UNIVERSITÉ DU LUXEMBOURG

PhD-FSTC-2018-54

The Faculty of Sciences, Technology and Communication

DISSERTATION

Defence held on 09/07/2018 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN SCIENCE DE L'INGÉNIEUR

by

Dario TORREGROSSA
Born on 19.04.1984 in Palermo (Italy)

A decision support system for energy saving in Waste Water Treatment Plants

Dissertation thesis committee

Dr Joachim Hansen
Professor, Université du Luxembourg

Dr Holger Voos
Professor, Université du Luxembourg

Dr Francesc Hernandez Sancho
Professor, University of Valencia (SPAIN)

Dr Georges Schutz
RTC4Water

Dr Gerd Kolish
Wupperverband

Declaration of Authorship

I, Dario TORREGROSSA, declare that this thesis titled, "A decision support system for energy saving in Waste Water Treatment Plants" and the work presented in it are my own. I confirm that:

- the content of this thesis is the result of the EdWARDS project, financed by FNR (National Research Fund of Luxembourg);
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- In this thesis, I have used the content of my original work produced in the framework of EdWARDS and already published in peer review journals. In particular, (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018);
- all the content of the thesis, the methodology, the software are the results of an original and autonomous work of the candidate.

Signed:

Date:

"I sincerely thank Dr. Alex Cornelissen, Dr. Georges Schutz and Dr. Prof. Joachim Hansen for having faith in my abilities since the writing of the PhD project proposal during my internship in Tudor (now LIST). I sincerely thank Dr. Cornelissen, Dr. Schutz, Dr. Prof. Hansen, Dr. Prof. Hernandez and Mr Ulrich Leopold for their supervision: thanks for the technical support, for the trust, for the large autonomy I had, and for the human support. Thanks to Dr. Prof.Hansen, Dr. Schutz and Dr. Prof. Holger Voos for their time as members of the CET committee. Thanks to Dr. Prof. Hansen, Dr. Schutz, Dr. Prof. Voos, Dr. Prof. Hernandez and Dr. Gerd Kolisch for their work as the members of review committee of my thesis. Thanks also to all the organizations involved in the INNERS project that contributed to data collection: LIST, the SIDEN, Waterschap Groot Salland (Now: Drents Overijsselse Delta), Waterschap Vallei en Veluwe and the Wupperverband.

I gratefully thank all the people that worked behind the scene, such us the doctoral school (in particular Dr. Andreas Bladt and Anke Zwetsch), the human resource staff and all the LIST administration. Many thanks also for the anonymous reviewers of my papers and to the journal editors.

I want also to mention the people that contributed to my previous work experiences in which I improved my skills in time management, organization, and conflict management.

A special thanks to all my colleagues that supported me during these years and that shared with me the work-life.

*Thanks to my big family: my father, my brother, my mother and all my Sicilian family. And thanks to my new international family, with my partner Marine and my daughter **Irma**, born few days before my thesis dissertation. "*

Dario Torregrossa

UNIVERSITY OF LUXEMBOURG

Abstract

Civil Engineer
Faculty of Science, Technology and Communication

Doctor of Philosophy

A decision support system for energy saving in Waste Water Treatment Plants

by Dario TORREGROSSA

Waste Water Treatment Plants (WWTPs) are complex facilities, in which an efficient energy management can produce relevant benefits for the environment and the economy. Today, big data can be used for a more efficient plant management, enabling high-frequency assessment and ultimately a more efficient use of resources. In order to achieve this, a computer-based support is necessary to analyse the enormous amount of data that WWTP sensors can produce. When this PhD project started, the literature review showed that, in the WWTP domain, the few available decision support systems (DSSs) were promising but still with large room for improvements; in fact, these tools were plant-specific, focussed mainly on process parameters and (most of them) working with low-frequency aggregated data (yearly data). This thesis instead proposes a cooperative decision support system called Shared Knowledge Decision Support System (SK-DSS).
SK-DSS is plant generic, i.e. able to simultaneously work with many WWTPs and based on key performance indicators. SK-DSS analyses the processes occurring in the plants and provide case-based solutions. Moreover, this DSS provides a platform to enable the plant managers to exchange information and cooperate. This thesis proposes the model of SK-DSS, a web-application, and applications to improve the energy performance of pump, blowers and biogas.

Acknowledgements

I gratefully acknowledge the financial support of the National Research Fund (FNR) in Luxembourg (7871388-EdWARDS) and additional funding from the Luxembourg Institute of Science and Technology.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

LEGAL DISCLAIMER: The present chapter partially reproduces research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of the EdWARDS project.

1.1 An overview of the topic

1.1.1 Energy and Waste Water Treatment Plants

In the last 20 years, an increasing interest in energy saving in the wastewater treatment plant domain can be observed from the number of available publications (fig. 1.1). Various studies demonstrated that WWTPs are voracious energy consumers and there is a relevant energy saving and production potential (Goldstein and Smith, (2002), Wett, Buchauer, and Fimml, (2007), Hernandez-Sancho et al., (2009), Mizuta and Shimada, (2010), Becker and Hansen, (2013), Krampe, (2013), Gude, (2015), Rehman et al., (2015), Castellet and Molinos-Senante, (2016), and Torregrossa et al., (2016)). For example, Goldstein and Smith, (2002) claim that WWTPs and the water sector consume 4% of the electric energy consumption in US. Torregrossa et al., (2016) show that in Europe the electric energy consumed by WWTPs corresponds approximately to the total energy consumption of a country like Serbia (27 TWh/year). In

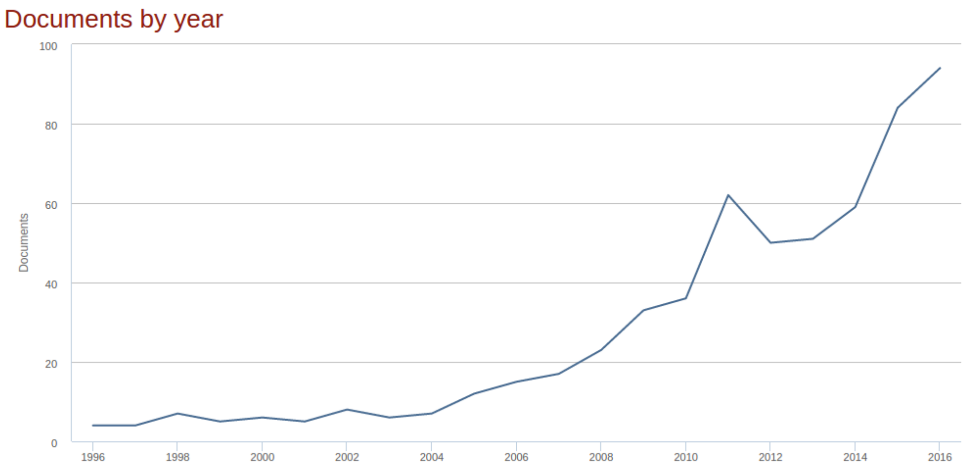


FIGURE 1.1: Number of publication in WWTP and energy. (Scopus)

Europe, different country-specific studies demonstrated that the electrical energy consumption contributes to the national energy consumption in a range between 1% and 4% of the total (Reinders et al., 2012; Foladori, Vaccari, and Vitali, 2015; Longo et al., 2016). Panepinto et al., (2016) show that the energy cost in WWTPs corresponds up to the 40% of the operational costs.

Despite the use of various datasets and methodologies, studies concerning the WWTP energy efficiency agree on a specific point: **the WWTPs are generally not efficient and a relevant energy saving potential can be exploited**. Hernández-Sancho and Sala-Garrido, (2009) and Hernández-Sancho, Molinos-Senante, and Sala-Garrido, (2011) applied data envelopment analysis to show that the small WWTPs in Spain can increase their input-output efficiency by 57%; Castellet and Molinos-Senante, (2016) used a non-radial weighted slack-based approach to demonstrate that in Spain the WWTP energy consumption can be decreased by 25%. Fig. 1.2, adapted from Gu et al., (2017), shows that the specific energy consumption[1] could vary between 0.21 kWh/m^3 and 0.34 kWh/m^3 depending on the technology. The INNERS project [2] has shown that the energy consumption in North West Europe, can be reduced by 3.14 TWh/year (INNERS, 2015). Moreover, in the energy balance, the biogas production plays a relevant role since it can cover between 40%-60% of global energy consumption of WWTPs (Hansen, 2018).

[1]expressed as energy consumption per volume of waste water treated (kWh/m^3)
[2]INTERREG NWE project 'INNERS'

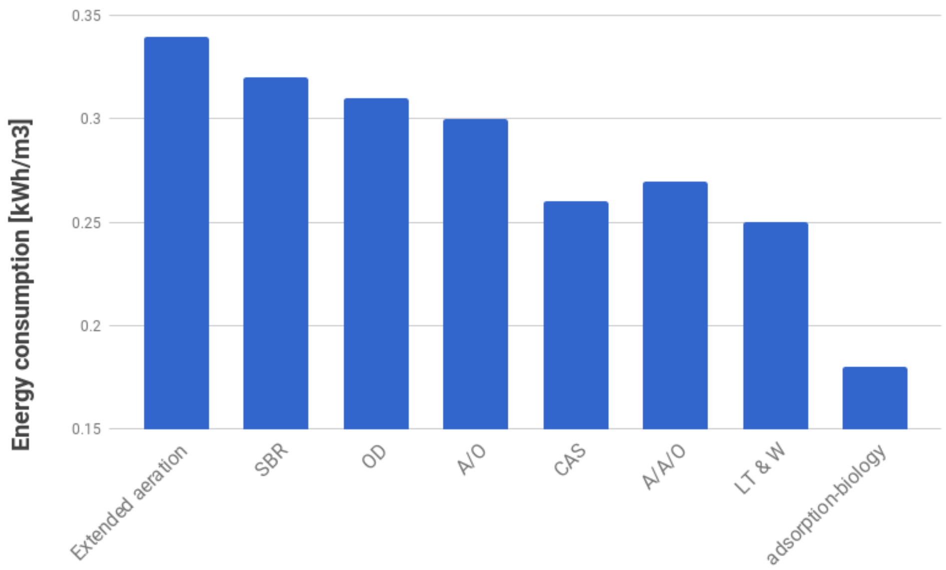


FIGURE 1.2: Energy consumption in WWTPs with different technologies. Data calculated from 599 Chinese WWTPs in 2006, Source: adapted from (Gu et al., 2017)
SBR=sequencing batch reactors, OD=oxidation ditch, A/O=anoxic–oxic systems, CAS=conventional activated sludge, AAO=anaerobic-anoxic-oxic, LT&W= land-treatment and wetlands.

1.1.2 Approaches to energy saving

This thesis proposes three categories to classify the approaches for energy saving in WWTPs:

1. first category: approaches based on technology optimization;
2. second category: approaches based on process assessment and strategic planning;
3. third category: approaches based on process management and decision support.

