DIGIT4WATER
REPORT 01

WP1–T1.4. Definition of key water quality parameters for feeding ML models

0.1 Abstract

This is the first report of the DIGIT4WATER project. In this report, we document the outcomes of Work package (WP) 1, Task 4.1. Specifically, we define and describe the key parameters for assessing the efficiency of alternative tertiary treatment methods (UV-based and Solar Advanced Oxidation Processes (AOP)). These variables will be used for evaluation in WPs 2 and 3, as well as for inputting data into Machine Learning (ML) models within WP 4. The variables accordance with their characteristics are classified as inputs and outputs of the AOP treatments. Moreover, the input variables are divided into intrinsic variables of the wastewater and adjustable parameters for the treatment.

0.2 List of abbreviations

AOP Advanced Oxidation Processes

dWeW Diluted Well Water

FCWW Fresh Cheese Whey Waste Water

IW Isotonic Water

ML Machine Learning

OMC Organic Microcontaminant

PAA Peracetic Acid

PMS Peroxymonosulfate

PS Persulfate

SUWW Simulated Waste Water

TOC Total Organic Carbon

UV-C Ultraviolet-C

UWW Urban Waste Water

WeW Well Water

WP Work package

Contents

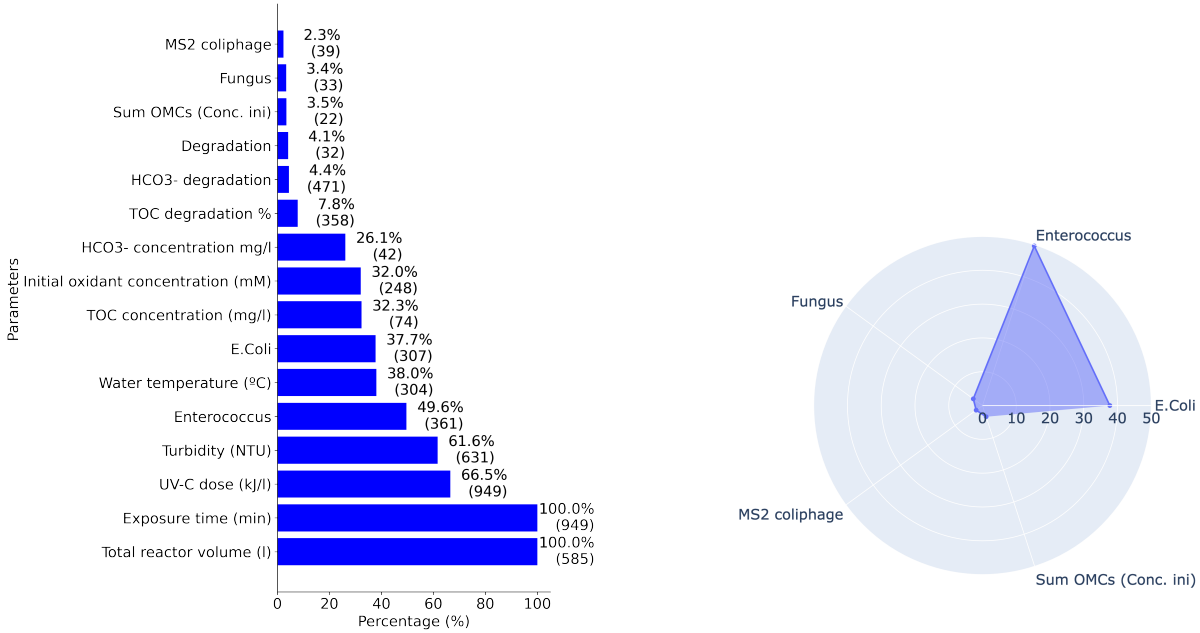
1 Introduction

In this report, we analyse two AOP databases, one incorporating the use of UV-C radiation and the other utilizing solar radiation, that include information on the treatment of diverse pollutants (e.g. *E. Coli, Enterococcus, Salmonella, Pseudomonas*, etc.). Moreover, we identify and extract the main parameters to implement the ML models and increase their robustness and versatility in addressing the multifaceted challenges of wastewater treatments.

