, and weather forecasts. Advanced predictive models are employed to generate the forecast, taking into account the complex interactions between rainfall, evapotranspiration, and groundwater flow within the SVH catchment. The 7-day forecast horizon enables proactive decision-making, allowing for timely adjustments to MAR system operations in response to anticipated flow conditions. The forecast is regularly updated to incorporate the latest available data, ensuring its accuracy and reliability. Specific objective: ***Streamflow predictions for the next 7 days at the Sant Vicenç dels Horts gauging station.***
- ***The long-term projection (up to 2100) of stored surface water volume*** will account for the complex and multifaceted impacts of climate change, including but not limited to rising temperatures, altered precipitation patterns, and increased frequency of extreme weather events. Additionally, the effects of global warming, such as glacial melt and sea-level rise, as well as demographic trends such as population growth and changes in water

usage patterns, will be incorporated into the model. These factors are critical for understanding the potential shifts in the water availability and drought status of the system. To enhance the accuracy and robustness of the prediction, advanced data engineering techniques will be applied to construct comprehensive datasets, including historical and projected climate and hydrological data.

The data will be processed through feature engineering, where key variables such as temperature, precipitation, runoff, evapotranspiration, and demographic shifts will be systematically transformed to generate meaningful features that enhance model performance. This may involve the creation of new variables based on existing data, such as seasonal trends, anomaly detection, and moving averages, which are essential for improving the predictive capacity of the model. By integrating this rich set of features with climate projections and socio-economic scenarios (using RCPs and SSPs), the model will forecast the annual stored surface water volume and drought status of the system with high accuracy.

To achieve this, AI models, including ensemble methods like Random Forests and Gradient Boosting Machines (GBM), will be employed for their ability to handle large datasets and capture complex non-linear relationships between variables. These AI-driven models will allow stakeholders to proactively develop and implement adaptive water management strategies, mitigating the potential impacts of future droughts, safeguarding water security, and ensuring the long-term sustainability of water resources in the face of climate change and shifting socio-economic conditions. Specific objective: ***Forecast of annual stored surface water volume (up to 2100) in relation to climate change, global warming, and demographic trends .→ Drought status of the system.***

- **DS2 - Comporta (Portugal)**

The MARCLAIMED project is developing an AI-based predictive model to forecast long-term surface water surpluses at basin and sub-basin scales up to the year 2100. This model will assess the feasibility of new MAR schemes, utilizing excess surface water for aquifer recharge. The project integrates insights from the daily sequential hydrological model of the land-water balance, BALSEQ, and applies the methodology from DS1-Llobregat for using RCPs and SSPs to the Sado basin in Portugal. This integration will provide a comprehensive evaluation of long-term surface water availability and its potential for MAR, supporting sustainable water management under future climate change and socioeconomic uncertainties.

To achieve the goal of predicting water availability, the BALSEQ model, , was employed to calculate potential surface water surpluses. This calculation utilized data projections

applied to 6,473 microbasins within the Sado basin for the historical period spanning from 1941 to 2019. The core variable of interest for deriving potential natural water surpluses will be computed assuming that generated surface runoff can be captured for infiltration using a formula akin to the water balance equation:

$$E = P - ET - I$$

Where E denotes direct runoff, P precipitation, ET denotes actual evapotranspiration,, and I stands for deep infiltration.

The calculated E variable for the historical period (1941-2019) will be enriched and validated with historical data sourced from CMIP6. This dataset encompasses crucial parameters such as precipitation, evapotranspiration, runoff, and temperatures, and will be integrated using the established methodology from the Llobregat demo site. This approach ensures a robust and reliable calibration of the model against observed data.

To project future water surpluses up to the year 2100, the model will leverage data derived from selected RCPs and SSPs scenarios. These scenarios provide a range of plausible future climate and socioeconomic conditions, enabling the assessment of potential impacts on water availability under different global development trajectories. By integrating these diverse data sources and methodologies, this comprehensive approach aims to provide a robust and accurate prediction of future water surpluses in the Sado basin, supporting informed water management and adaptation strategies.

After the dataset is assembled, incorporating both historical and projected data, AI models will be employed to predict the behavior of the target variable, which in this case is surface surplus as defined by BALSEQ. These AI models will be trained on the comprehensive dataset, enabling MARCLAIMED to forecast future surface water surpluses and provide valuable insights for developing sustainable water management strategies, particularly within the context of MAR schemes. By considering a wide range of climate change scenarios, encompassing varying temperature and precipitation patterns, as well as socioeconomic scenarios that account for shifts in population, land use, and water demand, the project's results will be robust and applicable to a multitude of potential future conditions. This approach ensures that the water security of the Sado basin can be maintained under an integrated management of surface and groundwater resources, making use of every potential water surplus, even in the face of unforeseen challenges. Furthermore, the insights and methodologies developed within this project have the potential to be applied to other regions grappling with similar water resource management issues, thereby promoting water security and sustainability on a broader scale.

- **DS3 - Texel (The Netherlands)**

The forthcoming irrigation season's water extraction volumes will be meticulously determined with a primary focus on optimizing the recharge system's recovery efficiency. This will be achieved by utilizing advanced forecasting models that leverage predicted precipitation levels for the recovery period. These models will integrate a multifaceted approach, incorporating historical climatological data and real-time values of relevant variables. Furthermore, the models will assimilate data from sophisticated physical models, such as the ECMWF's SEAS5, employing a modified methodology to enhance the precision and reliability of the resulting estimations. By synergistically integrating these diverse data sources and advanced modeling techniques, the overarching goal is to ensure that the extracted water volumes are harmonized with the anticipated precipitation patterns and recharge dynamics. This proactive and adaptive management strategy is designed to maximize the overall efficiency and sustainability of the irrigation system, ensuring both optimal agricultural yields and the long-term preservation of vital water resources.

3.3. MARAISER - Short and long-term MAR performance optimization tool

The **main objective of the MARAISER Tool** is to minimize the health, environmental and operational risk linked to the use of AWR in MAR systems. The specific objectives of the tool are:

- Quality control of the AWR from the delivery point in the infiltration/injection zone.
- Monitoring of the effects/impacts on the quality and hydrodynamics of the aquifer.
- Performance control of the MAR system (infrastructure).
- Serve as Early Warning Alert System and provide recommendations for implementing actions that minimize potential impact on groundwater and recharge infrastructure (MARCLAIMED IDST).

MARAISER Tool is composed of two modules:

- **Module 1 (short term):** Is the operational module for managing health, environmental, and operational risks linked with the use of AWR for MAR (based on Risk Management Plans for MAR. The objective is to provide different risk indicators to ensure optimal quality of groundwater resources and optimal performance of MAR systems.
- **Module 2 (long term):** Is the planning and optimization module for MAR system infrastructures based on predictions of available water resources and demands.

The **main objective of Module 1** is to deliver a set of environmental, health, and performance risk indicators (such as microbial, clogging, salinization, water exploitation and the recovery efficiency (ratio between the extracted volume and the infiltrated volume)) based on water quality and hydrodynamic data, improving the operation efficiency in terms of water quality monitoring while providing operational risk assessment. To improve the "day-by-day" system operation for different MAR schemes (infiltration ponds and deep injection wells; Aquifer, Storage and Recover (ASR) and

Aquifer, Storage, Transfer and Recover (ASTR)), these indicators and its behaviour and evolution will be used as an early warning system ones integrated in the MARCLAIMED IDST.

With this purpose the MARAISER is fed with the data registered by sensors (online/offline commercial sensors and MARVIS virtual sensors) strategically located in selected risk monitoring points of the MAR systems as well as with data based on the laboratory analysis from the water sampling campaigns of recharge water, soil, and groundwater. Nevertheless, selection of variables to be monitored may vary among demo sites as it will be based on regional policies and regulation and will follow previously defined instructions to ensure the health of the aquifer.

The **main objective of Module 2** is to estimate the volume available of AWR that potentially could be used for MAR according to the water resources availability forecasted at long term by PREDMAR Tool.

At this stage of the project, PREDMAR provides long term forecasting for DS1 and DS2 only, but the possibilities of providing long term forecasting for DS3 will be assessed during the implementation and validation period (WP4, T4.1). For DS2 the forecasted water surpluses at the basin/sub-basin scale will be directly an output to assess the feasibility of implementing new MAR schemes, but as this result is related with the conventional water resources availability, won't be on the scope of MARAISER long term module.

Therefore, this module will be developed only for DS1 integrating the forecasted status of drought in Ter-Llobregat System at long term provided by PREDMAR. The outputs will provide information on the reclaimed water available for MAR purposes.

The **MARAISER Module 1 (short term) methodology approach**, consisting in identifying the Risk Monitoring Points (RMP), the parameters to be followed in this RMP, the alert and critical thresholds and the main actions to be taken under a risk event, **is based on the Risk Management methodologies** described in the key guidelines **for water reuse**:

- **WHO** [34] uses a semiquantitative matrix to rank risk based on likelihood and severity. The organization prioritized the evaluation of microbial hazards in water due to their acute health risk, defining the Quantitative microbial risk assessment (QMRA) and also which steps are followed to determine it, like problem formulation, exposure and dose-response assessment, and risk characterization. WHO evaluates severity based on health impacts, which depend on concentration, toxicity, exposure frequency, and duration. Additional WHO methodologies include semi-quantitative risk ranking, hazard identification based on past contamination events, and the multi-barrier principle, which mitigates risks by applying multiple treatment layers.
- The **EPA** [35], employs two distinct risk assessment approaches: Human Health Risk Assessment, which centers on the potential health effects of water

contaminants, and Ecological Risk Assessment, which evaluates the environmental impacts of stressors like chemicals and land-use changes. Both approaches follow a structured process involving planning, hazard identification, exposure analysis, and risk characterization. The EPA's risk characterization process considers the probability and impact of various environmental stressors, including chemicals, land-use changes, and invasive species. Risk scoring is incorporated into monitoring systems to enable data-driven decision-making.

- The **CEDEX** methodology [36], aligned with WHO and EU regulations (2020/741[7] -minimum requirements for water reuse Directive- and 2024/1765 [23] -technical specifications of the key elements of risk management-), quantifies risks in reclaimed water reuse. It calculates overall risk probability (Pf) based on three factors: i) P_{os} : Likelihood of a hazardous event occurring (historical data review), ii) P_{pres} : Probability of contaminant presence (system configuration and operation) and iii) P_{exp} : Probability of human exposure (no exposure = no risk). By multiplying the probability and severity scores, a risk matrix (Table 3.3.a) can be developed in line with the guidelines of the three different methods. The frequency of monitoring and data collection influences the accuracy of risk characterization, ensuring that control measures respond effectively to changes in contamination levels.

Table 3.3.a Semi-quantitative risk assessment matrix (Source: Tool 3.4 of WHO Sanitation Safety Planning Manual, 2016) used in all three methodologies.

Likelihood (L)	Severity of consequences (S)				
	Insignificant (1)	Minor (2)	Moderate (4)	Major (8)	Catastrophic (16)
Rare (1)	1	2	4	8	16
Unlikely (2)	2	4	8	16	32
Moderate likely (3)	3	6	12	24	48
Likely (4)	4	8	16	32	64
Almost certain (5)	5	10	20	40	80
Risk Score $R=L \times S$	< 6	7-12		13-32	> 32
Risk level	Low Risk	Medium Risk		High Risk	Very High Risk

These approaches described above, considers the effect on the public health, but regarding MAR with AWR, both the environmental and operational related hazards and its likelihood and severity has to be also considered.

For this reason, for the development of MARAISER, the methodology has been defined as **MAR Safety Plans (MAR-SP) Risk Assessment including health, environmental and operational risks**. The objective is to conduct the Risk assessment for the **identification of RMP, parameters, thresholds and the definition of actions to be taken** under an hazardous event may occur.

Also, there are **two possible scenarios depending on the historical data available for the MAR-SP Risk Assessment implementation**. In case that the data available is enough to provide robust statistical significance the methodology applied will be the MAR Risk assessment itself, but if not, the identification of RMP, parameters and thresholds will be performed based on expert criteria.

The criteria for defining the likelihood and the severity on the MAR risk assessment (per potential RMP and parameter measured) is as follows:

- **Premise:** i) Preserve and improve the groundwater body's global status, ii) Comply with legal and permit requirements (if there is lack of legal or permit quality thresholds, expert criteria will apply) for recharge water quality and iii) take as a base line the native water quality status of the groundwater body or aquifer sector.
- **Likelihood:** how often the legal threshold has been overreached.
- **Severity:** to what extent the threshold defined (by legislation or expert criteria) has been exceeded.
- In case that exist RMP downstream in the aquifer, also will be considered if there is a significant decrease in the risk.

The general steps to be followed for this risk assessment are:

- Generate a **conceptual/functional scheme** of the main infrastructures involved in the MAR process and identify the main control points, the parameters measured and how (and how often) it is measured (on-line/offline sensors or laboratory analysis).
- Identify the **legal or expert criteria thresholds** and the **potential risk indicators** that apply in the specific case studio.
- Identify for which **parameters** there are **enough historical data** (at least 3 years) to perform the **probabilistic analysis** and if not apply the **expert criteria**.
- Identify the **RMP, the parameters and the mitigation measures to be taken**.
- Identify the **kind of data** provided in the RMP: i) **dynamic** data from online sensors, ii) **semi-dynamic** data from continuous monitoring sensor but offline, ii) **static** data from laboratory analysis.
- Afterwards, a flowchart will be produced for the codification and integration in the MARCLAIMED IDTS.

Regarding Key Risk components and indicators, ensuring the safety and sustainability of MAR and water reuse systems requires a comprehensive understanding of risk factors that could compromise water quality and public health, potentially impacting both human health and the environment. A primary concern is groundwater contamination, where pollutants infiltrate aquifers, degrading water quality and posing risks to human consumption and the environment. Changes in groundwater availability and quality can disrupt aquatic ecosystems, affecting species reliant on stable conditions. Climate-related events, such as droughts and floods, further exacerbate these risks by altering pollutant transport and water availability.

Even infrastructure reliability is critical. Issues like corrosion, clogging, and mechanical failures in storage and distribution systems can lead to supply disruptions and increased contamination risks. Process control failures, such as inadequate disinfection or filtration, may result in the release of contaminated water, threatening public health.

Therefore, the main indicators identified for MARAISER are shown in Figure 3.3.a.

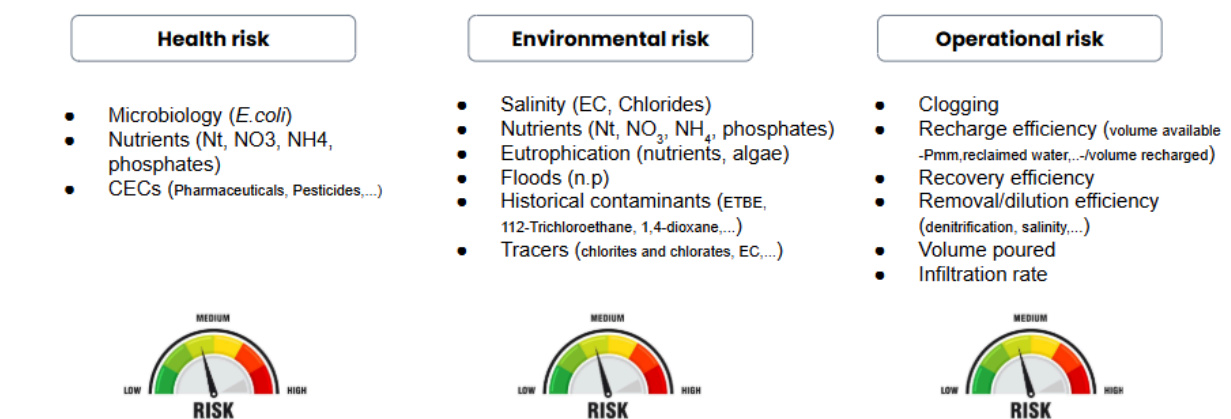


Figure 3.3.a. Health, environmental and operational risk indicators.

The **MARAISER Module 2 (long-term) methodology approach** consists in calculating the reclaimed volume available in the Ter-Llobregat system and therefore, the number of additional MAR systems that could be projected.

These estimations will be performed through an excel sheet considering the following equation:

$$V_A = V_{TC} - V_C$$

$$N_{MARS} = V_A / V_{GIC}$$

V_A = Reclaimed water volume available.

V_{TC} = Reclaimed volume total capacity on the Ter-Llobregat system (corrected by the real capacity).

V_C = Reclaimed water volume compromise depending on the System Drought Status.

The V_C is a function of the **PREDMAR long term output in DS1** and also of the setpoints provided by the Agència Catalana de l'Aigua, which dictate the volume of reclaimed water compromised for certain uses (such as ecological flow preservation, indirect reuse, among others). The PREDMAR long term output consists in the forecasted Ter-Llobregat System Drought Status which provides the expected value considering 5 SSP Scenarios. The characteristics of these SSP Scenarios are presented in section 4.1.2 (Table 4.1.2.a).

V_{GIC} = General infiltration/injection capacity of an average MAR operation in Ter- Llobregat system.

N_{MARS} = Number of MAR systems to be projected (from 1-2 hm³/year of infiltration/injection capacity -this has been ranged considering the current MAR systems recharge capacity operating in the area-).

This estimation will be provided annually from present to 2100, at the time that **PREDMAR long term in DS1** is executed.

4. Description of the Digital Tool

4.1. Conceptual scheme and main functionalities

4.1.1 MARVIS - Virtual sensing tool for near real-time water quality monitoring

MARVIS is a digital tool designed for continuous real-time monitoring of nitrate concentrations in water systems involved in MAR. This virtual sensor integrates data processing algorithms and ML techniques to provide accurate estimations of nitrate levels without the need for the physical sensor corresponding to that variable. The digital tool intakes and processes the data from physical sensors that account for the independent variables (e.g., pH, temperature, conductivity), including automated data validation, data cleaning and calibration of features if necessary. Afterwards, nitrate concentration predictions are performed at the same frequency at which sensors generate the data. Depending on the sensor and the demo site this is either 5 or 10 minutes. Finally, the results of such predictions will be delivered as reports taking into account the end-user requirements, which may vary from demo site to demo site. Additionally, the tool will allow the end-user to receive notifications when nitrate levels exceed predetermined thresholds. This will be specified in the MARCLAIMED IDST.

Conceptual Scheme

The general conceptual scheme for MARVIS, which is valid for both DS1 and DS3 is shown in Figure 4.1.1a. As previously mentioned, the solution integrates multiple data sources

that account for physicochemical parameters that ultimately feed the model capable of predicting the nitrate concentrations. Noteworthy, to develop MARVIS as an online sensor that provides the nitrates levels, it is also necessary to obtain this parameter in order to train the model. Once deployed, future inferences may not require this parameter. However, if present can serve to retrain the model as desired considering potential variations in the data distribution. This will also be specified within the IDST. Historical data for each demo site, if available, allows us to investigate the distribution of the physicochemical parameter levels involved in MARVIS. Finally, measures from the online sensors such as the nitrates parameters are validated with analytical tests in the laboratory. Such experiments can be used to validate and calibrate the measures observed in the online sensors.

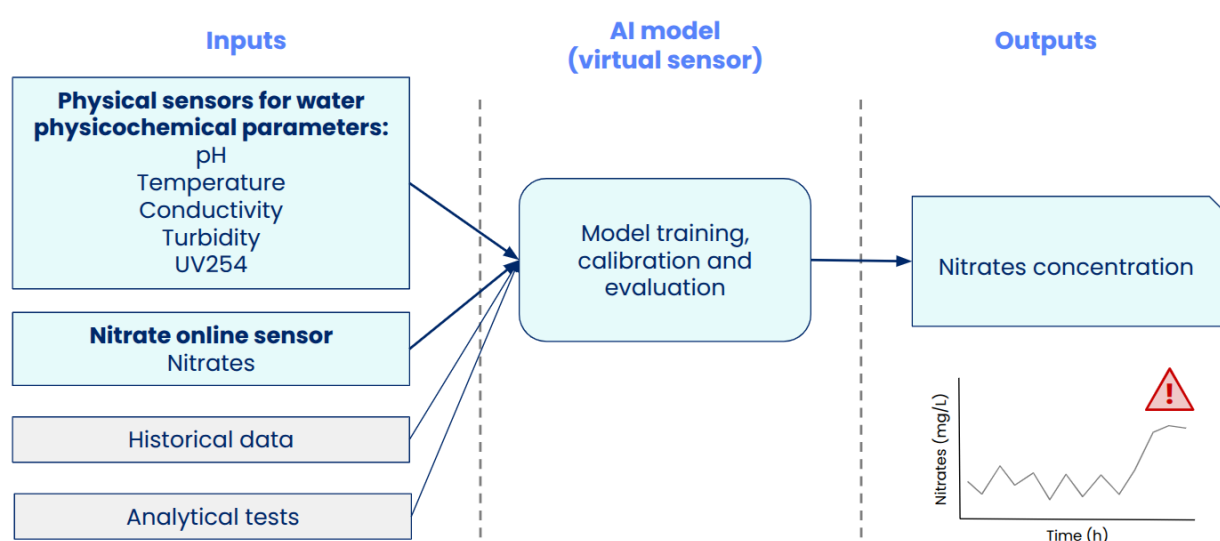


Figure 4.1.1.a Conceptual scheme of MARVIS tool for DS1 and DS3.

In the case of DS2, the general transfer learning method will follow the conceptual scheme shown in Figure 4.1.1b. At DS2, the same physicochemical parameters will be measured by means of physical sensors, obtaining representative data for this specific demo site. Such data will be employed to train a specific model for DS2 considering the models from DS1 or DS3 as starting points. These techniques can be applied both for XGBoost and neural networks models. In XGBoost, partial retraining works by freezing existing trees and adding new ones to refine predictions. Since trees are built sequentially, their structure cannot be modified once trained. Instead, you continue training by appending more trees or use incremental updates with new data, though this is less flexible than full retraining. In neural networks, you can freeze certain layers and update others, making them more adaptable. Lower layers, which capture general features, are often kept unchanged while fine-tuning higher layers with new data. Unlike XGBoost, neural networks adjust existing weights through backpropagation rather than

adding new structures. Optimally, transfer learning should be applied using neural networks, as sufficient data will be available at that stage of the project to train deep learning models effectively.

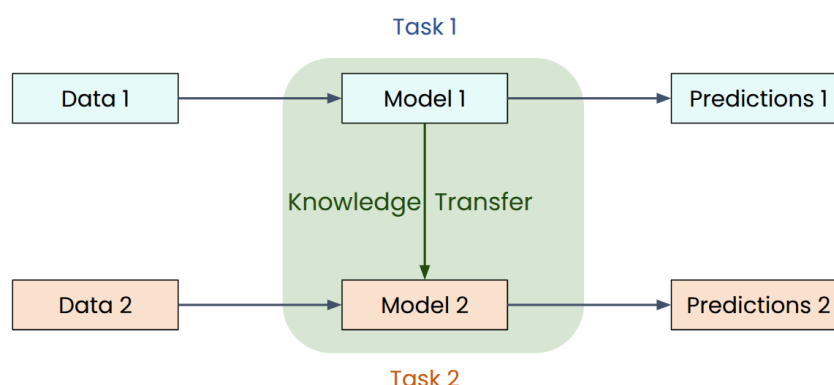


Figure 4.1.1.b General conceptual scheme of MARVIS tool development for DS2 applying transfer learning.

Main functionalities

The primary function of the virtual sensor is to **provide real-time prediction of nitrate concentrations based on multiple water quality parameters**. The system delivers predictions at regular intervals of 10 minutes, the minimum frequency at which the physical sensors are planned to measure. The expected accuracy ranges at least 80-90%, which is consistent with state-of-the-art ML applications in water quality monitoring. The implementation **includes a sophisticated alert system that triggers notifications when nitrate levels exceed a critical threshold** that depends on the demo site needs, enabling prompt response to potential water quality issues. The model undergoes monthly retraining cycles to maintain prediction accuracy and adapt to changing environmental conditions. Current system limitations primarily revolve around data availability constraints and the pending implementation of automated retraining procedures. The modular design allows for future enhancements and optimization of prediction accuracy as more data becomes available. Reporting will be integrated within the MARCLAIMED IDST, generating real-time nitrate concentration predictions, compiled detailed daily reports that include hourly predictions, as well as weekly summary reports for longer-term trend analysis. Reports can feature comprehensive tables of prediction values accompanied by intuitive visualization graphs that illustrate temporal trends and patterns in nitrate concentrations. The reporting system maintains a detailed history of alerts triggered by exceedances of the specified concentration threshold, providing valuable information for regulatory compliance and operational decision-making. This comprehensive reporting structure enables efficient monitoring and informed decision-making for water quality management, while maintaining flexibility for future enhancements based on user feedback and evolving requirements.

MARVIS interaction with other Tools of the MARCLAIMED IDST

MARVIS interacts with MARAISER by providing critical water quality status information based on nitrate concentration monitoring. Through its virtual sensing capabilities, MARVIS provides the real-time concentration of nitrates as potential contaminants accounting as a water quality parameter. Potentially, MARVIS could also generate simplified quality indicators that MARAISER uses to support decision-making in MAR operations. When MARVIS detects nitrate levels within acceptable ranges, indicating suitable conditions for recharge activities. Conversely, if nitrate concentrations approach or exceed threshold values, MARVIS issues an 'alert' status, enabling MARAISER to recommend appropriate adjustments to MAR operations or trigger necessary preventive actions.

4.1.2 PREDMAR - AI-based water availability and demand forecasting tool

The primary goal of this tool is to leverage AI techniques to predict the availability of water resources across both short-term and long-term time horizons. By providing accurate and timely forecasts of water availability in reservoirs, this AI-powered system aims to significantly enhance water resource management practices. The system will take into account a wide range of factors that can impact water availability, including precipitation patterns, temperature trends, evaporation rates, and human water usage. The ability to generate reliable short-term and long-term forecasts will enable water managers to make informed decisions about water allocation, reservoir operations, and drought management strategies, ultimately leading to more sustainable and efficient use of water resources.

Conceptual Scheme

The architecture of the system is designed to deliver accurate and reliable predictions of water availability for both short-term and long-term scenarios. This is accomplished through the integration of advanced AI-driven predictive models and a comprehensive data infrastructure that incorporates historical, real-time, and forecasted meteorological and hydrological data. The system employs a data fusion approach to combine diverse datasets from multiple sources, including precipitation, temperature, reservoir levels, and streamflow measurements. This methodology ensures that the model can account for all relevant variables, improving the accuracy and reliability of water resource predictions. In the short-term, the system focuses on predicting water availability within the next few days or months. The AI models process real-time data to forecast variables like reservoir volumes and precipitation events, providing actionable insights for day-to-day water management.

For long-term predictions, the system integrates climate projections using RCPs and SSPs, key frameworks for assessing future climate scenarios. RCPs represent different greenhouse gas concentration trajectories, ranging from low emissions scenarios (RCP2.6) to high emissions scenarios (RCP8.5), while SSPs combine various socioeconomic

development pathways that influence global emissions and water demand patterns. By applying these scenarios, the system forecasts the potential impacts of climate change on water availability through the end of the century. This allows for long-term water resource planning, taking into account factors such as population growth, urbanization, and climate-induced variability in water supply. Using AI, the system can analyze and integrate these complex models to produce projections that guide sustainable water management strategies under different climate and socioeconomic conditions.

In both short and long-term contexts, the AI-driven predictive models identify complex patterns and relationships within vast datasets, allowing for more precise and dynamic forecasting than traditional methods. By leveraging the computational power of AI, the system provides a forward-looking approach to water resource management, enhancing decision-making for the Ter-Llobregat basin under current and future climatic conditions.

Main Functionalities

1. Short-Term module in DS1 (Llobregat, Spain)

Objectives:

- Forecast the volume of water stored in the four reservoirs of the Ter-Llobregat system in the short term.
- Assess the current state of drought in the system based on water availability.
- Forecast streamflow at the SVH gauging station over the next 7 days

Functionality:

The fulfillment of the objectives involves the integration of autoregressive ML models to predict future water resources. The conceptual approach to this forecasting system is based on the following components (Figure 4.1.2.a & 4.1.2.b):

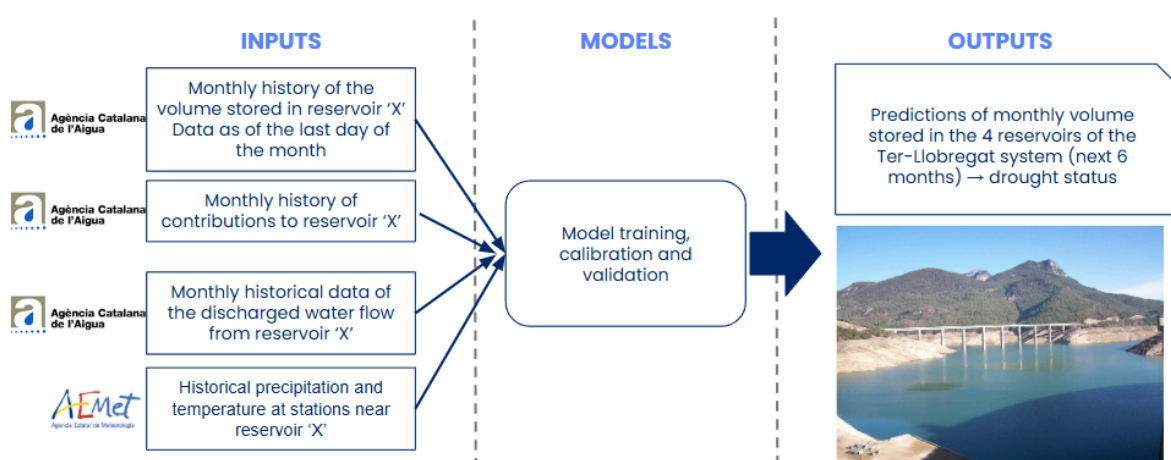


Figure 4.1.2.a Short-Term Prediction of Stored Water Volume conceptual scheme

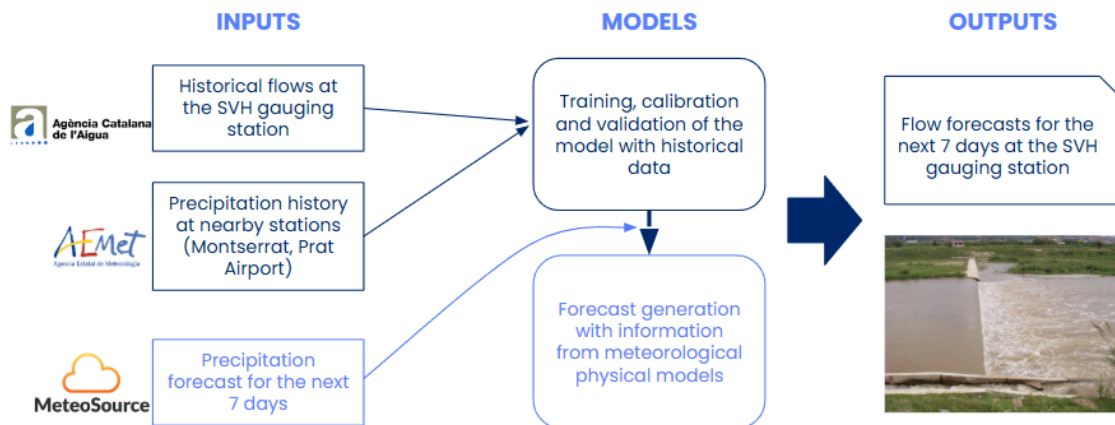


Figure 4.1.2.b Short-Term Forecast streamflow at Sant Vicenç dels Horts

Data Acquisition and Preprocessing:

To create a robust predictive model for streamflow and reservoir storage, it is essential to gather extensive historical data. This includes:

- **Streamflow Data:** Collect past streamflow measurements at relevant points within the watershed. This data should ideally span several years to capture seasonal and interannual variability.
- **Reservoir Storage Data:** Compile historical records of the volume of water stored in the reservoir over time. This data is crucial for understanding how the reservoir responds to inflows and outflows.
- **Meteorological Data:** Gather historical meteorological data for the watershed, including:
 - **Precipitation:** Rainfall and snowfall data at multiple locations within the watershed.
 - **Temperature:** Air temperature data to account for its influence on evapotranspiration and snowmelt.
 - **Other Relevant Variables:** Depending on the specific watershed and modeling objectives, additional meteorological variables may be necessary, such as humidity, wind speed, and solar radiation.

The collection of this historical data should ideally cover a period of several years, and the data should be of high quality and resolution to ensure the accuracy and reliability of the predictive model.

The data will be cleaned, ensuring no missing values, outliers, or inconsistencies. Missing data will be imputed using techniques such as linear interpolation or more advanced methods (e.g., Mice Forest Imputation).

The data will be normalized and standardized to ensure that all inputs are on the same scale for better performance of ML models.

Model Selection and Autoregressive Mechanism:

- Autoregressive ML models will be used to capture the temporal dependencies in the data and to account for complex nonlinear relationships in the data that traditional ARIMA models may not capture.
- Model Training and Validation:
 - Training the models will be performed using historical data.
 - Cross-validation techniques will be used to evaluate the models' generalization ability and avoid overfitting. A holdout method will reserve a portion of the historical data for validation.
 - Hyperparameter optimization will be performed using grid search or random search methods to maximize forecasting accuracy.

Forecasting:

Once trained and validated, the autoregressive models will be used to forecast the target variables in the corresponding prediction horizons. The models will generate flow predictions by incorporating the previous data points and adjusting the forecast using the temporal dependencies learned during training.

For each prediction, the models will output a prediction interval, indicating the level of certainty in the predictions.

2. Long-Term module in DS1 (Llobregat, Spain)

The Ter-Llobregat system, a critical water supply network for Catalonia, faces increasing stress from climate change, population growth, and water demand. To ensure long-term sustainability, a robust AI-driven predictive model should integrate climate scenarios (RCPs, SSPs), hydrological models, and adaptive water management strategies.

The long-term assessment involves two key components:

- Climate Projections: Based on CMIP6 SSP-RCP [37] frameworks, defining future climate trends.
- Hydrological Response: Using ML models to project water availability.

Using CMIP6 [9] projections, we classify future conditions based on RCPs and SSPs, for example:

Table 4.1.2.a SSP scenarios

Scenario	Climate Impact	Socioeconomic Context	Expected Water Stress
SSP1-1.9 (Sustainable)	Minimal warming (~+1.4-1.7°C)	Strong climate action, renewable energy transition, global cooperation	Low
SSP1-2.6 (Sustainable)	Moderate warming (+1.5-2°C)	Green economy, efficient water use	Low to Moderate
SSP2-4.5 (Intermediate)	Warming of +2.5-3°C	Business-as-usual, stable policies	Moderate
SSP3-7.0 (Regional Rivalry)	Warming of +3.5-4.5°C	Limited cooperation, high population growth	High
SSP5-8.5 (Fossil-Fueled Growth)	Extreme warming (+5°C)	High emissions, rapid economic growth	Severe

Each scenario influences precipitation patterns, temperature trends, and hydrological cycles, which directly affect the Ter-Llobregat reservoirs.

To enhance the precision and granularity of long-term water availability predictions, an advanced AI-driven model is being developed. This sophisticated model will leverage AI-based downscaling techniques, which are designed to refine the coarse spatial resolution of global climate models (GCMs). By dynamically adjusting and tailoring these global models to specific regional contexts, the AI-enhanced model can generate more accurate and localized projections of future climatic conditions. This, in turn, will enable more reliable and nuanced assessments of long-term water availability, taking into account the unique hydrological characteristics and potential climate change impacts of different regions (Figure 4.1.2.c).

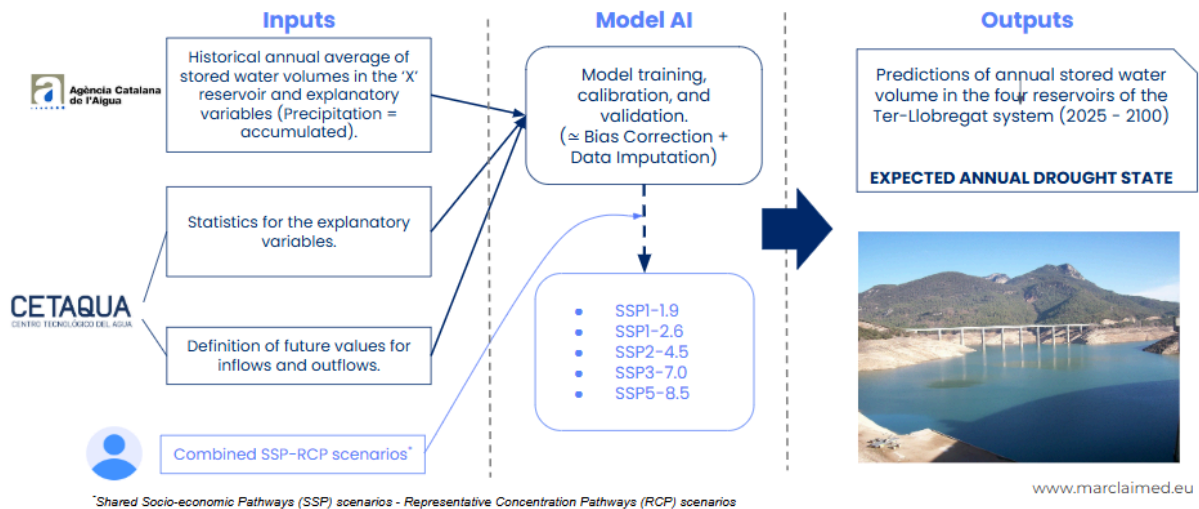


Figure 4.1.2.c Long-Term Forecast of Drought status of the Ter-Llobregat system

3. Long-Term module in DS2 (Comporta, Portugal)

The conceptual framework initially designed for the Llobregat demo site will serve as a foundation for our approach in the Comporta demo site. This framework, while robust in its design, will be meticulously adapted to align with the unique hydrological, ecological, and socioeconomic dynamics of the Sado River basin. This process of adaptation will ensure that the framework is not only theoretically sound but also practically applicable to the specific context of the Sado River basin, thereby maximizing the effectiveness and relevance of our interventions (Figure 4.1.2.d).

The integrated model will incorporate climate projections based on SSP-RCP scenarios, AI for water availability forecasting, and. Specifically, we will consider:

- CMIP6 Climate Projections: Evaluating climate change scenarios (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5) to assess trends in temperature, precipitation, and hydrological stress in the Sado basin.
- AI-Based Hydrological Modeling: Implementing ML algorithms to enhance prediction accuracy for water surpluses.

This approach will provide a robust and transferable methodology for sustainable water resource planning in Comporta, leveraging the insights gained from the Llobregat pilot implementation.

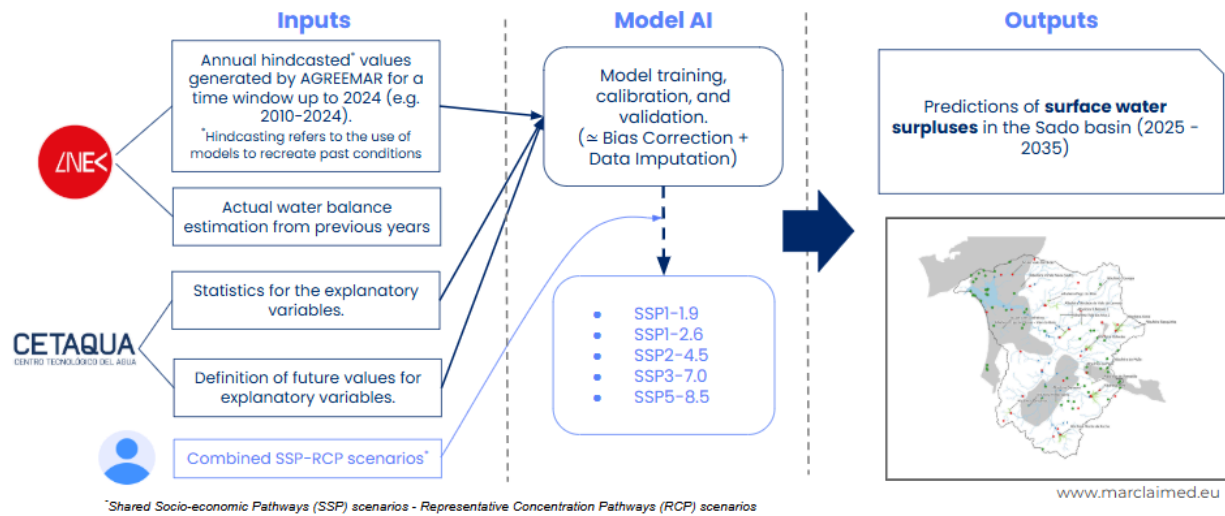


Figure 4.1.2.d Long-Term Predictions of surface water surpluses in the Sado basin

4. Short-Term module in DS3 (Texel, The Netherlands)

The proposed conceptual scheme (Figure 4.1.2.e) integrates physical climate models, data-driven AI techniques, and real-time observational data to enhance the forecasting of climatological variables. This framework is structured into five key components:

- ECMWF Climate Models: Utilizing SEAS5 for seasonal forecasting and ERA5 for reanalysis data.
- Meteorological Observations: Incorporating data from ground stations, ensuring localized corrections.
- Locational Interpolation: Adjusting coarse-resolution model outputs to site-specific conditions through spatial downscaling.
- Evaluate the recovery efficiency of the aquifer recharge system by conducting a comprehensive analysis of historical datasets from previous years.

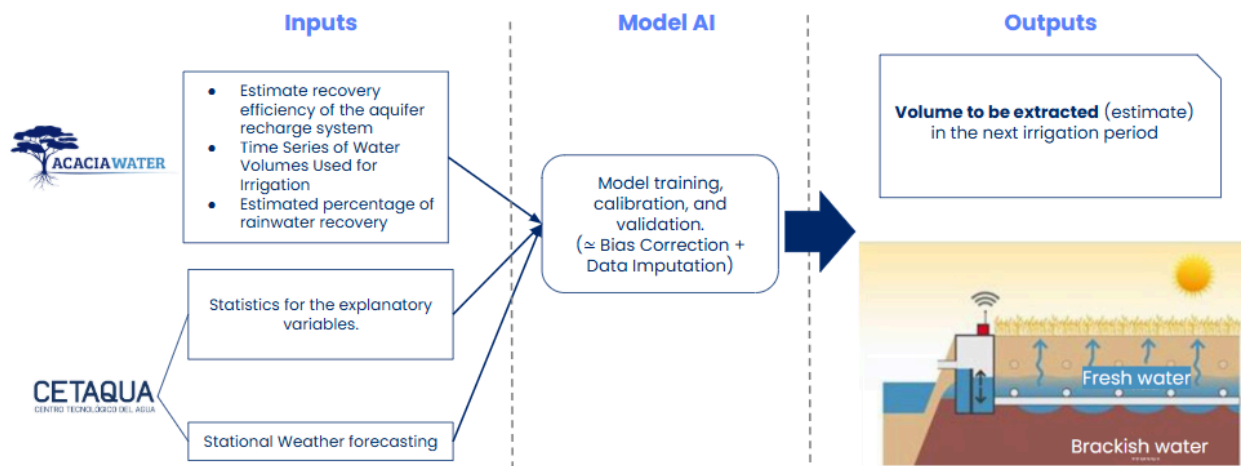


Figure 4.1.2.e Volume to be extracted (estimate) in the next irrigation period on Texel

PREDMAR interaction with other Tools of the MARCLAIMED IDST

- *MARAISER (short-term)* incorporates advanced functionalities for daily flow forecasting of the Llobregat River at the SVH station. This forecasting is powered by the PREDMAR Tool in DS1, which leverages real-time meteorological and hydrological data to predict daily water flow rates. The model uses historical data, current weather conditions, and predictive analytics to estimate future flow dynamics, providing operators with valuable insights into immediate water availability. These predictions allow for timely adjustments in water resource management, ensuring optimal allocation and mitigating potential risks associated with sudden fluctuations in river flow, such as those caused by storms or seasonal changes.
- *MARAISER and MARINSURE (long-term)*, the PREDMAR Tool plays a key role in forecasting the drought system status over extended periods. Using a combination of historical data, climate projections, and scenario-based modeling, PREDMAR is able to predict the evolution of drought conditions in the Ter-Llobregat system. The tool's ability to assess and monitor drought status over the long term is a critical component for supporting informed decision-making regarding water storage, usage restrictions, and investment in water infrastructure.

4.1.3 MARAISER - Short and long-term MAR performance optimization tool

The **conceptual scheme** of MARAISER is shown in Figure 4.1.3. In this Figure is represented both Module schemes (Module 1 (short term) in blue and Module 2 (long term) in orange), with the Input data needed, how is it processed and with the output obtained.