The **first category** includes all the publications which aim to increase the energy efficiency in WWTPs by the development of new technological solutions. Pretel et al., (2015), for example, demonstrated that it is possible to recover 0.1 kWh/m^3 from the WWTP process by using an anaerobic membrane bio-reactor (AnMBR). Campo et al., (2017) improved the energy production of biogas by introducing intermediate lysis treatments in the traditional plant layout.

Various approaches, belonging to the **second category**, contributed to the energy efficiency of WWTPs by comparing the best technologies and identifying the most relevant issues in WWTP energy management. For example, Friedler and Pisanty, (2006) focussed on the impacts of design parameters. Sala-Garrido, Hernández-Sancho, and Molinos-Senante, (2012) studied how the seasonality affects the efficiency of WWTPs, Ng et al., (2014) assessed WWTPs using a life-cycle management approach. Hernández-Sancho and Sala-Garrido, (2009), Hernández-Sancho, Molinos-Senante, and Sala-Garrido, (2011), Castellet and Molinos-Senante, (2016) proposed econometric approaches to evaluate the efficiency of WWTPs. Djukic et al., (2016) and Pablo-Romero et al., (2017) investigated the role of tariffs in the energy balance of WWTPs. The **third category** consists of research works that aim to improve the energy efficiency in WWTPs by monitoring and controlling operational parameters. This category contains benchmark approaches, expert systems, and decision support systems (DSSs). Benchmark approaches identify reference values for plant process parameters and make a comparison between desired values and measured one (more details in section 2.2.2). Expert systems monitor parameters and automatically react to undesired conditions. Baeza, Gabriel, and Lafuente, (1999), Ruano et al., (2010), Angulo et al., (2012), and Alex et al., (2015) demonstrated how expert systems can be used for the automatic control of plant processes.

Decision support systems monitor the parameter and assist plant operators in the decision making process. Krampe, (2013), INNERS, (2015), Lorenzo-Toja et al., (2016), and Doherty, (2017) propose approaches for WWTP benchmarking, showing interest in energy and eco-efficiency.

Decision support systems can further be classified according to their specific aim; Comas et al., (2004a), Gómez-López et al., (2009), Garrido-Baserba et al., (2015), Tomei et al., (2016), and Kalbar, Karmakar, and Asolekar, (2016) propose decision support system for the design of WWTPs. On the other hand, Paraskevas, Pantelakis, and Lekkas, (1999), Poch et al., (2004), Guerrero et al., (2011), Gibert, Conti, and Vrecko, (2012), Hakanen, Sahlstedt, and Miettinen, (2013), and Gisi et al., (2015) propose decision support tools for the management and upgrading of existing plants.

According to this classification, the present thesis belongs to the

third category: this works contributes to the development of a decision support system for the energy management of WWTPs.

1.2 Gaps in the literature

An extensive literature review, initially performed during the first six months of the PhD period and updated several times since then, showed room for improvements discussed in the following sub-chapters (1.2.1 → 1.2.4).

1.2.1 Plant specificity

In the WWTP domain, a common limitation of decision support systems (DSSs) was their plant specificity. In fact, DSSs are often built to serve a specific plant, with well-defined inputs, layouts and specific targets. For example, Corominas et al., (2008) proposed a decision support system to improve the management efficiency in the WWTP of Vall del Ges (Spain). De Gussem et al., (2011) proposed an LCA-based approach to minimize the operational costs of the WWTP of Bree, Belgium. These approaches were plant specific, i.e. they are designed to work with a specific plant (specific load, specific inflow, specific layouts, etc...). Instead, in (INNERS, 2015; Torregrossa et al., 2016), the author of this thesis and the researchers of INNERS demonstrated the advantages of using key performance indicators (KPIs) for the plant generic benchmarking of WWTPs; in fact, the use of KPIs enables the comparison between distinct and different plants and makes it possible to monitor them simultaneously. Decision support systems for energy efficiency in WWTPs, based on KPIs, were not found in literature.

1.2.2 Use of large database

Nowadays, a plant generic approach is a relevant advantage because of the availability of large databases containing data of hundreds of WWTPs. This kind of database starts to become popular; for example in the Netherlands, the Z-info database (managed by Croonwolter&dros) collects information from more than 300 WWTPs. Also in Spain, the Entitat de Sanejament d'Aigües (EPSAR) stores data for almost 200 WWTPs (Hernández-Sancho, Molinos-Senante, and Sala-Garrido, 2011). This data is generally collected, aggregated and analysed off-line, with a considerable time-delay between the collection and its assessment. Automatic, on-line assessment of high-frequency data is rare.

1.2.3 Specific applications for daily monitoring of devices

The literature review also showed a lack of decision support systems that consider simultaneously the operation of the most important WWTP energy consumers and perform a daily analysis. According to the current practice, the analysis of energy performance of WWTPs is executed once or twice per year with aggregated values (Torregrossa et al., 2016). This is due to the large amount of data that a plant operator cannot constantly manage without computer assistance. For instance, the sensors installed in the WWTP of Soligen-Burg produce almost 300.000 records per day (Torregrossa et al., 2016). The use of aggregated (i.e. by average or sum) data has a big disadvantage: the information about the dynamic of process parameters and the interaction between parameters is lost. For example, with yearly averaged values it is not possible to link the efficiency of pumps to the inflow conditions (an extensive investigation can be found in (Torregrossa et al., 2017b; Torregrossa et al., 2017d; Torregrossa et al., 2017c) .

1.2.4 Cooperative DSS

Moreover, recent works demonstrated that cooperative decision support systems are not used in the WWTP domain (Poch et al., 2004; Poch et al., 2014). The use and the success of cooperative platforms (as Dropbox, researchgate, linkedin and GoogleDrive) suggest that people have the predisposition to cooperate trough technological platforms. Moreover, in the WWTP domain, problems are generally recurring and the sharing of knowledge and solutions in a cooperative platform could be beneficial for plant managers.

In short, in the literature, at the time of the beginning of the EdWARDS project, there was a lack of decision support systems with the following four characteristics:

- specifically focussed on energy management of WWTPs;

- plant generic, i.e. able to simultaneously analyse multiple WWTPs;
- able to produce daily plant analysis reports using high-frequency information;
- provide plant managers with case-based suggestions for energy efficiency.

1.3 Research questions

The research questions of this thesis aim to address the literature gaps discussed in section 1.2. The starting point is the research question posed in the submission document to the National Research Fund of Luxembourg (FNR):

"Is it possible to develop a methodology that, based on benchmarking of on-line data from different WWTPs, enriched with expert knowledge, will be able to support a decision process aiming to increase the energy efficiency of WWTPs? Is it possible to apply this methodology to multiple plants simultaneously and at the same time to provide case-sensitive targeted advice?" (*FNR Application 7871388, 2014-03-20*).

In the strategic development of the project, additional elements to this research question were added:

- Is it possible to develop strategies to monitor, benchmark and provide decision support for the management of the most important energy consumers?
- Is it possible to monitor, benchmark and provide decision support for the optimization of biogas production in WWTPs?
- Is it possible to provide a plant generic DSS for energy saving in WWTP, able to combine case-based plant assessment with a cooperative platform to stimulate cooperation between plant managers?
- Is it possible to develop a methodology to estimate missing data for the daily calculation of key-performance indicators?

1.4 Seven specific objectives

During the development of this work, the research questions were reorganized in the following specific objectives:

1. identify a global model able to get information from several WWTPs, normalize the information, identify the operational scenarios and provide case-based solutions with daily frequency;
2. identify a methodology to normalize WWTP data;
3. identify a methodology to estimate missing data necessary for the daily KPI calculation;
4. identify a methodology to be applied to the main energy consumers (blowers and pumps);
5. identify a methodology for the biogas production in WWTPs;
6. identify a methodology to enable the cooperation between plant managers;
7. provide a user-friendly interface.

These objectives lead to identify some key features to be applied to the decision support system and discussed in the next section.

1.5 Key features of decision support system

The **first key feature** of this decision support system is its focus on the energy management of WWTPs. This is an important aspect, because an efficient energy management of WWTPs can play a relevant role in the strategy for the achievement of international environmental goals (Council of the European Union, 1991; Kallis and Butler, 2001; IPCC, 2007; Hernández-Sancho, Molinos-Senante, and Sala-Garrido, 2011; United States Environmental Protection Agency (USEPA), 2012). Moreover, in the last decades, the WWTP energy consumption increased because of different reasons, for example:

- an increase in effluent quality standards (Nakagawa et al., 2006; Naidu et al., 2016);

- emerging new contaminants (Petrie, Barden, and Kasprzyk-Hordern, 2014)

- an increase in connected population (Melorose, Perroy, and Careas, 2015)

Furthermore, Ko et al., (2003), Molinos-Senante, Hernández-Sancho, and Sala-Garrido, (2010), and Hernández-Sancho, Molinos-Senante, and Sala-Garrido, (2011) made the point that the energy management of WWTPs is a relevant economic issue.

This decision support system must be equipped with a set of tools and methodologies able to take into account the most important elements in the WWTP energy balance (**second key feature**). In the framework of this thesis, the most important plant energy consumers have been taken into consideration: the aeration systems (accounting for the 60% of total energy consumption) and the pump systems (approximately 12%) (Gu et al., 2017). Moreover, the decision support system has to include a methodology to monitor the biogas energy production and provide suggestions for its optimization. The optimization of biogas production can substantially contribute to reduce the energy demand of WWTPs and let them to interact with the electric grid as energy producers (Becker and Hansen, 2013; Venkatesh and Elmi, 2013; Metcalf and Eddy, 2014; Abuşoğlu et al., 2016).

This decision support system must be plant-generic (**third key feature**), i.e. able to manage simultaneously different WWTPs. This requires a system able to work with the information commonly available in WWTPs and able to calculate and process KPIs. Moreover, such a system should be able to estimate missing parameters and normalize the information (unit of measurement, nomenclature, Torregrossa et al., (2016)).

According to the **fourth key feature**, the proposed decision support system is required to exploit the potential of cooperative platforms. This feature, enabling the plant operators to cooperate and exchange information, must automatically enlarge and update the expert knowledge processed by the system. The idea at the basis of this aspect is that knowledge and experience can be shared and reused when a specific problem occurs for the second time. This approach, successfully adopted in medical decision support (Hudson and Cohen, 2012), at the beginning of this PhD project, was a novelty in the water domain.

With respect to the **fifth key feature**, this decision support system must produce daily assessment of the WWTP facilities. The current practise consists of plant assessments performed a few times per year, with aggregated data and consequent loss of information. This approach can be improved by adding on-line, high-frequency analysis of plant parameters; in fact, increasing the frequency of the analysis enables the reduction of the response time to failure, reduce inefficiencies and allows the retrieval of information about the impact of dynamic parameters. For example, in chapter 5, it is shown how the daily variation of inflow has an impact on the energy consumption of the pumps.

The achievement of these targets should guarantee a high practical tool for plant managers.