2 Overview of AOPs databases and identification of their key parameters

2.1 AOPs based on UV-C radiation

The database of AOPs based on UV-C radiation contains experimental data on the disinfection of five pollutants: *E. coli, Enterococcus, Fungus, MS2 Coliphage,* and *Sum Organic Microcontaminants (OMCs)*. Figure 1 shows the samples and percentage of the main parameters of the database of AOPs based on UV-C radiation. From Figure 1, we can notice that the predominant pollutants are: *Enterococcus* and *E. Coli* with 49.6% and 37.7% of the total of the samples, respectively.



(a) Bar graph depicting the main parameters by both percentage and number of samples.

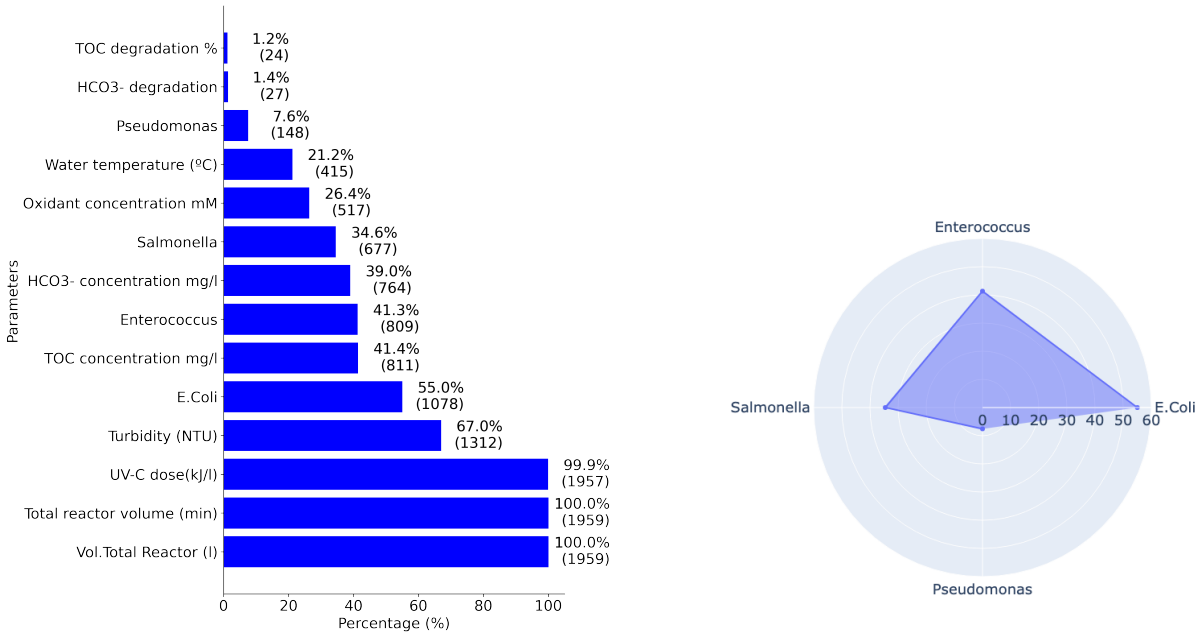
(b) Radar chart illustrating the distribution of pollutants by percentage.

Figure 1: Graphically representation of the database of AOPs–UV-C categorized by main parameters.

2.2 AOPs based on solar radiation

Similarly, The database of AOPs based on solar radiation contains experimental data on the disinfection of four pollutants: *E. coli, Enterococcus, Salmonella,* and *Pseudomonas*. Figure 2 shows the samples and percentage of the main parameters of the database of AOPs based on Solar radiation. From Figure 2, we can notice that

the predominant pollutants are: *E. Coli*, *Enterococcus* and *Salmonella* with 55%, 41.3%, and 37.7% of the total of the samples, respectively.