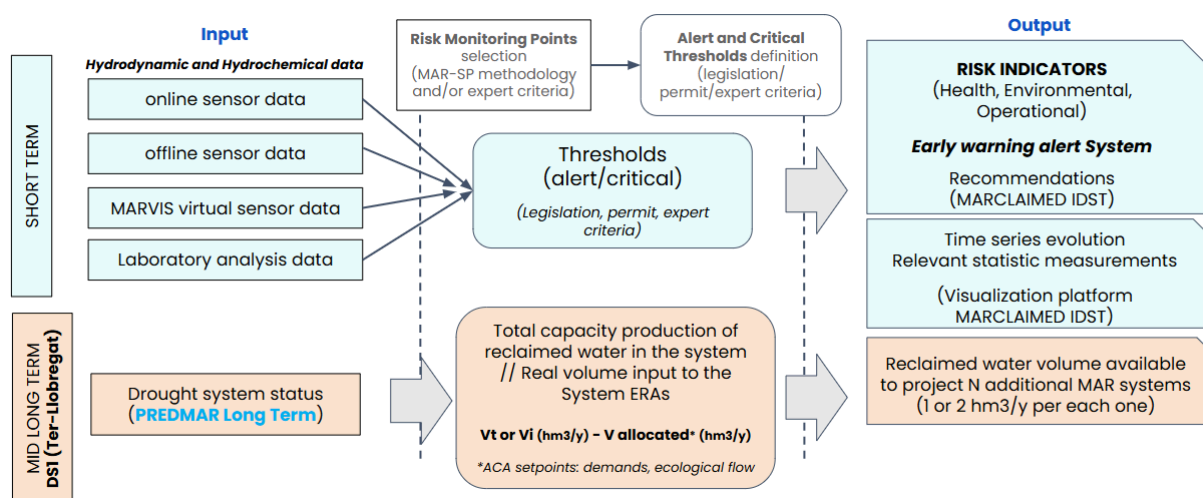


Figure 4.1.3. Conceptual scheme of MARAISER Tool.

MARAISER interaction with other Tools of the MARCLAIMED IDST

MARAISER Module 1 (short term) incorporates:

- The near-real time continuous measurements of NO_3 from **MARVIS Tool** in DS1 and DS3 (more detail is provided in section 5.3.1 and 5.3.3, respectively).
- The daily flow forecasting in Llobregat river by **PREDMAR Tool** in DS1 (more detail is provided in section 5.3.1).

MARAISER Module 2 (long term) incorporates:

- The drought system status forecasted in DS1 by PREDMAR Tool.

For the **Module 1 (short-term)**, once the MAR-SP Risk Assessment has been performed and the RMP, parameters and alert and critical thresholds have been identified, there is an important step in the process that is the identification of the kind of data that will feed the tool (inputs, Figure 4.1.3). As has been explained in the Methodology (section 3.3), there are 3 kinds of data (dynamic, semi-dynamic and static) that will determine the capacity of reaction in case an hazardous event will take place. All of this data has to be collected as much frequency as it could be possible for both detection of an overreachment of threshold and for control that the optimal conditions in the MAR system have been restored. This data collection will be explained in section 4.2.3.

- **Input data:** Hydrodynamic and hydrochemical data from:
 - Online, offline commercial sensors
 - Online virtual sensor (MARVIS, only in DS1 and DS3)
 - Laboratory analysis.
 - Forecasted daily flow output from PREDMAR short-term tool in DS1.

- **Setpoints:** Determined by legislation, permit or expert criteria, depending on the stage of the MAR system or downstream in the aquifer.
- **Basic Output information** (displayed in the visualization platform of the MARCLAIMED IDST):
 - Status of the system through the risk indicators (health, environmental and operational)
 - Time series evolution and relevant statistics measurements.
- **Elaborated Output information:** Recommendations on the actions that should be taken in thresholds overreachment situations. The MARAISER Tool integrated in the MARCLAIMED IDST will function as an Early Warning System to minimize the risk.

For the **Module 2 (long-term)** the conceptual scheme has been developed for DS1. In this module the input data, processing and output data have been conceptualized as follows:

- **Input data:**
 - The drought system status provided by PREDMAR Long term.
 - The setpoints provided by the Agencia Catalana de l'Aigua, which dictate the volume of reclaimed water compromised for certain uses (such as ecological flow preservation, indirect reuse, among others).
 - The total capacity on the Ter-Llobregat system of reclaimed water production (corrected if it is necessary by the real capacity, as this volume could be significantly reduced in drought periods).
- **Back end logic and calculations (MARCLAIMED IDST):** estimates the reclaimed volume available in the system annually from present to 2100.
- **Output data:** Number of additional MAR systems (from 1-2 hm³/year of infiltration/injection capacity -this has been ranged considering the current MAR systems recharge capacity operating in the area-).

Main functionalities

The MARAISER tool offers several potential functionalities and applications in three axes: i) water quality and operational risk management, ii) data management, reporting and stakeholder communication and iii) planning and design of MAR infrastructures.

i) Water quality and operational Risk Management:

- Real-time monitoring of health, environmental, and operational risks.
- Compliance tracking with regulatory standards.
- Early warning system for threshold overreachment: contamination prevention and control, infrastructure maintenance optimization, operational efficiency tracking and alert system for operational issues.
- Automated risk assessment and classification and decision support for risk mitigation actions (decision-making support).

ii) Data Management, Reporting and Stakeholder Communication:

- Automated data collection and processing, which enables real-time status and regulatory compliance reporting (transparency in operations).
- Historical data analysis and trending for performance reporting and visualization.

iii) Planning (medium-long term) of MAR infrastructures:

- Decision-making in investments for new MAR infrastructure based on the trends of the forecasted reclaimed water availability.
- Using this output joint with the output of RECOVER will be a key issue to determine the capacity of MAR systems cost recovery based on the new model tariff settings proposed.

Therefore, the MARAISER tool represents a comprehensive and innovative solution for MAR systems, integrating real-time monitoring, risk management, data analysis, and long-term planning capabilities. This integrated approach enhances the efficiency, sustainability, and economic viability of MAR operations through evidence-based decision-making and optimal system performance.

4.2. Data Collection and Processing

4.2.1 MARVIS - Virtual sensing tool for near real-time water quality monitoring

The system acquires data from physical sensors that provide the relevant physicochemical water parameters for MARVIS: pH, temperature, conductivity, turbidity, UV254 and nitrates. Physical probes have been installed in all demo sites. In the case of DS1, the multiparametric sensor is placed after the tertiary treatment of the WWTP of Baix Llobregat. A spectrometer probe V3 probe from s::can is employed to measure the desired water parameters [38]. The same sensor has been installed at Comporta (DS2), specifically within the infiltration ponds. Alternatively, at DS3 two different sensors are collecting data to monitor the required parameters for MARVIS: SIAP with several probes and TRIOS for nitrates [39, 40, 41, 42, 43]. This configuration has been set in two different plots as well (Broekman and Slot). Data collection occurs at regular intervals of at least 10 minutes, with secure transmission to an Azure virtual machine through FTP protocol or, when not possible, manual collection procedures. The data format for most of the cases is a .csv file. The preprocessing pipeline implements expert-knowledge-based quality checks to identify and remove unrealistic values specific to each parameter. The system accounts for regular weekly sensor maintenance to prevent data contamination. Missing values are handled through exclusion from training sets to maintain data integrity. Parameter ranges are validated based on site-specific characteristics, acknowledging the unique conditions of each deployment location. The current implementation utilizing XGBoost processes raw data without normalization, which simplifies the preprocessing pipeline while maintaining robust prediction capabilities. Tree-based models like XGBoost

naturally handle different scales by making splits based on relative ordering of values rather than their absolute magnitudes. Future implementation of neural networks will explore the potential benefit of a data normalization step, since it frequently improves training stability, speeds up convergence and enhances generalization. Common methods include min-max scaling or Z-score normalization.

4.2.2 PREDMAR - AI-based water availability and demand forecasting tool

DS1 - Llobregat (Spain)

Short term Module:

1. Forecast of stored surface water volume

The analysis focuses on the reservoirs of La Baells, Sau, Susqueda, and La Llosa del Cavall, collecting historical data on stored volume, precipitation, temperature, Normalized Difference Snow Index (NDSI), released volume, and reservoir inflows. An exploratory data analysis will be conducted to gain a deeper understanding of the dataset. This will involve identifying any missing values within the data, as well as detecting potential outliers that may skew the results. Additionally, the analysis will examine the relationships between the various variables present, with a particular focus on the 'stored volume' variable, due to its importance within the context of the project.

Data Collection and Processing

The reservoir inflow and outflow volume data, crucial for understanding the dynamics of the water body, were sourced from the Agència Catalana de l'Aigua (ACA), the Catalan Water Agency. This dataset encompassed a substantial timeframe, spanning from 2000 to 2024, and was characterized by its high temporal resolution, being available at a daily frequency. In addition to the hydrological data from ACA, meteorological data were also incorporated into the analysis. Precipitation and temperature data were gathered from meteorological stations operated by METEOCAT, the Meteorological Service of Catalonia, and AEMET, the State Meteorological Agency of Spain. The selection of specific stations for data retrieval was guided by their temporal coverage, ensuring a consistent and reliable dataset for the study period.

To access the climatological data from AEMET, an API KEY was requested, enabling programmatic and efficient data retrieval. This streamlined the process of acquiring vast amounts of historical weather data, including temperature, precipitation, wind speed, and other relevant meteorological parameters. In conjunction with this, hydrological data, such as river flow rates, water levels, and soil moisture content, were also collected from relevant sources.

The acquired meteorological and hydrological data were then subjected to a series of rigorous data cleaning and transformation methods. This crucial step involved identifying and correcting errors, handling missing values, and standardizing data formats. Additionally, data transformation techniques, such as scaling and normalization, were

applied to ensure data consistency and compatibility for subsequent analysis and modeling. These meticulous data preprocessing steps were essential to enhance data quality and reliability, ultimately leading to more accurate and robust predictive models.

Table 4.2.2.a data source

Origin	Variables	Period
ACA	Stored Volume, Outflow	2000-2024
METEOCAT	Temperature, Precipitation	2005-2024
AEMET	Temperature, Precipitation	2000-2024

Missing Values Analysis

The pervasive issue of missing data, a frequent challenge encountered in datasets, was observed across all series incorporated in this analysis. This issue was particularly noticeable within the meteorological station data. To mitigate this challenge and ensure the integrity of the analysis, a multifaceted approach to data imputation was adopted.

Simpler techniques, such as interpolation - which estimates missing values based on surrounding data points - and mean-based imputation - which replaces missing values with the mean of the available data for that variable - were utilized where appropriate. However, recognizing the potential limitations of these simpler methods, more advanced imputation strategies were also employed.

Specifically, the miceforest algorithm was leveraged to address missing data within a multivariate context. This sophisticated approach accounts for the interrelationships between variables, ensuring that the imputed values are consistent with the overall structure of the data. The selection of the most suitable imputation strategy for each instance was guided by a meticulous assessment of correlations between variables. This careful consideration ensured that the chosen method would preserve the inherent relationships within the data, thereby maintaining the integrity of the analysis and its subsequent findings.

Table 4.2.2.b missing values downloaded data

Reservoir	Missing Values
La Baells	864 missing records
Sau	2 missing records
Susqueda	70 missing records

La Llosa del Cavall	2
---------------------	---

The data was standardized into CSV files with a uniform format, and monthly series were created based on the following criteria:

- Last data of the month: Reservoir volume, NDSI.
- Monthly average: Temperature, discharged flow.
- Monthly accumulation: Precipitation, discharged volume, inflow volume.

These files enable a more robust analysis by reducing daily variability and facilitating modeling.

Exploratory Data Analysis

In the pursuit of unveiling hidden patterns and potential anomalies within the dataset, a multifaceted approach was adopted. The intricate relationships between variables were meticulously examined through correlation analysis, while the Granger Test [44] was employed to discern potential causality links, distinguishing mere correlation from genuine cause-and-effect relationships.

Leveraging the capabilities of advanced analytical tools such as SWEETVIZ¹ and Missingno², the distribution and interplay of variables were visually and statistically scrutinized. This comprehensive exploration facilitated the identification of potential outliers and anomalous data points, as well as the discovery of underlying patterns and structures within the data. The insights gleaned from these analyses not only deepened the understanding of the dataset's inherent characteristics but also paved the way for the development of robust and reliable predictive models.

Table 4.2.2.c correlation results

Variable	Correlation with Reservoir Volume
Precipitation	High (+)
Temperature	Medium (-)
Discharge, Inflow	High (-)

¹ <https://pypi.org/project/sweetviz/>

² <https://pypi.org/project/missingno/>

Study of the NDSI Index

The NDSI was obtained through ClimateEngine.org for mountainous areas near the reservoirs. Data from 1990 to 2024 were analyzed with a 1 km resolution and a monthly frequency. A significant relationship was observed between the NDSI and the stored volume in some reservoirs, but with limitations in temporal frequency.

Table 4.2.2.d NDSI areas analyzed

Reservoir	Study Area
Sau & Susqueda	Between the peaks of Puigmal and Canigó
La Baells & La Llosa del Cavall	Madriu-Perafita-Claror and Campcardós Valleys

Datasets

Dataframes were carefully constructed by selecting relevant variables that could be used to predict the stored volume in each of the reservoirs within the system. These variables were chosen based on their known or hypothesized influence on water levels, and they included critical climatological factors such as precipitation and temperature, along with hydrological variables like discharge and inflows. By systematically organizing these variables into structured dataframes, the dataset was prepared for further analysis and model development.

To assess the relationships between the selected variables and the stored volume, the Granger Causality Test was applied. This statistical test is commonly used to determine whether one time series can predict another, based on the assumption that if one variable 'Granger-causes' another, it contains valuable information for forecasting the latter. The application of the Granger test was a crucial step in identifying causal relationships, ensuring that only the most relevant predictors were incorporated into the predictive models.

The results of the Granger test revealed important insights into the factors influencing the stored volume in the reservoirs. It was found that precipitation and temperature play significant roles in explaining a portion of the variability in the stored volume. Specifically, precipitation directly affects the volume of water entering the reservoirs, while temperature influences evaporation rates, which in turn impacts the overall water balance. In addition to these climatological factors, hydrological variables, such as discharge and inflows, were also found to be important in explaining changes in the stored volume. Discharge directly influences the amount of water leaving the reservoirs, while inflows represent the water entering the system from various sources, further affecting the water storage dynamics.

These findings confirm that both climatic and hydrological factors are critical in predicting the stored volume, and they highlight the complex interplay between precipitation, temperature, discharge, and inflows. The relationship between these variables and the stored volume is essential for building accurate predictive models, as it allows for a comprehensive understanding of the factors driving water levels in the reservoirs. The insights gained from the Granger test helped refine the dataset and informed the development of ML models, which are now better equipped to forecast future reservoir volumes with higher accuracy.

Table 3.2.2.e Preliminary variables used

Reservoir	Selected Variables
La Baells	Precipitation, Temperature, Discharge, Inflow
La Llosa del Cavall	Precipitation, Temperature, Discharge, Inflow
Sau	Precipitation, Temperature, Discharge, Inflow
Susqueda	Precipitation, Temperature, Discharge, Inflow

Automation and Data Update

An automated update system was implemented in Google Sheets using App Scripts. The data is updated periodically:

- Daily: Reservoir volume, flow rates.
- Every 4 days: Meteorological data from AEMET.
- Monthly: Monthly data files.

In order to streamline the process and eliminate the need for manual intervention, automated triggers were implemented within Google Sheets. These triggers were designed to initiate updates to the data or formulas contained within the spreadsheet based on predefined conditions or events. By leveraging this functionality, the spreadsheet could be dynamically updated without requiring any direct action from a user, saving time and reducing the potential for errors.

The meteorological data has been downloaded from AEMET stations. The weather stations available from AEMET can be seen in the Figure 4.2.2.a (in blue, the reservoirs from left to right: La Llosa del Cavall, La Baells, Sau and Susqueda; in red, the weather stations from left to right: Berga Instituto, La Molina and Girona Airport).



Figure 4.2.2.a Location of the selected meteorological stations and reservoirs under study.

The final series for each reservoir have been selected taking into account the causality of the precipitation and temperature series with respect to the stored volumes at each point of interest. Since the Campdevanòl and La Seu d'Urgell stations have not been updated for several months as of August 2024, the series obtained from the La Molina and Girona Airport stations will be selected.

To select the algorithms and features that represent the points of interest in the most convenient way, the framework PREDMA³ developed by CETAND has been used. PREDMA © helps in the search for ML parameters and algorithms.

In order to enhance the predictive capabilities of the developed models, we have incorporated forecast information derived from European Centre for Medium-Range Weather Forecasts (ECMWF) physics-based models, such as SEAS5. A methodology has been developed to utilize these forecasts, following location-based interpolation, as future values for explanatory climate variables. This integration of physics-based forecasts and ML, referred to as a hybridization approach, has been further optimized through improvements in model learning. These improvements leverage real observations from weather stations associated with each point of interest, or alternatively, reanalysis data from reliable sources such as ERA5. This comprehensive process yields a hybrid predictive model, also known as PIML, which combines the strengths of both physics-based modeling and ML.

³ PREDMA© It is a framework designed for the selection of the optimal combination of parameters and hyperparameters, as well as the most suitable artificial intelligence algorithms. Additionally, it enables the identification of causal relationships between parameters for each analyzed water resource, ensuring a data-driven and optimized modeling approach.

2. Streamflow predictions SVH

For the daily flow of the Llobregat River, the flow gauging station of SVH was selected as the primary data collection point due to its strategic location and the comprehensiveness of its data set. Daily series were used for this analysis to ensure high temporal resolution and capture daily variations in river flow, which are essential for understanding the dynamics of the river and making accurate predictions.

The daily series data from this station provides a continuous and detailed record of the river's flow, allowing for a thorough analysis of daily fluctuations and trends. The data sources for the daily flow model are the same as those used for the surface water volume forecast, ensuring consistency and integration between the two models. These data sources include meteorological data, hydrological data, and other relevant environmental data. By utilizing the same data sources, the models can leverage shared information and provide a more comprehensive understanding of the water system.

The generated dataframe for this analysis includes the following key variables:

- ds: This column represents the date in the format yyyy-mm-dd, providing a temporal reference for each data point.
- y: This column represents the river flow rate, specifically the daily mean flow rate, measured in cubic meters per second (m^3/s). This is the primary variable of interest for understanding the river's hydrological dynamics.
- v1: This column represents the daily accumulated precipitation, measured in millimeters (mm). This variable is crucial for understanding the relationship between rainfall and river flow.
- v2: This column represents the mean daily temperature, measured in degrees Celsius ($^{\circ}\text{C}$). This variable can provide insights into how temperature influences river flow, especially through processes like snowmelt and evapotranspiration.

By incorporating these variables into a comprehensive dataframe, this analysis enables a thorough examination of the factors influencing the daily flow of the Llobregat River at the Sant Vicenç dels Horts gauging station. This information is fundamental for water resource management, flood forecasting, and ecological studies in the Llobregat River basin.

In order to forecast precipitation and enhance the precision of our predictive models, we have integrated an Application Programming Interface (API) provided by the meteorological company METEOSOURCE [45]. This API grants us access to a wealth of meteorological data, including real-time and historical precipitation data, which can be seamlessly incorporated into our existing models.

The specific methodologies and parameters utilized in the METEOSOURCE API will be thoroughly documented in subsequent sections of this report. This will include detailed information on data access procedures, API integration protocols, and the specific variables and data formats employed. Additionally, we will outline our strategies for data validation, quality control, and error handling to ensure the reliability and robustness of our precipitation predictions.

Long term Module “ Forecast of Drought status of the system”

To predict the reservoir volumes in four reservoirs of interest, future precipitation and temperature values are required. These values must consider both **RCPs** scenarios, which depend on future greenhouse gas emissions, and **SSPs** scenarios, combining socioeconomic development with climate projections. These data are obtained through the **Climate Data Store (CDS)** API, utilizing ten climate models employed in the **ICARIA** project, which are part of the CMIP6 initiative. This initiative aims to improve understanding of the climate system and support future assessments by providing a framework for comparing models across historical simulations and future projections.

Data Sources and Scenarios

The data for precipitation and temperature are downloaded from the CDS and supplemented by missing data from the **Earth System Grid Federation (ESGF)** website. The downloaded data includes historical simulations and future projections under five SSP scenarios (**SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5**), and the simulations run on ten selected models, including ACCESS-CM2 and CNRM-ESM2-1. Each model generates projections for key variables such as precipitation and temperature, allowing the evaluation of how different models respond to the same emissions and development scenarios.

The availability of precipitation and temperature data across various SSP scenarios and climate models is detailed in the table 4.2.2.f. Some data combinations could not be downloaded due to the unavailability of specific model-scenario combinations, which will only be available in future CDS updates.

Table 4.2.2.f variable availability precipitation + temperature

Model/ Scenario	SSP1-1.9	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
ACCESS-CM2	not available	available	available	available	available
BCC-CSM2-MR	not available	available	available	available	available
CanESM5	available	available	available	available	available
CMCC-ESM2	not available	available	available	available	available
CNRM-ESM2-1	available	available	available	available	available
EC-EARTH3	available	available	available	available	available
MPI-ESM1-2-HR	not available	available	available	available	available
MRI-ESM2-0	available	available	available	available	available
NorESM2-MM	not available	available	available	available	available
UKESM1-0-LL	available	available	available	available	available

Data Processing and Scenario Treatment

To ensure consistency across the data, a minimum study area is defined to encompass the grid points of the ten models that cover the coordinates of the four target reservoirs. Interpolations are then applied to obtain specific precipitation and temperature values for the centroid of each reservoir, ensuring the necessary spatial consistency for climate projections.

Once the data are downloaded and available, several processing steps are undertaken, including:

- Interpolation of model results to the centroid of each reservoir (Figure 4.2.2.b).
- Calculation of average, minimum, and maximum values across models for each time period.

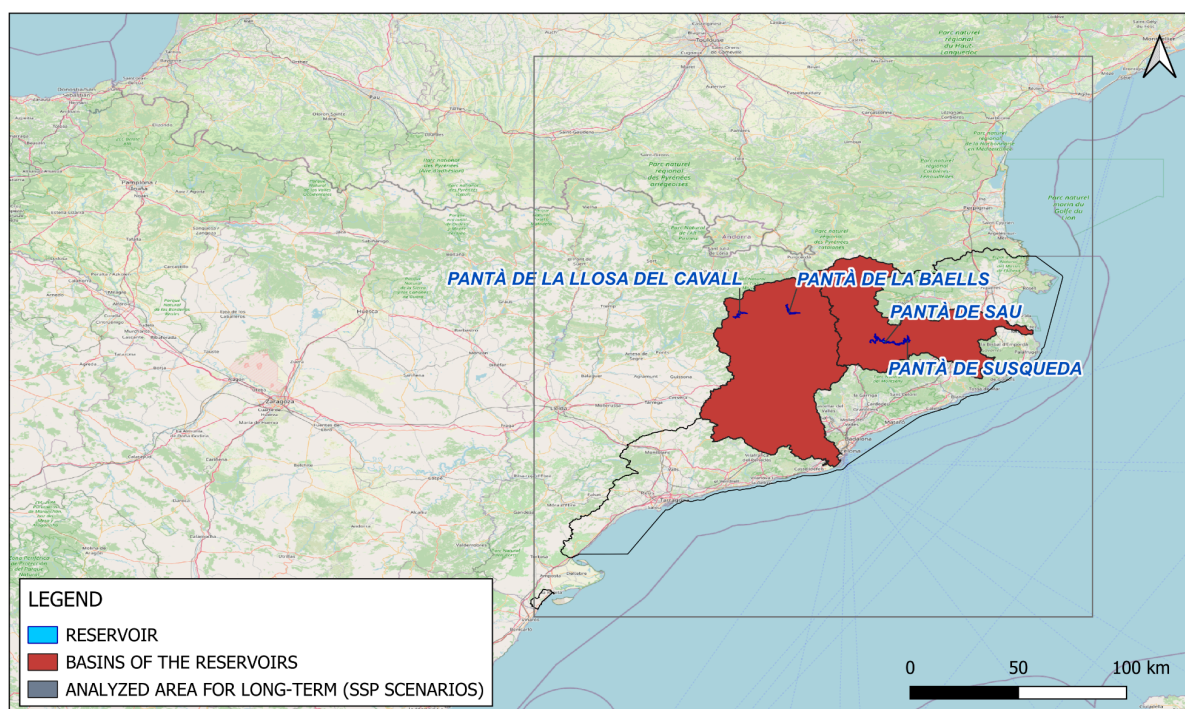


Figure 4.2.2.b Ter and Llobregat systems at the DS1 demo site.

These processed data are then compiled into high-level files for each reservoir, including annual values, the first and last month of each year, and seasonal values.

After interpolating and calculating the required values (mean, min, and max), final datasets are created for each reservoir, covering precipitation and temperature variables. These datasets include:

- Annual values.
- January and December values.
- Seasonal values.

For precipitation, the annual value corresponds to the total accumulated precipitation, not the mean, unlike temperature, which uses the mean value.

The next step in building the final datasets involves incorporating historical data, which were previously collected for each reservoir. The historical data for each reservoir are integrated based on their starting year:

- LA BAELLS: 2006
- LA LLOSA DEL CAVALL: 2008
- SAU Y SUSQUEDA: 2000

The final combined dataset contains:

- Date columns.
- Seven columns for precipitation data.
- Seven columns for temperature data. This data will span from 2015 to 2100, with values calculated for the required monthly, seasonal, and annual periods.

For each climate model, the data are interpolated to the centroids of the reservoirs, ensuring that the values are relevant to the geographical locations of interest. Each model has a specific resolution, and based on this, the grid points are selected for interpolation to provide accurate data at the required location.

Once the data downloaded from the CDS is processed, a scaling factor analysis is performed with the aim of capturing the projected trend of the climatic variables. The process begins with the selection of base year values, specifically the values for 2023, which will serve as the reference point for measuring changes in future years. Starting from this base value, the scaling factor will be calculated by considering the current value of each scenario and obtaining the difference with respect to the previous value, applying the following formula:

$$\text{Scaling Factor} = ((\text{Current Value} - \text{Previous Value}) / \text{Previous Value}) * 100$$

This calculation produces a percentage that reflects the increase or decrease in the variable compared to the value from the previous year. Positive values indicate an increase compared to the previous year's value starting from the base year, while negative values indicate a decrease. This scaling factor will be used to assign future values to the inflows and discharge variables according to the criteria of each studied scenario.

This process allows for a comprehensive approach to assess future reservoir conditions, considering the different emission scenarios and available climate models. The resulting datasets will provide critical information for water resource management under future climate conditions.

The work outlined above ensures that the data collected and processed from the CDS and ESGF platforms are of high quality and provide a robust foundation for reservoir volume prediction. By combining data from multiple models and scenarios, we aim to improve the accuracy of our long-term water resource forecasts, essential for future water management strategies.

DS2 - Comporta (Portugal)

The methodology employed for data collection and processing in this study mirrors the approach used in the long-term module of Demo Site 1 in Spain. This primarily involves

the utilization of RCPs and SSPs scenarios to project future climate conditions. However, the methodology has been tailored to accommodate the specific requirements of this research, which include the investigation of additional variables such as runoff and evapotranspiration, as well as the unique geographical characteristics of the Sado basin. This adaptation ensures that the data collected and processed is relevant and accurate for the specific context of the Sado basin, allowing for meaningful analysis and conclusions to be drawn. (Figure 4.2.2.c).

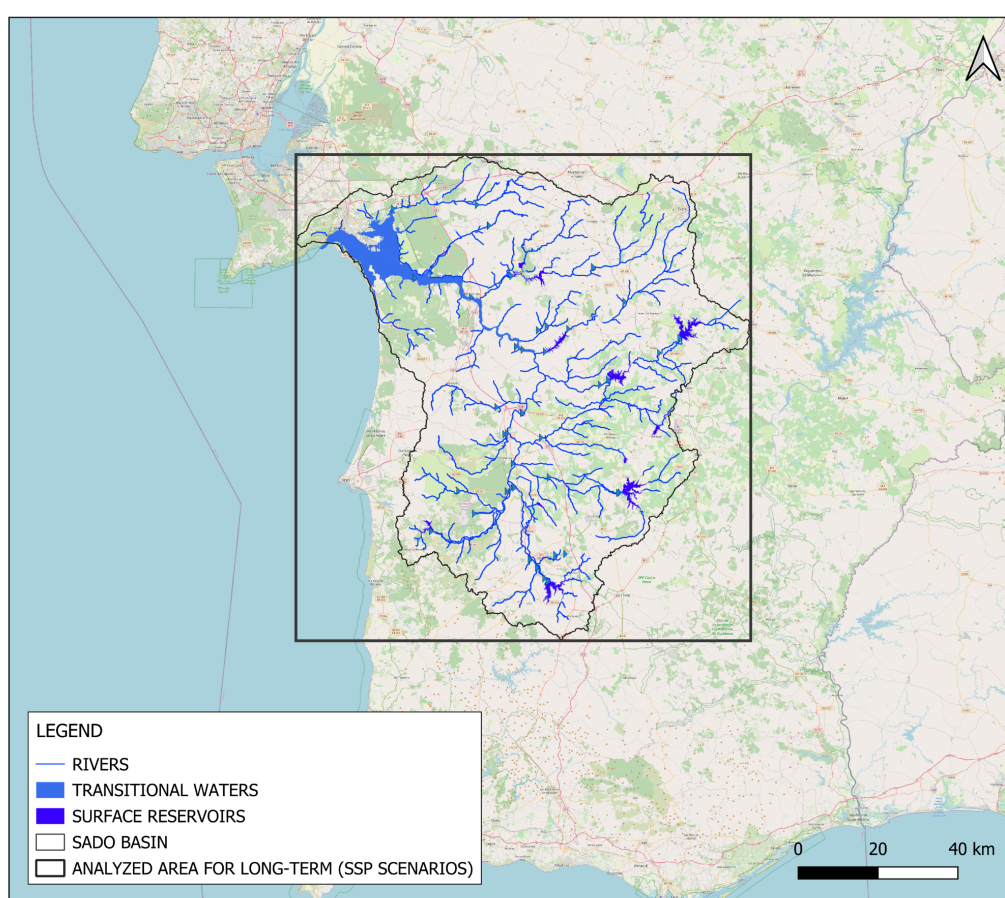


Figure 4.2.2.c Study area at the DS2 demo site.

DS3 - Texel (The Netherlands)

To project future values of climatological variables, data from ECMWF physical models, such as SEAS5, is processed using a specialized methodology. This approach applies spatial interpolation to forecasts to derive future values of explanatory climatological variables while incorporating real-time conditions at the moment of prediction. Statistical

analysis is conducted to assess the various ensembles, trajectories, and perturbations generated by the forecast models.

The hybridization process is further refined by integrating advanced model learning techniques that leverage real-world observations from meteorological stations near the target locations or, alternatively, reanalysis datasets from reliable ECMWF sources like ERA5.

The methodology for data acquisition and processing relies on ECMWF's Copernicus CDS models, specifically SEAS5 (seasonal forecasts extending up to seven months with daily resolution) and ERA5 (historical reanalysis datasets with hourly resolution). These datasets provide crucial input for PIML applications in hydrological and climatological modeling.

Variables and Unit Conversions

To ensure consistency and compatibility across models, the following variables and unit transformations are applied:

- Precipitation:
 - SEAS5: Converted to mm/day.
 - ERA5: Converted to mm/hour and then aggregated into mm/day for analysis.
- Mean Temperature:
 - Both SEAS5 and ERA5 provide values in Kelvin. These are converted to degrees Celsius by subtracting 273.15.

CDS Data Availability and Update Mechanism

CDS provides SEAS5 forecasts generated monthly, with records available since 2018. These forecasts are updated on the 10th of each month, following a systematic retrieval process:

- If the current date is between the 1st and 10th, the retrieval system fetches forecasts issued on the 1st of the previous month.
- If the current date is 11th or later, forecasts issued on the 1st of the current month are downloaded.

This ensures access to the most recent climate projections without delays in integration into predictive frameworks.

Automated Data Extraction and Spatial Filtering

The development of a new PIML-based predictive module necessitates the integration of both historical reanalysis data (ERA5) and seasonal forecasts (SEAS5). To optimize this process, a custom-built automated extraction system has been devised, incorporating the following key components:

1. Defining a Variable Grid:

- Flexible Spatial Grid (Figures 4.2.2.d and 4.2.2.f): An adaptable spatial grid, formatted in CSV, is employed to define the specific target locations for data extraction. This grid acts as a blueprint, guiding the system towards the desired geographical areas.
- Adjustable Grid Resolution: The resolution of the grid can be tailored to accommodate varying spatial granularities, ensuring that the data extraction aligns with the specific requirements of the study area. This adaptability is crucial for capturing the nuances of different geographical scales.

2. API-Based Data Retrieval:

- Copernicus CDS API: The extraction process is streamlined through the utilization of the Copernicus CDS API. This API serves as a gateway, facilitating efficient and controlled access to the vast climate datasets.
- Dynamic Query Construction: Queries are dynamically formulated to constrain downloads to the bounding box delineated by the maximum and minimum latitudes and longitudes of the study area. This dynamic approach ensures that data retrieval is focused and efficient, minimizing unnecessary data transfer.
- Targeted Data Extraction: The system is designed to extract only the pertinent data that falls within the defined spatial extent. This targeted approach significantly reduces computational overhead, optimizing the overall efficiency of the data acquisition pipeline.

This structured and methodical approach significantly enhances the efficiency of the data acquisition pipeline. By automating the extraction and integration of climate predictions and historical datasets into the PIML framework, the system provides a seamless flow of data for downstream modeling and analysis. The flexibility of the variable grid and the targeted data retrieval through the CDS API ensure that the data acquisition process is both adaptable and efficient, catering to the diverse requirements of climate research and predictive modeling.

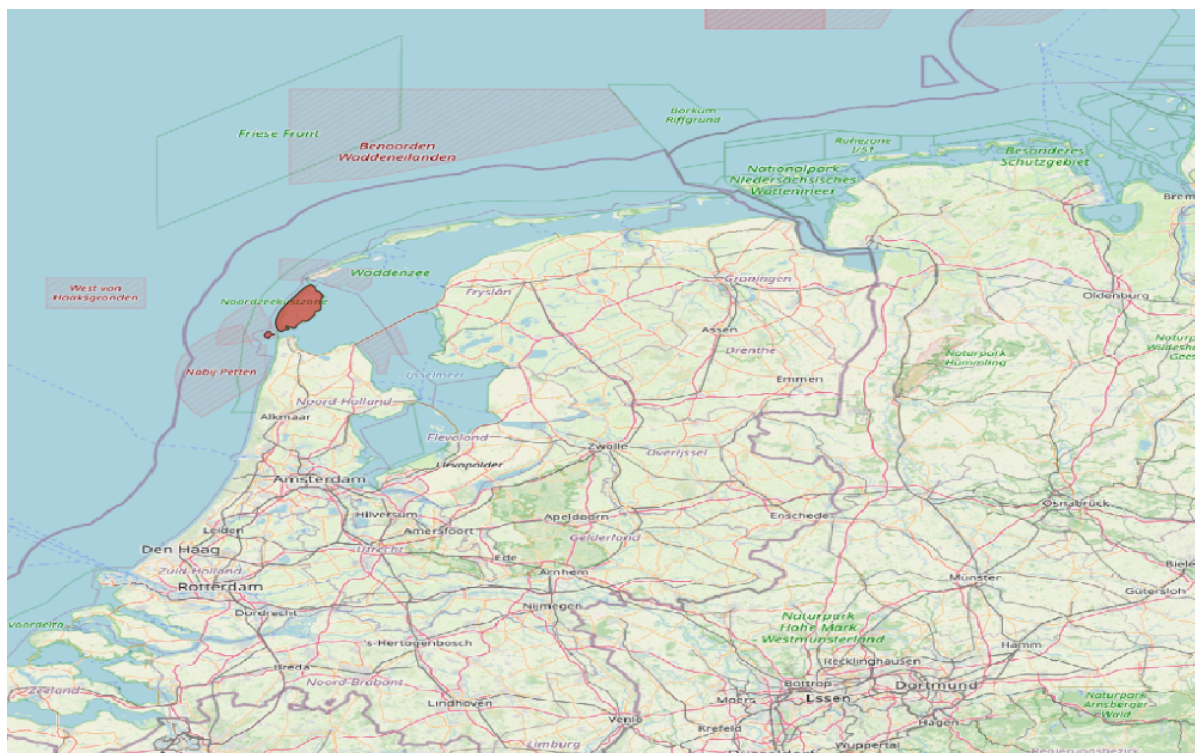


Figure 4.2.2.d General study area at the DS3 demo site.

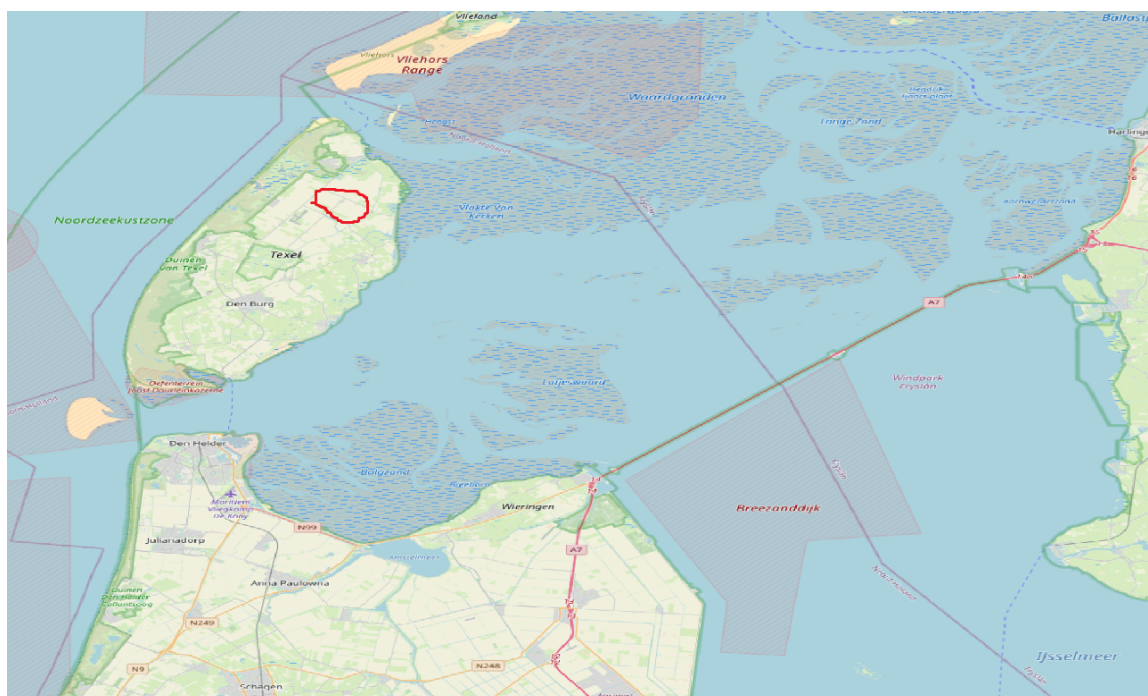


Figure 4.2.2.e Texel study area at the DS3 demo site.

4.2.3 MARAISER - Short and long-term MAR performance optimization tool

Module 1 (short-term)

As explained in the Methodology section (3.3), the first step is to conduct a MAR-SP Risk Assessment. For this purpose the data needed has to be provided by the Water Utility or MAR system operator and will consist in:

- Schematics of the MAR system itself (pdf).
- Permit requirements (pdf).
- Legislation framework and water quality thresholds (pdf).
- Recharged water quality historical data (csv): EC, pH, T, major ions, metals, CECs, microbiology.
- Groundwater quality historical data (csv): EC, pH, T, major ions, metals, CECs, microbiology.
- Hydraulic and Hydrodynamic groundwater historical data(csv): piezometric levels, pumping flow rate, infiltration rate, recharge flow.
- GIS layers for visualization: vector and raster.

This data will be processed as follows:

- The schematics of the MAR system will be digitalized for further integration in the visualization platform of the MARCLAIMED IDST.
- All the legal and regulatory documents will be analysed to extract legal thresholds and main conditions.
- All the historical data will be analysed to perform the MAR-SP Risk Assessment with a preprocessing to identify missing values and outliers. The statistical approach criteria of the MAR-SP Risk Assessment could differ between demo sites.

More detail regarding the **data collection and processing data per demo site has been described in section 5.1.3.**

The input data provided i) dynamic data from online sensors (commercial or virtual sensors; **MARVIS Tool NO₃ measurements will be integrated with different setpoints in DS1 and DS3**), ii) semi-dynamic data from continuous monitoring sensor but offline, ii) static data from laboratory analysis will be connected to a database, to read and save observed data and analyzed results.

The triggering mode is based on the execution of the flowchart of different setpoints depending on the risk, the indicator and the parameter (thresholds based regulation or expert criteria) to establish different status of alert. The different status of alert provides recommendations on actions to be taken to mitigate the identified risks.

Module 2 (long-term)

As explained in the Methodology section (3.3), long term MARAISER Module 2 calculates the number of MAR additional systems that can be projected based on the reclaimed water availability. For this purpose the data needed comes from:

- PREDMAR long term forecasts in the Ter-Llobregat System Drought Status.
- Reclaimed volume total capacity on the Ter-Llobregat system (corrected by the real capacity) provided by Aigues de Barcelona Water Utility.
- Reclaimed water volume compromise depending on the System Drought Status, provided by the Agència Catalana de l'Aigua.
- General infiltration/injection capacity of an average MAR operation in Ter-Llobregat system is considered as an average value from the last decade of performance.

This data will be processed in an excel sheet that includes the equations described in the Methodology (section 3.3). The estimations will be provided annually (from present to 2100), when the **PREDMAR long term in DS1** is executed in the MARCLAIMED IDST.

4.3. Digital Tool Development: AI components, algorithms and models used

In the realm of ML, the pursuit of accuracy often leads to the development of intricate and complex models. While these models may excel in predictive tasks, particularly when dealing with vast datasets, their inner workings can remain shrouded in mystery, even to seasoned experts. This opacity poses a challenge when attempting to understand the reasoning behind a model's specific predictions. To address this challenge, a range of interpretative methods have emerged, aiming to shed light on the decision-making processes of complex models. However, the landscape of these methods can be disorienting, with unclear relationships between them and a lack of guidance on when to favor one over another.

In response to this ambiguity, a unifying framework called **SHapley Additive exPlanations (SHAP)** [46] has been introduced. SHAP operates by assigning an importance value to each feature within a model for a particular prediction. This approach provides a quantifiable measure of the influence each feature exerts on the model's output, enabling users to discern which features play a pivotal role in driving specific predictions. At its core, SHAP is grounded in game theory, a mathematical framework for analyzing strategic interactions. It leverages the concept of Shapley values, which are used to distribute the "payout" or value created by a cooperative game among its players. In the context of ML, the "game" is the model's prediction, the "players" are the features, and the Shapley value represents the contribution of each feature to the prediction.

By adopting this game-theoretic perspective, SHAP offers a principled and consistent approach to interpreting model predictions. It allows users to move beyond the black box of complex models and gain actionable insights into the factors that shape their behavior. This enhanced understanding can foster trust in ML systems, facilitate debugging and improvement, and enable more informed decision-making based on model outputs (Figure 4.3.a).

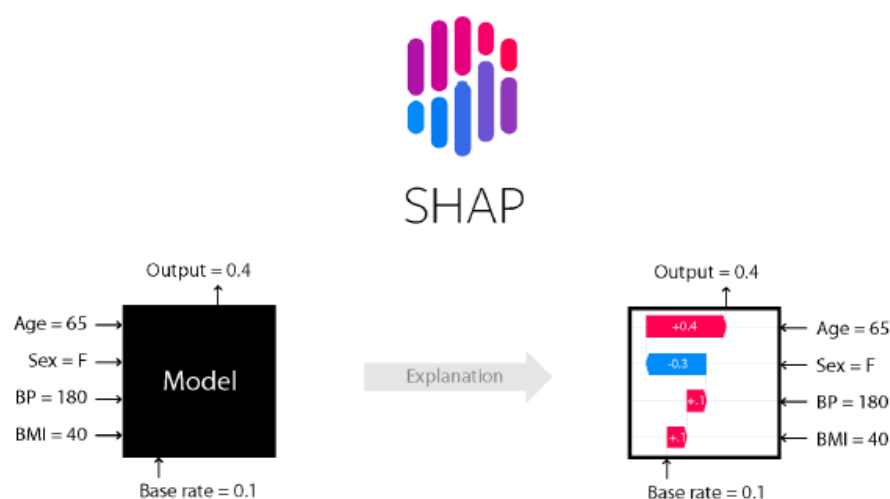


Figure 4.3.a SHAP - Feature importance⁴.

4.3.1 MARVIS - Virtual sensing tool for near real-time water quality monitoring

MARVIS exhibits a common framework for its AI components, algorithms and models, differing only on the data origin that ultimately depends on the demo site. Numerical variables are fed to a ML algorithm in order to predict nitrate concentrations, also numerical. Depending on the quantity of data, the first approach will consist in a XGBoost model and as data is being collected, neural networks will be considered. Explainability techniques such as SHAP will be employed to better comprehend the predictions obtained and the behaviour of the models.