1.6 Structure of the thesis

The thesis is composed of several chapters.

In the current chapter, the identification of the literature gaps and the definition of the objectives have been developed(1.1 $\rightarrow$ 1.5).

Chapter 2 reports general information about the working principles of WWTPs, energy issues, decision support system, random forest and the fuzzy logic algorithms. The reader can use chapter 2 to identify the key-concepts and become familiar with the methodologies later included in the decision support system.

Chapter 3 presents the model of the proposed decision support system. In particular, it is explained how this tool works and meets the objectives declared in section 1.4.

Chapter 4 explains how data is acquired from WWTPs and how the system deal with the most important data-related issues: nomenclature normalization, data validation, estimation of missing data and uncertainty management.

In chapter 5, it is explained how fuzzy logic can be combined with KPI calculation and benchmarking to produce a case-scenario analysis. Moreover, in this chapter, it is explained the process of sharing expert knowledge and how this information set can be stored, validated and continuously updated. Chapter 5 presents the proposed decision support system applied to some real problems:

- pump energy optimization;

- blower energy optimization;
- biogas production.

A user-friendly web-interface is proposed in chapter 6, that provides an overview about an interface prototype, in order to enable the reader to experience how the proposed methodology can be adopted by end-users.

Chapter 7 resumes the proposed methodologies and compares the achievement of the EdWARDS project with the original objectives.

Finally, chapter 8 discusses ideas and new research questions that were generated during the development of this PhD but not yet developed because of time limitations.

Chapter 2

State of the Art

LEGAL DISCLAIMER: The present chapter partially reproduces research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of the EdWARDS project.

This chapter provides the reader with an introduction in the main areas of interest covered in this thesis. Figure 2.1 clarifies the organization of this chapter and its relation to the Shared Knowledge decision support system. In particular, this chapter deals with:

- technical aspects of WWTPs; the section 2.1 provides information about wastewater characterisation, European legislation and the most common WWTP technologies;
- energy aspect in WWTPs (section 2.2), including the WWTP energy consumption, the potential energy savings and the benchmark approach;
- decision support system technology; in section 2.3, the concept of decision support system with the main theoretical definitions is provided. Advantages and disadvantages are discussed and particular attention is given to environmental decision support systems. Moreover, a literature review on decision support systems applied to WWTP management is presented;

- fuzzy logic methodology (section 2.5); in this section, the fuzzy logic methodology is presented, together with the mathematical definition and with a simple numerical example to introduce the reader to fuzzy logic reasoning;
- an introduction to random forest technology (section 2.6).

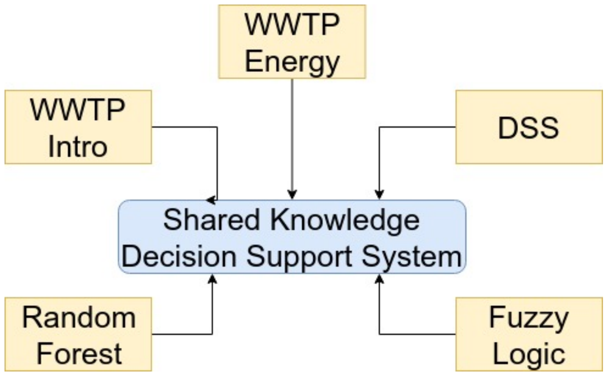


FIGURE 2.1: Logical organization of the chapter

The decision support system proposed in this thesis uses some key concepts extracted from the above-mentioned areas of interest.

2.1 WWTPs: working principles and possible configuration

In this section, the following topics are discussed: the legislation concerning WWTPs, the quality and the sources of wastewater, the technological solutions to accomplish with the legislation and the new challenges for an optimal management.

2.1.1 Definition of wastewater and characterization

The Council of the European Union, (1991) provided the following definitions with regards to the topic of wastewater treatment:

- "**'urban waste water'** means domestic waste water or the mixture of domestic waste water with industrial waste water and/or run-off rain water;"

- "**'domestic waste water'** means waste water from residential settlements and services which originates predominantly from the human metabolism and from household activities;"
- "**'industrial waste water'** means any waste water which is discharged from premises used for carrying on any trade or industry, other than domestic waste water and run-off rain water;"
- "**'1 p.e. (population equivalent)'** means the organic biodegradable load having a five-day biochemical oxygen demand (BOD5) of 60 g of oxygen per day." In scientific literature, the population equivalent can be expressed as 'p.e.', 'pe', or 'PE';
- "**'agglomeration'** means an area where the population and/or economic activities are sufficiently concentrated for urban waste water to be collected and conducted to an urban waste water treatment plant or to a final discharge point".

In this thesis, wastewater is used as a synonymous of urban waste water. As in the previous definitions, wastewater is the undesired output of human activities and rain water. The human activities can be domestic (metabolism and household activities) or industrial. Consequently, the quality and the quantity of wastewater produced by human agglomerations greatly vary depending on the mix of economic activities, domestic metabolism, and climatic conditions. The most influencing factors that affect the wastewater characteristics are:

- the connected population;
- the mix of activities in the agglomerations;
- daily and seasonal patterns;
- climatic conditions.

The connected population

The European Environmental Agency, (*Overview of electricity production and use in Europe — European Environment Agency*) has shown that the population connected to WWTPs increased over the last 20 years. In North, Centre and Southern Europe the connected population is approximately 90% of the actual population count, while in the Eastern and the South-Eastern Europe this percentage corresponds to approximately 60%. In the period 1991-2008 the connected population increased and wastewater technology shifted from primary and secondary treatments towards the use of tertiary treatments (*Overview of electricity production and use in Europe — European Environment Agency*).

The connected population is linked to the pollutant load. Table 2.1 reports some pollutants specific loads (Hansen, 2018). These values are used to express the load in term of population equivalent (pe). For example, 100 kg of Chemical Oxygen Demand (COD) correspond to the daily production of 833 persons that produce 120gCOD/day; if a waste water treatment plant receives 10,000 KgCOD/day, the connected population equivalent is 83,330 inhabitants.

TABLE 2.1: Specific load of pollutants.

Short name	Full name	Specific Load inhabitant
COD	Chemical Oxygen Demand	120 g/pe/day
BOD	Biological Oxygen Demand	60 g/pe/day
P	Phosphourous	1.8-2 g/pe/day
N	Nitrogen	11 g/pe/day

The mix of activities in the agglomerations

In (Metcalf and Eddy, 2014), the water consumption for activity typology is reported. Table 2.2 and table 2.3 (data estimated for the United States) show a great variability in water consumption not only between the categories but also inside the same class of human activities. Moreover, fig. 2.2, extracted by (Hansen, 2018), shows a great variation in water consumption between different countries. Furthermore, the mix of activities in each agglomeration can greatly change. Consequently, the variability and the uncertainty in the definition of wastewater production is high.

Daily and seasonal patterns

Daily and seasonal patterns influence the quality and the quantity of wastewater. For example, during the night, the quantity of produced waste water is less than that produced during the day. Also, seasonality can affect the quality and the quantity of waste water (Sala-Garrido,

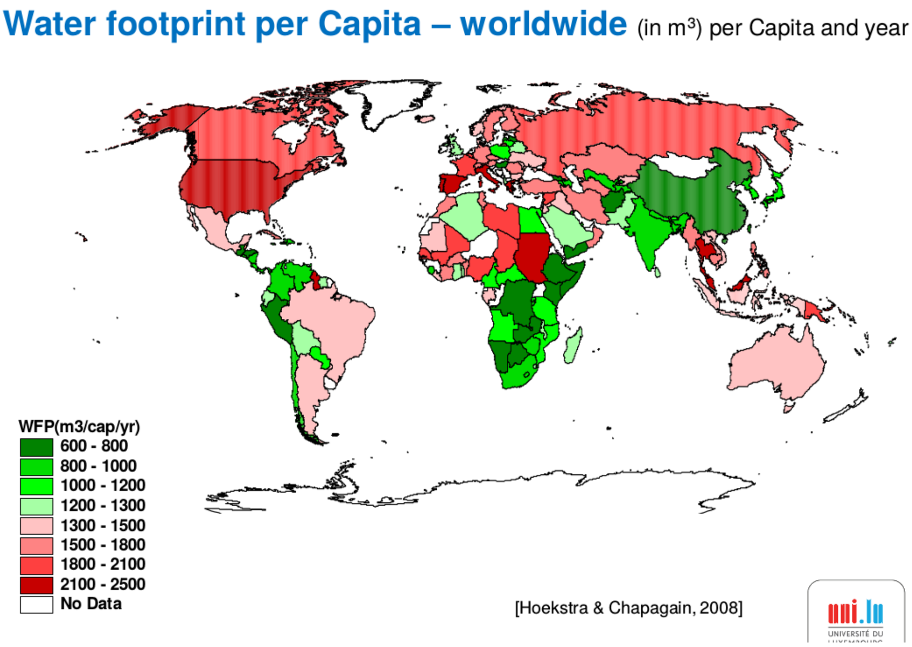


FIGURE 2.2: Water Footprint in the world. Extracted from (Hansen, 2018)

TABLE 2.2: Municipal uses of water and typical quantities in US. Extracted from (Metcalf and Eddy, 2014)

Use	Flow-rate [l/capita*d]
Domestic Indoor	150-300
Domestic Outdoor	60-340
Commercial	40-300
Public	15-25
Loss and Waste	60-100

Hernández-Sancho, and Molinos-Senante, 2012); for example, this happens especially in regions in which tourism that increases their population considerably during the summer.

Other elements

Weather conditions are unpredictable on the long term and these have a relevant impact on the operational conditions of WWTPs. Other elements can also play a role (for example changes in legislation). The contributions of all these factors together result in the large variability of wastewater characteristic observed.

Table 2.4 shows the large range in which the pollutant concentration may vary. Consequently, WWTP designers are required to produce

TABLE 2.3: Water consumption based on activity - typical values. Extracted from (Metcalf and Eddy, 2014)

Source	Unit	Flowrate L/unit*d
Hotel	Guest	200-215
Office	Employee	23-45
Restaurant	Customer	23-30
Shopping center	Employee	23-38

TABLE 2.4: Typical concentration values of various pollutants

Pollutant	Unit of measurement	Concentration at the inlet		
		low	medium	high
COD	mg/l	250	500	1000
BOD5	mg/l	110	220	400
TOC	mg/l	80	160	290
TSS	mg/l	145	300	525
N-tot	mg/l	20	40	85
P-tot	mg/l	4	8	15

COD= chemical oxygen demand, BOD= biological oxygen demand, TOC=Total organic carbon, TSS= Total suspended solids, N-tot= total Nitrogen, P-tot=total phosphorus

flexible facilities able to accomplish their tasks under a wide range of different operational conditions.