(a) Bar graph depicting the main parameters by both percentage and number of samples.

(b) Radar chart illustrating the distribution of pollutants by percentage.

Figure 2: Graphically representation of the database of AOPs–Solar categorized by main parameters.

Figure 3 shows all the parameters used in the treatment of wastewater based on the experimental data presented in both AOP databases, UV-C and Solar, respectively. In the following subsection, we enlist and categorize each of them as input or output.

2.3 Input parameters

The input parameters from the point of view of wastewater treatments could be classified into two categories, intrinsic parameters of wastewater and adjustable parameters for treatment.

Intrinsic parameters of wastewater are parameters of the initial phase of the experiments and described characteristics of the wastewater to use. Below, we enlist and describe each of them.

- **Turbidity** is a measure of the cloudiness or haziness of a fluid caused by large numbers of individual particles. In wastewater, turbidity is often an indicator of the presence of suspended solids, such as silt, clay, organic matter, and other particulate materials.
- **Inital temperature** refers to the temperature of the wastewater at the start of a treatment or monitoring process.
- **Initial concentration of Total Organic Carbon (TOC)** represents the amount of organic carbon compounds present in the water at the start of a treatment or analysis.
- **Initial concentration of HCO3-**. Bicarbonate ions are a component of the alkalinity of water and can influence the pH and buffering capacity of the wastewater.