AI Components

The MARVIS virtual sensor involves the development of ML models to estimate nitrate concentrations using data collected from physical sensors. Two main types of models are proposed: XGBoost and neural networks. The transition from XGBoost to neural networks is driven by data availability. Initially, XGBoost is used due to its efficiency with smaller datasets. As more data becomes available, neural networks are explored for their ability to capture complex patterns and improve predictive accuracy.

⁴ Source: <https://shap.readthedocs.io/en/latest/>

- XGBoost model:
 - Cross-Validation: 10-fold to ensure robustness and generalizability.
 - Hyperparameter Tuning: learning rate and maximum depth
 - Evaluation metrics: MSE and R^2
 - Transfer Learning: DS1 and DS3 models, adjusting hyperparameters.

- Neural Network model:
 - Architecture Proposal:
 - Input layer: 6 neurons corresponding to the input features (temperature, turbidity, conductivity, pH, UV254 absorbance).
 - Hidden layers: Multiple layers with neurons using ReLU activation functions (e.g. 3 layers of 64, 32 and 16 neurons).
 - Output layers: 1 neuron with linear activation for nitrate concentration prediction.
 - Regularization: Dropout layers to prevent overfitting (0.2 dropout rate).
 - Training: Uses backpropagation and optimization techniques like stochastic gradient descent.
 - Transfer Learning: Applied to DS2 using pre-trained models from DS1 and DS3. Initial layers are frozen, and later layers are adapted to DS2-specific data.

The methodology emphasizes data preprocessing, model training, validation, and deployment, ensuring the models are accurate and effective in predicting nitrate concentrations. Transfer learning is leveraged to efficiently adapt models to new sites, reducing the need for extensive retraining.

4.3.2 PREDMAR - AI-based water availability and demand forecasting tool

The development of the digital tool for climate and hydrological forecasting integrates advanced AI methodologies, ML algorithms, and physics-based models to enhance predictive capabilities. This section outlines the core components, data sources, and methodologies applied in the tool's development.

AI Components

The digital tool leverages PIML, which combines physical models with ML techniques to improve forecast accuracy. The primary AI components include:

- Supervised Learning Algorithms: Used for predictive modeling of hydrological and climatological variables.
- Unsupervised Learning Algorithms: Used for bias correction, anomaly detection, and clustering.
- Hybrid AI Models: Combining physics-based simulations with ML models.

- PIML: Integrating physical laws governing hydrological systems into ML models for enhanced accuracy.
- Hybrid Ensemble Models: Blending multiple AI models to reduce uncertainty and improve predictions.

Data Sources and Integration

The digital tool ingests multiple datasets to generate robust forecasts:

CDS Data

- SEAS5 (Seasonal Forecasts): Seven-month forecasts with daily resolution.
- ERA5 (Reanalysis Data): Historical climate data with hourly resolution.
- CMIP6 Scenarios (RCP, SSP): Long-term climate projections under different socio-economic scenarios.

Sensor and Hydrological Data

- Agència Catalana de l'Aigua (ACA): Real-time water inflow, outflow, and reservoir levels.
- AEMET/METEOCAT Weather Stations: Temperature, precipitation, and humidity data.
- Satellite Data (ClimateEngine, Copernicus): NDSI, evapotranspiration, and land surface temperature.

Machine Learning-Based Bias Correction

Bias correction is crucial for aligning climate model outputs with real-world observations. The following methods are applied:

- Quantile Mapping: Adjusts systematic biases in precipitation and temperature forecasts.
- Empirical Statistical Downscaling (ESD): Corrects discrepancies between large-scale climate models and local station data.
- AI Bias Correction: Uses ML models to dynamically adjust biases in climate model predictions based on historical errors.

Deployment and Automation

Automated Model Retraining: New data from sensors and CDS updates the models periodically.

- Google Sheets & Cloud API Integration: Automated data updates using App Scripts and cloud-based services.

This AI-powered digital tool effectively bridges ML, climate modeling, and real-time hydrological forecasting, ensuring robust decision support for water resource management under both short-term and long-term climate scenarios.

4.3.3 MARAISER - Short and long-term MAR performance optimization tool

MARAISER Tool is not an actual model, but the digitalization of the MAR-SP Risk Assessment outputs (RMP, parameters, data source identified -sensor, lab analysis-, kind of data -dynamic, semidynamic or static-, alert and critical thresholds and the actions to be taken in case of these thresholds are overreached) in a comprehensive flowchart. This flowchart will be translated in scripts deployed within the MARCLAIMED IDST.

The tool is fed with the MARVIS and PREDMAR short term forecasting outputs (continuous NO₃ measurements in DS1 and DS3, and daily flow forecasting in DS1) and PREDMAR long term drought Ter-Llobregat status forecasting in DS1, both based in AI approaches with the aim of improving the operational efficiency regarding risk management and infrastructure optimization.

4.4. Technical and Implementation Requirements for the integration in the MARCLAIMED IDST

The technical and implementation requirements for the integration in the MARCLAIMED IDST are described in deep detail in the following deliverables from WP2 “MARCLAIMED IDST: Integrated Decision Support Tool for MAR recommendations and early warning alert system” development:

- D2.1 Requirements and global architecture v1 (July 2024 - submitted to the Portal).
- D2.3 Data models (January 2025 - submitted to the Portal).
- D2.4 Open interfaces (January 2025 - submitted to the Portal).
- D2.5 MARCLAIMED Integrated Decision Support Tool v1 (July 2025 - to be submitted to the Portal).

5. Application and preliminary results in Demo Sites

5.1 MARVIS - Virtual sensing tool for near real-time water quality monitoring

In the following section the initial development of MARVIS is presented. Data collection has been successfully carried out at demo sites DS1 and DS3, accompanied by their corresponding exploratory data analysis (EDA). Subsequently, an initial XGBoost model has been trained for each site, with evaluation metrics provided to assess their current predictive performance. As more data will be fed to the model, an increment of predictive power is expected. These models are developed to account for site-specific characteristics and environmental conditions. In DS2 at the current state of the project, efforts have focused on verifying consistent data collection and planning for transfer learning in future steps (hence, the following preliminary results will be just presented for DS1 and DS3). This strategic approach aims to optimize data usage and enhance the

65

accuracy of nitrate concentration predictions across different sites considering their specific particularities. After EDA in each DS, Datasets were submitted and used through the following process in order to obtain a preliminary ML model.

Data Splitting

Datasets were split into training and test sets before model training with random splitting of 80/20 ratio, respectively. This accounted for 11.826 and 2.956 training and test samples, respectively in the case of DS1. Datasets from DS3 from the different wells were combined. As this dataset was imbalanced for the different wells, stratified splitting was implemented to maintain the proportional representation of both water sources, Slot and Broekman, in the resulting sets. The training set contained 7,559 samples, with 6,270 from Slot and 1,289 from Broekman, while the test set contained 1,890 samples, with 1,552 from Slot and 338 from Broekman.

Instead of using a separate validation set, 10-fold cross-validation was used during model training to ensure robust performance estimation, reduce the variance of model evaluation, and facilitate hyperparameter tuning.

Model training

Model fitting began with the implementation of three different ML algorithms to predict nitrate concentration. The selected techniques were Random Forest, XGBoost, and linear regression, chosen for their established performance and ease of interpretation in a feature importance study using SHAP values. Each algorithm's performance is evaluated using two metrics: MSE and R^2 . The best model was identified as the one with the lowest value of MSE and the highest of R^2 . In both demo sites, XGBoost exhibited better performance and thus was the model further employed

Hyperparameter tuning

The model hyperparameter tuning is a crucial process for optimizing the performance of the models to find the best combination of model configuration settings. This process was applied to the XGBoost model. To conduct this optimization, the GridSearchCV function from the Python package scikit-learn was utilized to identify the optimal combination of specified parameter values, and 10-fold cross-validation was employed to ensure robust results. The hyperparameters to be optimized are the learning rate (0.01, 0.05, 0.10 and 0.20) and the maximum depth of trees (3, 4, 5, and 6).

5.1.1 DS1 - Llobregat, Spain

Exploratory Data Analysis

The Barcelona dataset contains 22.918 observations collected from 24/10/2024 to 17/02/2025, with 10 minute frequency measurements. It includes seven relevant variables: conductivity in mS/cm, temperature in $^{\circ}\text{C}$, pH, total organic carbon equivalent (TOCeq) in mg/L, UV transmittance at 254 nm (UV254t), turbidity in NTU and nitrates

concentration (NO_3) in mg/L. While the latter is the target variable, the rest of the variables serve as predictors. The percentage of missing values for each variable is lower than 0,03%, except for UV254t, which accounts for 2,18%. Given that only one variable exhibits a percentage of missing values and those are few, rows with missing values were removed to prevent potential bias introduced by imputation. Afterwards, data filtering was performed based on predefined expert criteria for each variable to ensure the retention of all relevant information for modeling. Prior to model fitting, outliers were identified and removed using the Interquartile Range (IQR) method since none of the predictors followed a Gaussian distribution and the Z-score method could not be used. This process ensures the removal of extreme values and enhances the accuracy of the models. Following data cleaning, the final dataset contained 14.782 observations, which represented approximately 64% of the original dataset.

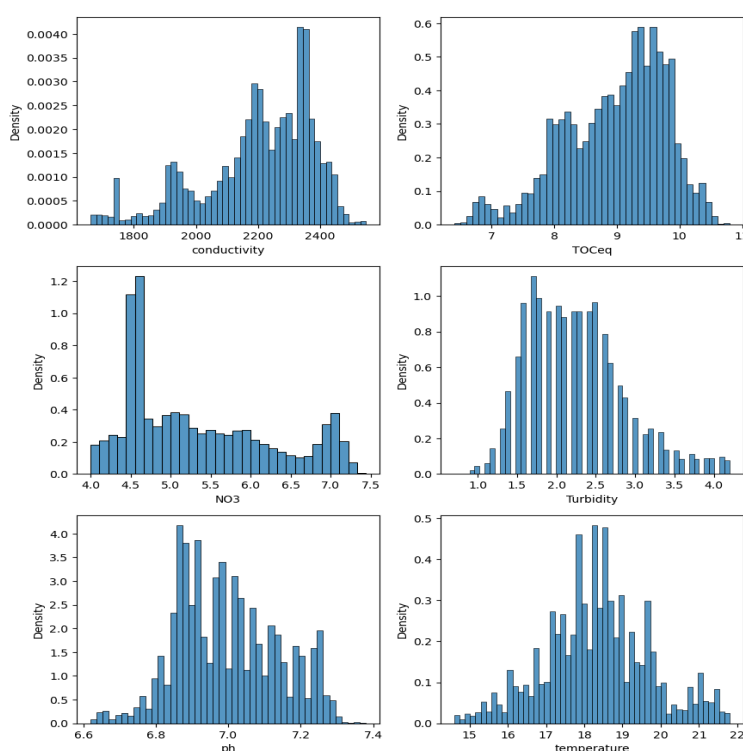


Figure 5.1.a. Frequency distribution of investigated variables for MARVIS at DS1.

The distribution of the variables after cleaning is plotted in Figures 5.1.a, showing that pH, TOCeq and NO_3 followed an asymmetric distribution, and some variables, such as conductivity and temperature, still presented mild extreme values after applying the IQR method.

The XGBoost's optimal parameter combination was a maximum depth of 6 and a learning rate of 0.10. The model exhibits a MSE of 0.017 and R^2 of 0.978. The correlation between variables is shown in the correlation matrix from Figure 5.1.b, which allows us to determine the potential influential predictors of nitrate concentration. According to the heatmaps, TOCeq, UV254t and conductivity are the most correlated variables with nitrate concentration, all showing negative associations. Moreover, TOCeq and UV245t are highly correlated; this means both features exhibit a strong relationship. Further data collection is essential to refine our understanding of the variable correlations and to validate the observed relationships, ensuring robust conclusions about the predictors of nitrate concentration.

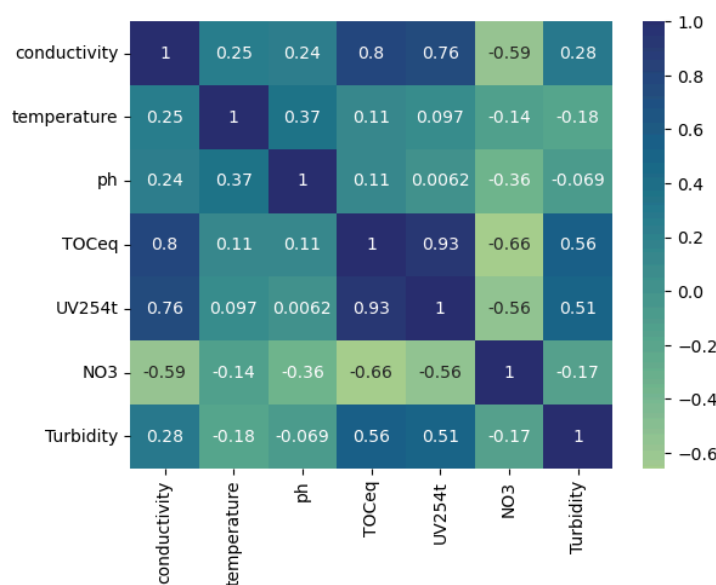


Figure 5.1.b. Correlation matrix for the investigated variables of MARVIS in DS1.

The SHAP values are a powerful tool for calculating each variable's contribution to the predictive model, taking into account the feature interactions. This beeswarm plot ranks the contribution of each variable from most to least important, with TOCeq as the most important feature in predicting nitrate concentration and UV254t as the least important (Figure 5.1.c). TOCeq and pH are the most influential features, with TOCeq having a predominantly positive impact and pH showing a mixed impact. Conductivity, turbidity, temperature, and UV254t have less influence, with conductivity showing a balanced impact around zero, indicating it may not significantly affect the model's predictions. High values of TOCeq and pH tend to increase the model output, while low values of temperature and UV254t have a slight negative impact.

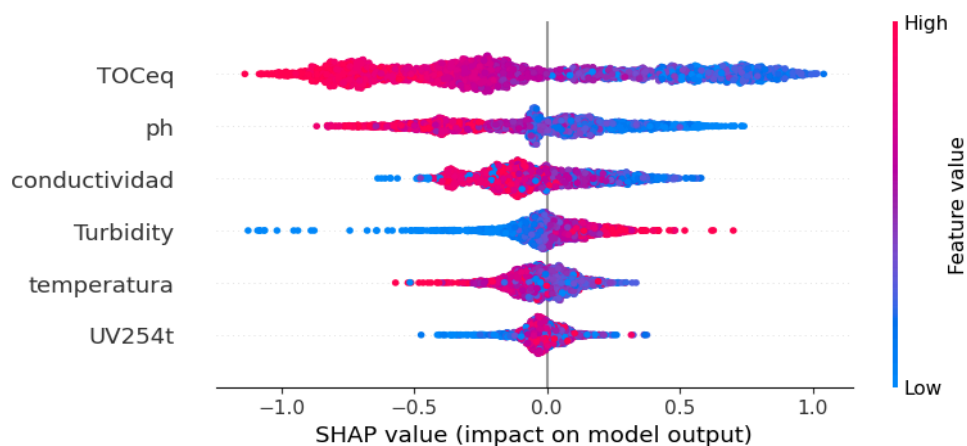


Figure 5.1.c. SHAP values for MARVIS in DS1.

5.1.2 DS3 - Texel, The Netherlands

In the case of DS3, data collection took place from two different monitoring wells: Slot and Broekman. We encountered significant challenges related to accessibility and maintenance. The remote location of DS3 posed logistical difficulties, making regular access to the sensor demanding. Consequently, the cleaning and maintenance schedule, which operates on a monthly basis, proved insufficient to ensure optimal sensor performance for all sensors employed, compromising data quality. However, this experience has been invaluable, as it led us to enhance our maintenance routine and access solutions. We have now implemented a more frequent cleaning schedule, ensuring that our sensors remain in top condition and continue to provide accurate data. This proactive approach will greatly improve the reliability of our sensor data from this site. The initial datasets contained 25.738 and 25.017 observations for Broekman and Slot wells, respectively, with six common predictor variables: TOCeq, turbidity, AU254, conductivity, pH, and temperature. Although data collection is currently ongoing, only data up to 19/12/2024 for the Broekman well and 15/01/2025 for the Slot well were included in the analysis to ensure data quality. The final datasets consisted of 3.466 and 10.514 observations for Broekman and Slot wells. Afterwards, outliers were identified and removed using the IQR method, following the same process as the DS1 dataset. The final datasets comprised 7.759 observations for Slot and 1.690 observations for Broekman. Subsequently, the two datasets were merged. Given the presence of two different water sources, a new site variable was created to indicate the water origin. Despite the current imbalance in the dataset due to differing amounts of data from the two wells, ongoing data collection efforts are expected to mitigate this issue, ensuring that model predictions remain robust and unbiased as the dataset becomes more balanced over time. The

distributions of variables in the entire dataset are illustrated through the histograms reported in Figure 5.1.d.

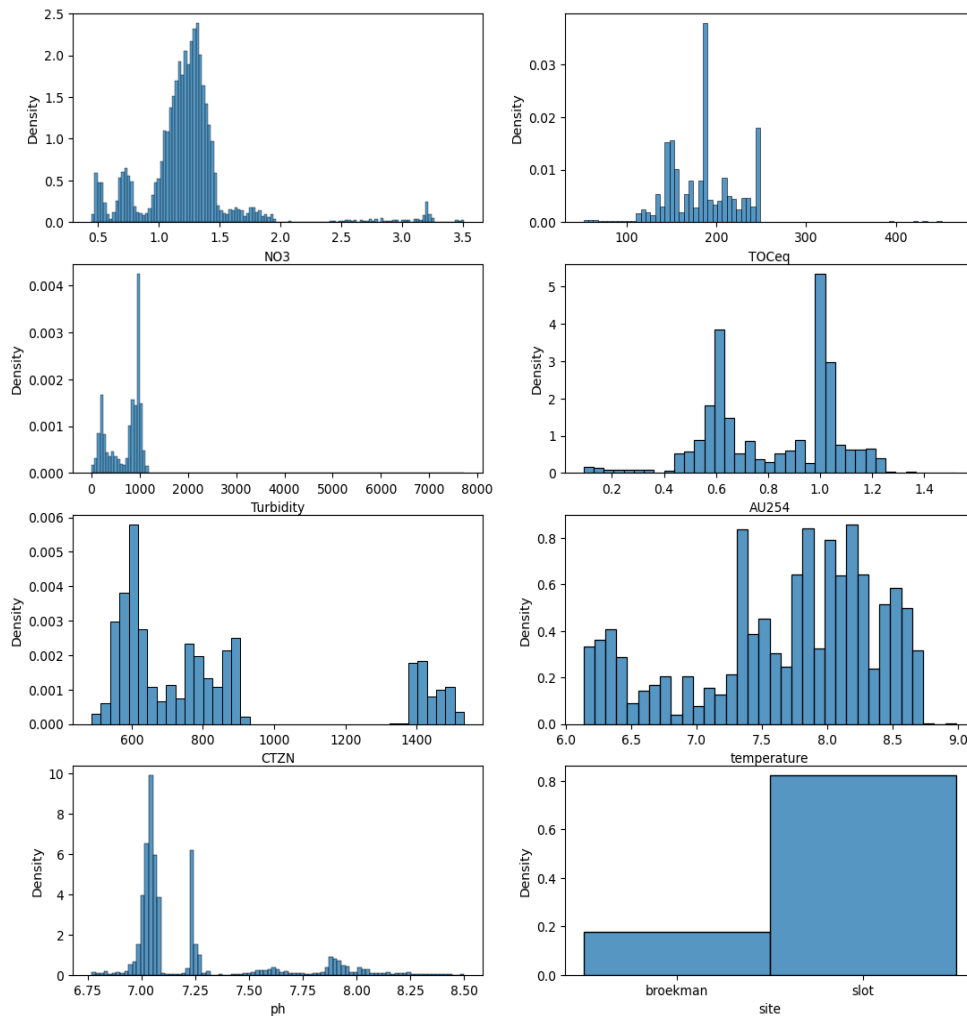


Figure 5.1.d. Frequency distribution of investigated variables for MARVIS at DS1.

Two sites with significant differences in nitrate concentration formed the basis of The Netherlands dataset. The amount of nitrate in water from Slot was higher than in water from Broekman, and there were different patterns for each predictor in both datasets, indicating these variables affect nitrate concentration to a certain extent. Considering the brief duration of the data collection period, caution is advised in making definitive conclusions about the influence of predictor variables on nitrate concentration, as extended data collection is necessary to validate these observations. Using the same method to train and test models on the Barcelona dataset, the XGBoost algorithm

worked best with a learning rate of 0.05 and a maximum depth of 4. This model shows a MSE of 0.002 and R^2 of 0.987.

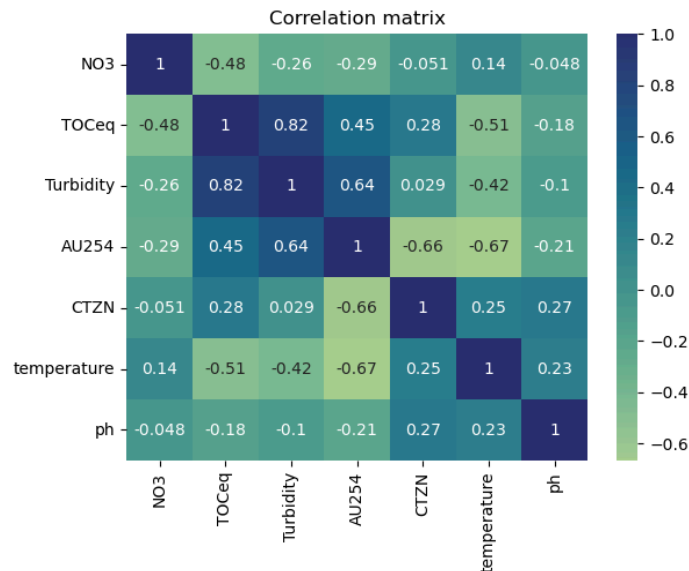


Figure 5.1.e. Correlation matrix for the investigated variables of MARVIS in DS3

The correlation matrix for the Netherlands dataset reveals that nitrate concentration is not strongly correlated with other variables (Figure 5.1.e). The only moderate correlation exists with TOCeq, suggesting that it may be the most significant predictor of nitrate concentration.

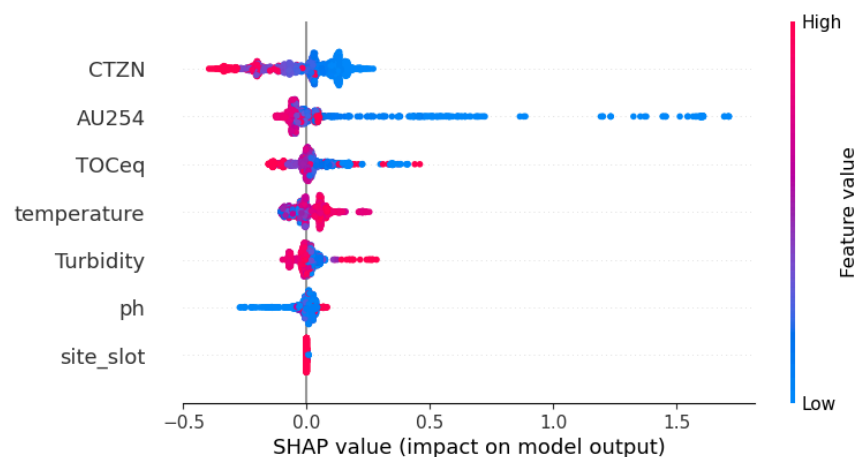


Figure 5.1.f SHAP values for MARVIS in DS3.

Also for DS3 the SHAP values were used to identify the most important features (Figure 5.1.f). The SHAP plot reveals that conductivity, AU254, and turbidity are the most influential features in the model's predictions. Conductivity stands out as a key factor, with high values significantly increasing the model output, indicating its crucial role in shaping the behaviour of the model. Similarly, AU254 is another critical feature, where high values lead to an increase in the model output, underscoring its importance in the prediction process. Turbidity also plays a significant role, with high values contributing to an increase in the model output, highlighting its relevance in the model's decision-making process.

In conclusion, while the preliminary results from the MARVIS project at demo sites DS1 and DS3 are promising, it is important to exercise caution in interpreting these findings due to the limited amount of data currently available. The models developed have shown strong predictive performance, with high R^2 values indicating a good fit to the existing data. However, the correlations observed between predictor variables and nitrate concentration, as well as the feature importance rankings provided by SHAP values, should be considered as initial insights rather than definitive conclusions. As more data is collected and fed into the models, we anticipate improvements in predictive power and a more comprehensive understanding of the factors influencing nitrate concentration. This optimism is grounded in the strategic approach taken, including site-specific model development and plans for transfer learning, which are expected to enhance the accuracy and robustness of predictions across different sites. Continued data collection and model refinement will be crucial in validating these preliminary observations and ensuring that the MARVIS tool can effectively support near real-time water quality monitoring in diverse environmental conditions.

5.2 PREDMAR - AI-based water availability and demand forecasting tool

In order to achieve the goal of creating probabilistic predictions of target variables through the utilization of ML methodologies, the standard phases involved in a supervised learning problem were undertaken. These phases encompass:

- **Data Collection and Pre-processing:** This initial stage involved the gathering of relevant data and its subsequent preparation for use in ML algorithms. This often includes cleaning the data, handling missing values, transforming features, and normalizing the data.
- **Learning:** In this phase, a ML model was trained on the pre-processed data. The model learns the underlying patterns and relationships between the input features and the target variable. This is typically done through an iterative process of adjusting the model's parameters to minimize the error between its predictions and the true values.

- **Evaluation:** Once the model has been trained, its performance is evaluated on a separate dataset that was not used during training. This is done to assess how well the model generalizes to unseen data and to avoid overfitting. Various metrics are used to measure the model's performance, depending on the specific task and the type of model.
- **Prediction:** Finally, the selected model is used to make predictions on new, unseen data (Figure 5.2.a).

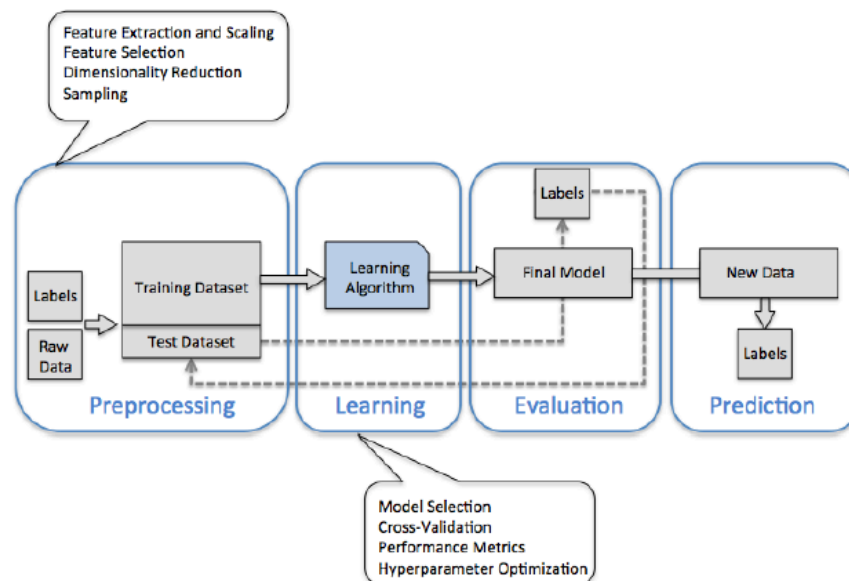


Figure 5.2.a General methodology for developing ML models. (Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017)).

Autoregressive modeling techniques, such as the sliding window method, are instrumental in transforming the challenge of forecasting future values within a time series into a more manageable supervised learning problem. This transformation is achieved by reframing the task as predicting the next value in the sequence based on a window of preceding values (Figure 5.2.b). By structuring the data in this manner, traditional ML algorithms can be readily applied to the task of time series prediction. The sliding window approach involves creating a moving window of fixed length over the time series data. Each window represents a training instance where the input features are the values within the window and the target variable is the subsequent value in the series. As the window slides along the series, new training instances are generated, capturing the temporal dependencies and patterns within the data. This technique effectively leverages the inherent sequential nature of time series data, enabling ML models to learn from past observations and make informed predictions about future values. The supervised learning framework provided by autoregressive modeling facilitates the application of a wide range of ML algorithms, including linear regression, decision trees, and neural networks, to the domain of time series forecasting.

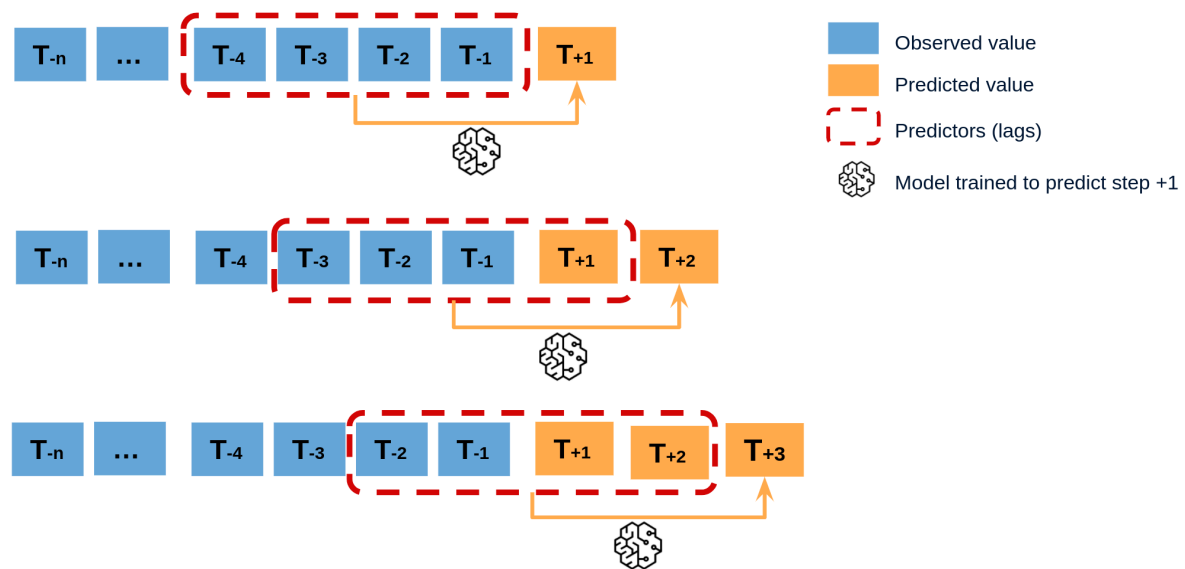


Figure 5.2.b Diagram of the recursive multi-step forecasting. Source: https://skforecast.org/0.11.0/user_guides/autoregressive-forecaster.

The initial stage in the development of robust and reliable predictive models involves a comprehensive search and definition process to identify the most suitable model architecture for the specific task at hand. This is followed by a series of semi-automated experiments, where numerous model variations are generated and rigorously compared based on key performance metrics such as Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (sMAPE), and Mean Directional Accuracy (MDA). This iterative experimentation and evaluation process continues until the models that demonstrate the highest validated predictive accuracy against historical data are identified.

The main metrics used to evaluate the performance of the models are as follows:

- MAE: It's the average of the absolute difference between the observed value and the predicted values. The mean absolute error is linear, meaning that all individual differences are equally weighted in the average.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

- MDA: It's a metric that evaluates the model's ability to correctly predict the direction of movements in the data. It's calculated by dividing the number of times the actual change direction matches the predicted change direction by the model, by the total number of predictions.

$$MDA = \frac{1}{n} \sum_j 1_{\text{sign}(X_j - X_{j-1}) == \text{sign}(F_j - X_{j-1})}$$

- sMAPE: It measures the relative accuracy of predictions by calculating the average absolute percentage difference between actual and predicted values. This metric will not be considered when the values fluctuate between -1 and 1 as it can lead to errors.

$$sMAPE = \frac{1}{n} \sum_{j=1}^n \frac{|\hat{y}_j - y_j|}{(|y_j| + |\hat{y}_j|)/2}$$

To ensure the selection of the model that best aligns with the desired objectives and generalizes well to unseen data, cross-validation, a widely used statistical technique, is employed. This method involves partitioning the available dataset into multiple subsets or folds. The model is then systematically trained on all but one of these folds and its performance is evaluated on the held-out fold. This process is repeated iteratively until the model has been tested on each of the folds, and the final performance metrics are calculated as the average of the scores obtained across all the folds. By utilizing cross-validation, the risk of overfitting, where the model learns the training data too well and fails to generalize to new data, is significantly mitigated. Additionally, this technique provides a more robust assessment of the model's performance compared to simple training and validation approaches, as it evaluates the model's ability to predict on multiple different subsets of the data.

In time series, the temporal dependence between observations must be preserved. Therefore, we cannot randomly assign samples to the validation or training sets. Instead, we use a small subset of data for initial training and predict subsequent data. Then, we evaluate the accuracy of the predicted data. The same forecasted data are then incorporated into the next training dataset, and subsequent data are forecasted (Figure 5.2.c).

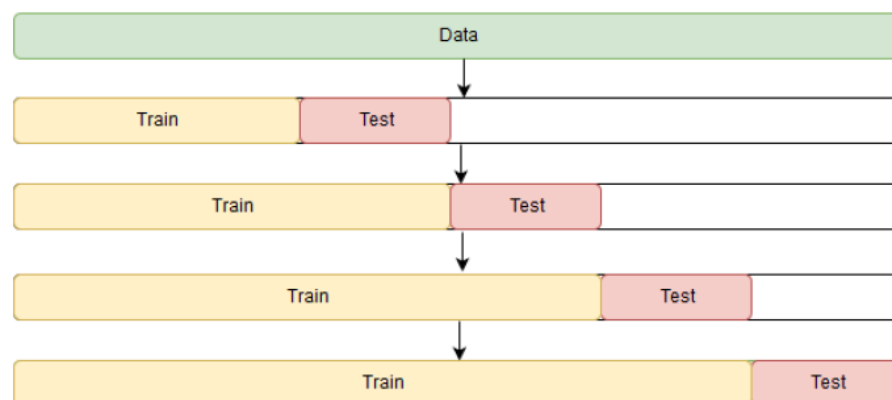


Figure 5.2.c Time series cross-validation.⁵

⁵ Source: <https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4>.

The list of algorithms being evaluated includes, among others, the Huber Regressor, Support Vector Regression, Random Forest, XGBoost, CatBoost, Linear Regression, Ridge, Stochastic Gradient Descent Regression, ElasticNet, Lasso, LassoLars, Orthogonal Matching Pursuit, Bayesian ARD Regression, Bayesian Ridge Regression, Quantile Regressor, Passive Aggressive Regressor, and K-nearest Neighbors Regressor.

To guarantee the dependability and robustness of the chosen predictive model, a rigorous validation process is implemented. This process encompasses a series of training procedures and predictions executed at various randomly selected points within the historical dataset. By simulating the model's behavior as if it had been operational at those specific moments in the past, we can retrospectively evaluate its performance and predictive accuracy. This approach, often referred to as "backtesting", allows us to assess how the model would have responded to different data inputs, market conditions, and environmental variables over time, providing valuable insights into its overall effectiveness, potential limitations, and susceptibility to overfitting. Furthermore, this temporal validation approach helps to identify any potential biases or inconsistencies in the model's predictive capabilities across different time periods. By examining the model's performance at various historical junctures, we can determine if its accuracy and reliability remain consistent or if there are specific periods where it exhibits underperformance or overfitting. This information is crucial for fine-tuning the model's parameters, adjusting its assumptions, and ensuring its robustness in the face of changing conditions. Additionally, the backtesting process allows us to compare the model's predictions against actual historical outcomes, enabling us to quantify its predictive accuracy and identify any systematic errors or biases. By analyzing the discrepancies between predicted and actual values, we can gain a deeper understanding of the model's strengths and weaknesses, and make informed decisions about its suitability for specific applications. Overall, the rigorous validation process, including backtesting at multiple historical time points, is essential for ensuring the reliability, accuracy, and robustness of the selected predictive model. By thoroughly evaluating the model's performance under diverse historical conditions, we can confidently deploy it for future predictions and decision-making, while remaining aware of its potential limitations and areas for improvement.

These predictions are probabilistic in nature, meaning that they provide an estimate of the likelihood of each possible outcome. In inferential statistics, specifically in predictive inference, a prediction interval is an estimation of a range of values within which a future observation will fall with a certain probability, given what has already been observed. Prediction intervals are commonly used when making predictions based on regression models, as is the case with the work conducted. This probabilistic approach has a solid and robust scientific basis [47].

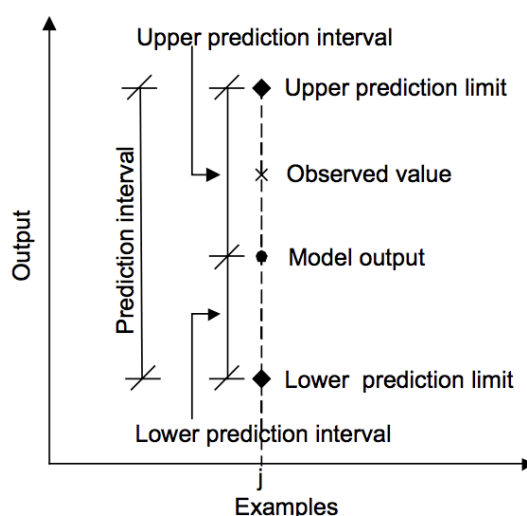


Figure 5.2.d Conceptual framework illustrating the relationship between predicted values (model output) and actual values (observed value), as well as the prediction interval (prediction interval), with its corresponding upper prediction limit and lower prediction limit.⁶

Preliminary results from DS1 and DS2 are presented in this report. However, results from DS3 have not been included at this stage, as the methodology for this demo site is still undergoing refinement. Final adjustments are currently being implemented, and the complete results for DS3 will be presented in upcoming deliverables, once data from the next growing season becomes available.

5.2.1 DS1 - LLobregat (Spain)

Short-Term module - DS1

The primary objectives of the Short-Term Module at the DS1 Llobregat demonstration site, as outlined in the conceptual scheme of Section 4.1.2, are defined as follows:

1. Forecast the volume of water stored in the four reservoirs of the Ter-Llobregat system in the short term.
2. Assess the current state of drought in the system based on water availability.
3. Forecast streamflow at the SVH gauging station over the next 7 days.

To address the first two objectives, a predictive system has been developed to project the monthly water storage volumes in the four reservoirs over a six-month horizon. This system is designed to deliver reliable forecasts of future reservoir levels, enabling proactive water resource management. Furthermore, through the integration of the Special Drought Plan (PES), the model will assess the system's drought-scarcity status, providing essential insights for early warning mechanisms, informed decision-making,

⁶ Source: <https://machinelearningmastery.com/prediction-intervals-for-machine-learning/>

and the implementation of mitigation strategies to minimize the impacts of drought conditions.

To address the second objective are three different scenarios on which the drought states will be obtained using the predicted volumes:

- Joint reserves Sau, Susqueda, La Llosa del Cavall, St. Ponç and La Baells)
- Ter Reserves (Sau and Susqueda)
- Llobregat Reserves (La Llosa del Cavall, St. Ponç i La Baells),

The PES is used to determine the thresholds and methodology for assigning statuses (Figures 5.2.e y 5.2.f) [15].

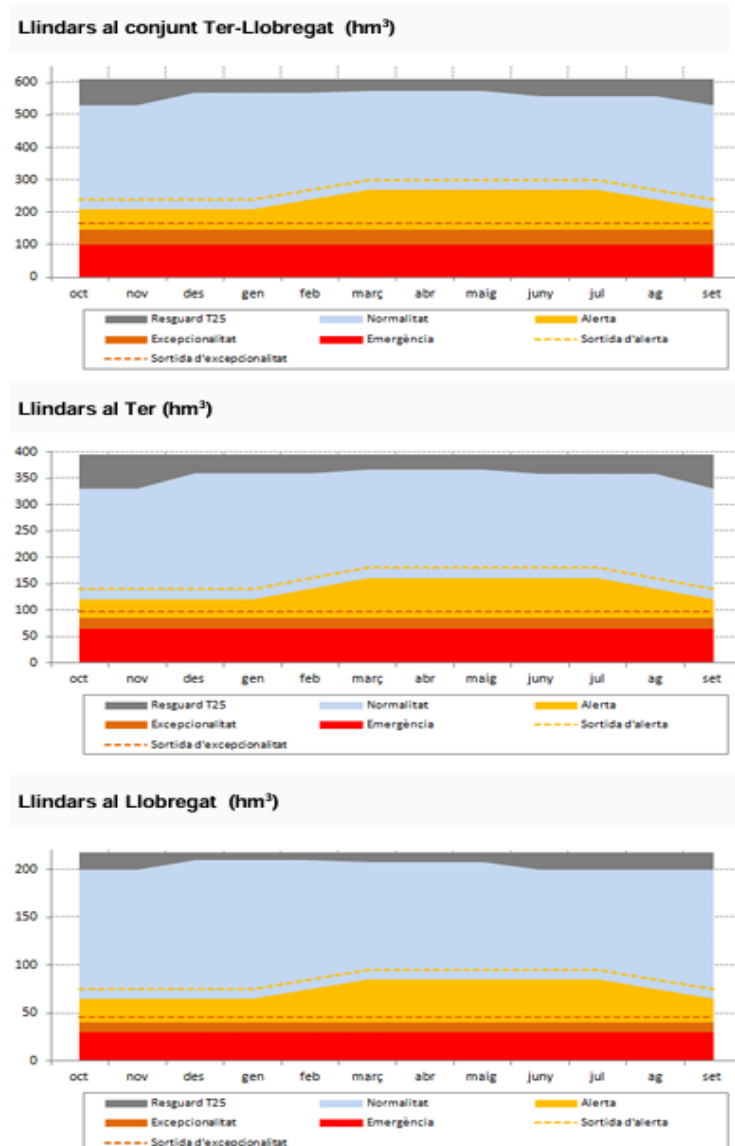


Figure 5.2.e Hydrological drought thresholds ("lindars" in Catalan) figure.

TER - LLOBREGAT	Oct	Nov	Des	Gen	Feb	Març	Abr	Maig	Juny	Jul	Ag	Set
Emb. màxim	612	612	612	612	612	612	612	612	612	612	612	612
Resguard avinguda 25 anys	530	530	569	569	569	574	574	574	558	558	558	530
Prealerta	367	367	367	367	367	367	367	367	367	367	367	367
Sortida Alerta	240	240	240	240	270	300	300	300	300	300	270	240
Alerta	210	210	210	210	240	270	270	270	270	270	240	210
Sortida Excepcionalitat	165	165	165	165	165	165	165	165	165	165	165	165
Excepcionalitat	145	145	145	145	145	145	145	145	145	145	145	145
Emergència I	100	100	100	100	100	100	100	100	100	100	100	100
Emergència II	67	67	67	67	67	67	67	67	67	67	67	67
Emergència III	33	33	33	33	33	33	33	33	33	33	33	33
TER	Oct	Nov	Des	Gen	Feb	Març	Abr	Maig	Juny	Jul	Ag	Set
Emb. màxim	395	395	395	395	395	395	395	395	395	395	395	395
Resguard avinguda 25 anys	331	331	360	360	360	367	367	367	359	359	359	331
Prealerta	200	200	200	200	200	200	200	200	200	200	200	200
Sortida Alerta	140	140	140	140	160	180	180	180	180	180	160	140
Alerta	120	120	120	120	140	160	160	160	160	160	140	120
Sortida Excepcionalitat	98	98	98	98	98	98	98	98	98	98	98	98
Excepcionalitat	85	85	85	85	85	85	85	85	85	85	85	85
Emergència I	65	65	65	65	65	65	65	65	65	65	65	65
Emergència II	43	43	43	43	43	43	43	43	43	43	43	43
Emergència III	14	14	14	14	14	14	14	14	14	14	14	14
LLOBREGAT	Oct	Nov	Des	Gen	Feb	Març	Abr	Maig	Juny	Jul	Ag	Set
Emb. màxim	217	217	217	217	217	217	217	217	217	217	217	217
Resguard avinguda 25 anys	199	199	209	209	209	207	207	207	199	199	199	199
Prealerta	100	100	100	100	100	100	100	100	100	100	100	100
Sortida Alerta	75	75	75	75	85	95	95	95	95	95	85	75
Alerta	65	65	65	65	75	85	85	85	85	85	75	65
Sortida Excepcionalitat	46	46	46	46	46	46	46	46	46	46	46	46
Excepcionalitat	40	40	40	40	40	40	40	40	40	40	40	40
Emergència I	30	30	30	30	30	30	30	30	30	30	30	30
Emergència II	20	20	20	20	20	20	20	20	20	20	20	20
Emergència III	7	7	7	7	7	7	7	7	7	7	7	7

Figure 5.2.f Hydrological drought thresholds (Table in Catalan) alerta : alert; sortida : exit; excepcionalitat :exceptionality; emergència : emergency.