2.1.2 European Legislation and implementation report 2016

Despite the great variability in the quality and quantity of wastewater produced, the discharge requirements defined by EEC Council, (1991) are fixed (table 2.5). The comparison between typical concentration values (table 2.4) at the inlet and the discharge requirements (table 2.5) helps to better understand the operational requirements of WWTPs: they must be able to treat the wastewater produced by several sources, and reduce the highly variable pollution load under the law requirements. The directive establishes a minimum number of samples that must be analysed to verify the discharge requirements and the maximum number of samples allowed to fail the requirements for each year; the minimal number of samples varies according to the plant size. The

TABLE 2.5: Requirements for discharges - adapted from table 1 and 2 of (EEC Council, 1991)

Parameters	Concentration	Minimum reduction [%]
BOD5	25 mg/l	70-90
COD	125 mg/l	75
Total suspended solids	35 mg/l	90
Total phosphorus (10 – 100.000 PE)	2 mg/l	80
Total phosphorus (>100.000 PE)	1 mg/l	80
Total nitrogen(10 – 100.000 PE)	15 mg/l	70-80
Total nitrogen(>100.000 PE)	15 mg/l	70-80

European Commission, 2016, as required by Article 17 of (EEC Council, 1991), produces every 2 years a report to evaluate the adherence to the Water European Directive. The last report " covers 19,000 towns and cities ("agglomerations") above 2,000 inhabitants, generating a pollution corresponding to 495 million so called population-equivalents (p.e.)" (European Commission, 2016).

Despite the improvements, there are relevant gaps to be filled as resumed in the report edited by European Commission, 2016:

- "11 million p.e. (2%) have to be connected and treated;"
- "48 million p.e. (9%) of the urban waste water already connected have to meet the performance of a secondary treatment;"
- "39 million p.e. (12%) of the urban waste water already connected have to meet the performance of a more stringent treatment."

The sections 2.1.1 and 2.1.2 have illustrated, on one side, the wastewater quality variability and, on the other side, the demanding legislative and environmental requirements (EEC Council, 1991; European Parliament, 2000; European Parliament, 2006). The section 2.1.3 shortly reports the most important processes and configuration of WWTPs.

2.1.3 An overview of WWTP: layout and technologies

The main objective of WWTPs is to reduce the pollution load from wastewater in order to meet the discharge requirements imposed by legislation. To do this, WWTPs generally require a combination of chemical, physical and biological processes.

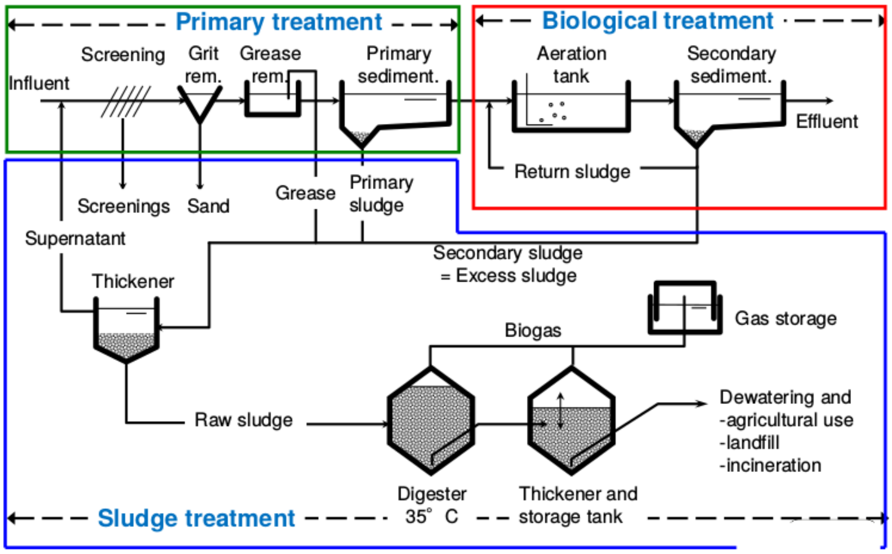


FIGURE 2.3: Conventional biological treatment. Figure extracted from (Hansen, 2018)

Various plant layouts and different treatments are proposed for different pollution removal targets. For example, fig. 2.3, extracted from (Hansen, 2018), shows a conventional wastewater treatment block layout. In this schema, it is possible to identify 3 main blocks:

- **primary treatments**, mainly based on physical processes. Screening, grit and grease removal process capture pollutants that can produce operational problems to the subsequent stages (such as rags, sticks, grease and floating parts). The primary sedimentation removes settleable solids and a part of the COD;

- **biological treatments** aim to remove the biodegradable fraction of the organic matter and it consists normally of an aerated treatment basin followed by a sedimentation tank although many alternative configurations are possible;

- **sludge treatments**; this process receives sludge from primary and secondary sedimentation with the aim to reduce the water content, stabilise and disinfect the sludge before disposal. Sludge stabilisation is a relevant aspect for energy balance; in fact, for plants with adequate capacity (>10 kPE (Hansen, 2018)) it is possible to operate an anaerobic stabilisation and produce biogas.

Out of the schema represented in fig. 2.3 it is necessary to consider the so-called 'advanced treatments' to remove contaminants of emerging concern, also often called 'micro-pollutants', and pathogenic or antibiotic-resistant bacteria and their genes (ARB&Gs). These are generally adopted for specific applications such as the pathogenic bacteria removal prior to potable water reuse, the adherence to low-concentration standards for nutrient in sensitive water bodies, or for the elimination of specific pollutants (such as pharmaceuticals).

The basic schema of fig. 2.3 can be adapted to the specific process requirements. For example, figures 2.4, 2.5 2.6 show pictures for a small, a medium and, a large WWTP. Generally, small WWTPs are not equipped with primary sedimentation and they perform aerobic digestion of the sludge. Medium and large WWTPs, instead, adopt a more complex solution with primary sedimentation, anaerobic digestion and biogas production (Hansen, 2018).

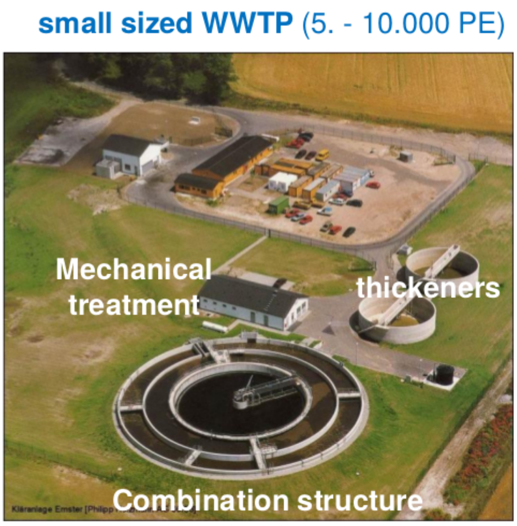


FIGURE 2.4: A small WWTP. Figure extracted from (Hansen, 2018)

2.1.4 Sludge line and biogas production

One of the by-products of waste water treatment is sludge coming from primary and secondary sedimentation. Sludge needs to be treated because it is largely composed of water (99%), contains pathogenic bacteria and produces a bad odour (Hansen, 2018). Small and big plants stabilize their sludge differently. Small plants are generally equipped with aerobic digestion, in which bacteria use the sludge biomass for

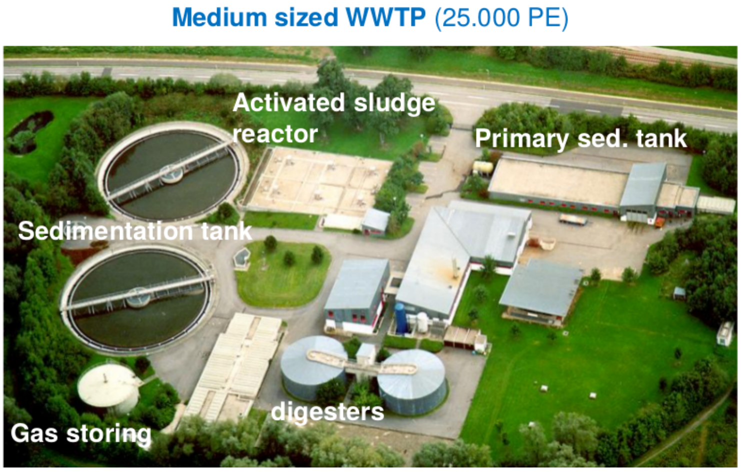


FIGURE 2.5: A medium sized WWTP. Figure extracted from (Hansen, 2018)

their metabolism. For plants with capacity larger than 10k PE, generally anaerobic digestion is performed. This process is important in the framework of this thesis because produces biogas that can be converted in energy. In the anaerobic digestion, many processes can be identified (Metcalf and Eddy, 2014, page 656):

- **hydrolisis**; hydrolytic enzymes degrade carbohydrates, proteins and lipids to basic monomers;
- **acidogens**; fermenting bacteria transform monomeric products into fatty acids;
- **acetogenesis**;in this process, longer fatty acids are transformed into acetate, CO_2 and H_2;
- **methanogenesis** in which the products of the previous steps are converted into methane.

Anaerobic processes are sensitive to environmental factors. In particular, $pH < 6.8$ can stop methanogenesis. Hydraulic retention time is also an important design parameter to guarantee that bacteria have enough time to develop their metabolic activities. Temperature influences the rate of hydrolysis; when the process temperature is in the range of 30-38 ^oC the process is called 'mesophilic', while when the temperature is in the range 50-57 ^oC, the process is defined as 'thermophilic' (Metcalf and Eddy, 2014, page 1504).

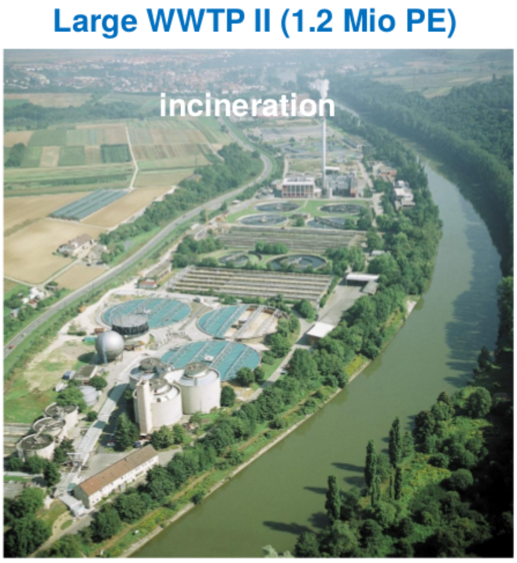


FIGURE 2.6: Large WWTP. Figure extracted from (Hansen, 2018)

Anaerobic digestion is discussed in detail in other parts of this thesis. Section 2.2.1 considers the role of biogas production in the energy balance and section 5.5 proposes an innovative methodology for biogas monitoring.