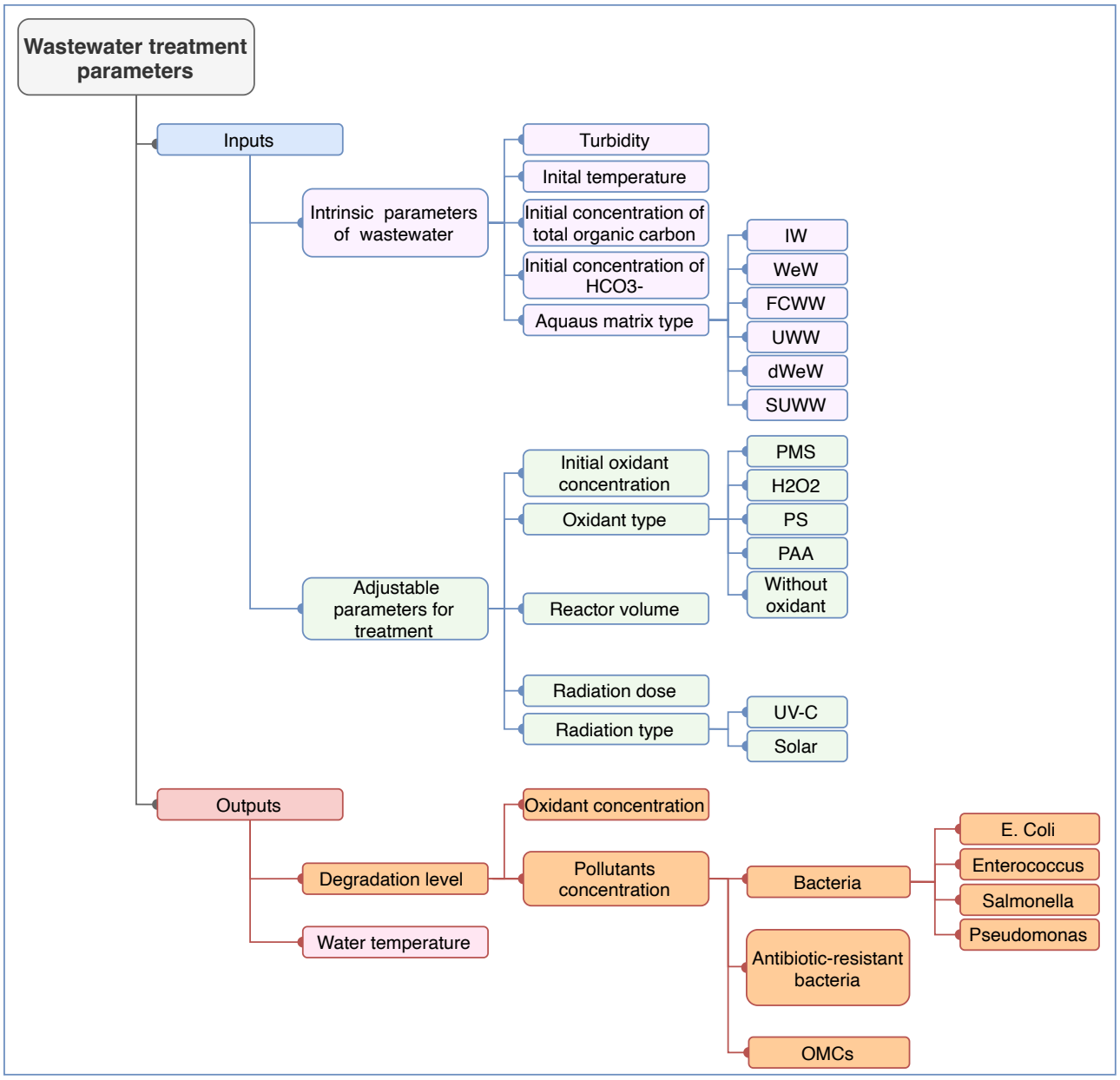


Figure 3: Parameters used in the treatment of wastewater based on AOP

- **Aquaus matrix type** refers to the composition and nature of the water sample being tested. It can include information about the source of the wastewater (e.g., industrial, municipal), the presence of specific contaminants or constituents, and any unique characteristics of the water that might affect the analysis or treatment process.
  - Isotonic Water (IW)[(Agua destilada + 9 g/L NaCl)]
  - Well Water (WeW)
  - Fresh Cheese Whey Waste Water (FCWW)[Agua simulada de cuarta gama]

  - Urban Waste Water (UWW) [Agua real de salida secundario de EDAR]
  - Diluted Well Water (dWeW)
  - Simulated Waste Water (SUWW) [Agua simulada de salida de secundario EDAR]
  - MQ
  - SW
  - UWW-AS
  - UWW-MBBR
  - UWW-CF
  - W

Adjustable parameters for treatment corresponds to those parameters that can be adjusted for each different treatment performed. Below, we enlist each of them.

- **Initial oxidation concentration**
- **Oxidant type**
  - Peroxymonosulfate (PMS)
  - Hydrogen Peroxide (H_2O_2)
  - Persulfate (PS)
  - Peracetic Acid (PAA)
  - Without oxidant
- **Reactor volume**
- **Radiation dose**
- **Radiation type**
  - Ultraviolet-C (UV-C)
  - Solar

2.4 Output parameters

- **Degradation level**
  - Oxidant concentration
  - Pollutants concentration
    - Bacteria
      - E. Coli
      - Enterococcus
      - Salmonella
      - Pseudomonas
    - Antibiotic-resistant bacteria
    - Organic Microcontaminants (OMCs)
- **Water temperature**

2.5 Key ML inputs/output parameters used in the treatment of wastewater

Based on the classification of the parameters presented in Figure 3, we selected the parameters useful for training and testing the ML models. Figure 4 displays the ML input/output parameters used in wastewater treatment. Additionally, Table 1 presents the types and key observations related to the parameters selected from the ML perspective.