The integration of the forecasted stored water volumes with the established hydrological drought thresholds generates a comprehensive output that effectively depicts the overall status of the water system. This integrated approach allows for a more accurate and timely assessment of the water system's vulnerability to drought conditions, enabling stakeholders to proactively implement mitigation measures and optimize water resource management strategies (Table 5.2.j).

Finally for the third objective, a predictive system was created to forecast volumes using autoregressive flow variables at the gauging station and specific meteorological variables. This system generates seven-day forecasts. These predictions are aligned with the selected future physical-based weather model [45], ensuring consistency with projected climatic conditions and enhancing the system's forecasting accuracy.

After the ML models have been thoroughly trained on relevant data, the next crucial step is to rigorously validate them using appropriate techniques. This validation process is essential to ensure the models' accuracy, reliability, and generalizability. Once the models have been successfully validated, they can then be deployed to generate predictions on new, unseen data. These predictions can be used to inform decision-making, drive insights, and support a wide range of applications.

Validation process

To validate the chosen model for the stored volume at each point of interest over the next six months, predictions have been carried out at 10 random moments during the last 10 years of data. The outcome consists of 60 predictions across different months of the period selected. In some instances, a date has appeared more than once among the 10 randomly selected dates. To maintain the randomness and avoid considering the same date twice, the decision has been made to remove the duplicate date without replacing it with an additional, different date.

In addition to the resulting metrics from comparing the actual values with the predicted values for each point, a representative graph of the set of predictions used in the validation is also provided.

The results for the reservoirs are shown below (Tables 5.2.a to 5.2.d and Figures 5.2.g to 5.2.j).

Table 5.2.a Validation metrics in La Llosa del Cavall volume of water stored

Date prediction	SMAPE (%)	MDA	MAE
2017-01	11.11	60	7.07
2018-04	27.32	80	18.48
2018-07	21.66	40	15.06
2018-10	24.46	60	16.76
2019-01	12.25	40	8.99
2019-02	3.89	40	2.94
2019-03	4.92	60	3.31
2019-05	17.81	100	11.38
2020-03	2.05	40	1.53
2020-04	5.95	60	4.2
AVERAGE	13.14	58	8.97

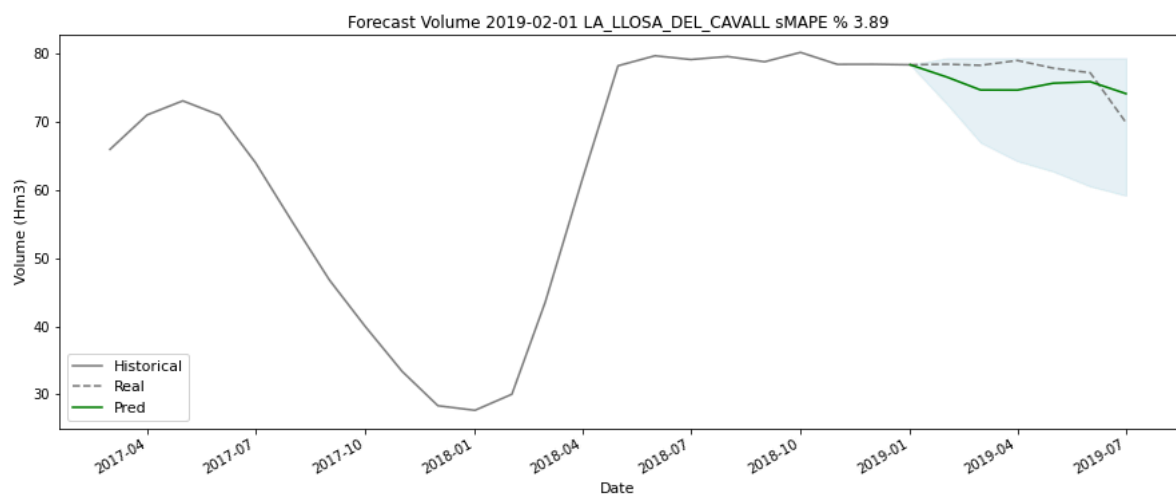


Figure 5.2.g La Llosa del Cavall volume of water stored validation example

Table 5.2.b Validation metrics in La Baells volume of water stored

Date prediction	SMAPE (%)	MDA	MAE
2014-01	6.94	80	6.07
2016-01	7.6	80	6.55
2016-05	6.75	80	5.98
2016-11	13.42	60	12.41
2017-04	19.84	40	18.48
2017-05	21.5	80	18.46
2017-07	23.81	60	17.37
2017-11	22.53	80	16.11
2019-01	7.19	40	7.39
2019-04	9.19	80	8.48
AVERAGE	13.88	68	11.73

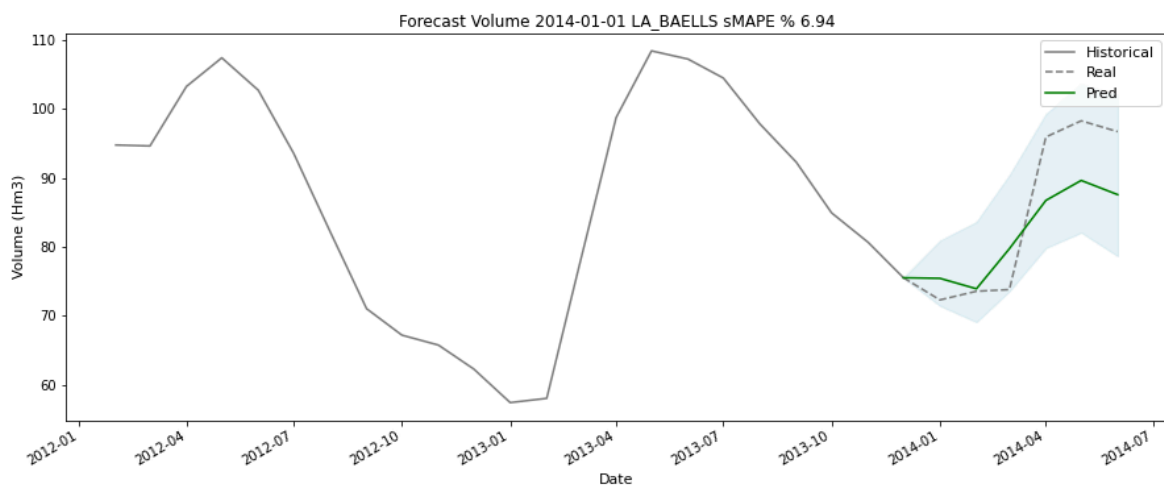


Figure 5.2.h La Baells volume of water stored validation example

Table 5.2.c Validation metrics in Susqueda volume of water stored

Date prediction	SMAPE (%)	MDA	MAE
2015-01	5.03	40	11.22
2015-04	13.52	0	29.07
2015-07	14.29	100	26.83
2016-06	24.32	20	36.83
2018-05	4.17	20	9.14
2019-09	23.52	20	39.49
2020-02	5.44	40	12.17
2020-12	3.4	60	7.25
2021-07	4.94	60	9.07
2021-09	8.39	60	14.39
AVERAGE	10.7	42	19.55

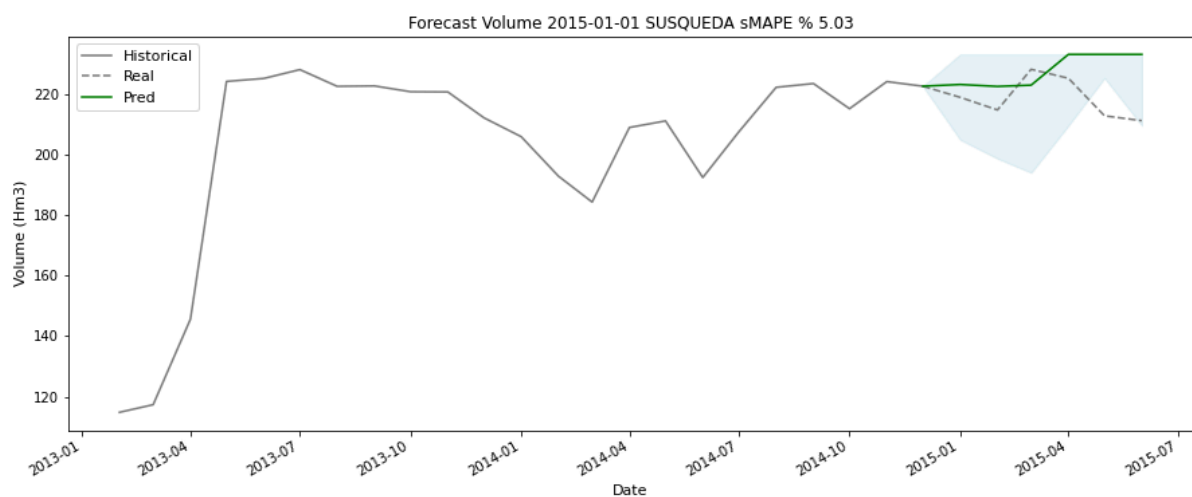


Figure 5.2.i Susqueda volume of water stored validation example

Table 5.2.d Validation metrics in Sau volume of water stored

Date prediction	SMAPE (%)	MDA	MAE
2014-01	2.97	40	3.67
2016-01	9.95	80	9.71
2016-05	11.97	80	9.99
2016-11	17.33	40	17.43
2017-04	16.55	100	17.38
2017-05	6.38	80	6.6
2017-07	19.24	60	15.93
2017-11	25.24	20	22.08
2019-01	19.06	80	24.91
2019-04	7.15	100	8.15
AVERAGE	13.58	58	13.59

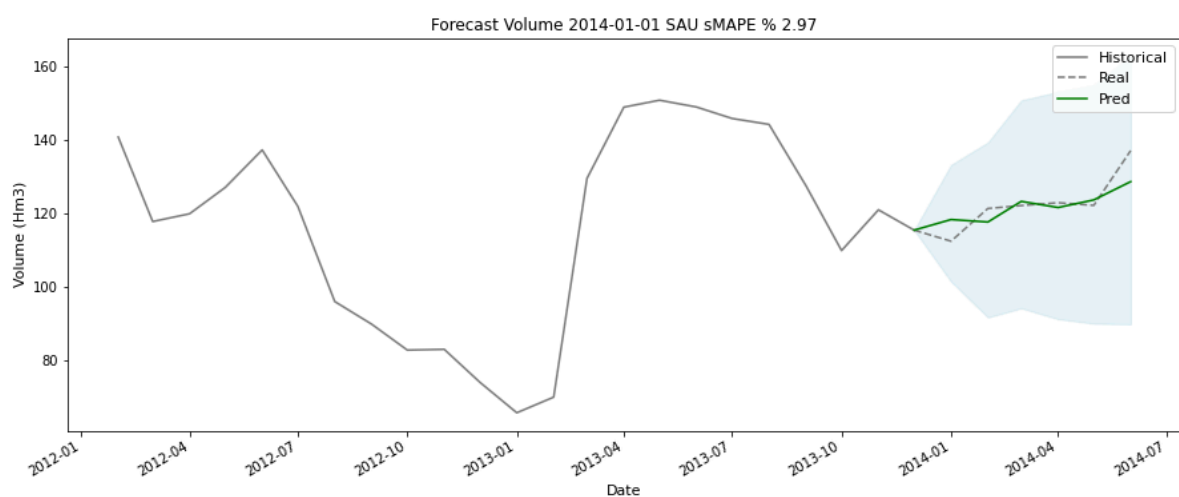


Figure 5.2.j Sau volume of water stored validation example

The validation results for the streamflow forecast at the Sant Vicenç dels Horts station are as follows (Table 5.2.e and Figure 5.2.i):

Table 5.2.e Validation metrics in SVH 7-day streamflow forecast

Date prediction	MDA	MAE
2024-10-10	50	3.30
2024-10-12	50	2.55
2024-10-21	20	12.25
2024-10-28	67	14.23
2024-10-31	67	20.51
2024-11-09	67	3.08
2024-11-14	50	3.62
2024-11-17	67	5.87
2024-11-30	50	3.09
2024-12-05	50	1.49
AVERAGE	55	7.00

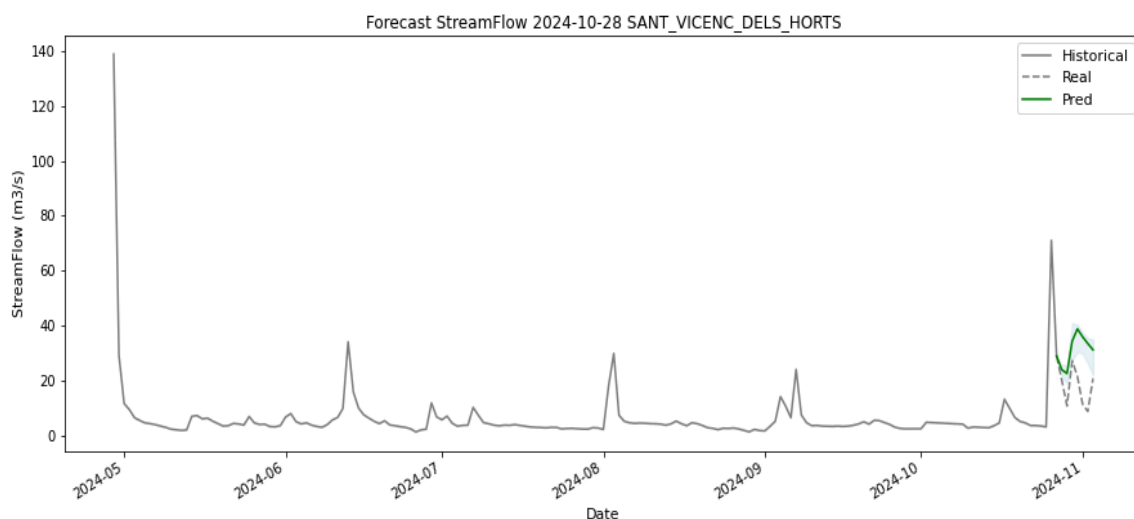


Figure 5.2.k SVH streamflow validation example

Feature importance

To assist the user in interpreting the predictions of the selected models, some plots have been generated using the unified framework for prediction interpretation, SHAP.

For each point of interest, two plots will be provided:

- Bar plot (left): the mean absolute value of the SHAP values for each feature is simply taken to produce a standard bar chart, indicating which variables are most significant in the prediction.
- Beeswarm plot (right): provides an overview of which features are most important for a model by plotting the SHAP values of each feature for each sample. The above chart arranges the features by the sum of the magnitudes of the SHAP values across all samples and uses the SHAP values to show the distribution of the impact of each feature on the model's outcome. The colour represents the feature value (high in red, low in blue).

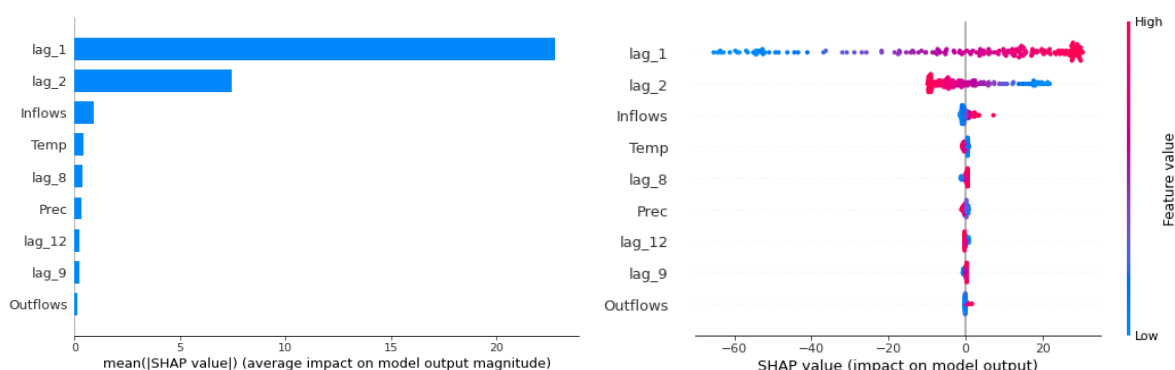


Figure 5.2.l La Llosa del Cavall feature importance

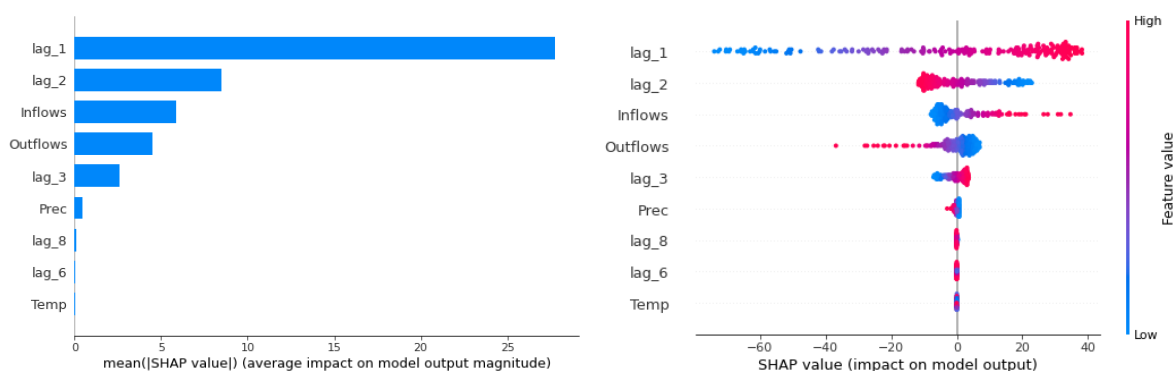


Figure 5.2.m La Baells feature importance

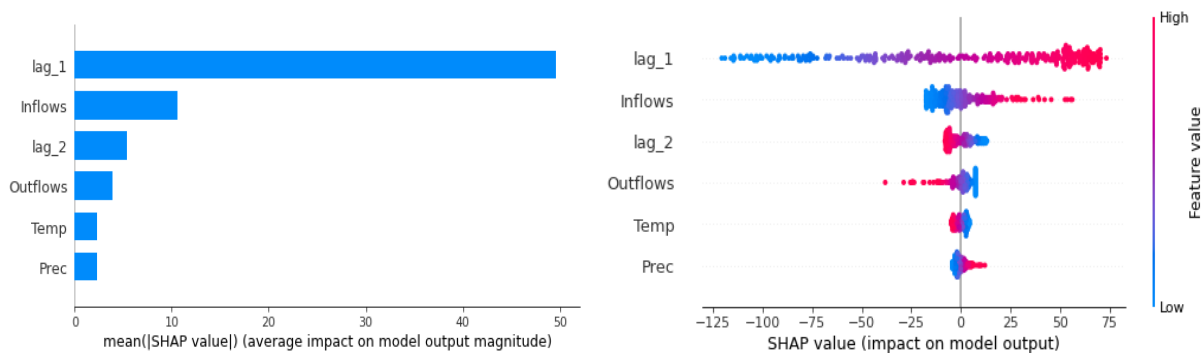


Figure 5.2.n Susqueda feature importance

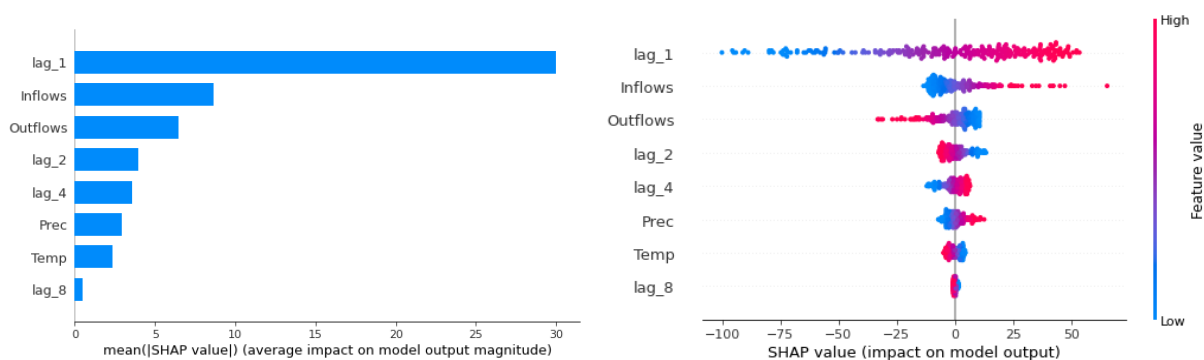


Figure 5.2.o Sau feature importance

Predictions

This section provides detailed future predictions for several key hydrological factors, including:

- **Modelled points of stored water** within each reservoir. These predictions will be based on the reservoir models developed and described earlier in this document. They will take into account various factors such as inflows, outflows, and evaporation.
- **Anticipated drought state** of the Ter-Llobregat region. This will be based on the drought prediction model developed earlier in this document. It will consider factors such as rainfall, temperature, and soil moisture.
- **Projected streamflow** for SVH (the Llobregat, Besòs and Ter river basins system). This will be based on the streamflow prediction model developed earlier in this document. It will consider factors such as rainfall, snowmelt, and antecedent moisture conditions.

These predictions are based on the models developed and described earlier in this document and will provide valuable insights for future water resource management and

planning in the Ter-Llobregat region. By understanding how these key hydrological factors are likely to change in the future, water managers can make more informed decisions about how to allocate water resources, mitigate the impacts of drought, and ensure the long-term sustainability of the water supply.

La Llosa del Cavall: Stored volume

Table 5.2.f La Llosa del Cavall: Stored volume forecast (hm^3)

Date prediction	Prediction	Lower bound	Upper bound
2024-10	19.06	14.34	22.31
2024-11	19.99	15.00	24.99
2024-12	21.74	16.30	27.18
2025-01	23.94	17.96	29.92
2025-02	25.81	19.36	32.26
2025-03	27.40	20.54	34.23

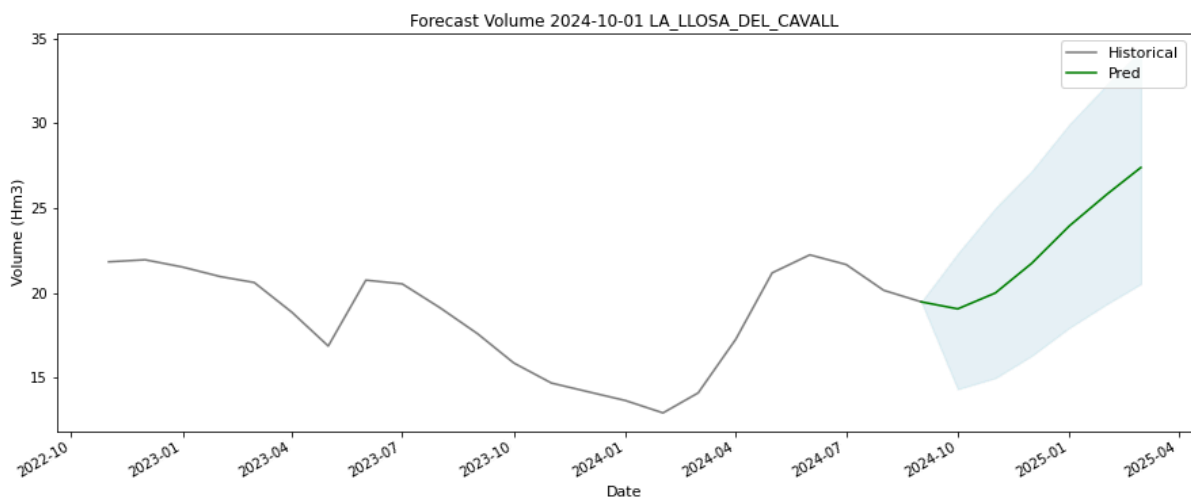


Figure 5.2.p La Llosa del Cavall: Stored volume forecast

La Baells: Stored volume

Table 5.2.g La Baells: Stored volume forecast (hm³)

Date prediction	Prediction	Lower bound	Upper bound
2024-10	42.48	38.48	47.18
2024-11	42.15	37.13	48.47
2024-12	44.79	38.65	54.01
2025-01	44.25	36.35	54.49
2025-02	44.05	34.95	54.66
2025-03	45.42	34.06	56.66

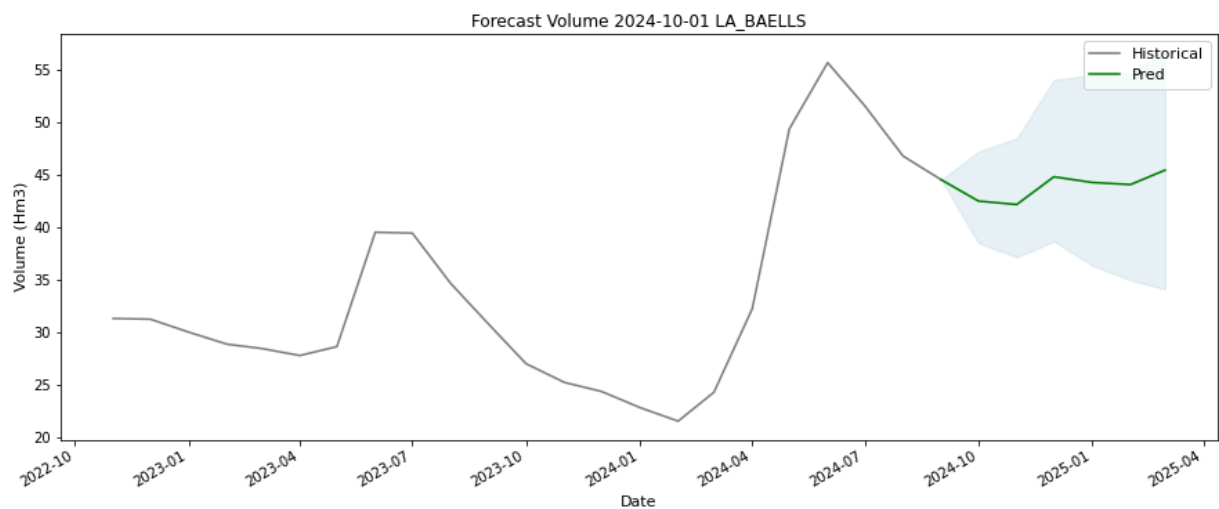


Figure 5.2.q La Baells: Stored volume forecast

Susqueda: Stored volume

Table 5.2.h Susqueda: Stored volume forecast (hm^3)

Date prediction	Prediction	Lower bound	Upper bound
2024-10	79.09	62.64	98.86
2024-11	79.95	59.96	99.94
2024-12	88.31	66.23	110.38
2025-01	95.46	71.59	119.32
2025-02	97.46	73.09	121.82
2025-03	99.62	74.71	124.52

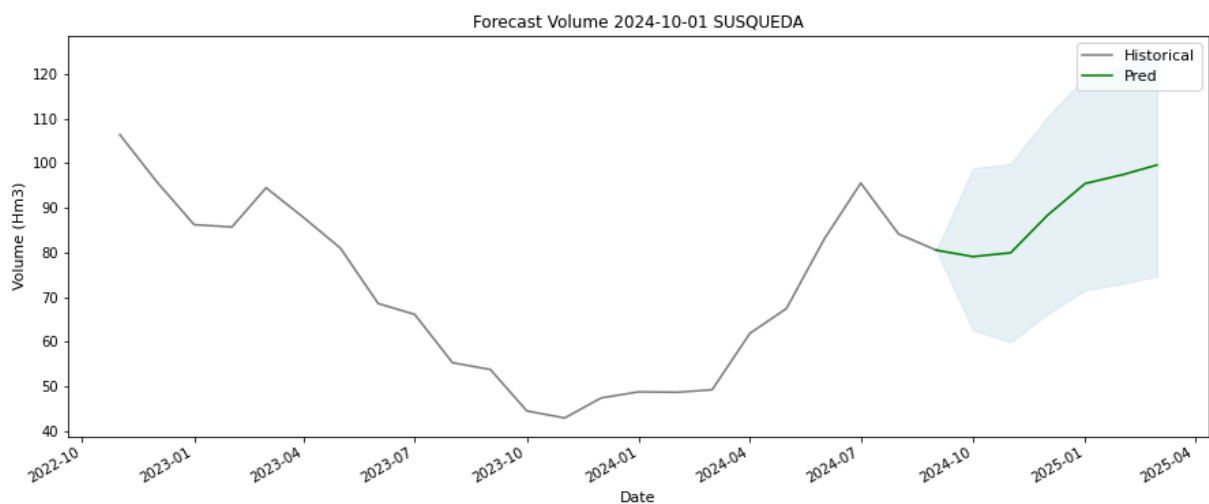


Figure 5.2.r Susqueda: Stored volume forecast

Sau: Stored volume

Table 5.2.i Sau: Stored volume forecast (hm^3)

Date prediction	Prediction	Lower bound	Upper bound
2024-10	26.52	19.89	33.15
2024-11	29.49	22.12	36.87
2024-12	30.54	22.90	38.18
2025-01	39.65	29.73	49.56
2025-02	41.60	31.20	52.00
2025-03	44.24	33.18	55.31

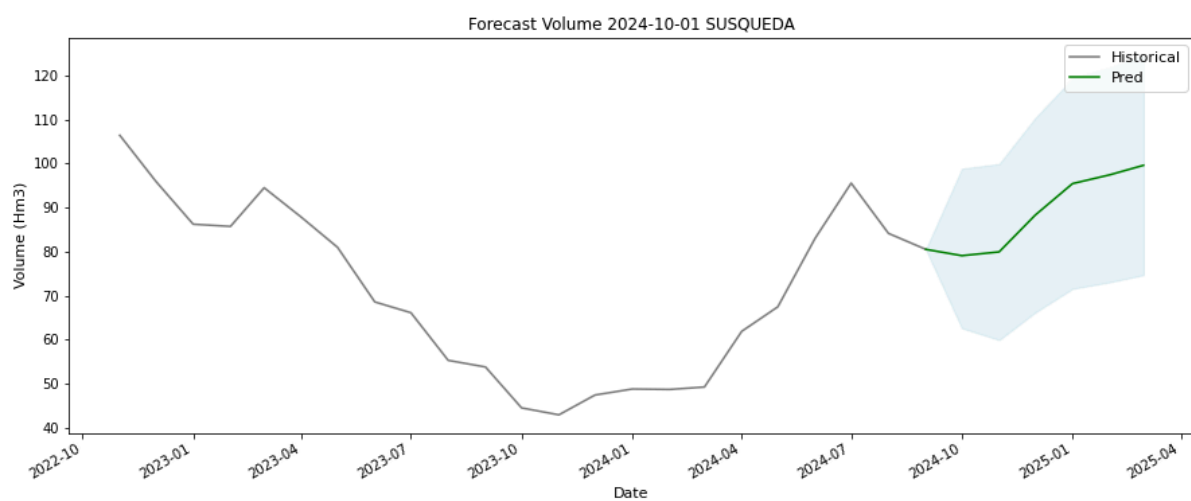


Figure 5.2.s Sau: Stored volume forecast

Ter-Llobregat: drought state

Table 5.2.j Ter-Llobregat drought state

Date prediction	Llobregat state	Ter state	Combined stated	Final state
2024-10	Alert	Alert	Alert	Alert
2024-11	Alert	Alert	Alert	Alert
2024-12	Alert	Alert	Alert	Alert
2025-01	Alert	Alert	Alert	Alert
2025-02	Alert	Alert	Alert	Alert
2025-03	Alert	Alert	Alert	Alert

Sant Vicenç dels Horts: Streamflow

Table 5.2.k SVH: Streamflow forecast (m³/s)

Date prediction	Prediction	Lower bound	Upper bound
2024-12-15	7.52	5.32	9.90
2024-12-16	7.39	3.77	11.41
2024-12-17	7.32	3.32	13.82
2024-12-18	7.34	2.83	15.62
2024-12-19	7.13	2.36	18.75
2024-12-20	7.11	2.33	17.32
2024-12-21	7.04	2.57	19.94

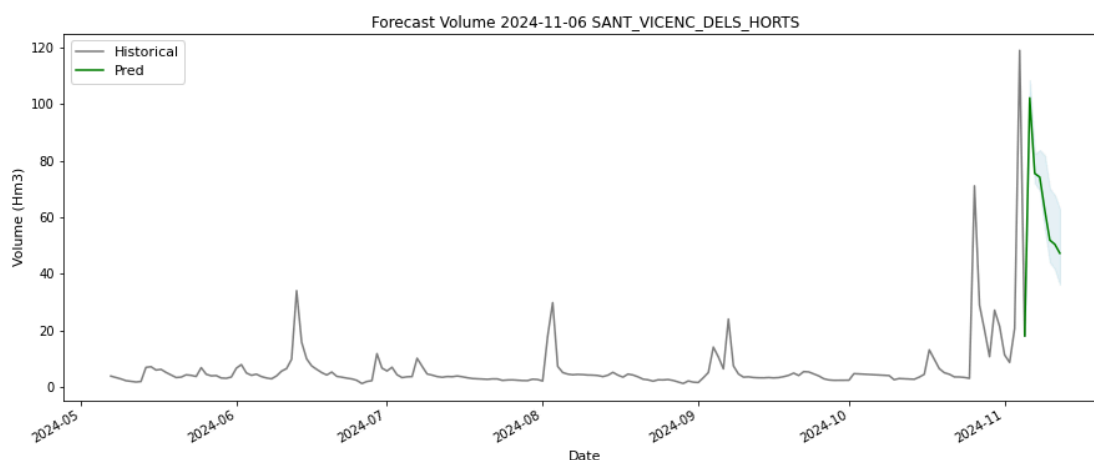


Figure 5.2.t SVH: Streamflow forecast

Long-Term module - DS1

The final datasets used for training and prediction in ML models were meticulously compiled from a diverse range of sources. A significant portion of the data was sourced from the CMIP6 initiative as explained in the 4.1.2 section, which provides high-quality climate data based on multiple global models. The raw data underwent a series of feature engineering processes to create a set of informative variables that complement the primary reservoir volume data, enhancing the model's predictive capabilities.

The engineered features generated during this process include seasonal, monthly, and yearly aggregates of essential climatological variables such as precipitation and temperature. Specifically, precipitation and temperature data were compiled for the entire year, with additional focus on the first and last months of the year, as well as each of the four seasons: spring, summer, autumn, and winter. These features allow the models to consider temporal variations and their potential impact on water volume fluctuations across different periods.

In addition to these climatological variables, the dataset also incorporated inflow and outflow variables, which are critical for understanding water volume dynamics in the short-term module. These inflow and outflow data points provide key insights into the movements of water into and out of the reservoirs, which are directly tied to the overall system's behavior. Another important aspect of the dataset construction was the integration of future climate projections. Based on the scenarios from the CMIP6 initiative, projections of future precipitation were incorporated into the dataset to assess how different climate scenarios might influence water availability. These projections were generated for each of the five primary RCP-SSP scenarios ranging from low to high emissions scenarios and were used to model the potential future behavior of the system.

For each of the five RCP-SSP scenarios, three separate datasets were generated to capture a range of possible outcomes. The first dataset represents the expected mean value derived from the ten climate models analyzed within the CMIP6 framework. The second and third datasets capture the extreme ends of the predicted range: the minimum and maximum mean values. This approach ensures that the models account for the inherent uncertainty in climate projections and provide a more comprehensive analysis of the potential future states of the system.

Upon obtaining the trained models for each reservoir and dataset, a total of 15 models (3 models per scenario) will be utilized to forecast the average annual volume of each reservoir through to the year 2100. These predictions will be based on the various climate scenarios defined by the respective RCP-SSP pathways. The forecasted data will provide valuable insights into the future water availability for each reservoir under the projected climate conditions. To further assess the system's resilience and vulnerability to future drought conditions, the predicted reservoir volumes will be integrated with the PES. This integration will enable a comprehensive evaluation of the drought status across the Ter-Llobregat system, as it will rely on the predicted water volumes in conjunction with established drought indicators outlined in the PES framework. The analysis of drought severity will inform water resource management strategies and support decision-making processes aimed at mitigating the potential impacts of future drought scenarios.

5.2.2 DS2 - Comporta (Portugal)

A methodology similar to the one applied in the DS1 demo site in Spain, particularly for the long-term module will be followed to construct a final dataset for making predictions for the target variable S , as defined in section 3.2.

The procedure for constructing the final dataset, which includes all the variables of interest for the Sado River basin, is outlined below.

1. Download CDS data.

- The initial step in comprehensively understanding the structure of the dataset will involve dividing it into two distinct sections to facilitate a clear separation between historical and future data. The first section will include historical data, spanning from 1942 to 2013. This period covers a substantial historical timeframe that provides valuable insights into past climate and environmental conditions. It will contain the core variables essential for understanding the historical context, such as past temperature trends, precipitation patterns, and other climatic or hydrological variables. The second section will be dedicated to future data, specifically focusing on the period from 2014 to 2100. This part of the dataset will contain projections for the explanatory variables, which are derived from climate models and represent future trends based on various socio-economic and environmental scenarios. These future data points will

provide critical information on how the studied system is expected to evolve under different climate scenarios, taking into account factors such as greenhouse gas emissions, technological advancements, and policy interventions. Projections for variables such as future temperature, precipitation, and other climate drivers will be used to generate forecasts and assess future challenges related to water availability, climate risks, and environmental sustainability.

- Data retrieval from the CMIP6 archive necessitates the specification of geographic coordinates, irrespective of the model chosen. This involves defining the minimum and maximum latitude and longitude values that delineate the desired region, thereby ensuring the acquisition of pertinent datasets only. Furthermore, the process often requires the establishment of supplementary criteria, such as temporal ranges (e.g., historical or projected data) and climate scenarios (e.g., RCP or SSP), to select the most suitable data aligned with the research objectives.

2. Preparation of datasets.

- The process involves integrating the 'S' column into the existing historical data and subsequently merging this enriched historical dataset with each of the individual datasets generated for each specific SSP scenario. This methodical combination will create a comprehensive dataset that incorporates both historical trends and projected future scenarios under different SSPs.
- Each dataset will have the following characteristics:
 - Data spanning from 1942 to 2100. Data from 1942-2014 represents historical data, while data from 2015-2100 corresponds to future values.
 - Each dataset includes seven columns per variable (annual, four for the stations, one for January, and one for December), the S variable column, and the date column, totaling 30 columns per file.
- The four represented variables are:
 - Runoff (abbreviated as "esc") in mm/month.
 - Evapotranspiration (abbreviated as "evapo") in mm/month.
 - Precipitation (abbreviated as "prec") in mm/month.
 - Surface temperature (abbreviated as "temp") in °C.

Upon the successful completion and development of the comprehensive datasets, the next critical phase of the project will be initiated. This stage is pivotal, as it will leverage advanced AI models to project and forecast the anticipated values for excess surface water within the Sado basin. These projections will be formulated for an extensive period, ranging from 2015 through to 2100, thereby encompassing both historical data trends and future climate scenarios.

The AI models employed in this phase will be meticulously trained on the datasets compiled in the preceding steps. These datasets will include a range of variables, such as precipitation patterns, temperature variations, seasonal shifts, and runoff data, all of which will be analyzed in conjunction with historical trends. The models will utilize ML techniques to identify patterns and correlations, which will then inform the future projections of surface water surpluses. By incorporating both past climate conditions and projected future scenarios, these models will create forecasts that extend into the next century.

5.3 MARAISER - *Module 1 (Short-term) MAR performance optimization tool*

The application of the **MAR-SP Risk Assessment methodology**, as well as the generation of the flowcharts in each demo site has been considered as firsts or preliminary results from MARAISER Tool development.

It is important to consider that the flowcharts are a dynamic file with the aim of optimizing the setpoints established during the implementation and validation phase (WP4 - T4.1), with special focus on the update of alert thresholds for contaminants that could suffer substantial concentration reduction in its way through the not saturated zone (NSZ) and during its travelling in the aquifer (such as microbiology, Contaminants of Emerging Concern (CECs) and nutrients, among others).

In this section is described:

- 1) The **specific objectives** in each demo.
- 2) The **data collection** from three main data sources:
 - a) Documents with quality requirements for MAR with AWR (legislation, permit requirements, among others)
 - b) Historical data series (laboratory analysis and sensors) from control and operation of the MAR system, the Waste Water Treatment Plant (WWTP), the Tail Drainage Water (TDW) system and groundwater.
 - c) Reference values from scientific articles and from the bibliography.
- 3) Application of **MAR-SP Risk Assessment** methodology, adapted depending on the specific demo site.
- 4) Selection of the **set of Risk Indicators**: health, environmental and operational.

Also, in this section is presented a screenshot (as an example) of the **flowchart generated** per demo site (full version in Annex 1).

5.3.1 DS1 - Llobregat (Spain)

Llobregat demo site introduction and specific objectives

In the Llobregat river basin there are two MAR systems implemented in both Llobregat riverbanks. On the right riverbank there is implemented the **Infiltration MAR system of Sant Vicenç dels Horts (SVH)** which is an ASTR-MAR system consisting of a decantation pond (5.500 m²) plus an infiltration pond (4.900 m²). The water recharge is reclaimed

water that comes from the El Prat RWTP. On the left riverbank there is implemented the **Deep injection MAR system**, which is an ASR-MAR system consisting of a deep injection well (P18) that infiltrates pre-treated surface water from the Llobregat river. This pre-treated surface water is a mix of natural surface water and reclaimed water that is poured upstream in the river (Figure 5.1.3.a). The annual infiltrated volume in both MAR systems is about 1hm³.

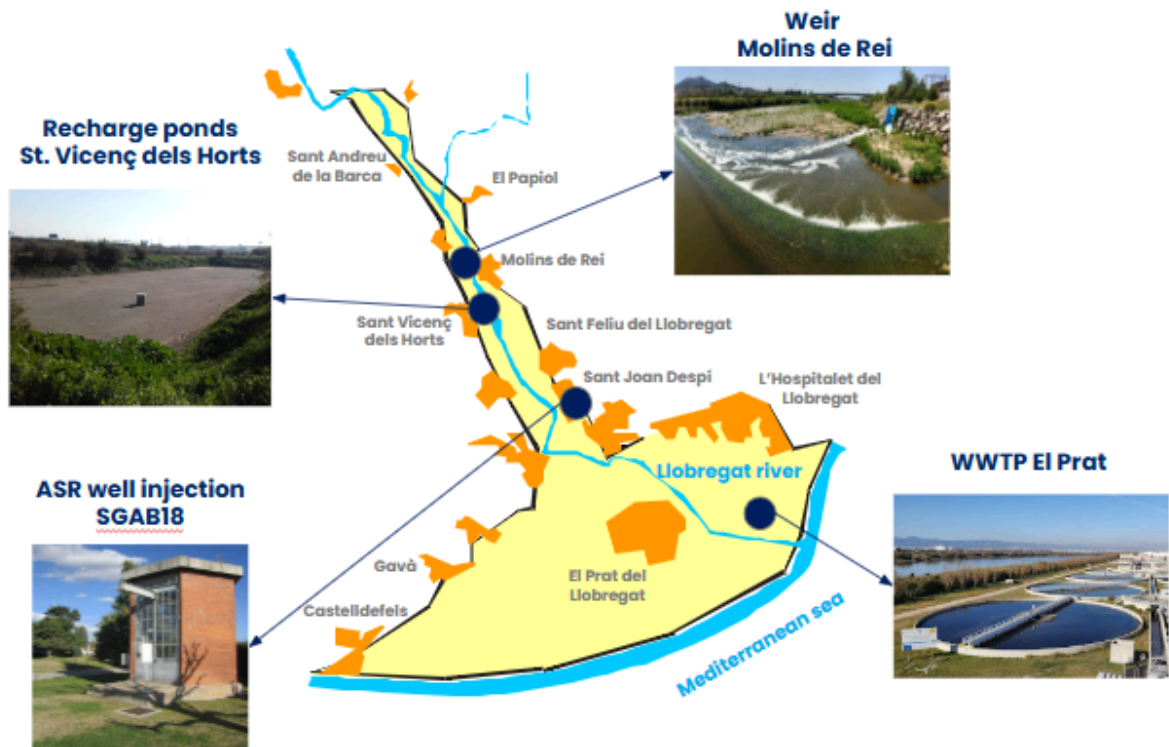


Figure 5.3.1.a. Llobregat aquifer MAR systems location.

The general objective in this demo site in the MARCLAIMED Project framework is to increase the availability of groundwater in the Llobregat aquifer, converting these two MAR operations into a structural supply system that could maximise its operability independent of the availability of surface water resources coming from Llobregat river. At the same time, **the specific objectives regarding MARAISER Tool implementation in both MAR systems** are:

- Comply with permit and legislation.
- Maximize the water volume recharged.
- Ensure the quality and hydrodynamic control to minimize the risk inherent to the reuse of reclaimed water through MAR.

In Figure 5.1.3.b and c is shown the location of both MAR systems, the control points and the closest drinking water wells.