2.1.5 Short introduction to main operational parameters

The parameters that influence the WWTP processes are various. In this section, some of the most important are briefly mentioned. The waste water inflow has a direct impact on the pollutant removal process (for example trough the hydraulic retention times) and the energy consumption (for example, efficiency of the pumps Torregrossa et al., 2017b). Pollutant concentrations has a direct impact on biological process (for example, toxic shocks can disturb the biological processes (WEF, (2008), chapter 20, page 166)). In section 2.1.4, the most important parameters of anaerobic digestion were presented. Restricting the focus on the biological process of the water line, the main parameters here are the dissolved oxygen concentration (DO) in the activated sludge tank, the mixed liquor suspended solids (MLSS), the sludge retention time (SRT) and the food-to-biomass ratio (Hansen, 2018). The dissolved oxygen concentration, depending on the process design, should be maintained in the range 0.5-2 mg/l (Hansen, 2018) in the aeration tank. The oxygen is one of the key element for the aerated biological processes and an insufficient oxygen amount can inhibit bacteria metabolism. An oxygen concentration above 4 mg/l does not provide any advantages to the process but results in an increased energy consumption (Metcalf and Eddy, 2014, page 729). The MLSS gives information about the concentration of bacteria in charge of BOD removal process. The higher the concentration of bacteria, the lower the time required for the process. Hansen, (2018) reports that in 1 hour:

- with MLSS = 0.5 g/l, the elimination rate is 30%;
- with MLSS = 3.0 g/l, the elimination rate is 95%;

Typical values of MLSS range from 2 to 4 g/l in conventional activated sludge systems.

The solid retention time (SRT) is the "average time the activated sludge solids are in the system" (Metcalf and Eddy, 2014, page 598). This quantity is calculated as the mass of solids in the aeration tank divided the mass that daily leave the system (via effluent or with the wasted sludge). In a WWTP like that in fig. 2.7, the equation 2.1 expresses the SRT. The optimal value of the SRT depends on many factors (such as temperature and process removal targets) but as a rule of trumps, The water Environmental Federation indicates that "the overall SRT typically ranges from 10 to 40 days" (WEF, (2008) chapter 20, page 177). However, considering that an overly large value of SRT increases energy consumption and reduces the dewaterability of the digested solids, SRT should be taken close to the lower part of the above indicated range.

$$SRT = \frac{\text{Volume-of-aeration-tank} * MLSS}{\text{Waste-Sludge}} \tag{2.1}$$

The food-to-biomass ratio expresses the ratio between the quantity of substrate to be decomposed and the quantity of bacteria. According to Metcalf and Eddy, 2014 (page 606), the food-biomass ratio assumes this form:

$$food - biomass - ratio = \frac{\text{total applied substrate rate}}{\text{total microbial biomass}} \tag{2.2}$$

The numerator is generally expressed as quantity of BOD (or COD) and the denominator as quantity of mixed liquor biomass. The optimal

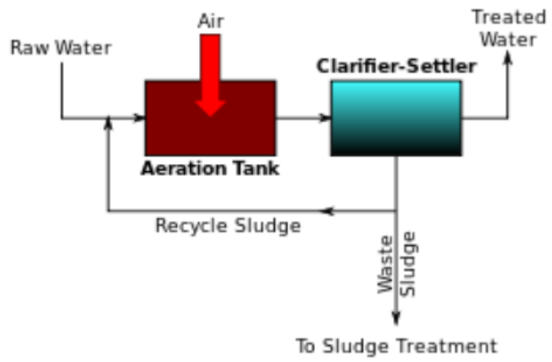


FIGURE 2.7: Aeration tank with recycled sludge. Figure extracted from (Hansen, 2018)

range for food-biomass ratio is connected to solid retention time (intuitively, the higher the number of bacteria for given substrate, the lower the degradation time). Metcalf and Eddy, 2014 (page 607) reports that for SRT in range 20-30 days, the corresponding food-biomass ratio is between 0.10-0.05 gBOD/gVSS*d. With a SRT in the range 5-7 day, the corresponding food-biomass ratio is between 0.3-0.5 gBOD/gVSS*d.

2.1.6 New opportunities: water reuse, nutrient recovery

The WWTP process produces a main product (clean water) and different by-products (such as energy and sludge). These outputs are of interest in WWTP management, because they can be considered as resource and treated using a circular-economy approach.

For example, an interesting application is the potable reuse of purified water from WWTPs (Gardoni, Catenacci, and Antonelli, 2015; Leverenz, Tchobanoglous, and Asano, 2011). Under extensive treatments and strict quality monitoring, the reuse of purified water offers various advantages:

- conservation of a primary source;
- economic benefit from water selling;
- potential use as reservoir or groundwater augmentation;
- the purified water is geographically close to the water demand;
- protection of receiving bodies.

Despite the advantages and the technical feasibility, the direct potable water reuse must still overcome some limitations (Leverenz, Tchobanoglous, and Asano, 2011) such as:

- improvement of design criteria;
- treatment reliability;
- development of adequate monitoring techniques;
- development of strategies to enhance the social acceptance.

Grant et al., 2012; Verstraete, Clauwaert, and Vlaeminck, 2016; Papa et al., 2017 propose to consider the WWTPs as a primary-source producers. In this direction, for example phosphate can be extracted frome sludge and the return flow and sell as agricultural product. According to (Metcalf and Eddy, 2014), in the future, phosphate recovery is going to be an economically viable activity because of the increasing price of this nutrient. Despite the high economic and environmental potential, Papa et al., (2017) demonstrated that nutrient recovery is still at a preliminary stage, and that, in Europe, 60% of the plants are not equipped for this task mainly because of economical (cost of technological implementation) and legislative (i.e. strict thresholds set by legislation) issues.

2.1.7 Short considerations about technical aspects of water cleaning processes

Section 2.1 presents the main definitions belonging to the WWTP domain. Moreover, it was discussed what wastewater is, how it is generated and its most important characteristics. In this section, the most common WWTP configuration was presented as well as the challenges and opportunities connected to a more efficient WWTP management. In the perspective of this thesis, it is important to underline some elements that justify the use of decision support tools. The first element is the great variability and uncertainty associated to wastewater production and its technological alternatives when treating it. The second element is the importance of the technological challenges that impact on wastewater regulation, environment, economic aspects, and more in general on society. The third element consists of the nature of the decisions taken by the stakeholders; these are based on multiple parameters and concern several, often conflicting, objectives (such as economic optimisation, environmental issues, technological limitations and the adherence to regulations).

2.2 Energy balance in WWTPs

As shortly discussed in section 1.1, wastewater treatment plants are considered interesting in an energetic perspective for the following reasons:

- wastewater contains a quantity of energy that can be recovered. For example, Hansen, (2018) reports that waste water has thermal energy content (estimated at 75 kWh/p.e./a), energy potential from organic matter (estimated at 153 kWh/p.e./a), and hydraulic potential energy (depending on inflow rate and available hydraulic height);
- wastewater energy consumption is relevant. In Europe, it corresponds to the 1% of national electric energy consumption (Longo et al., 2016);
- WWTPs account approximately for 30% of total energy consumption of municipalities (Hansen, 2018);
- the electric energy saving potential is high and, for example, in Spain, it is estimated at around 25% (Castellet and Molinos-Senante, 2016);
- the recovery from biogas production can generate 17kWh/p.e./a of electric energy and 27 kWh/p.e./a of thermal energy (Hansen, 2018).

In the following sub-sections, a more detailed analysis of energy consumption is carried out.

2.2.1 Impact of WWTP energy consumption in European countries and energy balance.

In Germany, the total electrical energy consumption is around $513 * 10^3 GWh/year$ (WebPage: Gross Inland energy consumption (Eurostat)). Reinders et al., (2012) reports that, in Germany, the total energy consumption of electric energy in WWTPs is 4.400 GWh which corresponds to the 0.85% of the total energy consumption.

Foladori, Vaccari, and Vitali, (2015) report that in Italy the WWTP energy consumption is around 3250 $GWh/year$. With a national energy consumption of around $281 * 10^3 GWh/year$ (WebPage: Gross Inland energy consumption (Eurostat)), in Italy, WWTPs account for 1,15% of the total energy consumption.

Although the ratio between WWTP electric energy consumption and national consumption can vary depending on the country, in Europe, a value of 1% can be considered a good approximation (Longo et al., 2016).

For plant size larger than 10k p.e., energy consumption is in the range 27.1-63.5 kWh/p.e./a (values extracted from table 1 of (Becker and Hansen, 2013), values referring to WWTPs located in NW Europe). Figure 2.8 shows the typical energy flows produced by anaerobic sludge digestion, that convert the chemical energy content of COD to electricity and thermal energy. Consequently, the biogas energy recovery can cover up to 62% of total energy requirements.

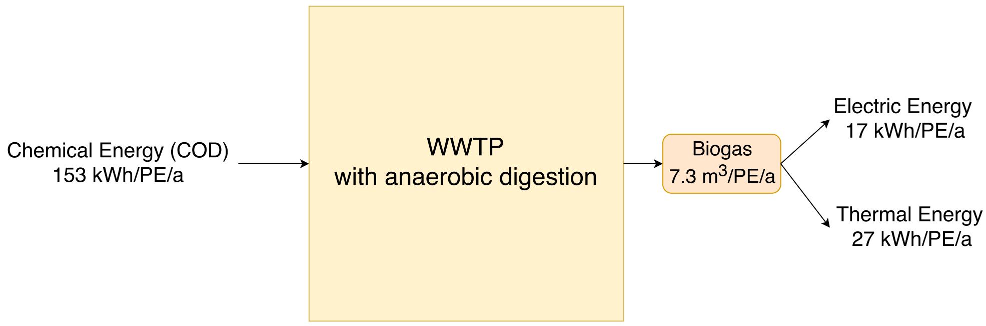


FIGURE 2.8: Typical Energy Intensity Flow in WWTPs. Information Sources: Hansen, 2018

As shown in fig. 2.9, the energy balance of conventional WWTPs shows that the most important energy consumers are the aeration system (60%), wastewater pumping (12%) and anaerobic digestion (11%), which together account for the 83% of global energy consumption of the plants (Gu et al., 2017).