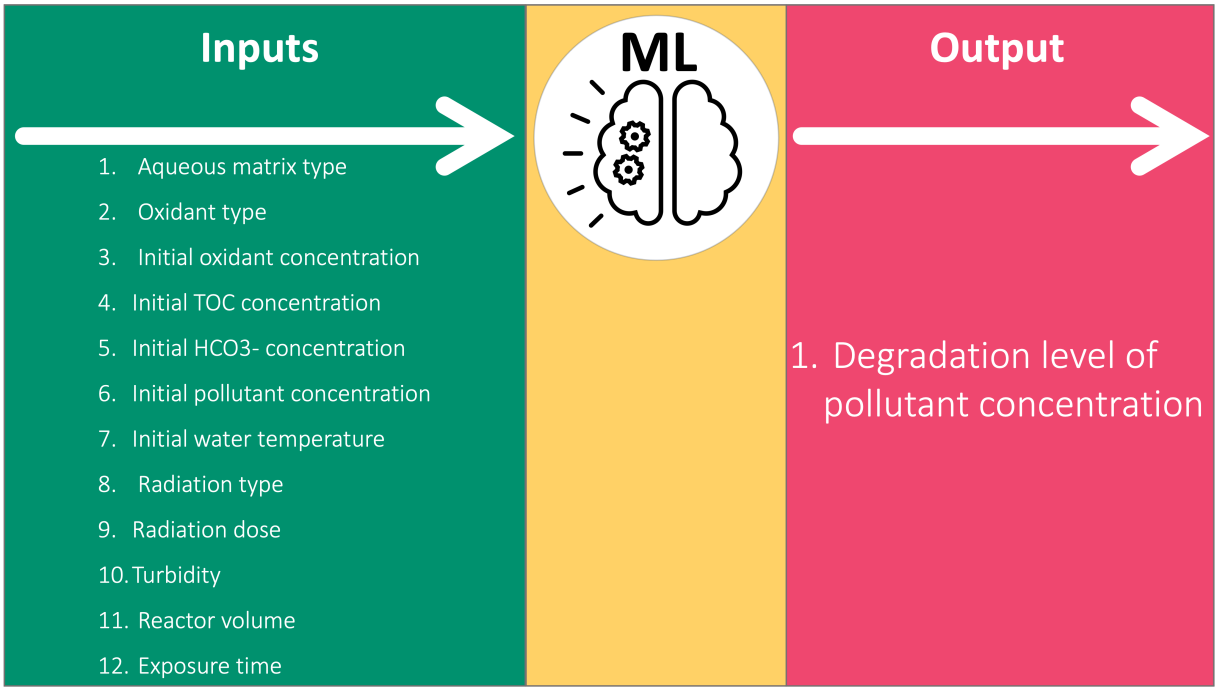


Figure 4: ML inputs/output parameters used in the treatment of wastewater

Table 1: Inputs and output considered for the ML model

Input/Output	Parameters	variable type	Observations
Input	Aqueous matrix type	Categorical	Solar: IW, FCWW; SUWW, UWW, dWeW, WeW, W UV-C: UWW, SUWW, IW, WeW, SW, MQ, UWW-CF, UWW-AS, dWeW, UWW-MBBR
	Oxidant type	Categorical	Both: PMS, H2O2, PS, PA, N/S or N/A
	Initial oxidant concentration	Numeric	Solar: $0.001 \geq$ value ≤ 21.896 UV-C: $0 \geq$ value ≤ 7.34
	Initial TOC concentration	Numeric	Both: "< 0.5" replaces by 0.5
	Initial HCO3- concentration	Numeric	Both: "< 0.5" replaces by 0.5
	Initial pollutant concentration	Numeric	Both: depend of the specific pollutant
	Initial water temperature	Numeric	Solar: $8.7 \geq$ value ≤ 50.2 UV-C: $12 \geq$ value ≤ 34.2
	Radiation type	Categorical	Solar: Solar UV-C: UV-C
	Turbidity	Numeric	Both: "< 0.5" replaces by 0.5
	Reactor volume	Numeric	Solar: $0 \geq$ value ≤ 90 UV-C: $0.3 \geq$ value ≤ 80
	Exposure time	Numeric	Solar: $0 \geq$ value ≤ 240 UV-C: $0 \geq$ value ≤ 180
Output	Degradation level of pollutant concentration	Numeric	Both: depend of the specific pollutant