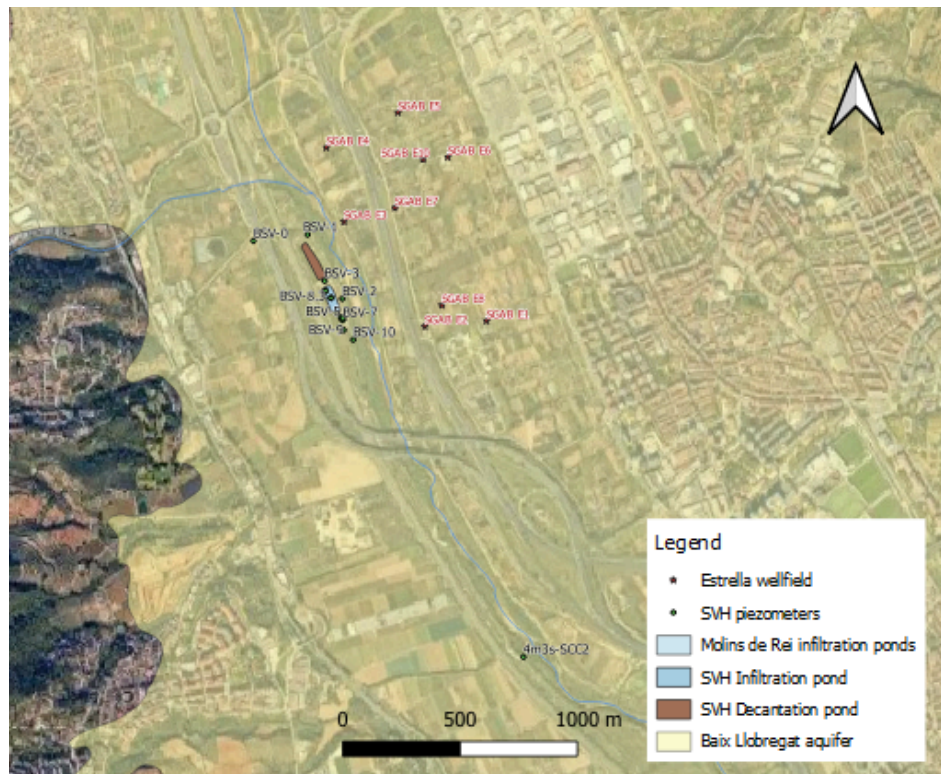


Figure 5.3.1.b. Infiltration MAR system location (SVH); natural groundwater flux direction NW-SE.

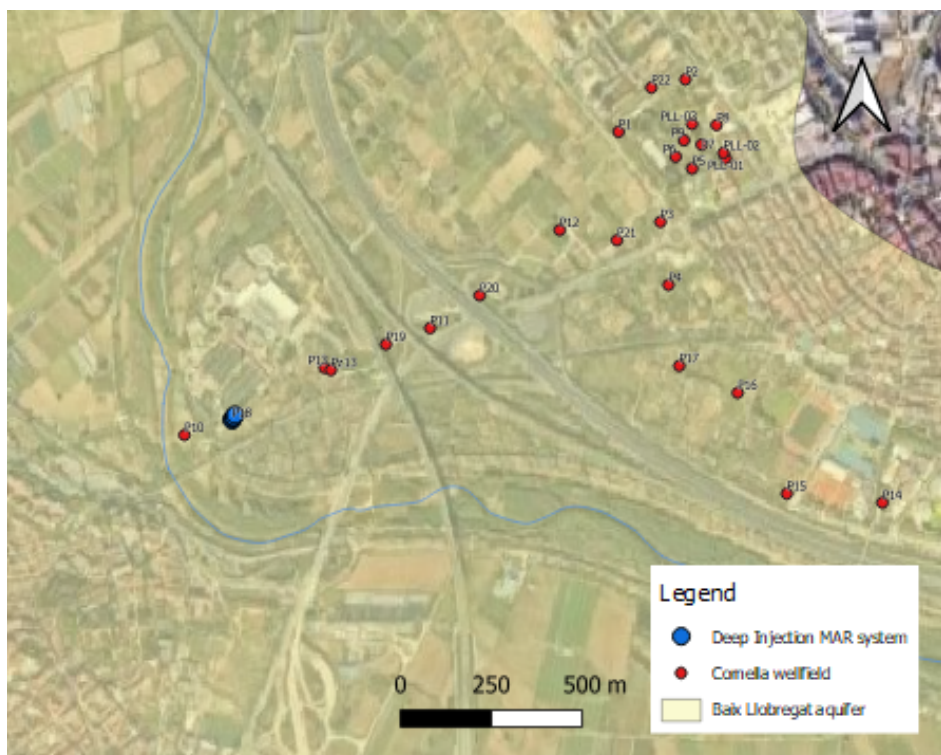


Figure 5.3.1.c. Deep injection MAR system (P18) location; natural groundwater flux direction NW-SE.

Llobregat's data collection

Infiltration MAR system (SVH)

The water quality thresholds and standards considered have been gathered from the following regulation:

- Reuse of reclaimed water for MAR through infiltration Permit. Granted by the Agència Catalana de l'Aigua in 2024. Which is based mainly in RD 1620/2007⁷ "Establishing the legal criteria for the reuse of treated water" (repealed by RD 1085/2024⁸, but not considered in the permit) and in the RD 3/2023⁹ "Technical and health criteria for the quality of drinking water".
- RD 1085/2024 "new legal criteria for the reuse of treated water", in which it is described as mandatory for the WWT Operator to design a Risk Management Plan for the specific use of the reclaimed water, in this case for MAR with reclaimed water.
- The Llobregat river basin is part of the Catalanian Internal Basins, which means that in addition to the EU and Spanish regulation, it is needed to consider the water quality thresholds and requirements established in the "Plan de gestión del distrito de cuenca fluvial de Cataluña (PGDCFC)" (Catalan River Basin Risk Management Plan) 2022-2027, Annex X (Threshold values of groundwater bodies), Tables 4 and 5.
- RD 817/2015¹⁰ "Criteria for monitoring and evaluating the status of surface waters and environmental quality standards".
- RD 1514/2009¹¹ "Protection of groundwater against pollution and deterioration" transposition of the Directive 2006/118/EC.

Other documents and data gathered have been:

- WHO drinking-water quality guidelines [21]
- Historical data of Llobregat aquifer' groundwater quality.

Deep injection MAR System (P18)

In this specific case, the focus has been to i) comply with the hydraulic requirements established in the permit granted by Agència Catalana de l'Aigua, and ii) automatize the setpoints related with the cleaning maintenance to avoid the loss of yield due to clogging.

The data gathered has been:

- Historical concentration of metals during the cleaning operations (Aluminium, Iron, Manganese, Total Organic Carbon and Turbidity).

⁷ <https://www.boe.es/buscar/doc.php?id=BOE-A-2007-21092>

⁸ <https://www.boe.es/buscar/act.php?id=BOE-A-2024-21701>

⁹ <https://www.boe.es/buscar/act.php?id=BOE-A-2023-628>

¹⁰ <https://www.boe.es/buscar/doc.php?id=BOE-A-2015-9806>

¹¹ <https://www.boe.es/buscar/act.php?id=BOE-A-2009-16772>

Llobregat's MAR-SP Risk assessment

Infiltration MAR system (SVH)

In Figure 5.3.1.d and e. it is presented the conceptual scheme with the monitoring points, the permit and legal requirements as well as the qualitative and water quality monitoring plan.

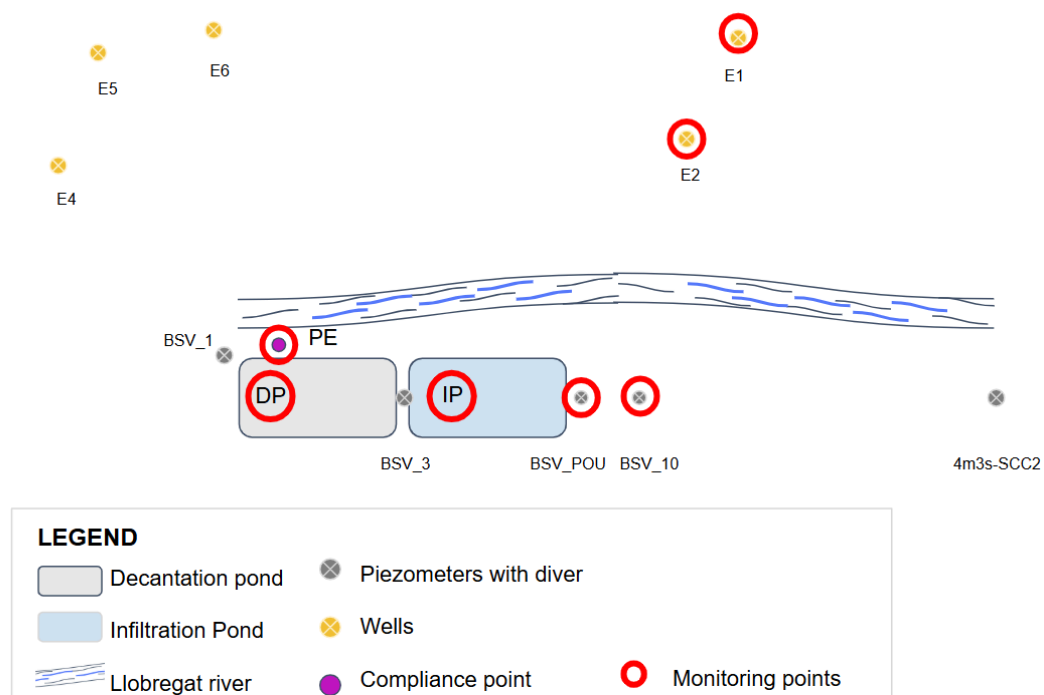


Figure 5.3.1.d. Conceptual scheme of the Infiltration MAR system (SVH).

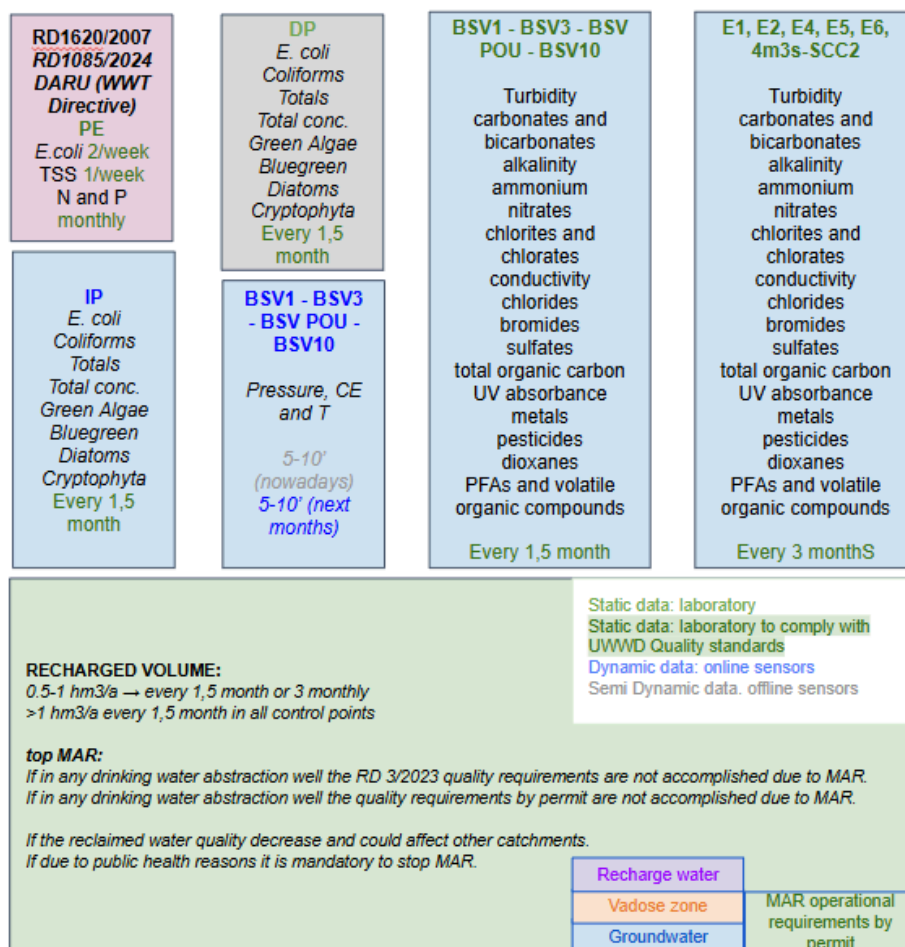


Figure 5.3.1.e. Monitoring plan and permit requirements of the Infiltration MAR system (SVH).

The premises considered for the MAR-SP Risk Assessment have been:

- Focus the assessment on wells which its use is for human consumption.
- Consider the existence of post-treatment, like in the health risk assessment methodology [34].
- To avoid groundwater deterioration (compared to aquifer pre-MAR operation average quality values), it has been considered that if the average groundwater quality values are below $\frac{2}{3}$ of the guidance value (GV), it is accepted a deterioration of $\frac{1}{3}$ of GV.
- It has been considered the Threshold Values (TV) of the GWB39 RBMP¹²: “quality standards used to assess the chemical status of water bodies”.

¹²

https://info.aca.gencat.cat/ca/aca/informacio/geco/plans-programes/PDG/CA/01-10_Annex10_Valors_Llindars_MASubt.pdf

Table 4. Threshold Values established for water bodies taking into account the supply use criterion.

CODI MAS	NOM MAS	Amoni (mg/L)	Arsenic (µg/L)	Bor (µg/L)	Cadmi (µg/L)	Clorurs (mg/L)	Conductivitat (µS/cm)	Crom (µg/L)	Fosfats (mg/L)	Nitrats (mg/L)	Nitrits (mg/L)	Plom (µg/L)	Sulfats (mg/L)	Plaguicides Compost (µg/L)	Plaguicides Suma (µg/L)	Suma PCE-TCE (µg/L)	Tetraclorur de Carboni (µg/L)	Benzè (µg/L)	PFASs - Total (µg/L)
MAS39	Vall baixa i delta del Llobregat	0,5	19	1000	4	440	2300	20	-	50	-	10	250	0,1	0,5	10	5	1	0,5

Table 5. Threshold Values established for water bodies taking into account the environmental criterion of salinization and other intrusions

CODI MAS	NOM MAS	Clorurs (mg/L)	Conductivitat (µS/cm)	Sulfats (mg/L)
MAS39	Vall baixa i delta del Llobregat	450	4100	240

- Also, has been considered the demonstrative test of reclaimed water reuse at the El Prat WWTP of 2019¹³. In which were identified two levels of pollutants: i) regulated compounds by the RD 3/2023 and non regulated compounds with high toxicity or persistence: Carbamazepine and lomeprol, and ii) non regulated compounds that are considered in the international health guidelines (WHO, EPA, among others): Hidroclorotiazide and Sulfametoxazole.

The MAR-SP Risk assessment has been performed for 3 stages: Tertiary Treatment (El Prat WWTP), Non Saturated Zone and Saturated Zone. The different phases of the risk assessment (i) identification of non compliance, ii) risk assessment and iii) decision three for RMP, Operational Requirement Point (ORP) or Pre-Requirement Point (PRP). This exercise has been performed during several working sessions with the DS1 Leaders (Aigues de Barcelona). The tables with the results are presented in Annex 1.

In this demo site MARAISER includes **outputs from the PREDMAR and MARVIS tools** (Table 5.3.1.a). From PREDMAR the daily flow forecasting in the catchment of Sant Joan Despí DWTP is considered to set up an alert regarding the preservation of minimum flow in that point of the Llobregat river, that implies that at least a flow of 3 m³/s has to be running. This means that the reclaimed water used for pouring in Molins de Rei Azud has to be prioritised over its use for MAR in SVH (this is important in the case that there won't be enough flow to cover both uses. In the case of MARVIS NO₃ measurements, these are considered in the RMP BSV-POU setpoint as the reference value to quantify the denitrification in the aquifer.

Here below it is presented the Risk Management table (Table 5.3.1.a) and the location of the Risk Monitoring Points (Figure 5.1.3.f):

¹³

https://aca.gencat.cat/web/.content/10_ACA/J_Publicacions/07-estudis-informes/35-prova-reutilitzacio-ERA-Prat-Llobregat.pdf

Table 5.3.1.a. Risk Management table (SVH)

ID	STAGE	RISK	RISK	CONTROL PARAMETER	UNITS	LEGAL THRESHOLD EXPERT CRITERIA	REGULATION EXPERT CRITERIA	ALERT THRESHOLD	CRITIC THRESHOLD
ST	Secondary outflow	All	Nutrients/TOTAL NITROGEN	Total Nitrogen	mg N total/L	10	RD1620/2007	>12	>14
Pump	Reclaimed water Pump	Operational	MAR operation compromised	m3/s in the river (link with PREDMAR)	m3/s	3	PGDCFC	2,5	3
PE	Delivery point	Operational	Not compliance with permit volumes	Volume poured	hm3/a	1	Permit	-	1
PE	Tertiary outflow	Operational	Turbidity/MES	Online Turbidity	NTU	>10 for 2 hours	Expert criteria	>5 for 2 hours	>10 for 2 hours
PE	Tertiary outflow	Health/Environmental	Microbiology/E. coli	Total chlorine online	mg Cl2/L	0.1< for 2h	Expert criteria	0.2< for 2h	0.1< for 2h
PE	Tertiary outflow	Health/Environmental	Chlorides	Chlorates	ug/l	700	Expert criteria	500	700
PE	Tertiary outflow	Health/Environmental	Salinity/Chlorides	Online conductivity ERA input	µs/cm	>4000 for one hour	Expert criteria	>3000 for one hour	>4000 for one hour
DP	Decantation pond	All	Eutrophication	Total concentration (Green algae, bluegreen, diatoms, cryptophyta)	µg/L	25	MITECO (Surface Water Indicators Guideline)	-	25 (Ind Calidad AASUP)
DP	Decantation pond	All	Microbiology	E. coli	MPN/100mL	1000	RD1620/2007 RD1085/2024	-	1000
IP	Infiltration pond	All	Eutrophication	entre DP e IP	µg/l	75 (max during field sampling)	MITECO (Surface Water Indicators Guideline)	-	25 (Ind Calidad AASUP)
IP	Infiltration pond	All	Microbiology	E. coli	MPN/100mL	1000	RD1620/2007 RD1085/2024	observació: 280 (media últimos muestreos Colilert)	1000
IP	Infiltration pond	Operational	Clogging	Discharge coefficient (deposit) from the Pond WL and Flow	-	TBD	Expert criteria	TBD	TBD
PE - IP	Infiltration pond	Operational	MAR efficiency***	Volume authorized (1hm3/yr) vs real volume recharged	%	TBD	Expert criteria	TBD	TBD
BSV_POU BSV_10	Non saturated zone	Health/Environmental	Nutrients/Nitrates	Nitrates ERA output (MARVIS)	mg NO3/l	40 (9,03NO3-N)	Expert criteria	25 (5,64NO3-N)	40 (9,03NO3-N)
		Operational		Nitrate reduction between tertiary output and unsaturated zone	%	<10% removal	Expert criteria	<15% removal	<10% removal
BSV_POU BSV_10	Saturated zone	Health/Environmental	CECs/Pesticides	Imidacloprid reduction percentage	%	<35% removal	Expert criteria	<40% removal	<35% removal

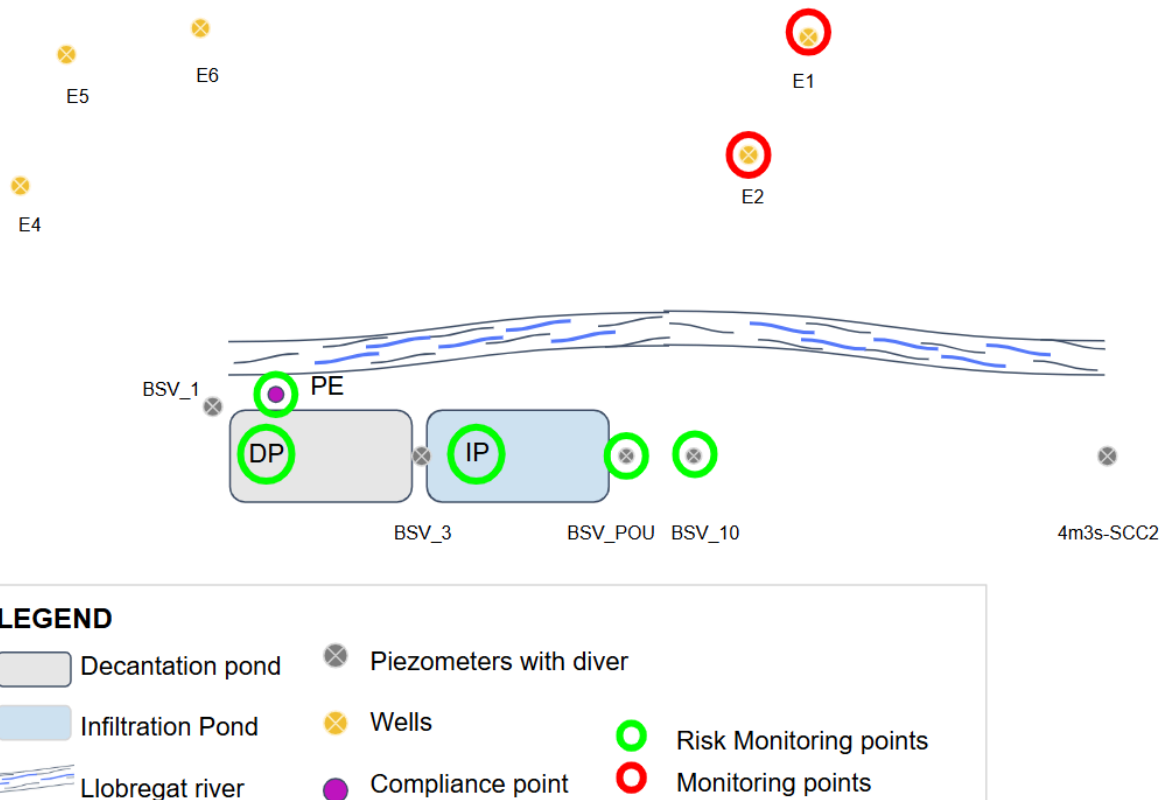


Figure 5.1.3.f. Location of the RMP (SVH)

Deep injection MAR System (P18)

In Figure 5.3.1.g the conceptual scheme of the deep injection MAR system (P18) with the monitoring points, the permit and legal requirements are presented, as well as the qualitative and water quality monitoring plan.

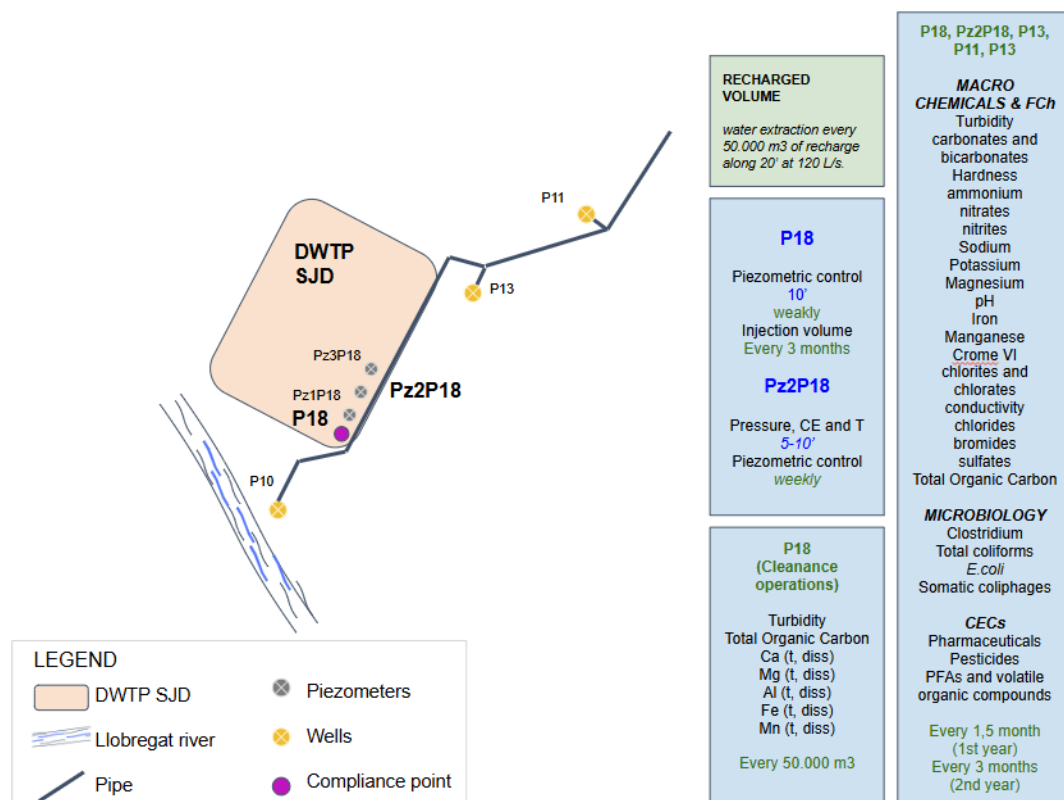


Figure 5.3.1.g. Conceptual scheme, monitoring plan and permit requirements of the Deep injection MAR system (P18).

The premises adopted for the MAR-SP Risk Assessment have been:

- Focus the assessment on real operational historical issues in P18: clogging. The alert and critical thresholds have been established by expert criteria analysing the historical data from the last two years (from when the indirect potable reuse of reclaimed water by pouring a certain volume upstream of the Llobregat river started). The control parameter defined is the turbidity.
- Comply with injected and extracted volumes dictated by the permit.

Here below (Table 5.3.1.b) it is presented the Risk Management table for the deep injection MAR system (P18)

Table 5.3.1.b. Risk Management table (P18)

ID	STAGE	RISK	HAZARD	CONTROL PARAMETER	UNITS	LEGAL THRESHOLD EXPERT CRITERIA	REGULATION EXPERT CRITERIA	ALERT THRESHOLD	CRITIC THRESHOLD
P18	Injection control	Operational ▼	Clogging - cleaning process required	Injection volume	m3	50000	Permit		50000
P18	Injection control	Operational ▼	Annual volume recharged	Annual injection volume	m3/a	950000	Permit	855000	950000
P18	Cleaning	Operational ▼	clogging	Turbidity	NTU	400	Expert criteria	245	400
P18	Cleaning	Operational ▼	clogging	Turbidity	NTU	70% turbidity reduction	Expert criteria	80% turbidity reduction	70% turbidity reduction
P18	Injection control	Operational ▼	MAR efficiency	Volumen disponible vs volumen recargado	h3/a	1	Permit		1
P18	Injection control	All ▼	Clogging and groundwater pollution	Mn	microg /L	TBD	Expert criteria	TBD	TBD

Llobregat's set of Risk indicators

The Health (**H**), Environmental (**E**) and Operational (**O**) Risk Indicators identified have been:

Infiltration MAR system (SVH)

PE

- Microbiology (*E.coli*) [**H**, **E**]
- Volume poured [**O**]
- MES (Turbidity)[**O**]
- Salinity (Chlorides) [**H**, **E**]

- Recharged efficiency (volume authorized vs volume recharged) [**O**]

BSV_POU

DP

- Microbiology (*E.coli*) [**H**, **E**]
- Eutrophication (algae) [**H**, **E**, **O**]

- Nutrients (NO₃) [**H**, **E**, **O**]
- CECs (pesticides) [**H**, **E**]

BSV_10

IP

- Microbiology (*E.coli*) [**H**, **E**]
- Eutrophication (algae) [**H**, **E**, **O**]
- Clogging (discharge coef.) [**O**]

- Nutrients (NO₃) [**H**, **E**, **O**]
- CECs (pesticides) [**H**, **E**]

Deep injection MAR System (P18)

P18

- Daily volume recharged [O]
- Annual volume recharged [O]
- Clogging (turbidity) [O]
- Pre-treatment Permanganate dosage [H, E, O]

Llobregat's Flowchart

Infiltration MAR system (SVH)

The Risk Management table has been adapted and converted to a MARAISER flowchart using the WebApp Draw.io as a base document for the integration in the MARCLAIMED IDST establish the rules and logic behind the different actions to be taken in case a hazardous event occurs.

In Figure 5.3.1.f. is presented a screenshot of the Llobregat's SVH MAR system MARAISER Flowchart, representing the different RMPs. The full version of the flowchart is presented in the Annex 1.

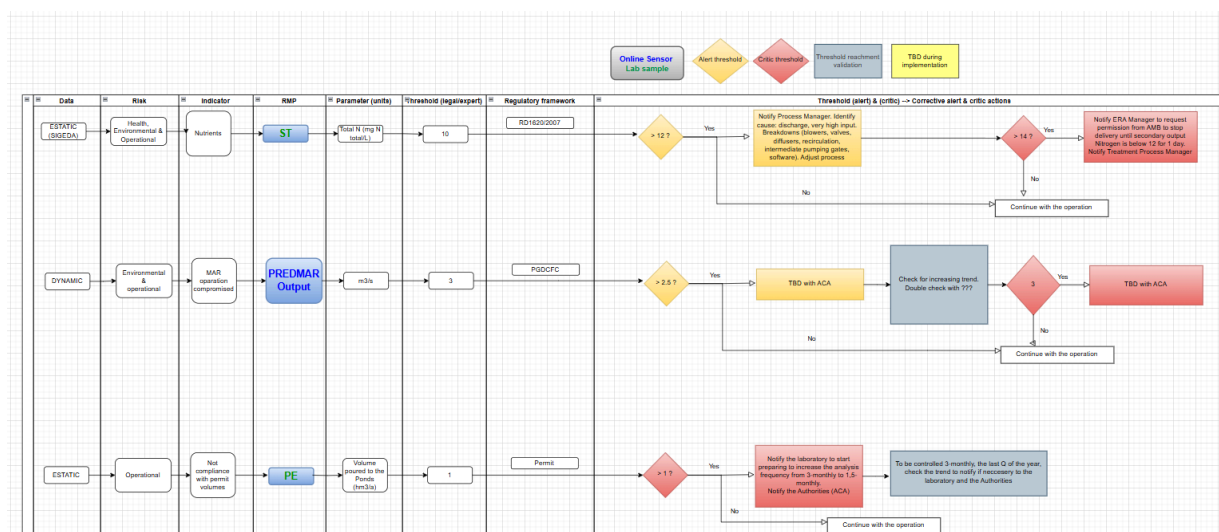


Figure 5.3.1.f. Llobregat's SVH MAR system MARAISER Flowchart screenshots .

Deep injection MAR System (P18)

The Risk Management table has been adapted and converted to a MARAISER flowchart using the WebApp Draw.io as a base document for the integration in the MARCLAIMED IDST establish the rules and logic behind the different actions to be taken in case a hazardous event occurs.

In Figure 5.3.1.g. is presented a screenshot of the Llobregat's P18 MAR system MARAISER Flowchart, representing the RMP P18. The full version of the flowchart is presented in the Annex 1.

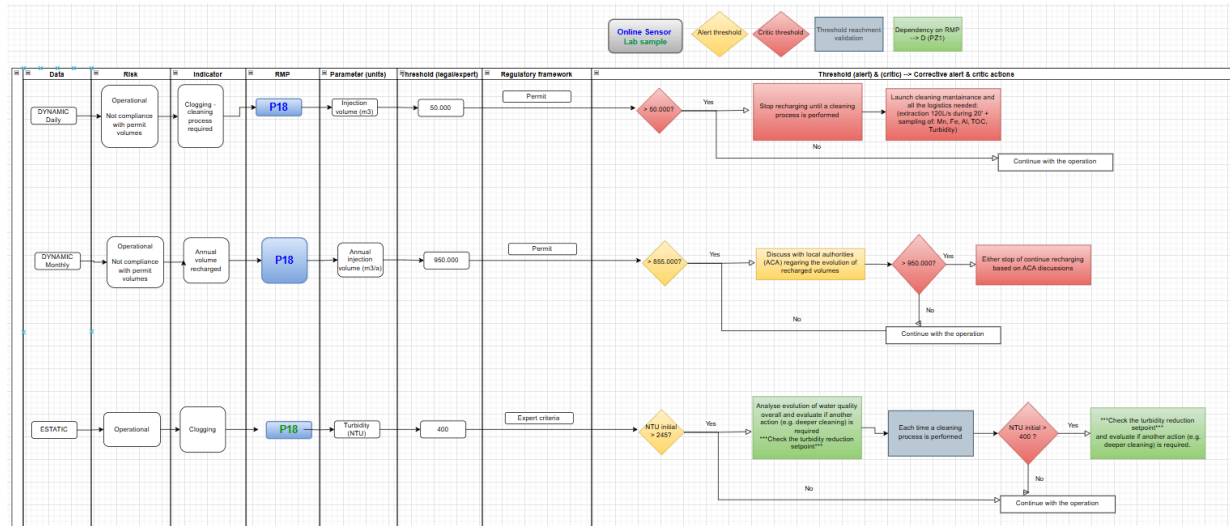


Figure 5.3.1.g. Llobregat's P18 MAR system MARAISER Flowchart screenshots .

5.3.2 DS2 - Comporta (Portugal)

Comporta demo site introduction and specific objectives

Comporta WWTP, located in the Alentejo region, is the 1st example of a SAT-MAR system running in Portugal since October 1st, 2021. The WWTP is equipped with a technologically advanced treatment system, including primary, biological treatment and additional disinfection to produce water for reuse and treatment of sludge by dehydration. The MAR system is composed of four infiltration basins, and it is presented as an alternative to the commonly implemented method of direct discharge of treated effluents into surface water bodies. The region in which it operates is Sado estuarine ecosystem, a sensitive area classified as a **Nature Reserve, and is over the Left-Margin Tejo-Sado aquifer** (the largest aquifer of the Iberian Peninsula with 6,875.44 km²) (Figure 5.3.2.a and b).

The high risk of water scarcity in the Alentejo region has brought forward the need of using AWR, namely secondary treated wastewater, to increase the groundwater reserves. However, the replication of SAT-MAR systems requires further research and scientific knowledge to assess the risk of groundwater contamination. Within the MARCLAIMED project, we aim to analyse the performance of the Comporta SAT-MAR system, not only considering the existing solution but also testing other solutions, namely with reactive organic layers (ROL). Besides, a complementary monitoring and sampling program was set to provide data that will serve as a basis for replication in areas with similar water availability and site intrinsic physical characteristics (e.g., lithology, permeability, soil, groundwater table).

The main objectives in this demo site in the MARCLAIMED Project framework are to 1) improve the recharged water quality, optimise infiltration rates, and increase local piezometric levels; 2) improve the understanding of groundwater pollution risk deriving from SAT-MAR; 3) improve the awareness and practices among water utilities; 4) improve the social and policy acceptance, reliability and trust of reclaimed to promote SAT-MAR with AWR as a structural solution for small WWTP in the whole region.

The **specific objectives regarding MARAISER Tool implementation in Comporta's MAR systems** are:

- Comply with permit and legislation.
- Maximize the water volume recharged to avoid pouring the secondary treated water in surface flows.
- Ensure the quality and hydrodynamic control from the Biological reactor, effluent treatment, infiltration ponds and aquifer.

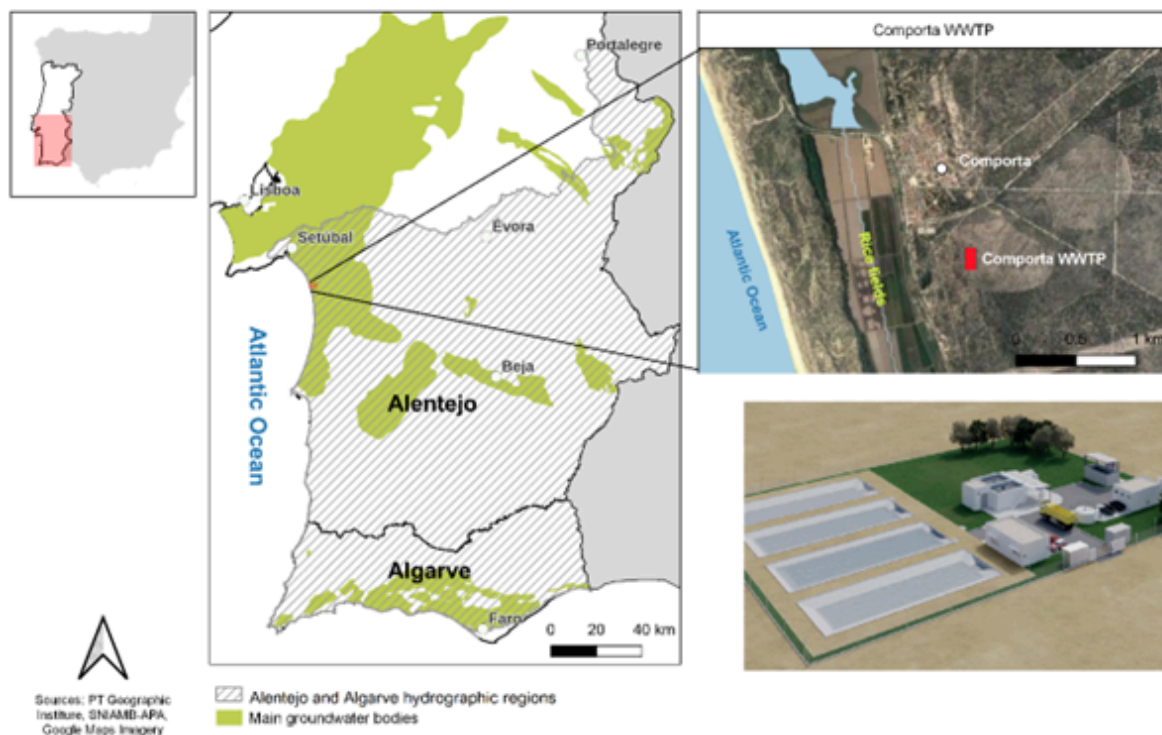




Figure 5.3.2.b. Comporta's zoom-in MAR system.

Comporta's data collection

The water quality thresholds and standards considered have been gathered from the following regulation:

- VLE p/Comporta: MAR with secondary treated water in Comporta's WWTP infiltration ponds Permit granted by the Portuguese Water Authority (Agência Portuguesa do Ambiente (APA)).
- Directive (UE) CD 91/271/EEC¹⁴ and Directive (EU) 2024/3019 [20] of the European Parliament and of the Council of 27 November 2024 concerning urban wastewater treatment (recast) 2024/3019 (WWT Directive (recast)). And also, the CD 98/15/EC (amending CD 91/271/EEC)¹⁵.
- Directive 2006/118/EC (Groundwater Directive) [21]
- DL 119/2019¹⁶ "production of water for reuse, obtained from wastewater treatment, as well as its utilization".

¹⁴ https://environment.ec.europa.eu/topics/water/urban-wastewater_en

¹⁵ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31998L0015>

¹⁶ <https://diariodarepublica.pt/dr/detalhe/decreto-lei/119-2019-124097549>

Other documents and data gathered have been:

- "Limiares" defined by the Agência Portuguesa do Ambiente.
- Water quality historical data from the inflow and outflow of Comporta WWTP.
- Groundwater quality historical data from the control piezometer (located downstream of the infiltration ponds in groundwater flow direction (S-N) (Figure 5.3.2.b.).

Comporta's MAR-SP Risk assessment

In Figure 5.3.2.c and .d it is presented the conceptual scheme with the monitoring points, the permit and legal requirements as well as the qualitative and water quality monitoring plan.

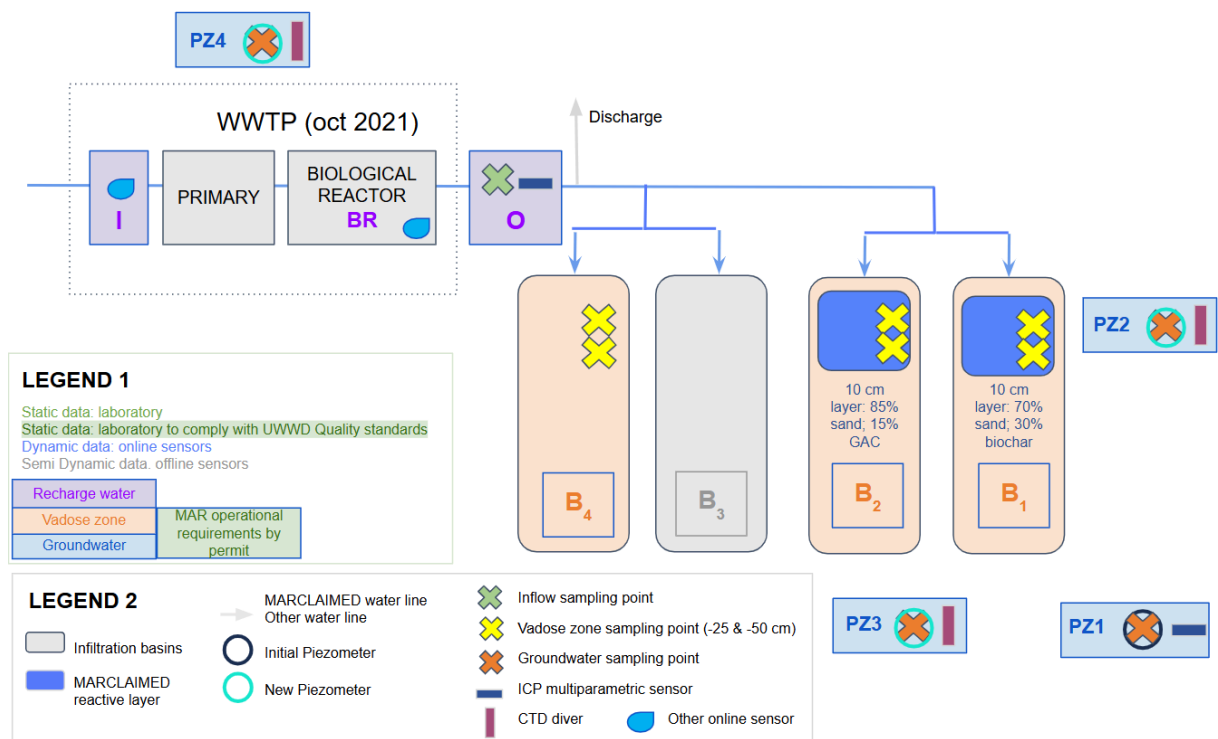


Figure 5.3.2.c. Conceptual scheme of the Infiltration MAR system.

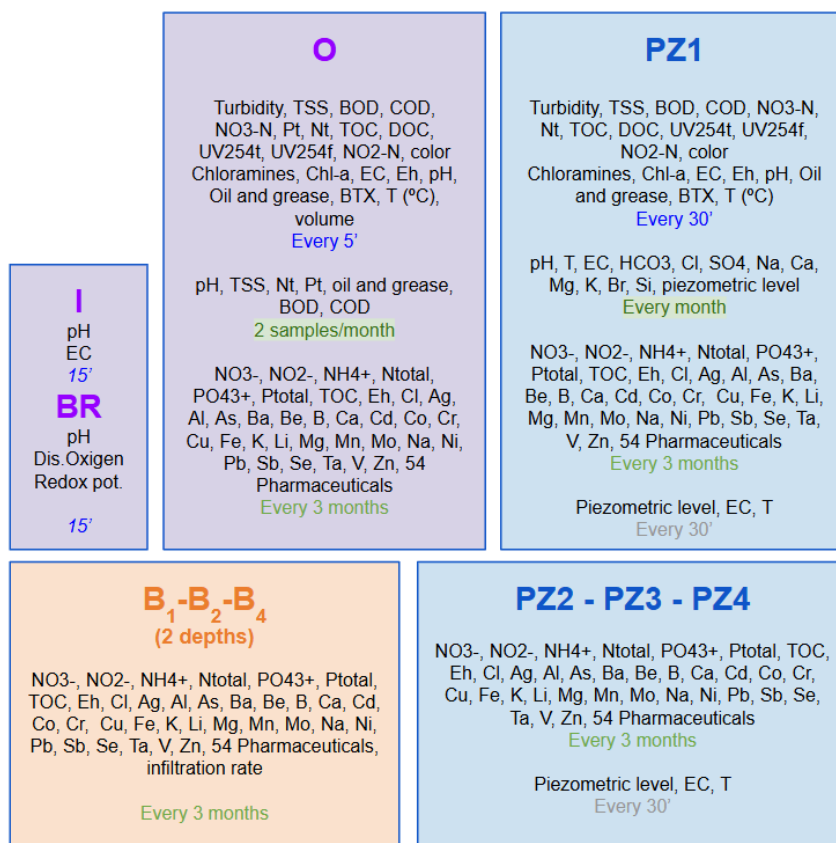


Figure 5.3.2.d. Monitoring plan and permit requirements of the Infiltration MAR system.