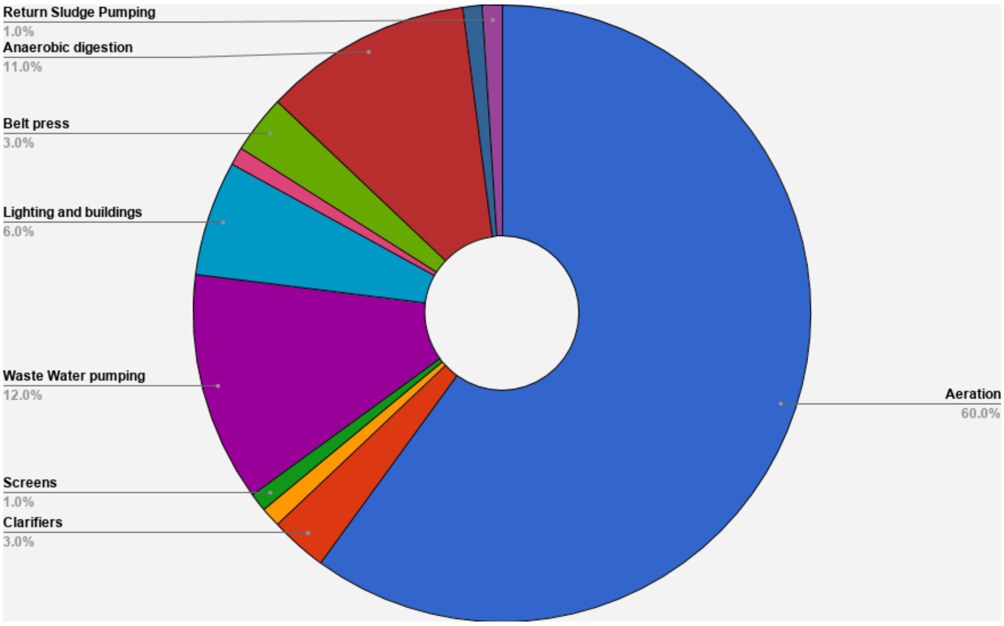


FIGURE 2.9: Energy consumption in WWTPs. Adapted from (Gu et al., 2017)

2.2.2 Energy benchmarking

In order to assess the energy consumption of WWTPs, different authors proposed benchmark-based approaches, because benchmarking enables the detection of inefficiencies and supports the development of detailed strategies for energy savings (Krampe, 2013).

Benchmarking includes a large set of methodology and can be performed with different approaches. The common part of all benchmarking approach consists of two steps:

- calculation of key parameters;
- comparison with reference values.

Often the key parameters are expressed as key performance indicators, that are intuitive and common in engineering fields: for example, a car fuel consumption is generally expressed in l/km , in order to allow the comparison of vehicles that run under different conditions. In the case of the car, the reference unit to which refer the calculation of fuel consumption is the km. In WWTP domain the reference unit could change according to the preferences of the researchers.

It is important to remark here that the benchmark analyses and the implementation of efficiency saving measures assume that the quality of process performance must not decrease. Otherwise, *ad absurdum*, the most efficient suggestion to reduce the energy consumption would consist in the complete switching-off of the plants.

In the next parts of this section, Swiss and German regulations are presented because they were a source of inspiration for this PhD project.

Swiss regulation

The first, well-known, impulse to energy efficiency benchmarking in WWTPs was provided by (Muller, Thommen, and Stahli, 1994; Muller, Kobel, and Schmid, 2010) with the support of Swiss environmental agencies. The resulting manuals identify two benchmark operations: 1) energy checks, 2) energy analysis. The first energy assessment consists of a comparison between measured values and references values. The second assessment is suggested for special conditions, such as plant updates and/or after the detection of relevant energy inefficiencies. This second assessment should include energy, financial and cost-benefits assessments in order to identify effective actions to reduce energy consumption. In the Swiss manuals, the plant reference values are based on two specific model-plants:

- plant-model 1. Population equivalent: 20k PE; daily specific inflow: 300 l/PE; anaerobic stabilization, full nitrification (Muller, Thommen, and Stahli, 1994).

- plant-model 2. Population equivalent: 130k PE; daily specific inflow: 380 l/PE; anaerobic stabilization, full nitrification (Muller, Kobel, and Schmid, 2010).

In (Muller, Thommen, and Stahli, 1994; Muller, Kobel, and Schmid, 2010), the reference values for benchmarking are expressed as key performance indicators: kWh/p.e./a and kWh/m^3. The benchmarks are available for the global energy consumption and for the consumption of single stages (such as activate sludge reactor, pumping stations). The main limitation of this approach consists in the plant-model specific benchmark values; in fact, it is inappropriate to use these benchmarks for plants which differ from the given models.

German regulation

The German Association for Water, Wastewater and Waste (DWA) published the first national regulation focuses on energy benchmarking of WWTPs. The updated version of this document (DWA, 2015) incorporates the results and extends previous works (such as (Muller, Thommen, and Stahli, 1994; Haberkern, Maier, and Schneider, 2008)). DWA, (2015) contains some interesting elements:

- energy assessments of WWTPs are required to be performed annually;
- a list of parameters to be monitored (for example, the power consumption of the pumping station, the pump static head and the total energy consumption);
- benchmark values are generated through statistical procedures.

In the German approach, for each parameter and for a given plant size a target value and a guide value are provided, which are the results of a statistical analysis on a group of approximately 10k WWTPs. In particular, as in (Baumann, Roth, and Maurer, 2014), the guide value for energy consumption corresponds to the 50^{th} percentile, while the target value is calculated as the 10^{th} percentile.

The German legislations recommends seven steps:

1. evaluation of the current state, in order to compare the operational parameters with their reference and identify the energy saving potential;
2. calculation of the energy balance flows for electric and heat energy consumption;
3. benchmark calculation of individual process units;
4. comparison of current-state values and benchmarks calculated at the previous step;
5. calculation of energy saving potential and financial analysis;

6. identification of actions and priorities.

The German and Swiss regulation can be considered as a milestone in energy benchmarking of WWTPs and they inspired the INNERS project (INNERS, 2015) and this PHD project 'EdWARDS'.

Approaches in scientific literature

According to (Longo et al., 2016), in the WWTP domain, the classic key performance indicators (KPIs) for global energy consumption are those reported in equations 2.3, 2.4 and 2.5.

$$KPI_1 = \frac{\text{electrical energy consumption}}{\text{volume of treated wastewater}} \text{ [kWh/}m^3\text{]} \tag{2.3}$$

$$KPI_2 = \frac{\text{electrical energy consumption}}{\text{served PE}} \text{ [kWh/p.e./a]} \tag{2.4}$$

$$KPI_3 = \frac{\text{electrical energy consumption}}{\text{COD load removed}} \text{ [kWh/kg } COD_{rem}\text{]} \tag{2.5}$$

These KPIs express a plant generic value for the energy consumption that makes possible the comparison between WWTPs. Each KPI has advantages and disadvantages. For example, KPI_1 is the easier to calculate because of data availability but it does not take into consideration the pollutant load. By contrast, KPI_2 and KPI_3 consider the pollutant load but they require a large amount of water quality data to be measured by laboratory analysis or through expensive on-line samplers. During, the development of this PhD thesis, a methodology was proposed based on the estimation of missing water quality parameters to calculate the KPI_2 at a daily resolution (Torregrossa et al., 2016).

Once a key performance indicator is calculated, it is necessary to compare it with reference values. For example, a typical range of the value for energy consumption per cubic meter of treated wastewater, corresponds to KPI_1 is $0.10 - 0.18 kWh/m^3$ [Metcalf and Eddy, 2014, page 1812]. In [Shi, 2011, page 53], the author proposes a set of benchmark values for the energy consumption of WWTPs depending on the size of the plants:

- $20 - 26$ kWh/p.e./a for large WWTP (p.e. > 100.000);
- $23 - 30$ kWh/p.e./a WWTP with 30.000<p.e. < 100.000;
- $26 - 34$ kWh/p.e./a for WWTPs in which 5.000<p.e. < 10.000;

- $35 - 46$ kWh/p.e./a for WWTPs in which 2.000<p.e. < 5.000.

In (Longo et al., 2016), the author reports typical values for KPI_3:

- 0.69 kWh/CODrem for large WWTP (p.e. > 100.000);
- 0.82 kWh/CODrem WWTP with 50.000<p.e. < 100.000;
- 1.02 kWh/CODrem WWTP with 10.000<p.e. < 50.000;
- 1.54 kWh/CODrem WWTP with 2.000<p.e. < 10.000;
- 3.01 kWh/CODrem WWTP with p.e. < 2.000.

These classes of benchmark values show that large WWTPs are generally more efficient because of several factors such as economies of scale, stability of operational condition and automatic systems (Longo et al., 2016). During the development of this thesis and in (Torregrossa et al., 2017c; Torregrossa et al., 2017d), it is discussed how operational conditions affect the energy pump consumption and a methodology was proposed for pump online monitoring and decision support.

Another alternative approach to calculate a benchmark value for energy consumption is proposed in [Metcalf and Eddy, 2014, page 1815]. In this case, instead of a key performance indicator, the value to be benchmarked is the value for energy consumption [kWh] and the benchmark value is calculated with equation 2.6.

$$\begin{aligned} ln(E_s) = 15.8471 + 0.8944 * ln(I_{ww}) + 0.4510 * ln(BOD_i) - \\ 0.1943 * ln(BOD_e) - 0.4280 * ln(I_{ww}/I_{wwD} x 100) - \\ 0.3256 * TF + 0.1774 * NR \end{aligned} \tag{2.6}$$

In this equation, the terms have the following meanings

- E_s is the energy consumption in kBtu/year;
- I_{ww} is the average inflow expressed in Mgal/day;
- BOD_i and BOD_e are the BOD concentration [mg/l] at the inlet and at the outlet;
- I_{wwD} is the designed inflow expressed in Mgal/day;

- TF is a binary factor equal to 1 if there is a trickling filter,0 otherwise;
- NR is a binary factor equal to 1 if there are nutrient removal processes, 0 otherwise.

The approach proposed in [Metcalf and Eddy, 2014, page 1815] enables the calculation of an ideal value of energy consumption for an individual WWTP and to compare this theoretical value with the observed value. This approach could be considered as a benchmark methodology, even if not based on a key performance indicator.

Another class of benchmark approaches is composed by non-parametric techniques like those proposed by Hernández-Sancho, Molinos-Senante, and Sala-Garrido, (2011) and Molinos-Senante, Hernandez-Sancho, and Sala-Garrido, (2014). These approaches are based on the analysis of data and on the identification of the best Pareto frontier, taken as a reference to evaluate the efficiency of similar facilities. Since the benchmark techniques adopted in this thesis are based on pre-defined parameters, there is no need to explain in detail the non-parametric approaches, but the reader should be aware of their importance for energy assessment in WWTPs.

2.2.3 Results of energy assessment

Various studies focused their attention on the assessment of the energy performance of WWTPs [for example Shi, (2011), Hernández-Sancho, Molinos-Senante, and Sala-Garrido, (2011), Becker and Hansen, (2013), Foladori, Vaccari, and Vitali, (2015), Gude, (2015), and Castellet and Molinos-Senante, (2016)]. These studies claim that there is room for an improved energy efficiency in the WWTP domain; for example, Hernández-Sancho, Molinos-Senante, and Sala-Garrido, 2011 assessed 117 WWTPs in the region of Valencia (Spain) with a data envelopment analysis approach and found that almost 90% of WWTPs have a sub-optimal efficiency index . Foladori, Vaccari, and Vitali, 2015 found similar results by focussing on small WWTPs in Italy. Castellet and Molinos-Senante, 2016 show that the average energy saving potential in WWTPs is around 25%. In [Shi, 2011, page 74], the author claims that “an energy efficiency of between 30 and 80% is achievable based on the Best Available Practices (BAP) of full-scale application." Improving the energy efficiency of WWTPs could produce several benefits listed by Gude, 2015:

- positive environmental impact through the reduction of air pollutant emissions,
- economic gain produced by a reduced energy consumption,
- economic growth and creation of jobs,
- enhancement of political leadership of local government through the application of development strategies,
- water security.