To perform the MAR-SP Risk Assessment in this demo site, the two approaches (described in Methodology section - 3.3-) have been applied. The semi-quantitative risk assessment matrix (likelihood and severity) for the Treated effluent and for the Piezometer (PZ1) stages, as the length of the historical data series for the parameters measured is bigger than 3 years, and expert criteria based for the rest of the stages. This exercise has been performed during several working sessions with the DS2 Leaders (Águas Publicas do Alentejo and LNEC as technical DS support). The tables with the results are presented in Annex 1.

The thresholds conditions have been defined as follows:

- The critical threshold matches the legal threshold or the value defined by expert criteria.
- The alert thresholds: i) if there are more than 3 years of data it has been applied the sum of the *Median + Variance* of the data series, unless this value is higher than the critical threshold. In this case the alert threshold is 90% of the critical threshold. And ii) if there are less than 3 years of data, the alert threshold equals 90% of the critical threshold.

Here below it is presented the Risk Management table Table 5.3.2.a and the location of the RMP (Figure 5.3.2.e):

Table 5.3.1.a. Risk Management table.

CCP	PROCESS STAGE	RISK	RISK INDICATOR	CONTROL PARAMETRE	DATA	UNITS	LEGAL THRESHOLD/EXPERT CRITERIA	REGULATORY FRAMEWORK	ALERT THRESHOLD	CRITIC THRESHOLD
B	Biological reactor BR	Environmental Operational	Yield	Diss.Ox.	Online sensor	mg/l	0.5 /2	Expert criteria	0.5 /2	0.5 /2
		Environmental Operational	Yield	Redox Potential	Online sensor	V	-150 / 150	Expert criteria	-150 / 150	-150 / 150
C	Treated effluent O	Environmental Operational	GWT contamination Clogging	BOD	Both	mg/L	25	VLE p/Comporta	25	25
		Environmental Operational	GWT contamination Clogging	pH	Both	pH units	6-9	VLE p/Comporta	6-9	6-9
		Environmental Operational	GWT contamination Clogging	TSS	Both	mg/L	35	VLE p/Comporta	35	35
		Environmental Operational	GWT contamination Clogging	COD	Both	mg/L	125	EU 2024/3019	125	125
		Environmental Operational	GWT contamination Clogging	Oil and grease	Both	mg/L	15	VLE p/Comporta	15	15
		Environmental Operational	Nutrients	P total	Both	mg/l P	5	EU 2024/3019	5	5
		Environmental Operational	Nutrients	N total	Both	mg/l N	15	EU 2024/3019	15	15
		Environmental Operational	GWT contamination Clogging	TOC	Probe	mg/L	37	EU 2024/3019	37	37
		Environmental Operational	CECs	UV254t	Probe	Abs/m	26	Expert criteria	26	26
		Environmental Operational	Physical Clogging	Turbidity	Probe	NTU	5	EU 2024/3019	5	5
		Environmental Operational	Salinity	EC	Probe	µS/cm	1200	Expert criteria	1200	2500
		Environmental Operational	REDOX conditions	Eh	Probe	mV	0	Expert criteria	-150 / 150	-150 / 150
		Environmental Operational	Volume poured	Volume	Probe	m3	TBD	Expert criteria	TBD	TBD
IP	Infiltration ponds B1, B2, B4	Operational	Infiltration rate	TBD	TBD	TBD	TBD	Expert criteria	TBD	TBD
D	Piezometer PZ1	Environmental Operational	Nutrients	P total	Lab analysis	mg/L	0,13	GWT thresholds	0,12	0,13
		Environmental Operational	Clogging	TOC	Both	mg/L	10	Expert criteria	9	10
		Environmental Operational	REDOX conditions	Sulphates	Lab analysis	mg/l SO4	250	GWT thresholds	57	250
		Environmental Operational	Salinity	Chorides	Lab analysis	mg/L Cl	250	GWT thresholds	250	250
		Environmental Operational	Salinity	EC	Both	µS/cm	1200	Expert criteria	1200	2500
		Environmental Operational	REDOX conditions	pH	Both	pH units	5.5-9	GWT thresholds	5.5-9	5.5-9
		Environmental Operational	CECs	UV254t	Probe	Abs/m	19	Expert criteria	17	19
		Environmental Operational	Nutrients	N total	Both	mg/L	11,73	GWT thresholds 11,73N(50NO3+0,5NH4+0,5NO2)	10,6	11,73
		Environmental Operational	REDOX conditions	Eh	Both	mV	0	Expert criteria	0	0

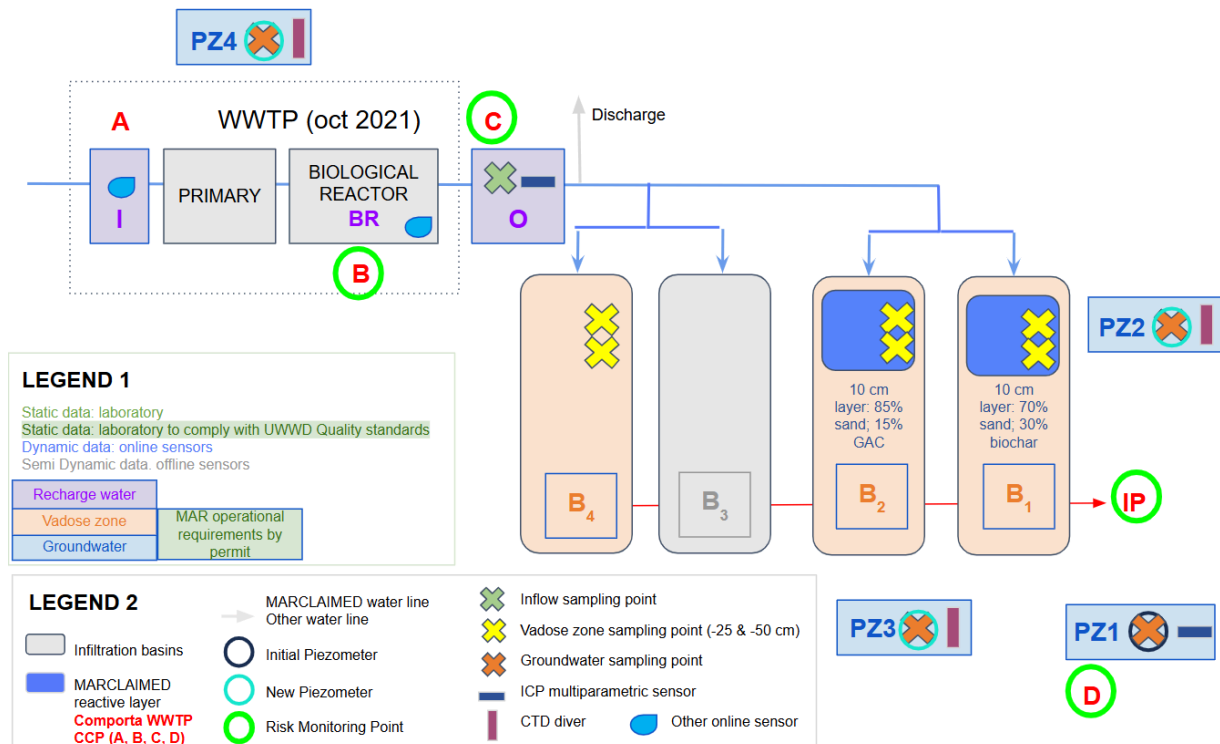


Figure 5.3.2.e. RMP location in Comporta's MAR system.

Comporta's set of Risk indicators

The Health (H), Environmental (E) and Operational (O) Risk Indicators identified have been:

B

- Biological Reactor Yield (*Diss. Oxygen, Redox potential*) [E, O]

C

- Volume poured [O]
- Groundwater contamination (Oil & grease) ([E, O])
- Ponds bioclogging (Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), pH, Total Organic Carbon (TOC)) [E, O]
- Ponds physical clogging (Oil & grease, Total Suspended Solids (TSS), turbidity, pH) [O]
- Ponds chemical clogging (REDOX, pH) [E, O]
- Nutrients (Nt, Pt) [E, O]
- CECs (UV254f, Pharmaceuticals) [E]
- Salinity (Electro Conductivity (EC)) [E]

IP

- Infiltration rate [O]

D

- REDOX conditions (SO₄, pH, Eh) [E, O]
- Nutrients (Pt) [E, O]
- Clogging (TOC) [E, O]
- Salinity (EC, Chlorides) [E]
- CECs (UV254f, Pharmaceuticals) [E]

Comporta's MARAISER Flowchart

The Risk Management table has been adapted and converted to a MARAISER flowchart using the WebApp Draw.io as a base document for the integration in the MARCLAIMED IDST establish the rules and logic behind the different actions to be taken in case a hazardous event occurs.

In Figure 5.3.2.f. are presented two screenshots of the Comporta's MARAISER Flowchart, representing the RMP C (Treated effluent) and D (Piezometer PZ1) setpoints. The full version of the flowchart is presented in the Annex 1.

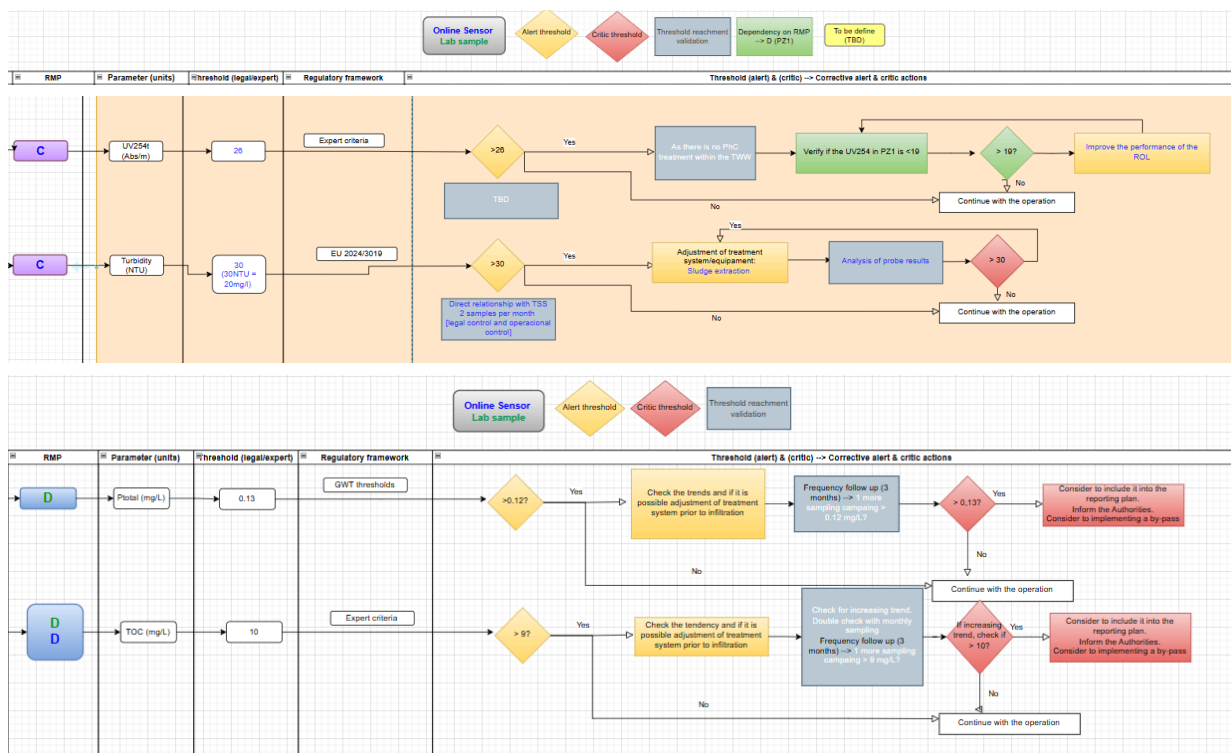


Figure 5.3.2.f. Comporta's MARAISER Flowchart screenshots (above: RMP C (Treated effluent) and below: RMP D (Piezometer PZ1) setpoints).

5.3.3 DS3 - Texel (The Netherlands)

Texel demo site introduction and specific objectives

The DS3 is located on Texel, one of the Dutch Wadden islands in the province of North Holland, the Netherlands (Figure 5.3.3.a). The island consists mainly of farmland surrounded by sand dunes and dykes along the coasts. The municipality of Texel encompasses the whole island, which is approximately 586 km². There is an ASR MAR system which collects water from a TDW network approximately 1 m below a 30-ha agricultural parcel. This drainage network normally discharges water to the surface water, but in this system all tile drains end up in one of three collection wells, which discharges into the MAR system, a Slow Sand Filtration (SSF) in combination with Granular Activated Carbon (GAC) schema running through two horizontal injection wells into a confined sandy aquifer (Hoofdweg plot in the north, which have other two plots: Broekman and Slot plot (Figure 5.3.3.b).

Securing freshwater availability in the area is of great importance for the economic position of the agricultural sector. Furthermore, groundwater is considered highly vulnerable to pollution, needing to be protected in this agricultural area.

Specific DS3 MARCLAIMED objectives are to: 1) preserve the aquifer water quality via nutrient and pesticide control in the TDW; 2) make TDW available for large-scale ASR-MAR scheme, improving its cost-effectiveness, optimising the maintenance, and achieving a small footprint; 3) prolong the lifetime of the infiltration wells by reducing the clogging potential of infiltration water; 4) increase the reliability and acceptance of MAR with TDW in the agricultural and water sector and thus contribute overall, to its feasibility and sustainability; 5) promote farmers' engagement to recognize the necessity of MAR with TDW and willingness to invest in these systems and to know how to operate them for the sake of potentially more stringent future requirements on the quality of TDW.

The **specific objective regarding MARAISER Tool implementation in Comporta's MAR systems** is to monitor the water quality to inform and influence Administration (law and regulation). Controlling:

- Nutrients and pesticides of the in- and outflow SSFW.
- To be recharged water quality (outflow).
- The volume of water that is infiltrated and recovered in the ASR system.
- The regional impact on the head in the aquifer and its fluctuations.
- Behavior of the freshwater body during the recovery period. This will be monitored through continuous EC-measurements in the observation wells using divers.

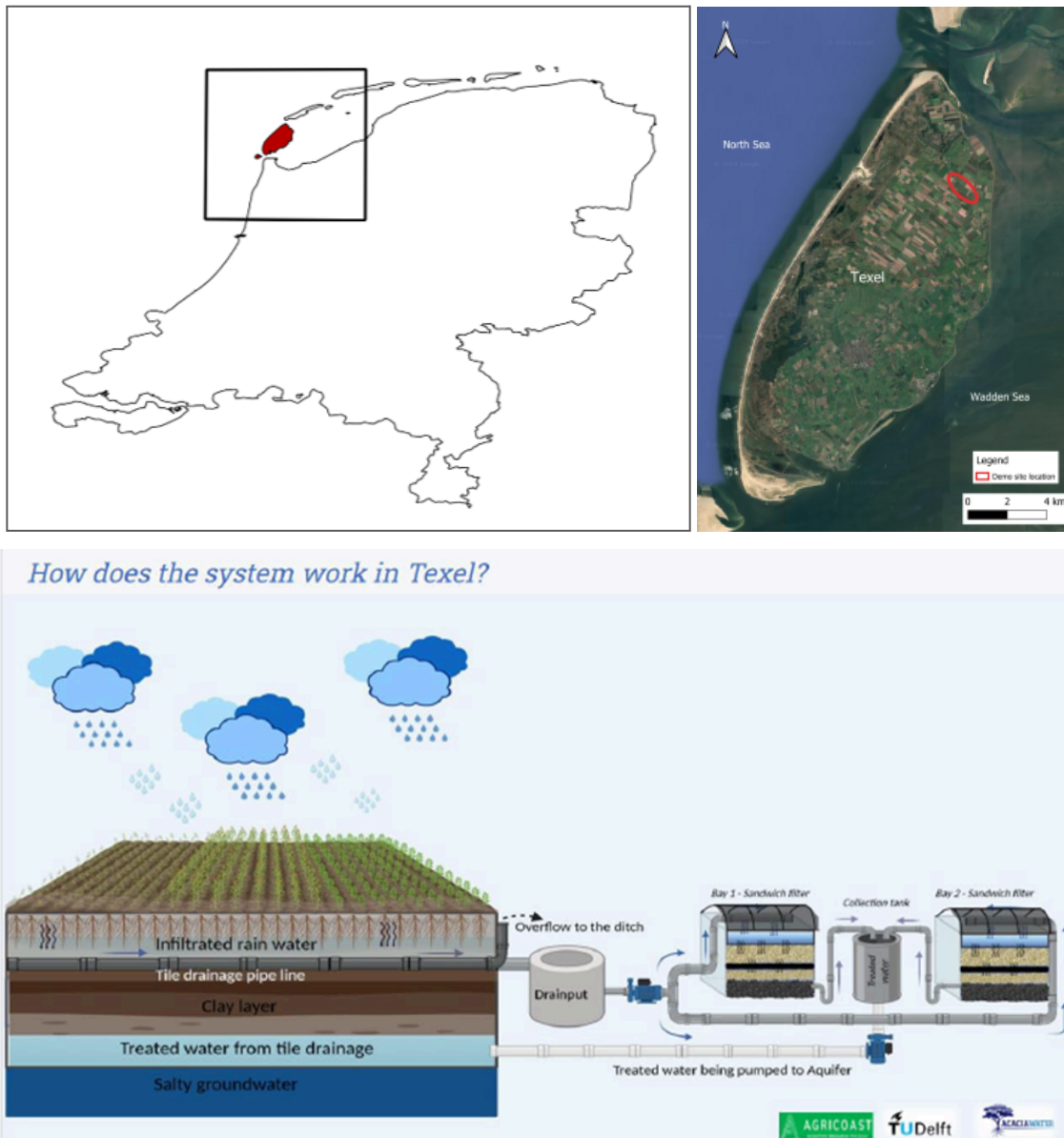


Figure 5.3.3.a. Texel demo site: regional insertion (red circle) and ASR-MAR scheme.

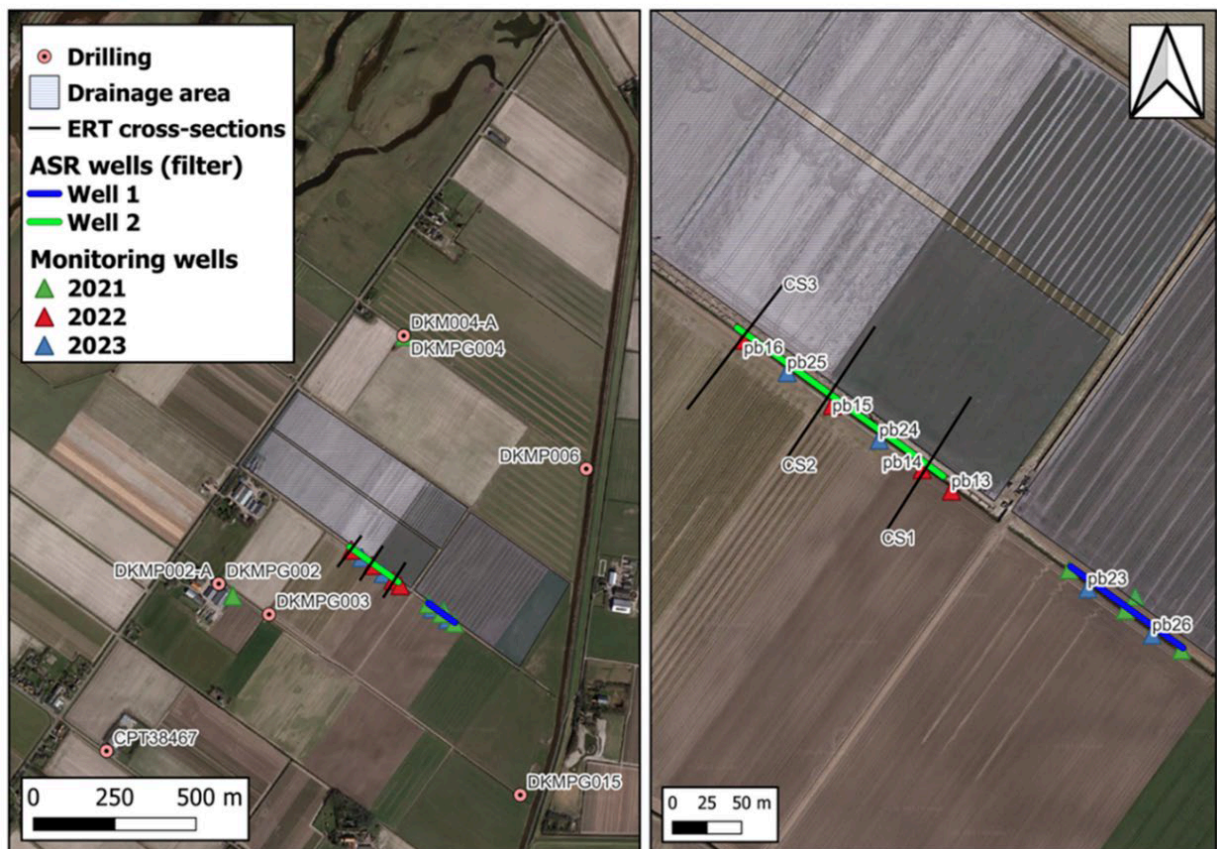


Figure 5.3.3.b. Hoofdweg MAR system (northeastern plot).

Texel's data collection

The water quality and hydrodynamic thresholds and standards considered have been gathered from the following regulation:

- Quality legal requirements: Huishoudelijk reglement (*House rules*) Freshwater Cooperative Hoofdweg/Postweg Zoete Toekomst Texel U.A. Concept versie 5 - 18 January 2024, Ledenovereenkomst Zoetwatercoöperatie (*Membership Agreement Freshwater Cooperative*) Hoofdweg/Postweg Zoete (Future Texel U.A. Version 3 - November 1, 2023) and Oprichting Cooperatie.
- MAR with Tail Drainage Water (TDW) in Hoofdweg horizontal well temporary 5-year permit granted by Hollands Noorderkwartier Hoogheemraadschap (*Water board of Northern Holland*), permit number HHNK/21004983. Following the conclusion of the pilot testing, a permanent permit will be issued.

Texel's MAR-SP Risk assessment

In Figure 5.3.3.c and .d it is presented the conceptual scheme with the monitoring points, the permit and legal requirements as well as the qualitative and water quality monitoring plan.

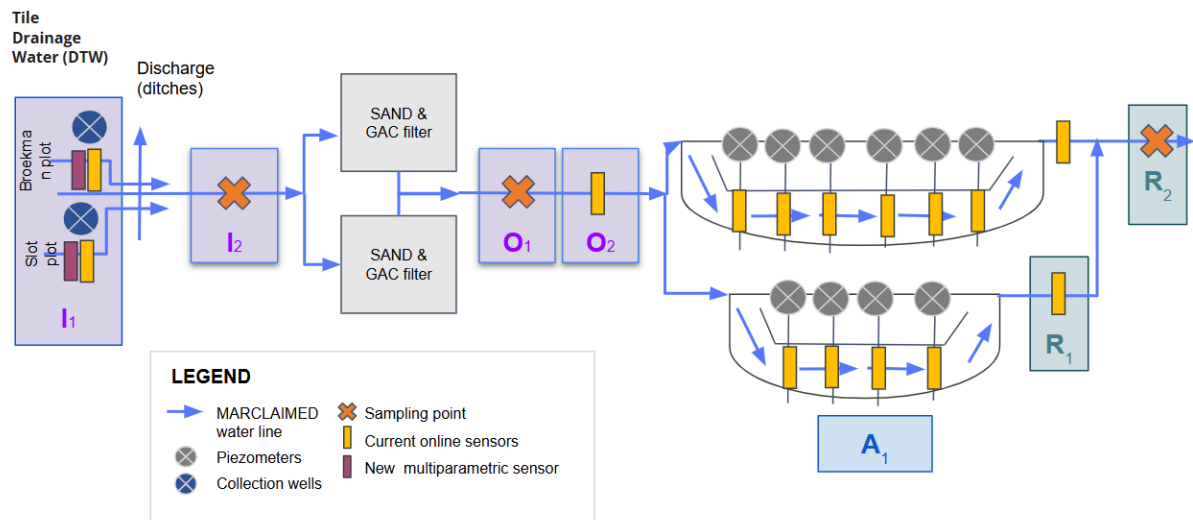


Figure 5.3.3.c. Conceptual scheme of the TDW-MAR system.

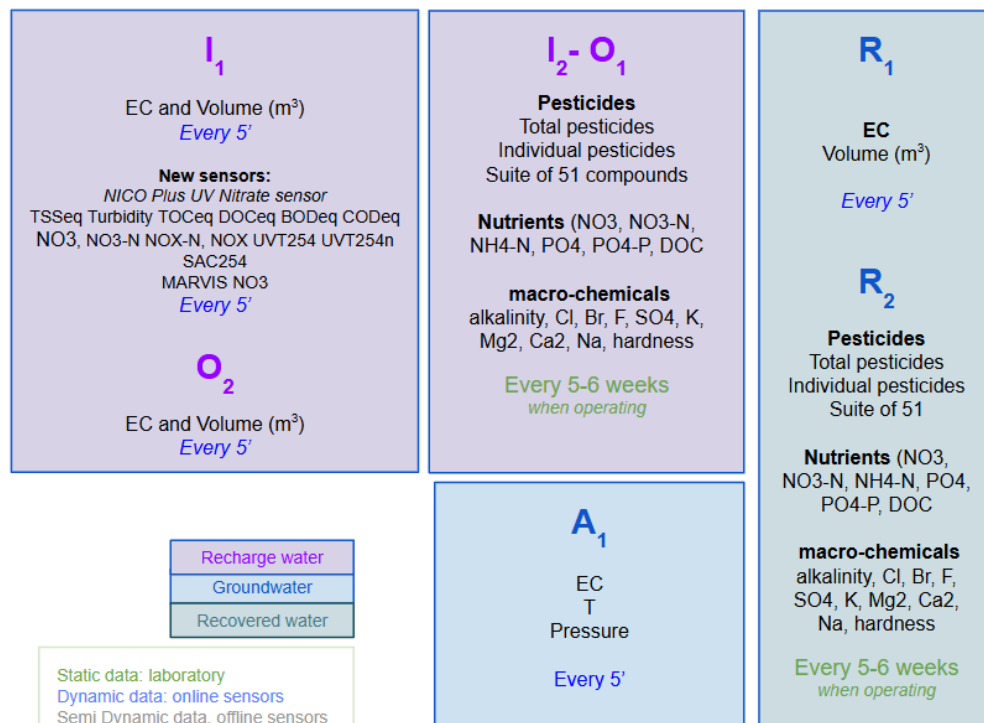


Figure 5.3.3.d. Monitoring plan and permit requirements of the TDW-MAR system.

Texel demosite started collecting data from december 2024, so the MAR-SP Risk assessment has to be applied considering the expert criteria both to determine the RMP and the alert and critical thresholds (based on permit/legal water quality requirements if any). This exercise has been performed during several working sessions with the DS3 Leaders (Acacia Water).

Here below it is presented the Risk Management table Table 5.3.3.a and the location of the RMP (Figure 5.3.3.e). In this demo site MARAISER includes **outputs from the MARVIS tool** (Table 5.3.3.a). In this case MARVIS NO₃ measurements are considered in the RMP I1 setpoint as the reference value to quantify the denitrification capacity in the SSF/GAC pre-treatment.

Table 5.3.3.a. Risk Management table.

CCP	PROCESS STAGE	RISK (H - Health risk; O - Operational)	RISK INDICATOR	CONTROL PARAMETRE	UNITS	THRESHOLD/THRESHOLD BASED IN	REGULATORY FRAMEWORK	ALERT THRESHOLD	CRITIC THRESHOLD
I1	Inlet/inflow before ditch discharge	Environmental Operational	Salinity	EC online sensor	mS/cm	1,5	Expert criteria		1,5
		Environmental Operational Health	Nutrients	MARVIS NO3-N	mg/L	5,6	Expert criteria	6,72	10,1
O1	Outflow	Environmental Operational Health	CECs	Total pesticides (51 compounds)	µg/L	0,5	EU Groundwater Directive		0,5
		Environmental Operational Health	CECs	Individual pesticide	µg/L	0,1	EU Groundwater Directive		0,1
		Environmental Operational Health	Nutrients	NO3-N	mg/L	5,6	Dutch Environmental Quality Decree		11,2
		Environmental Operational Health	Nutrients	PO4-P	mg/L	0,4	EU Groundwater Directive		0,4
		Operational	Macrochemicals	Fe (II)	mg/L	3	Expert criteria	2	3
		Operational	Macrochemicals	DOC	mg/L	2	EU Groundwater Directive	2	2
		Operational	Pressure Head	Piezometer	bar	0,3	Expert criteria	0,25	0,3
A1	Aquifer	Operational	Pressure Head	Piezometer	bar	0,3	Expert criteria	0,25	0,3
R1	Recovered water	Environmental Operational	Salinity	EC online sensor	mS/cm	Depending on crop	Expert criteria	1,5	General: 1,6 - 2,0
		Environmental Operational	Salinity	EC online sensor	mS/cm	Depending on crop	Expert criteria	1,5	Onions: 2,5
		Environmental Operational	Salinity	EC online sensor	mS/cm	Depending on crop	Expert criteria	1,5	Potato: 4,0
		Environmental Operational	Salinity	EC online sensor	mS/cm	Depending on crop	Expert criteria	1,5	General: 1,6 - 2,0
		Environmental Operational	Salinity	EC online sensor	mS/cm	Depending on crop	Expert criteria	1,5	Onions: 2,5
		Environmental Operational	Salinity	EC online sensor	mS/cm	Depending on crop	Expert criteria	1,5	Potato: 4,0
		Environmental Operational	Salinity	EC online sensor	mS/cm	Depending on crop	Expert criteria	1,5	Potato: 4,0

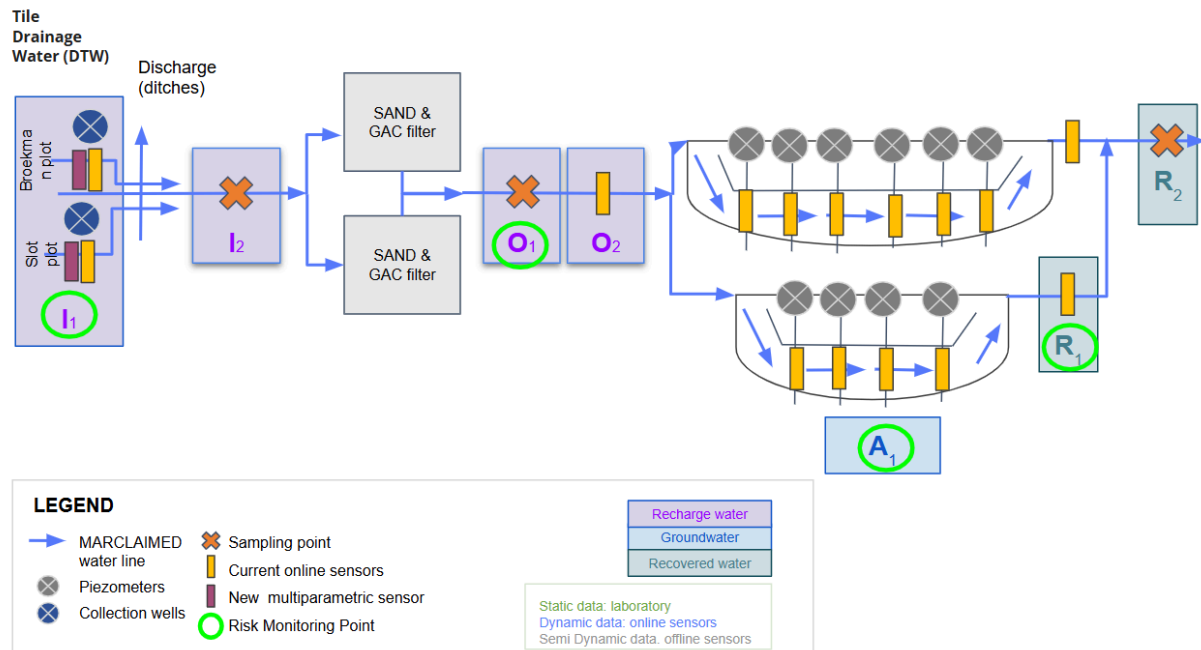


Figure 5.3.3.e. RMP location in Texel's TDW-MAR system.

Texel's set of Risk Indicators

The Health (**H**), Environmental (**E**) and Operational (**O**) Risk Indicators identified have been:

I_1

- Salinity (CE) [**E**, **O**]
- Nutrients (NO_3) [**H**, **E**, **O**]

O_1

- CECs (Pesticides) [**H**, **E**, **O**]
- Salinity (EC) [**E**]
- Nutrients (NO_3) [**H**, **E**, **O**]
- Clogging (NO_3 , $\text{PO}_4\text{-P}$, Fe (II), DOC) [**H**, **E**, **O**]

A_1

- Pressure head (pressure) [**O**]

R_1

- Salinity (CE) [**E**, **O**]

Texel's Flowchart

The Risk Management table has been adapted and converted to a MARAISER flowchart using the WebApp Draw.io as a base document for the integration in the MARCLAIMED IDST establish the rules and logic behind the different actions to be taken in case a hazardous event occurs.

In Figure 5.3.3.f. is presented a screenshot of the Texel's MARAISER Flowchart, representing the RMP I1 (SSF/GAC Inlet) and O1 (SSF/GAC outlet) setpoints. The full version of the flowchart is presented in the Annex 1.

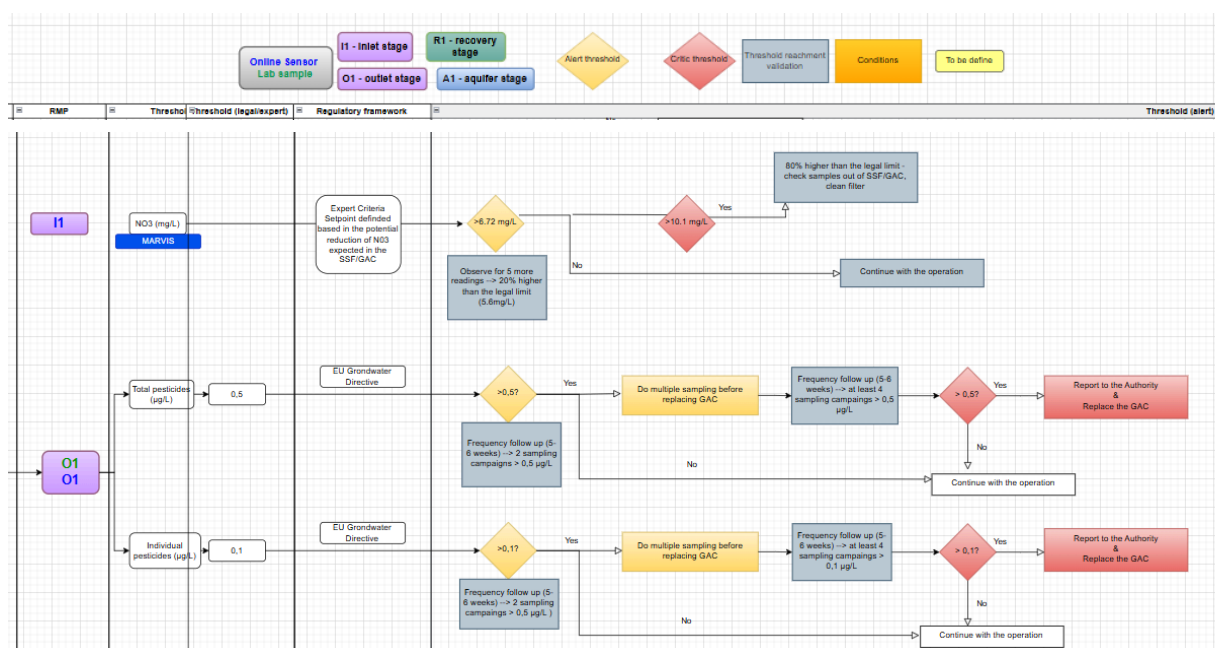


Figure 5.3.2.f. Comporta's MARAISER Flowchart screenshots (RMP I1 (SSF/GAC Inlet) and O1 (SSF/GAC outlet) setpoints).

6. Discussion, Conclusions and Next steps (follow up, implementation, and validation)

6.1 MARVIS - Virtual sensing tool for near real-time water quality monitoring

The MARVIS tool represents a significant advancement in the field of water quality monitoring through the application of virtual sensing technology. By focusing on nitrate concentration measurement, MARVIS addresses critical needs related to compliance with EU Directive 91/676/EEC¹⁷ and the mitigation of penalties due to nitrate pollution at MARCLAIMED demo sites. The innovative approach of using virtual sensors to correlate easily measurable parameters with nitrate concentrations offers a cost-effective and

¹⁷ <https://eur-lex.europa.eu/eli/dir/1991/676/oj/eng>

scalable solution compared to traditional nitrate sensors, which are either prone to interference or prohibitively expensive. MARVIS has demonstrated the potential of virtual sensing technology to revolutionize water quality monitoring, particularly in MAR operations. The use of XGBoost and plans to transition to deep learning models highlight the tool's commitment to leveraging advanced ML techniques for improved accuracy and adaptability. The adaptive calibration methodology further ensures the reliability of the virtual sensor in dynamic water quality conditions. To drive the aforementioned transition the followed criteria is based on obtaining sufficient data to develop a robust ML model. The amount of data considered enough for training a ML model, particularly a neural network, depends on several factors including the complexity of the model, the dimensionality of the data, and the variability within the dataset. Generally, more data is better, as it allows the model to learn more robust patterns and generalize better to unseen data. For simple models, thousands of data points might suffice, whereas complex models like deep neural networks may require tens of thousands to millions of data points to perform effectively. Not only the quantity but also data quality is crucial; it should be representative of the problem domain and contain sufficient variability to capture the underlying patterns. Cross-validation techniques such as the ones employed here can help assess whether the dataset size is adequate by evaluating the model's performance stability across different data splits. Our preliminary results indicate that the accuracy of XGBoost models on DS1 and DS3 not only matches state-of-the-art ML methodologies but also offers insights into feature importance in MARVIS predictions.

The next steps for MARVIS involve ongoing data collection to enhance the robustness of the ML models. In turn, this process will also reinforce the conclusions drawn from the explainability AI methods employed (SHAP), strengthening the novel insights provided. Data acquisition will facilitate the transition between models and their refinement, improving accuracy and generalizability. Ultimately, the goal is to implement, deploy, and validate the virtual sensor across various MAR contexts, ensuring its effectiveness in real-world applications. This will involve rigorous testing and validation to confirm the sensor's performance and reliability, paving the way for widespread adoption and integration into water management strategies. By enabling real-time monitoring and adaptive management of water quality, MARVIS supports the sustainable use of alternative water resources for aquifer recharge, thereby contributing to water scarcity mitigation and enhancing the resilience of water supply systems.

The subsequent phases essential for the realization and substantiation of MARVIS encompass:

- **Establishment of Data Source Connectivity with the MARCLAIMED IDST Server:** This entails the configuration and integration of diverse data sources, ensuring a seamless and continuous flow of data into the MARCLAIMED IDST server for subsequent processing and analysis.

- **Uninterrupted Data Acquisition (DS1 and DS3):** This involves the implementation of mechanisms and protocols to facilitate the ongoing collection of data from designated sources DS1 and DS3. This uninterrupted data stream is pivotal for maintaining an up-to-date and comprehensive dataset for the MARVIS system.
- **Model Transition from XGBoost to a Neural Network:** This signifies a strategic shift in the underlying model architecture from the current XGBoost model to a neural network-based model. This transition is aimed at potentially enhancing the model's predictive capabilities and performance by leveraging the inherent strengths of neural networks in handling complex patterns and relationships within data.
- **Transfer Learning to DS2:** This refers to the application of transfer learning techniques to extend the knowledge and insights gained from one dataset or domain (potentially DS1 or DS3) to a new dataset or domain (DS2). This approach can expedite the model development process for DS2 and potentially improve its performance by leveraging pre-existing knowledge.

6.2 PREDMAR - AI-based water availability and demand forecasting tool

The development and application of AI tools in the context of MAR and broader AWR management represent a significant advancement in the transition toward data-driven, adaptive, and resilient water systems. The integration of AI into MAR planning and operations facilitates both strategic long-term assessments and tactical short-term responses, contributing to improved groundwater sustainability and climate adaptation.

One of the primary implications of the AI tool lies in its ability to assimilate and process large volumes of heterogeneous data including climate projections (e.g., CMIP6 via CDS), sensor networks, hydrological inputs, and historical recharge events to support predictive decision-making. Through ML models trained on multivariate datasets, the tool enables forecasting of recharge potential, aquifer response, and system performance under varying climatic and operational conditions. This is especially relevant for identifying windows of opportunity for recharge, assessing recovery efficiency, and optimizing the coordination between surface and groundwater use.

Furthermore, the AI tool supports the integration of MAR into AWR management by providing scenario-based assessments that align with RCP-SSP climate trajectories. These scenarios allow water managers to explore the implications of different socio-economic and climate futures on aquifer recharge rates and demands, informing robust planning strategies. In particular, the tool's hybrid learning architecture (e.g., PIML) ensures that forecasts are physically plausible while capturing nonlinear interactions and uncertainties inherent to groundwater systems.

Another key implication involves the enhancement of operational resilience. By linking short-term meteorological forecasts (e.g., SEAS5) to recharge opportunities, the AI tool

enables near-real-time guidance on when and how much water to recharge, based on predicted inflows, storage capacity, and drought status as defined by regulatory instruments like the Special Drought Plan (PES). This anticipatory capacity strengthens the ability of MAR schemes to buffer against extremes such as prolonged droughts or intense rainfall events thereby stabilizing water availability.

Finally, the AI-driven insights can be embedded within digital dashboards or decision-support interfaces, offering stakeholders from utilities to basin agencies an accessible and dynamic tool for managing MAR within the broader water portfolio. As data quality and availability continue to improve (e.g., through IoT deployments and satellite-based monitoring), the performance of these tools will also advance, reinforcing their role as critical enablers of integrated and sustainable AWR strategies.

In conclusion, the application of AI tools to MAR and AWR management offers transformative potential. It enhances the capacity to forecast, optimize, and adapt groundwater recharge operations in a changing climate, while aligning with policy frameworks and contributing to long-term water security goals.

The PREDMAR tool is a flexible support system intended to assist water resource management through a set of preliminary functionalities that aim to enhance forecasting efforts. For operational purposes, provides short-term monitoring capabilities, helping to improve the management of water systems. In addition, the tool is being developed to offer functionalities for longer-term planning and infrastructure assessment. PREDMAR provides a practical tool designed to support various aspects of water resource management. Its data handling and reporting features aim to improve communication with stakeholders and assist in supporting data-informed decision-making. Additionally, its long-term planning functions are being developed to offer useful insights for infrastructure assessment, including the potential availability of reclaimed water.

The general next steps for the implementation and validation will be:

- Continuous feedback from stakeholders and potential end users through the CoPs workshops (WP3).
- Validate tool results in operational environments of each demo site (WP4, T4.1).

Regarding the development and update for each demo site, the specific next steps will be:

- **DS1 - Llobregat (Spain) next steps**

Potential enhancements to each of the short-term modules will be thoroughly explored. This comprehensive investigation will encompass an in-depth analysis of the existing variables and an exploration of incorporating additional parameters that could significantly enhance the precision, accuracy, and overall robustness of the generated results. This may include incorporating factors such as seasonality, weather patterns, or other relevant environmental data.

In the long-term module, the primary objective will be to finalize the development of advanced predictive models based on the generated datasets. These models will leverage a range of sophisticated statistical and ML techniques, such as time series analysis, regression models, neural networks, and ensemble methods, to forecast future trends and outcomes related to maritime activities, environmental impacts, and socioeconomic factors.

In addition to the aforementioned, the technical study will conduct a thorough investigation into the potential advantages and enhancements that could be achieved by incorporating supplementary data sources into the PREDMAR tool. This comprehensive analysis will encompass an assessment of the extent to which these additional data streams can augment the predictive capabilities and overall robustness of the tool, thereby contributing to its efficacy and dependability in fulfilling its intended purpose.

- **DS2 - Comporta (Portugal) next steps**

The development of predictive models for surface water surpluses, utilizing the datasets generated during the data acquisition and preprocessing phases, will be completed as a core component of the analytical framework. These models will be designed to identify, quantify, and forecast the occurrence of surface water excesses that may be available for alternative uses, including MAR, agricultural supplementation, or flood mitigation strategies.

To achieve this, the datasets—comprising historical observations and future climate scenario projections—will be processed using advanced feature engineering techniques. These include the extraction of seasonal trends, anomaly detection, precipitation-runoff relationships, and time-lagged correlations between rainfall events and surface accumulation.

The modeling approach will integrate statistical learning and ML algorithms, including regression-based models (e.g., Random Forest Regression, Gradient Boosting Machines etc.), depending on the temporal and spatial granularity of the data. The integration of climate projections—based on CMIP6 RCP-SSP scenarios—will allow for long-term forecasts of surface water surpluses under varying emission and socioeconomic pathways. This enables water managers to anticipate shifts in water availability due to climate change and to plan accordingly for periods of both surplus and deficit.

By leveraging the knowledge and insights gained from the AGREEMAR project, the predictive models can be refined and optimized, ensuring their accuracy and effectiveness in real-world scenarios. This collaborative approach will not only enhance the performance of individual MAR systems but also contribute to the overall advancement of MAR technology, paving the way for its widespread adoption and implementation in various contexts.

- **DS3 - Texel (Netherlands) next steps**

The forecasting framework, as outlined in the conceptual model presented in Section 4.1.2, is to be constructed upon the integration of seasonal meteorological predictions, serving as the foundational input for predictive modeling. This structure will be specifically tailored to reflect the operational characteristics and constraints of the MAR system currently implemented on the island of Texel. The primary objective is to enable the generation of robust, seasonal forecasts for aquifer recharge volumes, facilitating improved planning and adaptive resource management under variable climatic conditions.

To achieve this, the framework will incorporate climate model outputs—such as those from SEAS5 or equivalent seasonal forecasting systems—subjected to spatial and temporal downscaling and bias correction methods. These climate variables, particularly precipitation and evapotranspiration, will serve as explanatory inputs for AI-based predictive models capable of estimating the seasonal performance of the MAR system in terms of infiltration and recovery efficiency.

In parallel, a technical feasibility study will be undertaken to evaluate the potential extension of the forecasting capabilities toward long-term horizons. This assessment will explore the application of RCP and SSP scenario frameworks, as already implemented at other demonstration sites such as DS1 in the Llobregat Basin. The analysis will determine the extent to which such long-term planning tools can be calibrated and adapted to Texel's hydrogeological profile, infrastructure design, and regulatory context.

Particular attention will be given to the alignment between the predictive tool's temporal resolution and the operational cycles of the MAR system, ensuring actionable insights for both infrastructure operation and strategic planning. Additionally, data availability—both historical and real-time—will be assessed to support model training, validation, and scenario development. The inclusion of feedback loops using observed recharge performance and sensor data will enhance model accuracy and enable iterative improvement of the forecasting system over time.

Ultimately, the customized framework aims to contribute to the optimization of water storage strategies on Texel, supporting climate resilience, efficient aquifer use, and the broader objectives of sustainable AWR integration.

6.3 MARAISER - Short and long-term MAR performance optimization tool

The **MARAISER tool demonstrates significant potential in enhancing the operational efficiency of MAR systems using AWR**. The integration of short and long-term capabilities provides a robust framework for decision-making in MAR operations.

Key strengths of the MAR-SP Risk Assessment methodology include its comprehensive approach, **combining quantitative and qualitative risk assessment methods** while **successfully integrating established frameworks** (WHO, EPA, CEDEX).

The MAR-SP Risk Assessment methodology combines quantitative and qualitative approaches. The quantitative risk assessment involves numerical analysis of specific parameters and their probability of occurrence, while qualitative evaluation considers broader aspects that may not be easily quantifiable. This includes systematic hazard identification, detailed exposure assessment, and comprehensive risk characterization, as has been described specifically in DS1 with the specification of premises to take into consideration to perform the assessment.

Continuous monitoring is an essential component of risk management, encompassing regular measurement and analysis of water quality parameters, hydraulic performance indicators, environmental impacts, and system efficiency metrics. This ongoing monitoring allows for **early detection of potential issues and enables timely intervention when necessary**.

The tool's **flexibility in accommodating different MAR infrastructure types** (infiltration ponds and deep injection wells) and data availability levels is particularly noteworthy. The preliminary results reveal several critical insights: i) the necessity of balancing quantitative data with expert criteria for effective risk assessment, ii) the importance of site-specific parameter customization, and iii) the value of continuous monitoring for early risk detection.

The success of risk management in MAR systems with AWR depends on the integration of these various components into a **cohesive framework that can adapt to changing conditions while maintaining consistent protection of public health and environmental values**. This requires regular review and updating of risk assessment protocols, management strategies, and monitoring programs to ensure their continued effectiveness and relevance.

Nevertheless, the tool's **adaptability across different contexts**, combined with its comprehensive risk assessment capabilities, positions **MARAISER as a valuable asset for improving MAR system operational efficiency**, despite existing implementation challenges.

Regarding the demo site applications, in **DS1 (Llobregat -Spain-)**: the methodology has been successfully developed in two different MAR systems, i) the infiltration system (SVH) with focus on water quality and recharge volume and in ii) deep injection system (P18) with emphasis on clogging prevention. In DS1, MARAISER includes the integration with other tools (MARVIS for NO₃, PREDMAR for flow forecasting in the short term and PREDMAR for System Drought Status in the long term). In **DS2 (Comporta -Portugal-)**: the tool focuses on maximizing treated wastewater recharge, attending potential clogging issues and ensuring the maximum water quality improvement. In **DS3 (Texel - The Netherlands-)**: it has been demonstrated the application in agricultural context focusing on nutrient and pesticide control showing the tool's usefulness in stakeholder engagement (farmers).

As key learnings, it is worth noting that: i) the approach needs enough historical quantitative data and expert criteria (variability of the data availability across demo sites), ii) the importance of site-specific threshold definition (different regulatory requirements by country), iii) the value of continuous monitoring and early warning capabilities.

The MARAISER tool encompasses a comprehensive range of functionalities and applications that make it a valuable asset in MAR operations. From a risk management perspective, the tool provides continuous monitoring capabilities for health, environmental, and operational risks, incorporating an early warning system that alerts operators when thresholds are exceeded. This proactive approach enables automated risk assessment and supports informed decision-making for risk mitigation actions.

In terms of operational management, MARAISER enables real-time monitoring of MAR system performance and quality control of alternative water resources. The tool optimizes infrastructure maintenance schedules and tracks operational efficiency, while maintaining a vigilant alert system for any operational issues that may arise. This ensures smooth and efficient day-to-day operations of MAR facilities.

The tool's long-term planning and design capabilities are particularly valuable for infrastructure development. It assists in capacity planning for MAR systems, forecasts infrastructure expansion needs, predicts resource availability, and optimizes demand management. These features are essential for sustainable system growth and development.

Water quality management is another crucial aspect of MARAISER's functionality. The tool provides continuous monitoring of water quality parameters, tracks compliance with regulatory standards, and monitors chemical and biological parameters. This comprehensive approach to quality control helps prevent contamination and ensures regulatory compliance.

In the realm of data management and reporting, MARAISER automates data collection and processing, facilitates historical data analysis, and generates performance reports with clear visualizations. This makes it an invaluable tool for regulatory compliance reporting and system optimization.

The tool also serves as a powerful platform for research and development, supporting performance analysis of MAR systems, technology validation, and scientific research. It aids in the development of best practices and contributes to the advancement of MAR technology and operations.

Finally, MARAISER plays a crucial role in stakeholder communication by providing real-time status reporting and ensuring transparency in operations. This feature supports informed decision-making and facilitates effective public information sharing, making it an essential tool for maintaining stakeholder engagement and trust.

Therefore, the MARAISER tool represents a comprehensive and innovative solution for MAR systems. **The tool excels in water quality and operational risk management through its real-time monitoring capabilities and early warning system, ensuring regulatory compliance and optimal system performance.** Its robust data management and reporting features **facilitate transparent communication with stakeholders while enabling evidence-based decision-making through automated data processing and historical analysis.** Furthermore, the **tool's long-term planning capabilities provide valuable insights for infrastructure investment decisions,** particularly when considering reclaimed water availability forecasts. When combined with RECOVER outputs (WP4, T4.1 and T4.2), MARAISER becomes an essential instrument for determining MAR systems' cost recovery potential based on proposed tariff models. This integrated approach to MAR system management demonstrates the tool's significant potential to enhance the efficiency, sustainability, and economic viability of MAR operations.

The next steps needed for the implementation and validation of MARAISER are:

- Connection of data sources with the MARCLAIMED IDST server.
- Continuous feedback from stakeholders and potential end users through the CoPs workshops (WP3).
- Refinement of thresholds and adaptation of mitigation actions according to the results of each demo site (WP4, T4.1).
- Excel file for Module 2 (long term) set up in DS1 with the first results of PREDMAR long term (WP4, T4.1).

6.4 General discussion, conclusions and next steps

The integration of AI-based tools such as MARVIS and PREDMAR, in combination with the risk management module MARAISER, offers a promising approach to improving the management of MAR systems. While still under validation, these tools show potential in supporting various aspects of MAR operations, including water quality monitoring, forecasting of key variables, and structured risk assessment. Further testing and application across different scenarios will be essential to confirm their effectiveness and operational robustness.

MARVIS Tool represents a significant step forward in water quality monitoring using virtual sensing technology. Its focus on nitrate concentration measurement addresses critical compliance needs related to EU Directive 91/676/EEC. The tool's innovative approach of correlating easily measurable parameters with nitrate concentrations provides a cost-effective and scalable solution compared to traditional sensors. The implementation of XGBoost models, with plans to transition to deep learning approaches, demonstrates MARVIS's commitment to leveraging advanced ML techniques for improved accuracy and adaptability. The tool's success in providing real-time monitoring

capabilities while maintaining cost-effectiveness positions it as a valuable asset for MAR operations.

PREDMAR Tool is a valuable tool in water resource management, providing support for both short-term monitoring and long-term planning needs. Its ability to integrate diverse data sources, including climate projections (CMIP6), sensor networks, and historical events, enables sophisticated predictive decision-making for water allocation. The tool's hybrid learning architecture, incorporating Physics-Informed Machine Learning, ensures physically plausible forecasts while capturing complex system dynamics. PREDMAR's capacity to link long-term meteorological forecasts with recharge opportunities is particularly valuable for operational resilience, especially in managing extreme weather events.

MARAISER Tool offers a useful approach to improving MAR system operational efficiency by incorporating a structured risk assessment framework. The tool successfully combines quantitative and qualitative approaches, integrating established frameworks from WHO, EPA, and CEDEX. Its flexibility in accommodating different MAR infrastructure types and varying levels of data availability has proven particularly valuable across demonstration sites. MARAISER's continuous monitoring capabilities and early warning system for health, environmental, and operational risks provide crucial support for real-time decision-making and risk mitigation. The tool's ability to optimize infrastructure maintenance, track operational efficiency, and ensure regulatory compliance makes it an essential component of modern MAR operations.

Despite these overall breakthroughs, several challenges will emerge during implementation, including variations in data availability and quality across sites, the necessity for site-specific customization, regulatory differences between countries, and the complexity of integration with existing systems.

Moving forward, the focus will be on technical implementation, including the connection of data sources to the MARCLAIMED IDST server and the continuation of data acquisition to enhance machine learning models. The transition from XGBoost to neural network models and the implementation of transfer learning across demonstration sites will further improve system capabilities. Validation efforts will concentrate on enhancing predictive models with additional parameters and refining site-specific thresholds, while strengthening the integration between the different tools. Stakeholder engagement will remain crucial, with ongoing feedback through CoPs workshops, improved communication features, enhanced user interfaces, and the development of comprehensive training programs to ensure effective system utilization and adoption by the end users.

7. Bibliographic References

- [1] Referential data regarding water scarcity and droughts is accessible on the European Commission's official website:
https://environment.ec.europa.eu/topics/water/water-scarcity-and-droughts_en.
- [2] Jaywant, S. A., and Arif, K. M. (2024). "Remote Sensing Techniques for Water Quality Monitoring: A Review." *Sensors*, vol. 24, no. 24, 8041.
<https://doi.org/10.3390/s24248041>.
- [3] Ransom, K. M., Nolan, B. T., Stackelberg, P. E., Belitz, K., & Fram, M. S. (2021). Machine learning predictions of nitrate concentrations in groundwater utilized for potable supply within the contiguous United States. *Science of the Total Environment*, 151065.
<https://doi.org/10.1016/j.scitotenv.2021.151065>
- [4] Diana Yaritza Dorado-Guerra, G. Corzo, J. Paredes-Arquiola, and Miguel Ángel Pérez-Martín. (2022). Machine learning models for the prediction of nitrate concentration in a river basin. *Environmental Research Communications*, 4(12), 125012.
<https://doi.org/10.1088/2515-7620/acabb7>
- [5] Divas Karimanzira, J. Weis, A. Wunsch, L. Ritzau, T. Liesch, and M. Ohmer (2023). "Application of machine learning and deep neural networks for spatial prediction of groundwater nitrate concentration to improve land use management practices." *Frontiers in Water*, 5. <https://doi.org/10.3389/frwa.2023.1193142>
- [6] Chaves, A. J., Martín, C., Llopis Torres, L., Díaz, M., Fernández-Ortega, J., Barberá, J. A., & Andreo, B. (2025). A soft sensor open-source methodology for the cost-effective monitoring of water quality: A case study on NO₃⁻ concentrations. *Journal of Computational Science*, 85, 102522.
<https://doi.org/10.1016/j.jocs.2024.102522>
- [7] Baena-Miret, S., Puig, M. A., Rodes, R. B., Farran, L. B., Durán, S., Martí, M. G., Martínez-Gomariz, E., & Valverde, A. C. (2024). Miglioramento dell'efficienza e del controllo di qualità: L'impatto dei gemelli digitali nelle reti di acqua potabile. *Water Environment Research*, 96 (10). <https://doi.org/10.1002/wer.11139>
- [8] Saenger, V., García, L., Díaz, M. A., Ganzer, M., Montes, S., & González, S. (2022). "A reliable real-time virtual trihalomethane sensor solution for drinking water facilities." In *Proceedings of the 39th IAHR World Congress*.
<https://doi.org/10.3850/IAHR-39WC2521711920221666>.
- [9] The Coupled Model Intercomparison Project (CMIP) website is accessible at <https://wcrp-cmip.org/>.
- [10] GOTHAM is a project under the GOTHAM-PRIMA initiative. Website is accessible at: <https://www.gotham-prima.eu/>
- [11] MAR2PROTECT project, funded under the European Union's Horizon Europe research and innovation programme, identification number 101082048, can be accessed at the following URL: <https://cordis.europa.eu/project/id/101082048>.
- [12] The AGREEMAR platform, accessible at <https://agreemar.inowas.com/>.
- [13] The ICARIA project, with the identifier:
<https://cordis.europa.eu/project/id/101093806>

- [14] Kashinath K. et al. (2021), "Physics-informed machine learning: case studies for weather and climate modelling," Philosophical Transactions of the Royal Society A, <http://doi.org/10.1098/rsta.2020.0093>.
- [15] Reference:
https://info.aca.gencat.cat/ca/aca/informacio/geco/plans-programes/PDG/ES/01-00_Plan_gestion_2022-27.pdf
- [16] Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy, commonly known as the Water Framework Directive, can be accessed at the following URL: <https://eur-lex.europa.eu/eli/dir/2000/60/oj/eng>.
- [17] Directive 2008/32/EC, can be accessed at the following URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32008L0032&qid=1741591962089>
- [18] Council Directive 91/676/EEC, can be accessed at the following URL: <https://eur-lex.europa.eu/eli/dir/1991/676/oj/eng>
- [19] Directive 91/271/EEC, can be accessed at the following URL: <https://eur-lex.europa.eu/eli/dir/1991/271/oj/eng>
- [20] Directive (EU) 2024/3019, can be accessed at the following URL: <https://eur-lex.europa.eu/eli/dir/2024/3019/oj/eng>
- [21] Directive 2006/118/EC of the European Parliament and of the Council of 12 December 2006 on the protection of groundwater against pollution and deterioration, hereinafter referred to as the EU Groundwater Directive, can be accessed at the following URL: <https://eur-lex.europa.eu/eli/dir/2006/118/oj/eng>.
- [22] Regulation (EU) 2020/741, as available at: <https://eur-lex.europa.eu/eli/reg/2020/741/oj/eng>.
- [23] Commission Delegated Regulation (EU) 2024/1765 serves to empower responsible parties engaged in water reuse projects to develop comprehensive risk management strategies and guarantee the safety of said projects, can be accessed at the following URL: https://eur-lex.europa.eu/eli/reg_del/2024/1765/oj/eng
- [24] The Safe Drinking Water Act, as defined by the Environmental Protection Agency and accessible via the following link: <https://www.epa.gov/sdwa>.
- [25] The Clean Water Act, as referenced by the following hyperlink: <https://www.epa.gov/laws-regulations/summary-clean-water-act>
- [26] The Global Water Tool, accessible via the following link: <https://shift.tools/resources/835>.
- [27] The Water Risk Filter, accessible via the following URL: <https://riskfilter.org/water/assess>
- [28] The Strategic Project for Economic Recovery and Transformation (PERTE) Plan of Spain, as documented at the following URL: <https://www.mdpi.com/2071-1050/14/20/13384>.
- [29] The MARSOL Project, supported by the European Union's FP7 research and innovation programme under grant agreement number 619120, is documented and further information may be accessed at the following URL: <https://cordis.europa.eu/project/id/619120/reporting>.

- [30] LIFE Matrix Project, funded under the European Union's LIFE research and innovation programme, identification number: LIFE20 ENV/ES/000788, as referenced by the following link: <https://www.life-matrix-project.eu/>.
- [31] MARSOLUT Project, an initiative funded by the European Union's Marie Skłodowska-Curie Actions, operating under Grant Agreement ID: 814066, and further detailed at <https://www.marsolut-itn.eu/>.
- [32] Brown, C. M., Lund, J. R., Cai, X., Reed, P. M., Zagona, E. A., Ostfeld, A., Hall, J., Characklis, G. W., Yu, W., & Brekke, L. (2015). The future of water resources systems analysis: Toward a scientific framework for sustainable water management. *Water Resources Research*, 51, 6110–6124.
- [33] O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9, 3461–3482.
<https://doi.org/10.5194/gmd-9-3461-2016>.
- [34] The World Health Organization's Drinking-water Quality Guidelines are accessible via the following URL:
<https://www.who.int/teams/environment-climate-change-and-health/water-sanitation-and-health/water-safety-and-quality/drinking-water-quality-guidelines>.
- [35] EPA Drinking Water Regulations, as referenced at the following URL:
<https://www.epa.gov/dwreginfo/drinking-water-regulations>.
- [36] Las directrices de CEDEX para la gestión de riesgos del agua regenerada en el riego, a las que se hace referencia en el siguiente enlace:
<https://www.cedex.es/centros-laboratorios/centro-de-estudios-hidrograficos-ceh/proyectos/guias-para-la-gestion-del>.
- [37] Katarzyna B. Tokarska et al., "Past warming trend constrains future warming in CMIP6 models," *Science Advances*, vol. 6, eaaz9549 (2020). DOI: 10.1126/sciadv.aaz9549.
- [38] Information regarding the Spectrolyser V3 spectrometer probe, a water quality monitoring product, can be found at the following URL:
<https://www.badgermeter.com/products/water-quality-monitoring/spectrometer-probes/spectrolyser-v3/>.
- [39] Please find the MicroSUM datalogger documentation at the following URL:
<https://www.siapmicros.com/wp-content/uploads/2022/12/MicroSUM-ENG-2.pdf>
- [40] The Trios Niko Plus product is described at the following web address:
<https://trios.de/en/nico-plus/>.
- [41] DO/T: <https://www.aqualabo.fr/en/produit/optod-stainless-steel-sensor/>
- [42] Specifications for digital sensor PHEHT, measuring pH, Oxidation-Reduction Potential (ORP), and Temperature (T) are available at:
<https://www.aqualabo.fr/en/produit/digital-sensor-pheht/>
- [43] The electronic conductivity and temperature sensor, CTZN, can be found at the following URL: <https://www.aqualabo.fr/en/produit/digital-sensor-ctzn/>.
- [44] Shojaie, A., & Fox, E. B. (2022). Granger Causality: A Review and Recent Advances. *Annual Review of Statistics and Its Application*, 9(1), 289-319.

<https://doi.org/10.1146/annurev-statistics-040120-010930>. PMID: 37840549; PMCID: PMC10571505.

[45] Meteosource technology is accessible at the following URL:

<https://www.meteosource.com/meteosource-technology>.

[46] SHAP library documentation can be found at:

<https://shap.readthedocs.io/en/latest/>.

[47] Gneiting, T., & Katzfuss, M. (2014). Probabilistic Forecasting. Annual Review of Statistics and Its Application, 1(1), 125-151.

<https://doi.org/10.1146/annurev-statistics-062713-085831>.

ANNEX 1: MARAISER - Short and long-term MAR performance optimization tool

DS1 - Llobregat (Spain) - MAR-SP Risk Assessment

Here below can be found screenshots of the different assessment phases: i) identification of non compliance, ii) risk assessment and iii) decision three for RMP, Operational Requirement Point (ORP) or Pre-Requirement Point (PRP) identification; Figure A1.a, b and c)).

ETAPA	Unitats	Agent perillós	Result. Resu It	Promedio de Result. Resu It2	Más. de Result. Resu It3	Min. de Result. Resu It4	Limit	Go valor n	Limit acceptant empejorament	Número Incomplime nts	% Complemen t	MOSTRA	NORMATIVA
TRACTAMENT TERCIARI	ng/l	1,4-dioxà	9	15126	30075	6785	6032,8	50000	22699,47	1	11%	Puntual	Blanc de l'aquífer + marge VP o VG
TRACTAMENT TERCIARI	µg/l	2,4-D*	4	0,01925	0,021	0,018	<0,015	0,1	0,03	0	0%	Puntual	Blanc de l'aquífer + marge VP o VG
TRACTAMENT TERCIARI	ng/l	2,5,5-trimetil-1,3-dioxà	5	2,325	9,3	0	<5,0			1	20%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	ng/l	2-i-propil-5,5-dimetil-1,3-dioxà	5	9,92	49,6	0	<10			1	20%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	ng/l	5,5-dimetil-1,3-dioxà	5	159	288	30	38			1	20%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	µg/l	6:2 fluorotelomer sulfonic acid (6:2FTS)	5	0,046	0,092	0,019	<0,015			5	100%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	Abs/100cm	Absorció a l'UV	6	20,783333	25,8	10,7	1,94			6	100%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	mg CaCO3/l	Alcalinitat	1	213	213	213	334,8			0	0%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	µg/l	Alumini	12	599,4	1696,3	229,77	268	600	468	7	58%	Puntual	Blanc de l'aquífer + marge valor no aptitud RD3/2023
TRACTAMENT TERCIARI	mg/l	Amantadina	3	155	197	127	96			3	100%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	µg/l	AMPA	5	6,58	10	4,9	0,1			5	100%	Puntual	RD 3/2023
TRACTAMENT TERCIARI	ng/l	Azitromicina	3	299,666667	899	0	<25	100	33,33	1	33%	Puntual	Blanc de l'aquífer + marge VP o VG
TRACTAMENT TERCIARI	µg/l	Azoxystrobin*	4	0,074	0,113	0,056	<0,015	0,1	0,03	4	100%	Puntual	Blanc de l'aquífer + marge VP o VG
TRACTAMENT TERCIARI	ng/l	Bisoprolol	3	156,466667	162	147,4	<10			3	100%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	mg SO4/l	Sulfats	74	150,147297	182	89,1	250			0	0%	Puntual	Valor lliandar MAS39 PGDCFC
TRACTAMENT TERCIARI	mg Br/l	Bromurs	74	1,3125	1,8	1	0,566		1,16	74	51%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	mg/l	Calci	12	104,333333	121	88	137			0	0%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	ng/l	Carbamazepina	3	125	148	108	58		100000	0	0%	Puntual	ACA N1 Prova demostrativa 2019
TRACTAMENT TERCIARI	mg C/l	Carboni Orgànic Total (COT)	11	8,95454545	11	4,5	1,33333333	7	3,67	11	100%	Puntual	Blanc de l'aquífer + marge valor no aptitud RD3/2023
TRACTAMENT TERCIARI	ng/l	Ciprofloxacina	3	836,333333	2129	126	100			3	100%	Puntual	Analogia watch list
TRACTAMENT TERCIARI	ng/l	Cis-2-etil-4-metil-1,3-dioxolà	5	1,92	9,6	0	<8,0			1	20%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	µg ClO3/l	Clorats	74	956,102564	4600	201	96,925	700	330,26	38	51%	Puntual	Blanc de l'aquífer + marge VP o VG
TRACTAMENT TERCIARI	mg Cl/l	Clorurs	74	505,162162	710	244	440			61	82%	Puntual	Valor lliandar MAS39 PGDCFC
TRACTAMENT TERCIARI	Ufp/100ml	Colifags somàtics	6	383,333333	1400	140	<20			6	100%	Puntual	Blanc de l'aquífer
TRACTAMENT TERCIARI	NMP/100ml	Coliformes totals	39	2,4494983	3,38381537	0	1,06445799	2		38	58%	Puntual	Límit no aptitud RD3/2023
TRACTAMENT TERCIARI	µS/cm	Conductivitat a 20 °C	41	2109,636	2512,0894	1092,21128	2300			9	22%	Puntual	Valor lliandar MAS39 PGDCFC

Figure A1.a Identification of non compliance (screenshot).

Paràmetre	TT	ZNS	ZS	Atenuació
1,4-dioxà	4	1	1	
2,4-D*	1	1	1	
2,5,5-trimetil-1,3-dioxà	2	1	1	
2-i-propil-5,5-dimetil-1,3-dioxà	2	1	1	
5,5-dimetil-1,3-dioxà	16	1	1	
6:2 fluorotelomer sulfonic acid (6:2FTS)	8	2	2	
Alcalinitat	1	1	1	
Alumini	48	1	1	
Amantadina	64	48	1	
AMPA	64	16	1	
Azitromicina	16	1	1	
Azoxystrobin*	64	4	1	
Bisoprolol	8	2	2	
Bor	8	1	1	
Bromurs	28	12	4	
Calci	1	1	4	
Carbamazepina	1	1	1	
Carboni Orgànic Total (COT)	64	16	1	
Ciprofloxacina	64	16	1	
Cis-2-etil-4-metil-1,3-dioxolà	2	1	1	
Clorats	48	24	4	
Clorurs	24	12	12	Descens d'un 8% entre AR i ZS. Càlcul fet amb VG 440 mg/l. Autorització marge fins a 800 mg/l
Colifags somàtics	8	1	1	
Coliformes totals	24	4	1	
Conductivitat a 20 °C	4	1	1	
Diazinon	8	4	1	
Diclofenac	64	1	1	
Diuron	64	16	12	Descens de més del 30% entre AR i ZS
Doxiciclina	16	1	1	

Figure A1.b. Risk assessment results screenshot (green: low, yellow: medium and red: high).

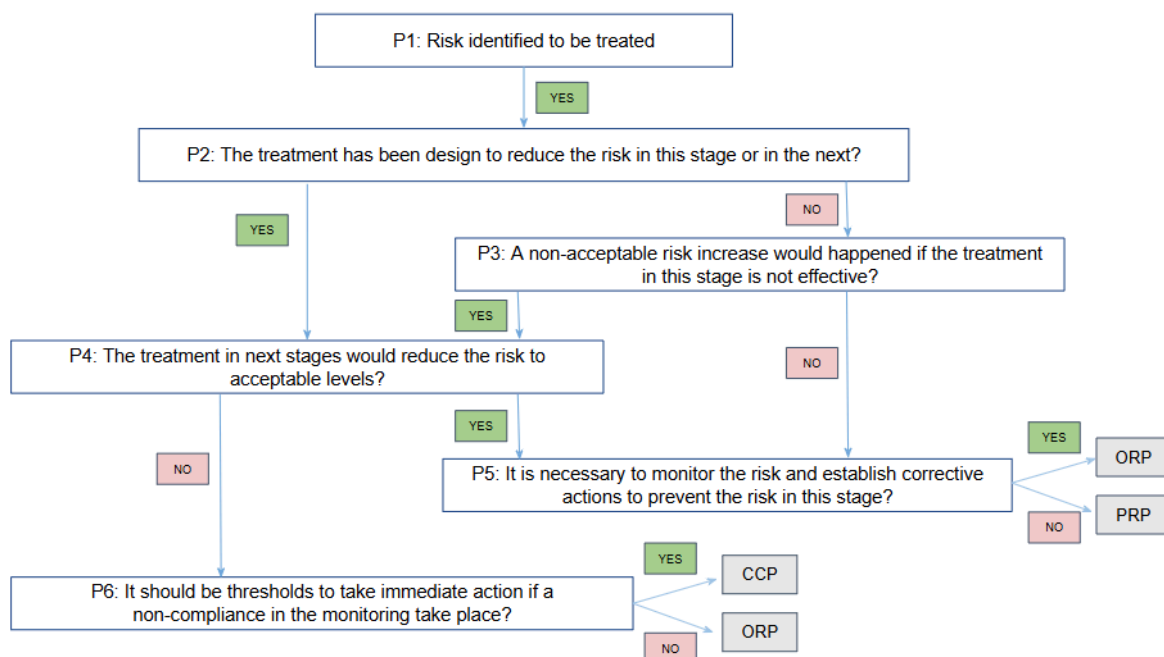


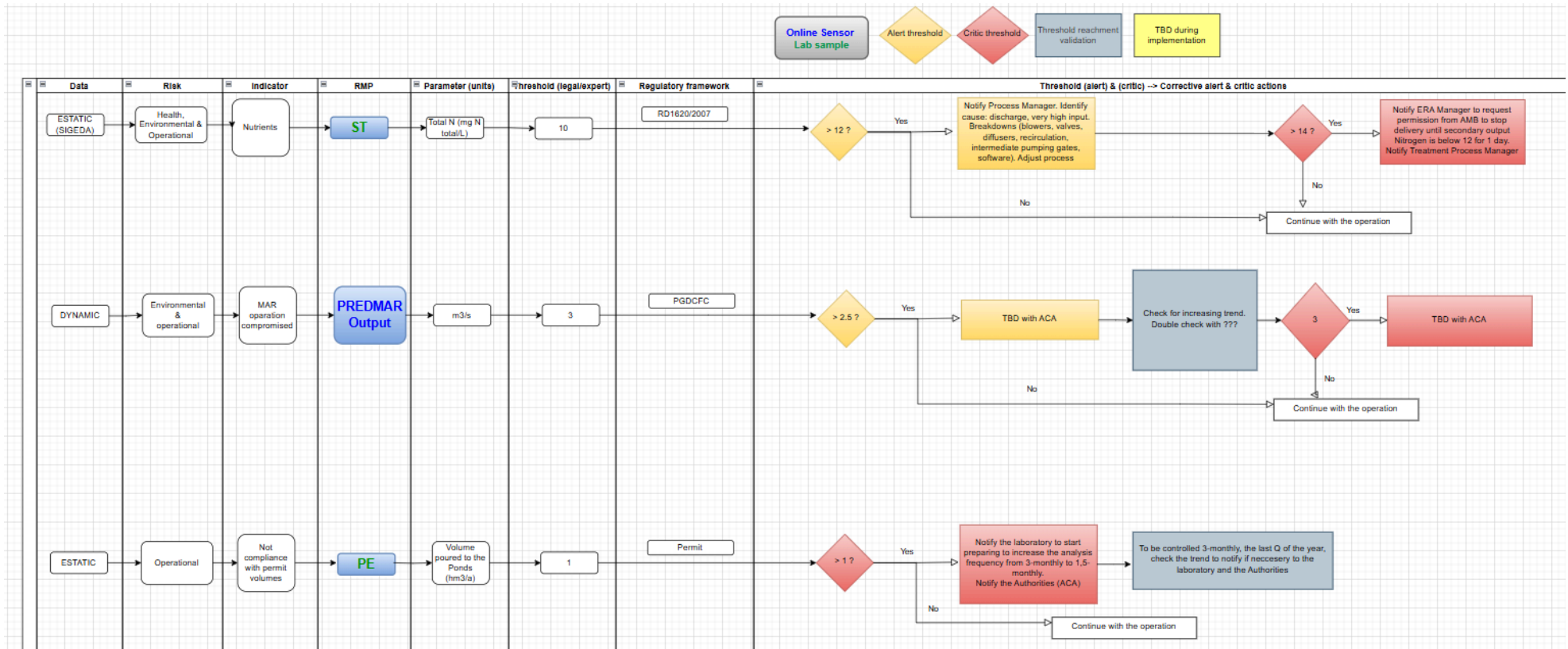
Figure A1.c. Decision tree to identify the RMP (CCP: critical control point, ORP: operational requirement point and PRP: pre-requirement point).

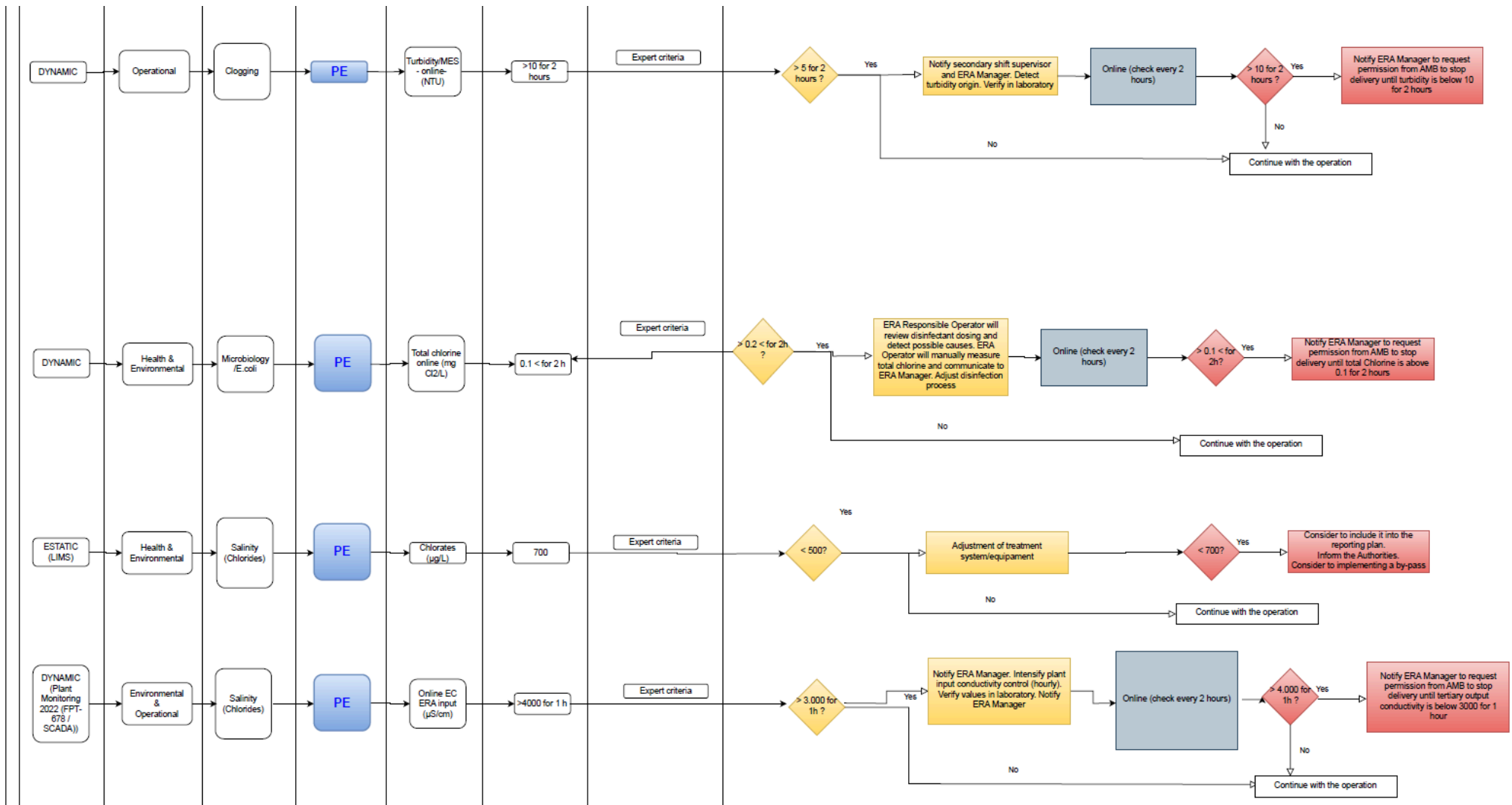
DS2 - Comporta (Portugal) - MAR-SP Risk Assessment

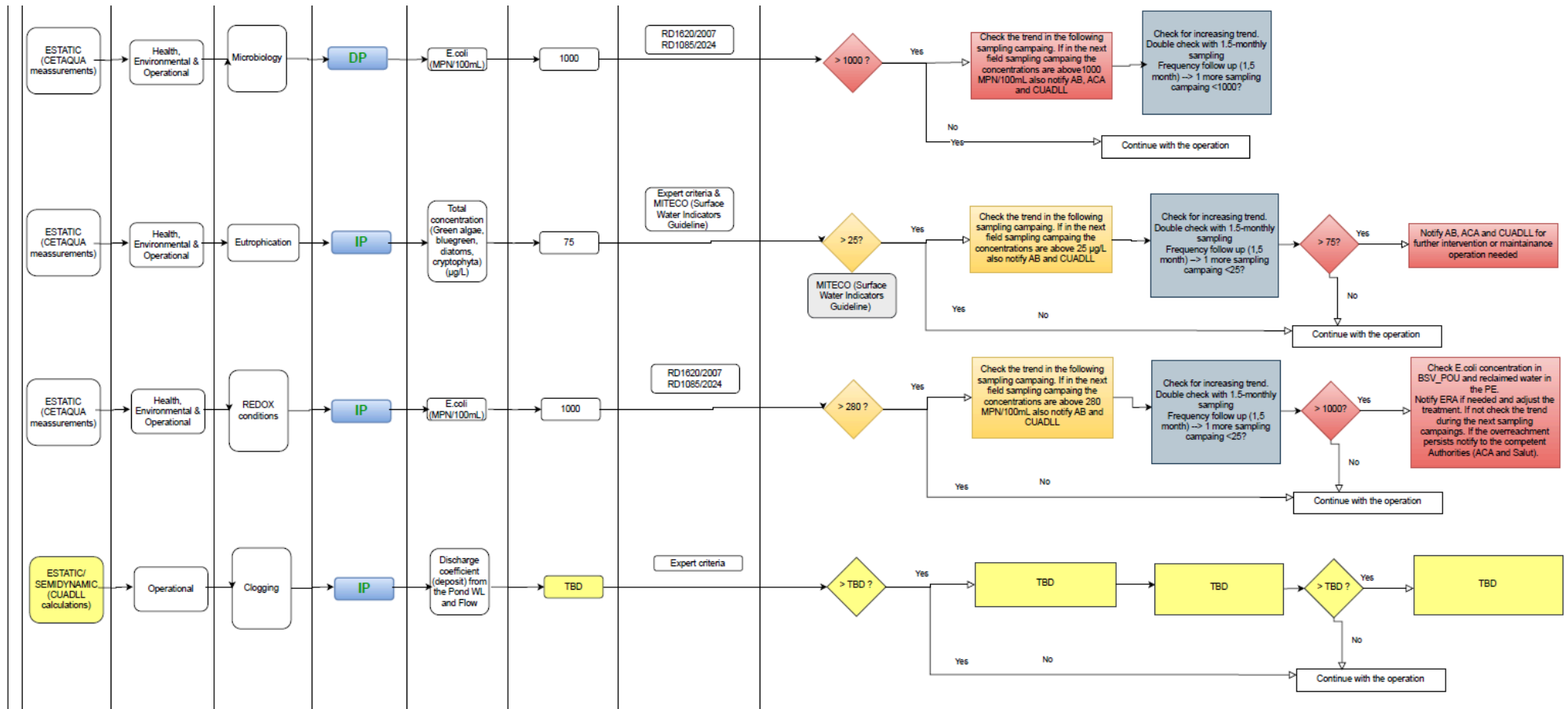
CCP	STAGE	HAZARDOUS AGENT	UNITS	N Data	Med	Max	Min	Threshold	Regulation	N non compliances	% compliance	PR	S	RISC
C	Treated effluent	P total	mg/l P	34	6,965	24	0,52	5	EU 2024/3019	21	38	6	8	48
		N total	mg/l N	34	13	70	5	15	EU 2024/3019	13	62	6	8	48
		TSS	mg/L	34	8,55	111	2,1	35	VLE p/Comporta	2	94	2	4	8
		pH	pH units at 20°C	35	7	7,5	6,5	6-9	VLE p/Comporta	0	100	1	1	1
		COD	mg/L	34	38,5	446	10,9	125	EU 2024/3019	2	94	2	8	16
		Oil and grease	mg/L	29	1,2	26	0,22	15	VLE p/Comporta	1	97	2	8	16
D	Piezometer PZ1	Sulphates	mg/l SO4	35	43	62,8	8	9-6	GWT thresholds	0	100	1	1	1
		Cholides	mg/L Cl	35	180	310	23	9-6	GWT thresholds	6	83	4	2	6
		EC	µS/cm	7	932	1192	751	4-14	Expert criteria	0	100	1	1	1
		pH	pH units at 20°C	7	6,3	6,5	6,2	5.5-9	GWT thresholds	0	100	1	1	1

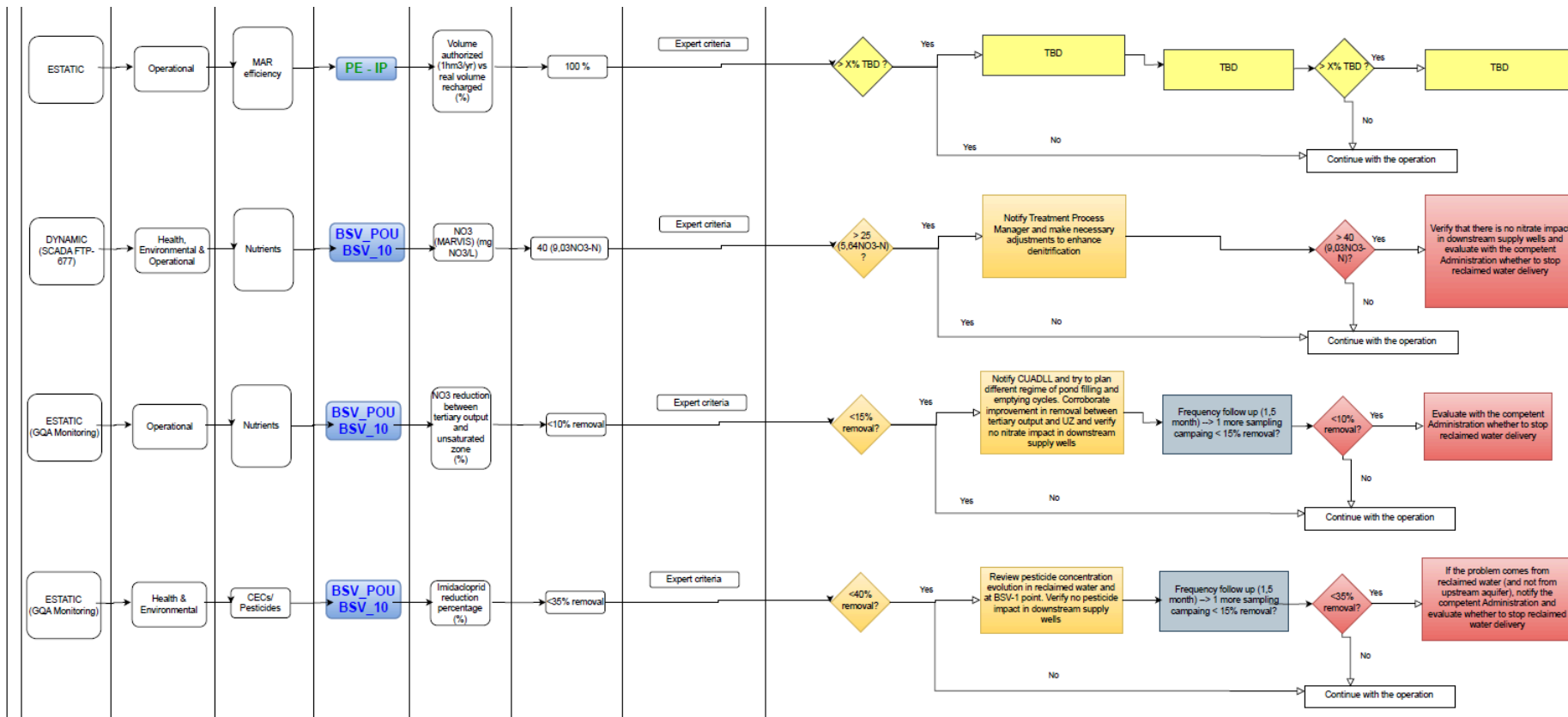
Figure A1.d Risk assessment results screenshot (green: low, yellow: medium and red: high) (N Data: number of data entries considered; PR: likelihood and S: severity).