2.2.4 Short considerations about energy aspects in WWTPs

In Section 2.2, it was shown that WWTP energy consumption is relevant from both an environmental and economic perspectives. A great energy saving potential is available alongside a great potential to generate energy by biogas. Together, these elements make energy management in WWTP an important and challenging topic for the next years.

2.3 Decision support system technology

According to (*DSSresource.com*):
"a Decision Support System (DSS) is an **interactive computer-based system** or subsystem intended to **help decision makers** using communications technologies, data, documents, knowledge and/or models to **identify and solve problems, complete decision process tasks, and make decisions**."
In this definition, there are some key elements:

- the DSSs are interactive computer-based system. A decision support system has a software component that processes information. Other decision support tools, such as static decision tree, per se, cannot be considered DSSs;
- decision support systems help decision-makers and do not replace them. The tools able to analyse information and automatically react are generally called 'expert systems'. In DSSs, human interaction is fundamental. For example, a software for the automatic control of air conditioning system is not a DSS;
- decision support systems have not an a-priori structure, neither a specific application fields and they are flexible tools able to be adapted to various domains.

In the decision-science, the problems can be classified as structured, semi-structured and unstructured(*Introduction to Ill-Structured Problems - Wikiversity*; McIntosh et al., 2011).

In structured problems, there is a clear definition of initial conditions, goals and constrains. Structured problems are not affected by uncertainty and they can be fixed with well-defined procedures. For example, the calculation of the area of a rectangle is a structured problem: there is a standard procedure and for given input there is just one solution.

An unstructured problem is characterised by high uncertainty, conflicting objectives and often ethic issues. For example, a government could deal with an unstructured problem when required to decide between security and the privacy of citizens; the choice of the colour of a T-shirt is also an unstructured problem; in both cases, the solution to the given problem is complex and it depends on the personality of the decision-maker.

Semi-structured problems are situated between unstructured and structured problems. For example, maximizing the efficiency of a car factory is a semi-structured problem because:

- there are conflicting objectives (*save money or increase the production rate?*);
- it is possible to have many optimal solutions to the same problem;
- it is still possible to have a structured approach that simplifies the problem by selecting a sub-set of potential optimal solutions.

Decision support systems are specifically concerned with *semi-structured problems* (Power, 2000; Poch et al., 2004). This characteristic makes them suitable to deal with environmental problems (Rizzoli and Young, 1997; Poch et al., 2004; Poch et al., 2014). In fact, most environmental problems present multiple conflicting objectives (*in the WWTP domain, saving energy or improve water quality?*), data uncertainty (Torregrossa et al., 2016) and several suitable solutions.

In (Power, 2000), decision support systems are classified according to various categories:

- *Data-Driven DSS*, based on large databases and able to manipulate information and produce reports. A geographic information system (GIS) is a typical example of a data-driven DSS;
- *Model-based DSS*, based on models or algorithms that elaborate input-information. Although it is possible to produce data-model hybrid DSSs, generally model-based DSSs do not use large datasets;
- *Knowledge-driven DSSs* are those that incorporate expert knowledge and provide a list of solutions to the end-user;
- *Document-based DSS* are those that support the decision process by organising and making documents easily accessible ;
- *Group decision support systems* are those that support the decision-making process by enabling collaboration between people in order to solve a given problem. A mailing list, Skype or Dropbox could be also classified as basic Group DSS.

- *Web based DSSs* associated to the use of web applications.

Often DSSs fit within several categories. For example, a document based decision support system could be developed with a web-interface. The decision support system presented in this thesis has features of data-driven, model-based, group and web-based decision support systems.

2.3.1 Use of DSS: Pro and cons

The literature widely discusses of the advantages and disadvantages of the use of decision support systems (Rizzoli and Young, 1997; Power, 1997; Power, 2000; McIntosh et al., 2011; Management Study Guide, 2017). The main advantages are:

- **cost reduction** of the decision-making process; the use of the DSSs enables operators to reduce the time consumed, to improve the quality of the decision and to reduce the use of an external consultancy. Consequently, the decision-making process has lower costs;
- **decision quality**; the decision quality is not affected by human weakness such as fatigue, boredom, fear or state of mind. A computer can analyse the dataset in a more impartial way than a human;
- **decision process continuity** ; DSSs can monitor and process data 7/7 and 24h/24h;
- **experience and knowledge store**; compared with a human-expert memory, a well-structured and well-maintained decision support systems can store large sets of information and can access them efficiently to provide results.

The main disadvantages are:

- **investment cost**; the implementation of a DSS could be expensive and this cost could make the investment uneconomic. An adequate cost-benefit analysis should be done before implementing a DSS. Generally, the investment is warranted when there is a large amount of data to be analysed, a high uncertainty and a large number of operational parameters to be set-up;

- **difficulties to manage not-quantitative data**; the DSS are formidable in analysing quantitative information. Nevertheless, some information is difficult to process correctly. For example, the happiness of employers could have an impact on the performance of companies but this is not easy to measure and process for a DSS;

- **lack of creativity**; decision support systems are not suitable to create innovative solutions. They are able to re-use expert knowledge stored in the system or efficiently run a model, but the outputs are limited to the machine set-up;

- **End-user awareness**; the end users may not be aware of the limitations of the DSS, such as the model constrains, the hypotheses adopted or the data process routines. A non-expert user could be tempted by the non-critical use of the results and this could be a problem if the DSS experiences an error.

The advantages of a DSS can be increased and disadvantages decreased by a correct design of the decision support system, which must be as close as possible to the real needs of the end-users. A good communication between designers and end-users is necessary during the design phase, the prototype tests and the final implementation. A lack of communication during these steps can generate extra costs and create conflicts. For example, the introduction of a DSS in a company could be accepted or obstructed by employees depending on their technical ability or their willingness to change.

In few words, decision support systems perform well in dealing with complex problems, analysing a large amount of data and information, and guarantee an efficient decision-making process. Nevertheless, in order to obtain useful results, their adoption should be carefully evaluated according to the potential benefits and the end users should be informed about their potential limitations.

In the specific case of WWTPs, the adoption of a decision support system is beneficial, because the disadvantages are minimised. For example:

- the efficient management of WWTPs can produce relevant environmental and economic benefits (Castellet and Molinos-Senante, 2016). Therefore, the investment cost of a decision support system should be paid-back in a convenient time;

- the large amount of data makes IT support necessary because a human operator cannot efficiently analyse the full parameters set (including for example, the energy consumption of devices at 15s time-resolution): in the INNERS project, each WWTP generates up to 300.000 values per day INNERS, (2015) and Torregrossa et al., (2016);

- the plant managers can generally be considered expert end-users.

2.3.2 Environmental Decision Support Systems

An interesting category in the decision support domain is the Environmental DSS (EDSS). The definitions of Rizzoli and Young, (1997), Cortés and Sànchez-Marré, 2001,Elmahdi and McFarlane, 2009, McIntosh et al., 2011, suggest these main characteristics of an EDSSs:

- they are dedicated to environmental issues;

- they integrate models, data and tools in a user-friendly framework;

- they improve the consistency of the decision;

- they reduce the time of decision-making process.

Table 2.6 reports some recent contributions in the field of EDSS. These tools are applied to various domains and problems. Conventionally, in this thesis, these EDSSs are classified according to their task: environmental management, environmental planning, and risk management. The environmental management EDSSs mainly deal with resource management optimization, the environmental-planning EDSS are focussed on problem of design and the last category of EDSSs aims to manage efficiently the risk associated to human activities. Moreover, EDSSs are applied to different domains, such as water, industry, agriculture and urban planning.

In other words, table 2.6 shows EDSSs are flexible and effective tools that can be used for a wide range of practical applications. Section 2.4 reports a detailed presentation of decision support systems applied to the waste water domain.

TABLE 2.6: Recent contributions in EDSS domain

Author	Reference	Field of application	Task
Navarro-Hell?n et al., (2016)	A decision support system for managing irrigation in agriculture	Agriculture	Environmental management
Duah and Syal, (2016)	Intelligent decision support system for home energy retrofit adoption	House	Environmental management
Rose et al., (2016)	Decision support tools for agriculture: Towards effective design and delivery	Agriculture	Environmental planning
Castillo et al., (2016)	Validation of a decision support tool for wastewater treatment selection	Water domain	Environmental planning
Bottero, Mondini, and Oppio, (2016)	Decision Support Systems for Evaluating Urban Regeneration	Urban Planning	Environmental planning
Zodiatis et al., (2016)	The Mediterranean Decision Support System for Marine Safety dedicated to oil slicks predictions	Water domain	Environmental management Risk assessment
Zulkafli et al., (2017).	User-driven design of decision support systems for polycentric environmental resources management	Water domain	Environmental management
Mustajoki and Marttunen, (2017)	Comparison of multi-criteria decision analytical software for supporting environmental planning processes	Various – Review article	Environmental planning
Little et al., (2017)	Decision support for environmental management of industrial non-hazardous secondary materials: New analytical methods combined with simulation and optimization modeling	Industry	Environmental management
Rahmanpour and Osanloo, (2017)	A decision support system for determination of a sustainable pit limit	Open pit mining	Environmental planning
Yang et al., (2017)	A flexible decision support system for irrigation scheduling in an irrigation district in China	Water domain	Environmental management
Caeiro et al., (2017)	Environmental risk assessment in a contaminated estuary: An integrated weight of evidence approach as a decision support tool	Water domain	Environmental management Risk assessment
Argyris and French, (2017)	Nuclear emergency decision support: A behavioural OR perspective	Industry	Environmental management Risk assessment
Aiello et al., (2017)	A decision support system based on multisensor data fusion for sustainable greenhouse management	Agriculture	Environmental management

2.4 DSSs applied to the wastewater domain

The WWTP domain is suitable for the application of environmental decision support systems because the decision-makers are generally required to deal with a great amount of information, uncertainty and multi-parameter, complex, conflicting objectives. During the literature review, various manuscripts concerning environmental decision support systems applied to WWTPs were founded. Table 2.7 shows more than twenty applications of decision support systems in WWTP domain. The author of this thesis classified these applications according to their main function:

- WWTP design; these decision support systems support the operator in the selection of treatment processes during the design stage;
- WWTP management; these decision support system aim to reduce the costs, optimise the use of resources or improve the plant performance.

An important function of a decision support system is the knowledge discovery; for example, Comas et al., 2001 shows how it is possible to extract information from data and support decision making processes. Knowledge discovery should be considered a fundamental element of decision support systems. Nevertheless, according to the definition of decision support systems (section 2.3), tools limited to statistical analysis and data discovery cannot be considered decision support systems because they miss the interaction in the decision making process. Hence, this kind of tools are not considered as DSS.