DS1 - Llobregat (Spain) - Infiltration pond (SVH) MAR system Flowchart

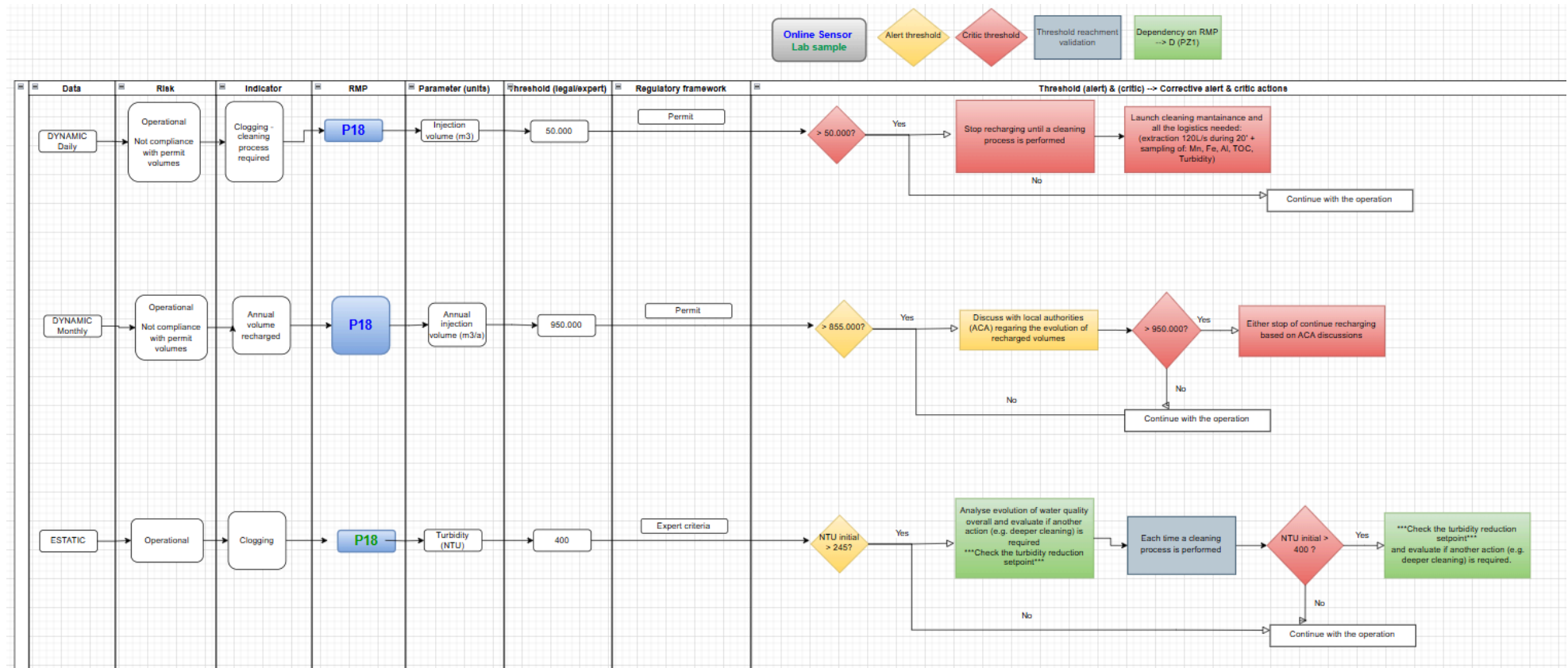


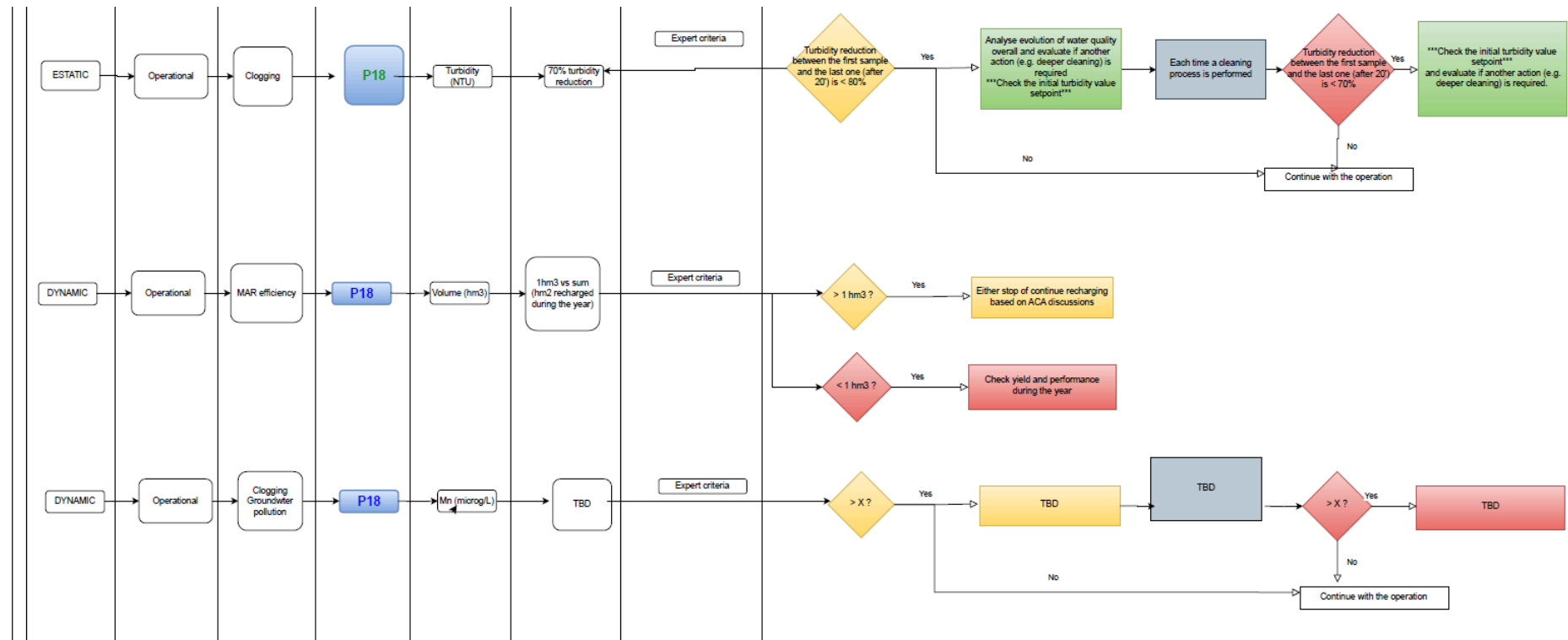




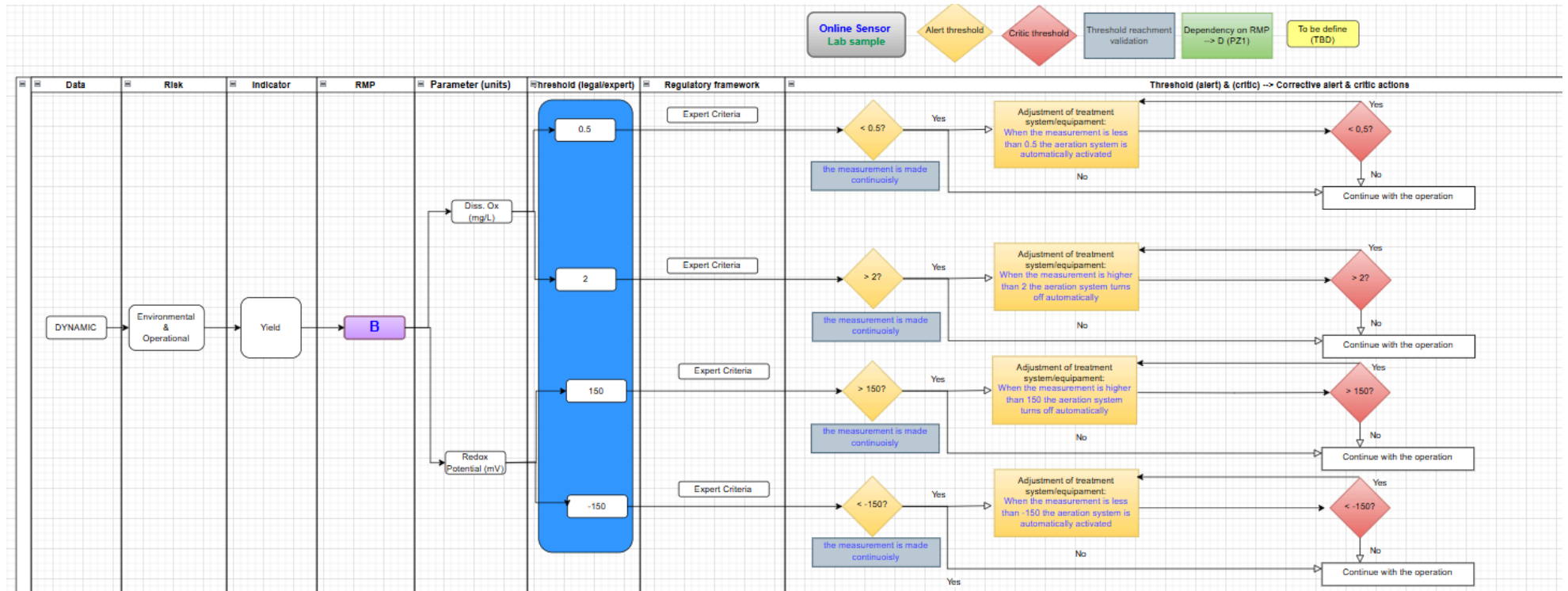


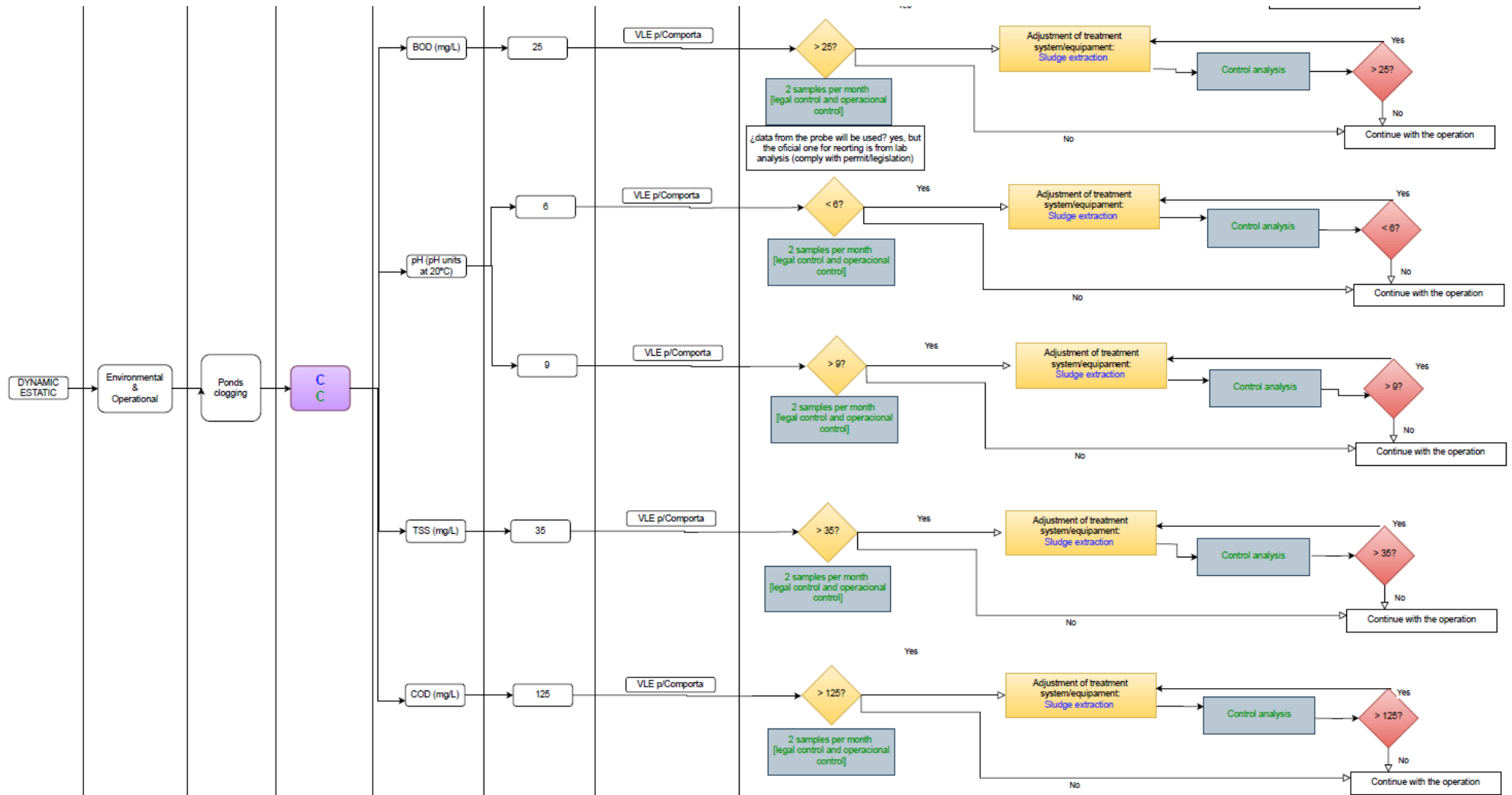
DS1 - Llobregat (Spain) - Infiltration pond (P18) MAR system Flowchart

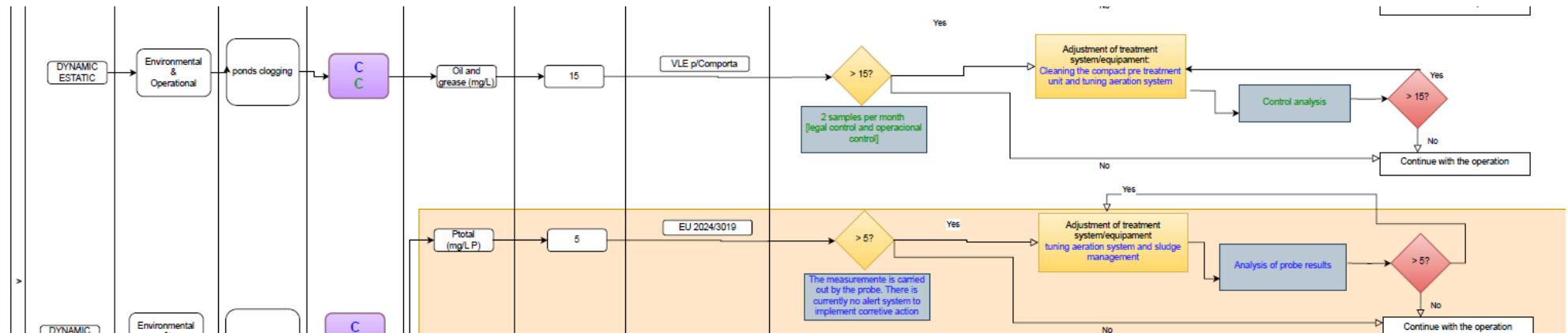


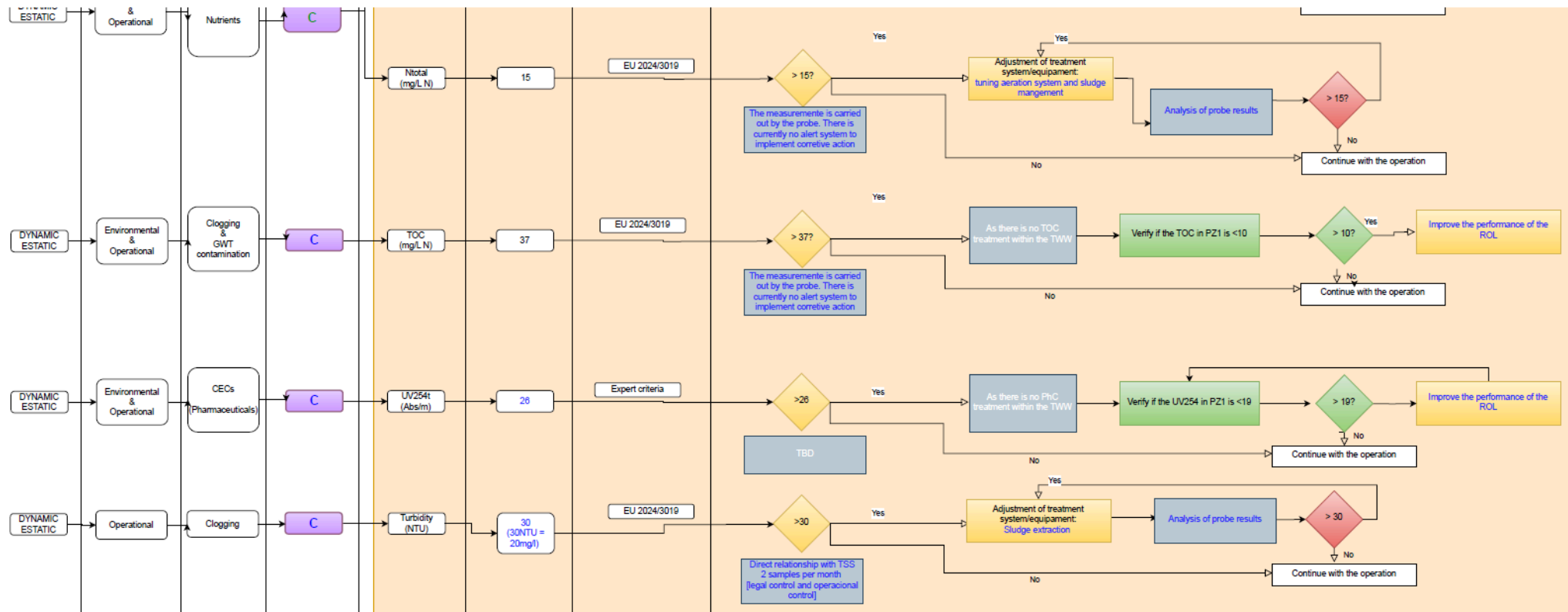


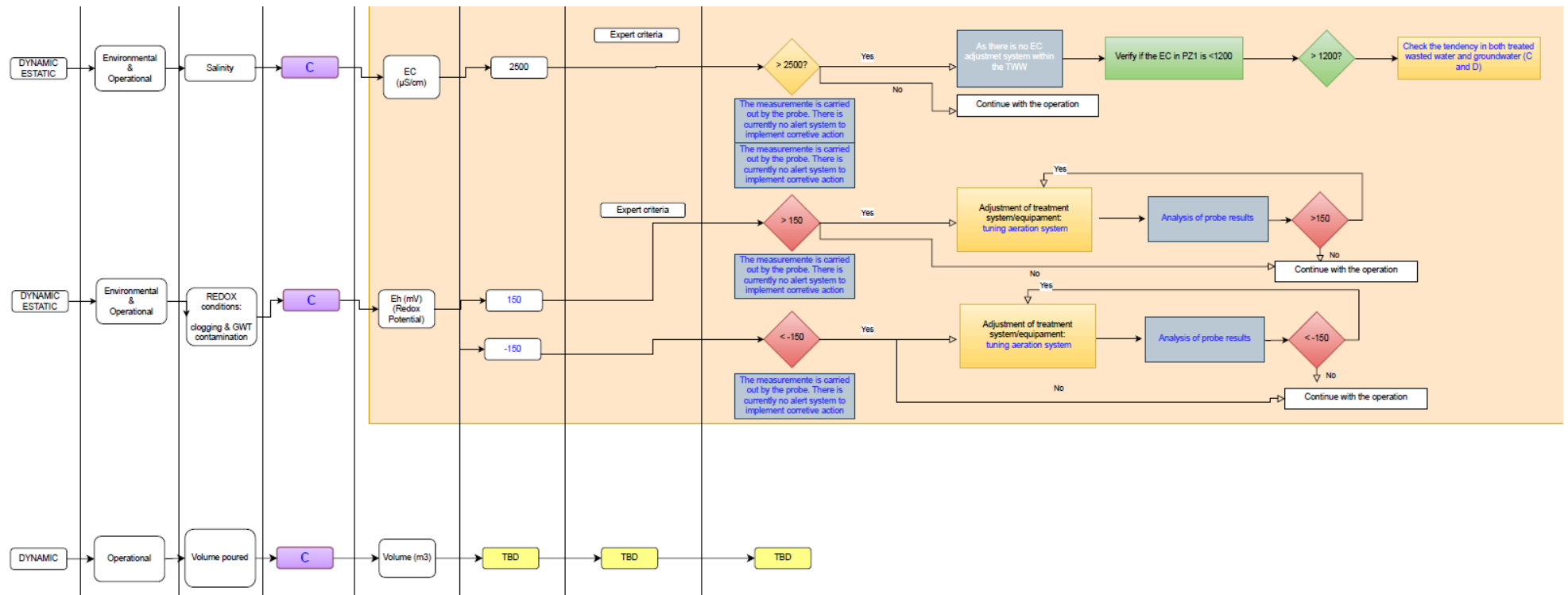
DS2 - Comporta (Portugal) - MAR system Flowchart

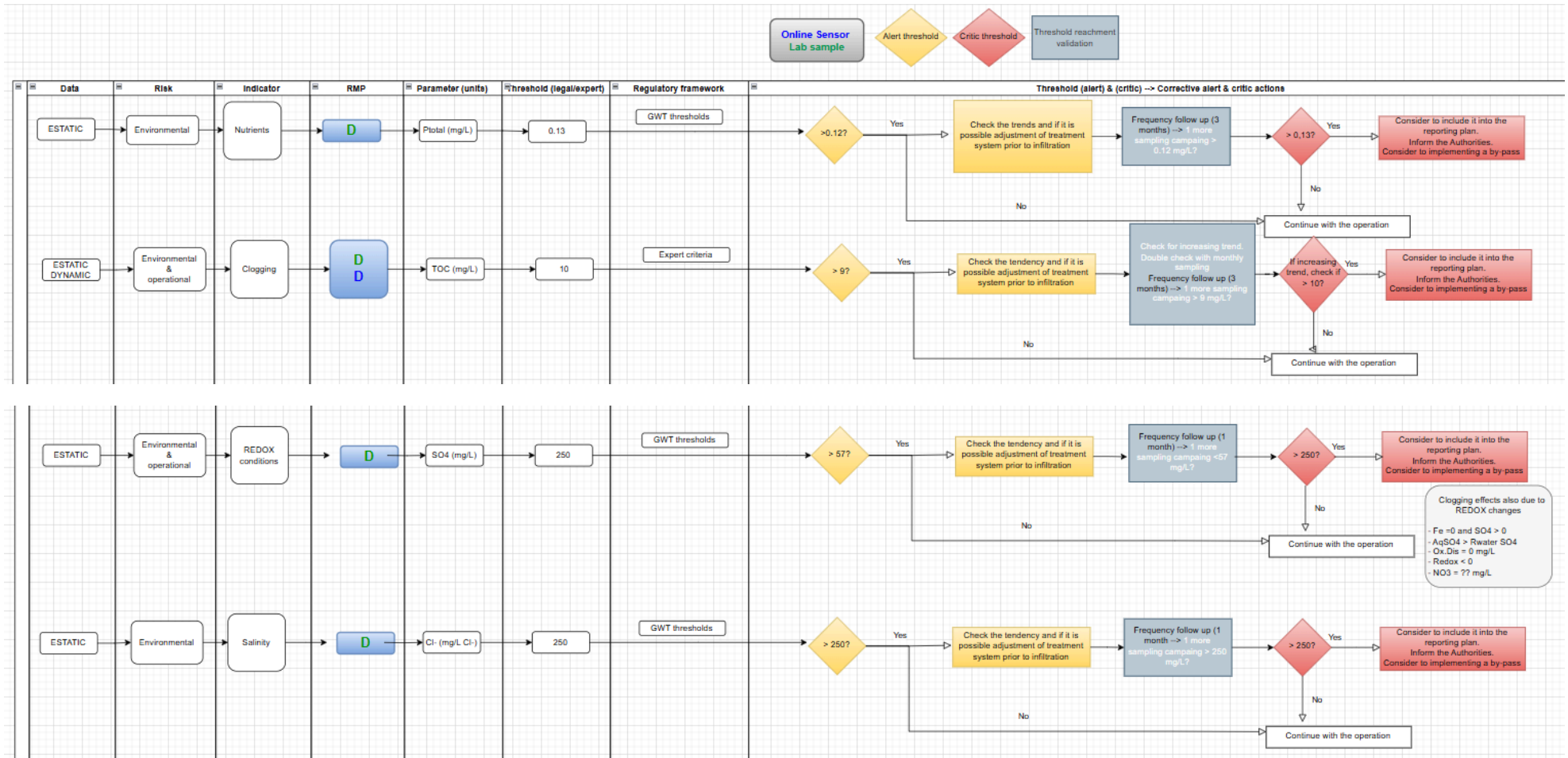


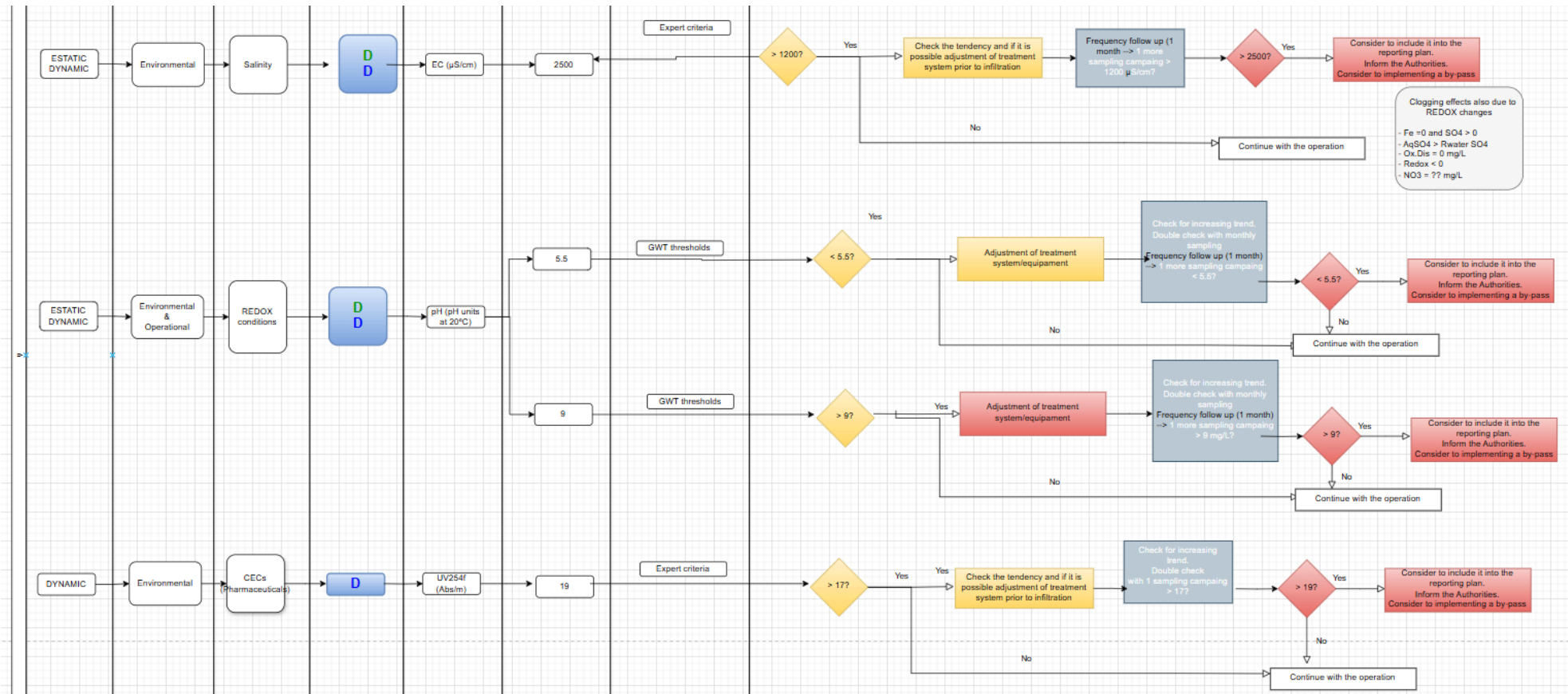


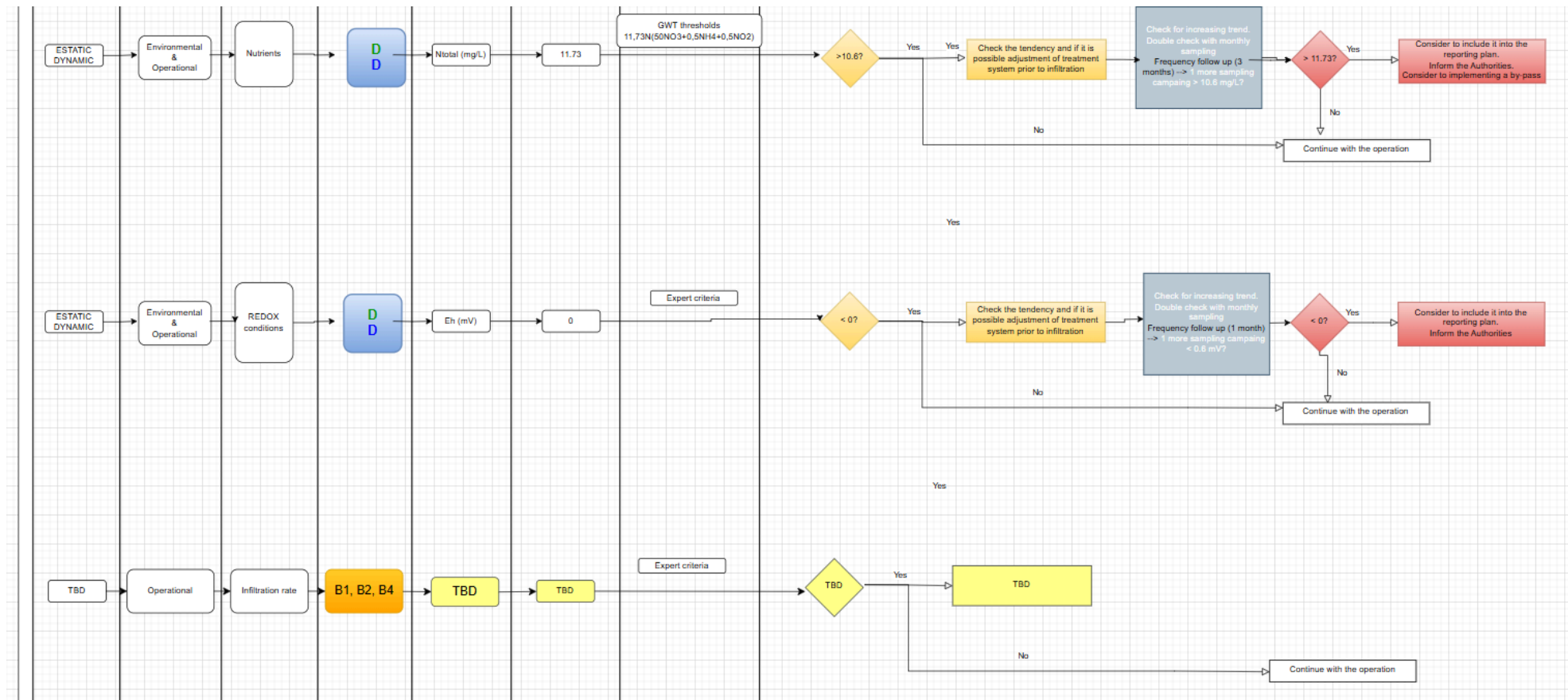




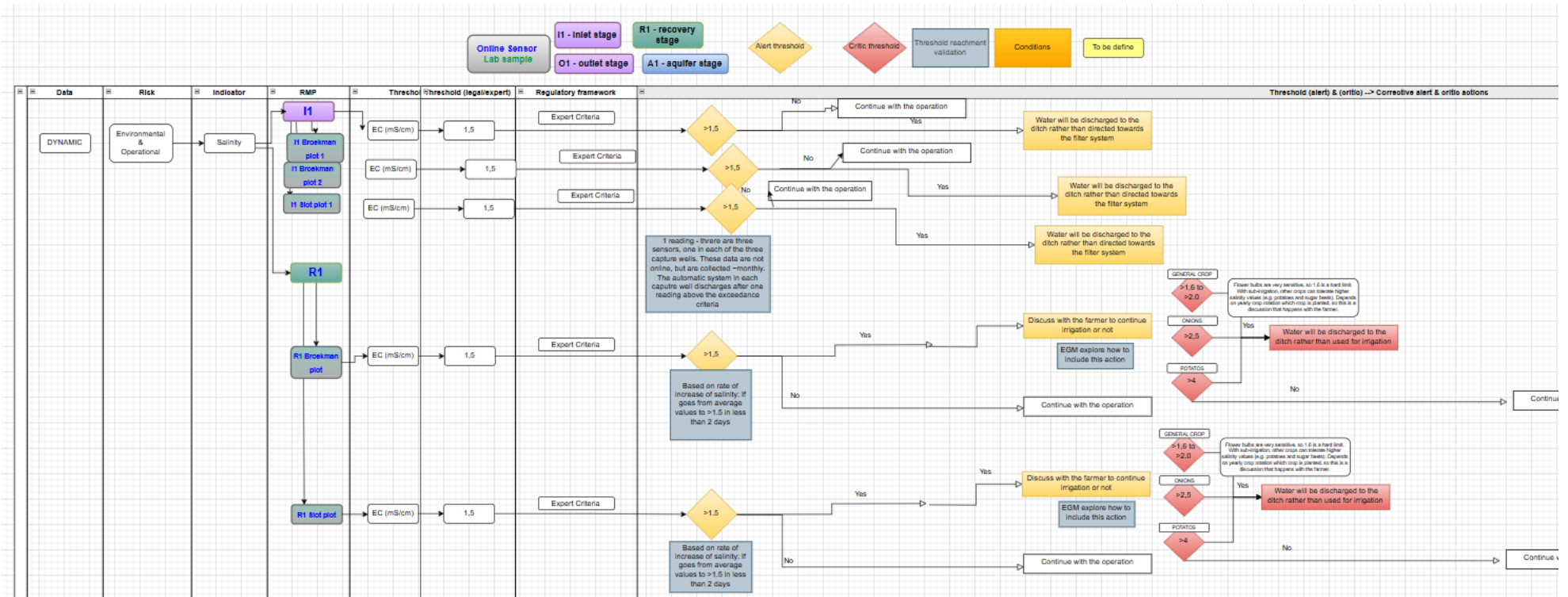


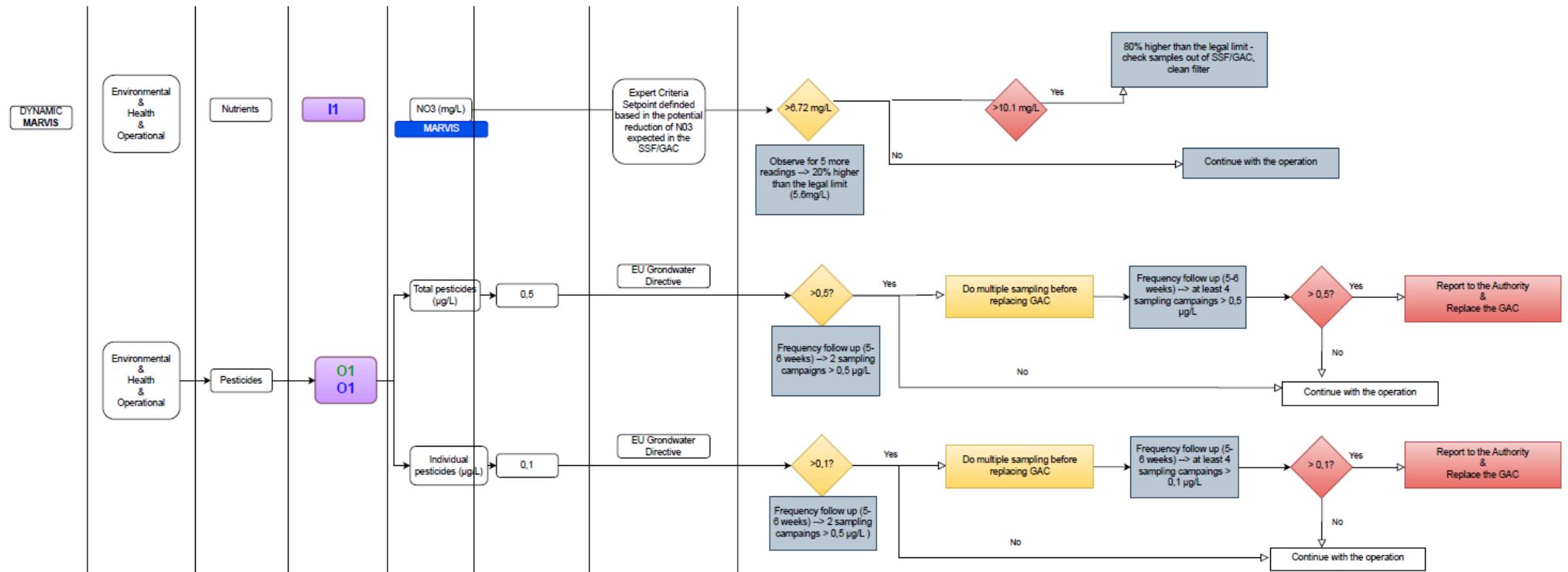


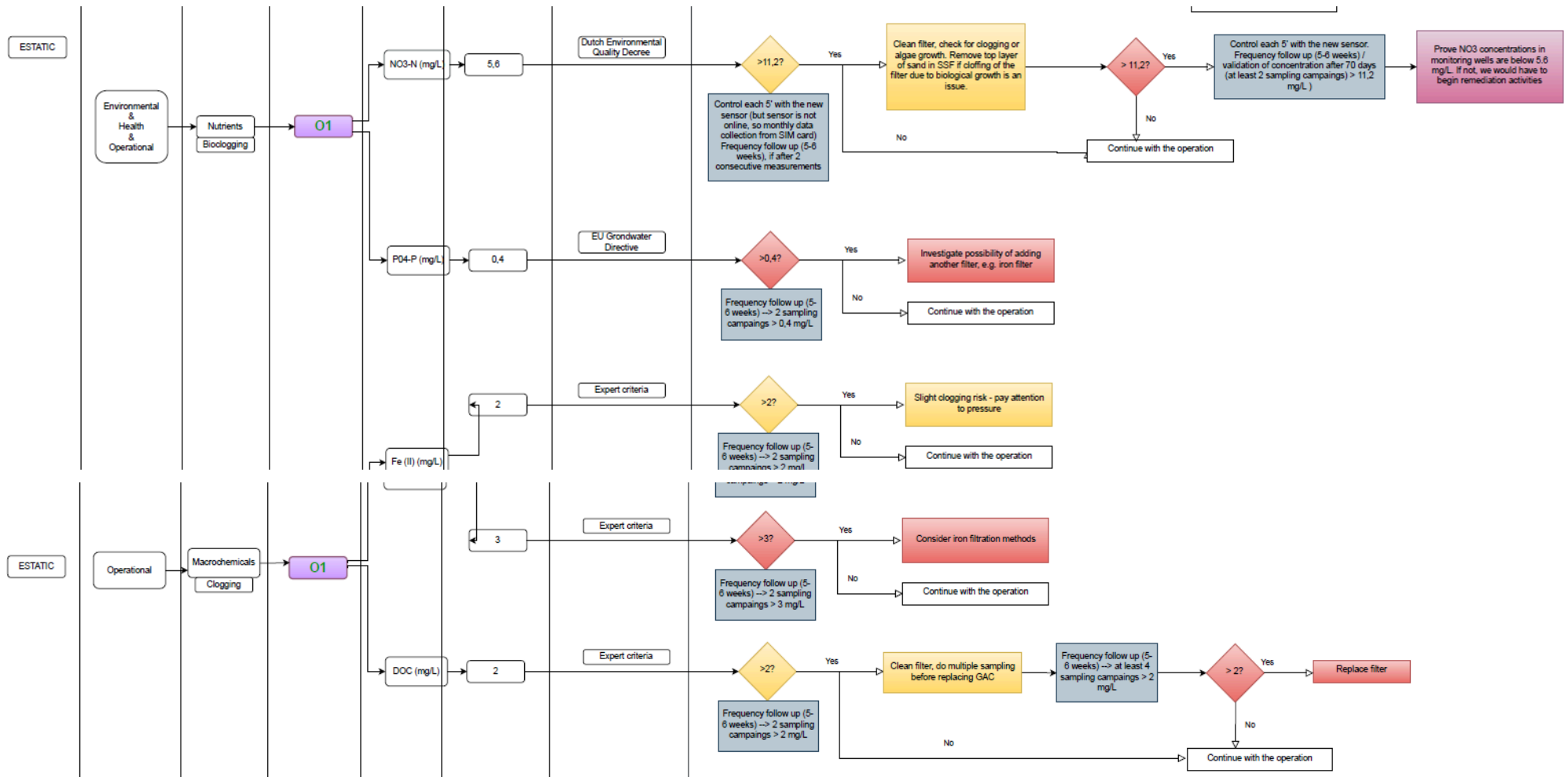


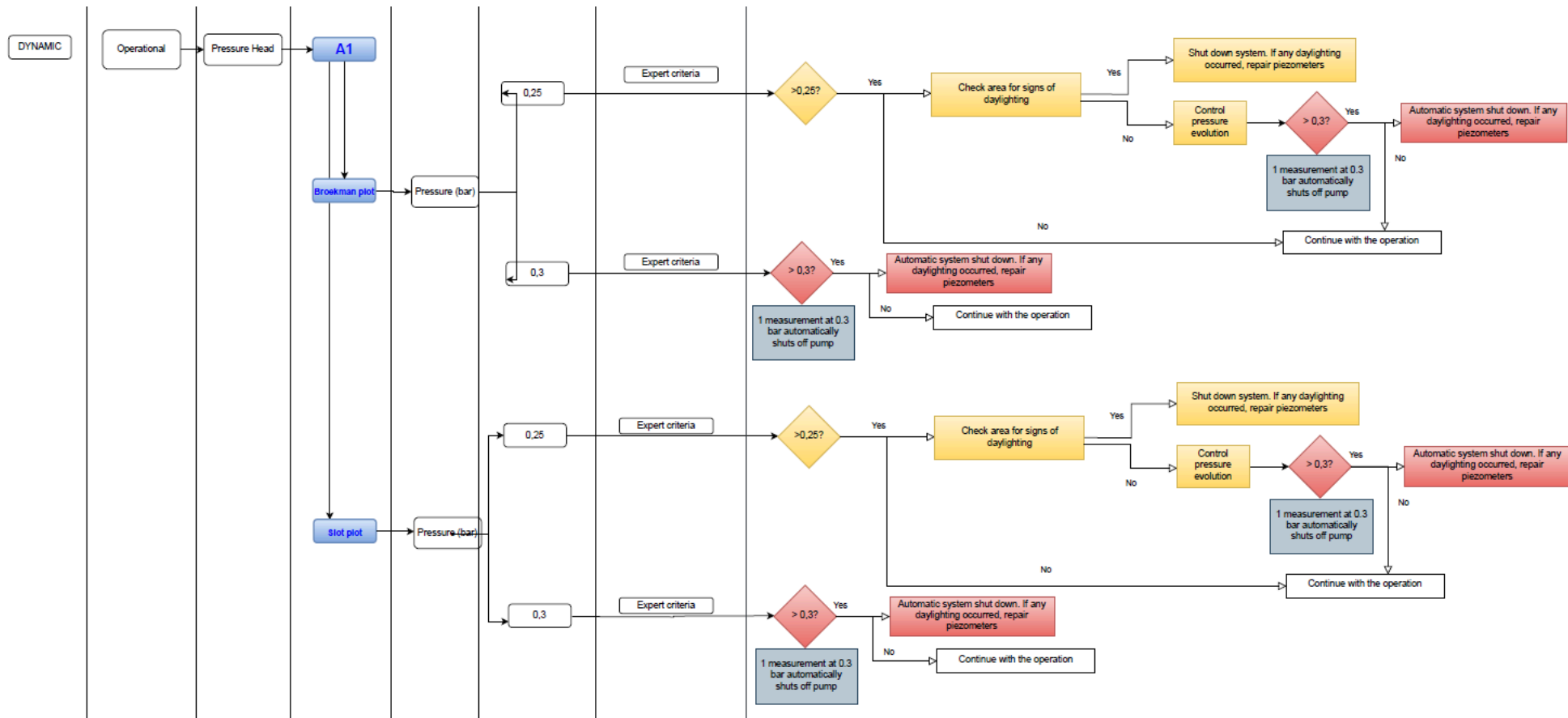


DS3 - Texel (The Netherlands) - MAR system Flowchart









MARCLAIMED will support decision makers to integrate Managed Aquifer Recharge (MAR) with Alternative Water Resources (AWR) in River Basin and Drought Management Plans supporting the adaptation and resilience of structural supply systems in the context of climate change. The main outcome will be the MARCLAIMED Integrated Decision Support Tool (IDST) that will provide solutions to tackle existing and emerging threats and support decision making and adaptation policies.