Table 2.7 also reports two review papers by (Hamouda, Anderson, and Huck, 2009) and Poch et al., (2014).

2.4.1 DSS for WWTP design

The decision support systems for the WWTP design have various objectives and use multiple methodologies. For example, Poch et al., (2004), Hakanen, Sahlstedt, and Miettinen, (2013),Garrido-Baserba et al., 2015, Garrido-Baserba et al., (2016),Kalbar, Karmakar, and Asolekar, (2016), Castillo et al., (2016), and Rawal and Duggal, (2016) focussed their work on the selection of the most suitable processes for their WWTPs. Papa,

TABLE 2.7: WWTP and Decision support system

Author(s) and year	Title	Application
Paraskevas, Pantelakis, and Lekkas, 1999	Advanced integrated expert system for wastewater treatment plants control	Management
Comas et al., 2001	Knowledge discovery by means of inductive methods in wastewater treatment plant data	Knowledge discovery
Comas et al., 2004b	Development of a knowledge-based decision support system for identifying adequate wastewater treatment for small communities.	Design
Poch et al., 2004	Designing and building real environmental decision support systems	Design
Fiter et al., 2005	Energy saving in a wastewater treatment process: an application of fuzzy logic control.	Management
Gómez-López et al., 2009	Decision support in disinfection technologies for treated wastewater reuse	Management
Hamouda, Anderson, and Huck, 2009	Decision support systems in water and wastewater treatment process selection and design: A review	Design – Review
Guerrero et al., 2011	Improving the performance of a WWTP control system by model-based setpoint optimisation	Management
Guerrero et al., 2012	Multi-criteria selection of optimum WWTP control setpoints based on microbiology-related failures, effluent quality and operating costs	Management
Hakanen, Sahlstedt, and Miettinen, 2013	Wastewater treatment plant design and operation under multiple conflicting objective functions	Design
Bertanza et al., 2014	How can sludge dewatering devices be assessed? Development of a new DSS and its application to real case studies	Management
Poch et al., 2014	Where are we in wastewater treatment plants data management? A review and a proposal	Design – Review
Garrido-Baserba et al., 2015	Selecting sewage sludge treatment alternatives in modern wastewater treatment plants using environmental decision support systems	Design
Caniani et al., 2015	Towards A New Decision Support System for Design, Management and Operation of Wastewater Treatment Plants for the Reduction of Greenhouse Gases Emission	Design and management
Gisi et al., 2015	An integrated approach for monitoring efficiency and investments of activated sludge-based wastewater treatment plants at large spatial scale	Management
Thürlimann, Dürrenmatt, and Villez, 2015	Energy and process data processing and visualisation for optimising wastewater treatment plants	Management
Castillo et al., 2016	Validation of a decision support tool for wastewater treatment selection	Design
Garrido-Baserba et al., 2016	Application of a multi-criteria decision model to select of design choices for WWTPs	Design
Kalbar, Karmakar, and Asolekar, 2016	Life cycle-based decision support tool for selection of wastewater treatment alternatives	Design
Kim et al., 2016	Operator decision support system for integrated wastewater management including wastewater treatment plants and receiving water bodies	Management
Papa, Bertanza, and Abbà, 2016	Reuse of wastewater: a feasible option, or not? A decision support system can solve the doubt	Design
Rawal and Duggal, 2016	Life Cycle Costing Assessment-Based Approach for Selection of Wastewater Treatment Units	Design
Tomei et al., 2016	Techno-economic and environmental assessment of upgrading alternatives for sludge stabilization in municipal wastewater treatment plants	Design
Tran, Schwabe, and Jassby, 2016	Wastewater reuse for agriculture: Development of a regional water reuse decision-support model (RWRM) for cost-effective irrigation sources	Design

Bertanza, and Abbà, (2016) and Tran, Schwabe, and Jassby, (2016) focussed their work on water reuse and Tomei et al., (2016) worked on sludge stabilization. These decision support systems are based on different methodologies. For example, Poch et al., (2004) used artificial intelligence techniques, Hakanen, Sahlstedt, and Miettinen, (2013) multi-objective optimisation methodologies,Garrido-Baserba et al., (2015) used a cost-benefit assessment approach, Kalbar, Karmakar, and Asolekar, (2016) and Rawal and Duggal, (2016) used life cycle approaches.

2.4.2 DSS for WWTP management

In WWTP management, decision support systems have been designed for different tasks such as: sludge dewatering (Bertanza et al., 2014), energy saving (Fiter et al., 2005; Thürlimann, Dürrenmatt, and Villez, 2015), plant control (Paraskevas, Pantelakis, and Lekkas, 1999; Guerrero et al., 2011; Guerrero et al., 2012; Kim et al., 2016). It was observed that a great variety of methodologies were adopted for these tasks. For example, Paraskevas, Pantelakis, and Lekkas, (1999) used artificial intelligence techniques, Fiter et al., (2005) used fuzzy logic, and (Guerrero et al., 2012) used the ASM2d in combination with multi-criteria functions.

2.4.3 Short considerations about decision support systems and gaps in the literature

The applications discussed in this section show that decision support systems are suitable to be adopted in WWTP domain with great flexibility and great effectiveness. There is not a dominant methodology or a dominant approach in DSSs applied to wastewater processes. Moreover, table 2.7 shows that these DSSs pursue various targets. Nevertheless, despite the relevant number of applications, apart from SK-DSS, there is not in literature a DSS applied to wastewater domain, which:

- enables the cooperation between end-users;
- is specifically focussed on energy management of WWTPs;
- is plant generic, or, in other words, is able to simultaneously analyse multiple WWTPs;

- is able to produce daily analysis reports;
- provides plant managers with case-based suggestions for energy efficiency.

2.5 Fuzzy Logic

In this section, the fuzzy logic methodology is presented because the fuzzy approach is the core of the DSS developed in this thesis. This method was chosen to deal with knowledge expressed as human-like language and for the high-performance in dealing with uncertainty (Starczewski, 2013). Fuzzy logic is a relatively recent branch of mathematics. It has been developed by Zadeh, (1965) with the aim to deal with parameters affected by uncertainty and ambiguities. The classical set theory, developed since Aristotle, assumes that an element can belong to a class or not (Sivanandam, Sumathi, and Deepa, 2006); for example, a cat belongs to the class 'Animals', while a pen does not. When approaching real problems, this classical crispy set theory encounters some difficulties. In reality, some classes are affected by vagueness and uncertainty; for example, let T be the class 'Tall men' and S the class 'Short men'. In this case, each person has a different definition of 'tall' or 'short', and classification difficulties can arise. Let's assume that to overcome the vagueness by defining a class T, the class that includes men taller than 1.80m and a class S, the class that includes men shorter than 1.20m. In this case, in order to classify a 1.70m-tall man, it is necessary to create a new class, M that includes all the elements not in S and T. The main limitation is that the class M is too large and includes all the man with $height > 1.20$ and $height < 1.80$. It is possible to be more precise by iteratively add new classes, but the class definition could be complicated, affected itself by uncertainty and vagueness, and the loop of class creation could be potentially infinite.

The classification under uncertainty and vagueness is not only a linguistic game or a philosophic problem but it has also practical engineering implications. For example, a pump could not only be 'old' or 'new', but it could be defined with several (potentially infinite) adjective classes to express an intermediate condition between 'old' and 'new'. A street could be 'obstructed' when traffic does not let the cars to move and 'free' when no car occupies the road, but it is possible to identify infinite classes to describe intermediate conditions. Zadeh, (1965), therefore, proposes a methodology, called fuzzy logic, that successfully deals with this kind of problems and this is presented in the following subsections.

2.5.1 Fuzzy sets and Crispy sets: general concepts

Starczewski, (2013) proposed an enlightening mathematical description of fuzzy logic and, in this section, the definitions are extracted from his book while the examples are original.

There are five steps to use the fuzzy logic (Sivanandam, Sumathi, and Deepa, 2006, page 121):

1. fuzzify the inputs;
2. identify the fuzzy rules;
3. combine the inputs and the rules to calculate the truth degree of each rule;
4. identify an output distribution;
5. defuzzify the output distribution to obtain a crisp value (optional).

In figure 2.10, a classical block-schema of fuzzy logic is reported. The input fuzzification is performed by the fuzzifier, the rule-based knowledge storage consists of the rule base, the inference engine combines fuzzified input and identifies the output distribution, while the defuzzifier produces a synthetic crisp output.

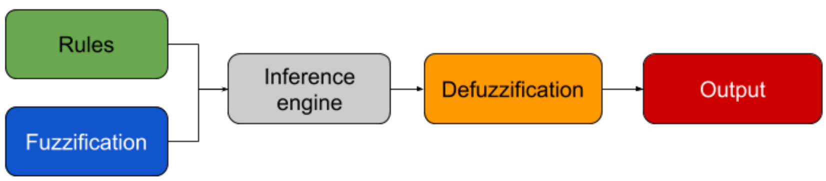


FIGURE 2.10: Fuzzy logic system, adapted from (Starczewski, 2013), page 138

These aspects will be discussed with a rigorous mathematical approach in this section and better explained with a numeric example in section A .

Membership Degree and fuzzification

Definition: "Let X be a non-empty set. A fuzzy set A in X is characterised by its membership function" (Starczewski, (2013), page 1)

This means that, for each element of A, there is a membership function that defines its degree of membership μ to the classes, also called linguistic variables (equation 2.7).

$$\begin{cases} \mu \in [0:1] \\ \mu = 0, \text{the element does not belongs to the class} \\ \mu = 1, \text{the element entirely belongs to the class} \\ 0 < \mu < 1, \text{the element partially belongs to the class} \end{cases} \tag{2.7}$$

For example, let's assume 2 fuzzy sets: 'Young' and 'Old'. A newborn belongs to the class 'Young' with $\mu = 1$, and he belongs to the class 'Old' with $\mu = 0$. A 80-year-old man belongs to the class 'Young' with $\mu = 0$, and to the class 'Old' with $\mu = 1$. A 35-years old man belongs partially to the class 'Young' and partially to the class 'Old' (for example $\mu_{Young} = 0.65$ and $\mu_{Old} = 0.35$). The attribution of the membership values depends on the membership function definition. Figure 2.11 shows an example of membership function in which there are 2 classes (Young and Old) and the age (on the x-axes) enables , for each class, the calculation of the corresponding value of the membership factor (on the y-axis).

Definition: "A kernel of A, being a fuzzy subset of X, denoted by ker (A), is the ordinary subset of X whose all elements have membership grades equal to unity in A" (Starczewski, (2013),page 2).

$$ker(A) = \{x \in X | \mu_A(x) = 1\} \tag{2.8}$$

In the example of figure 2.11, the kernel of the class Young corresponds to the region between 0 and 10 years of age. In this region of age, the individuals are 100% belonging to the class 'Young'.

Definition: "A support of A, being a fuzzy subset of X, denoted

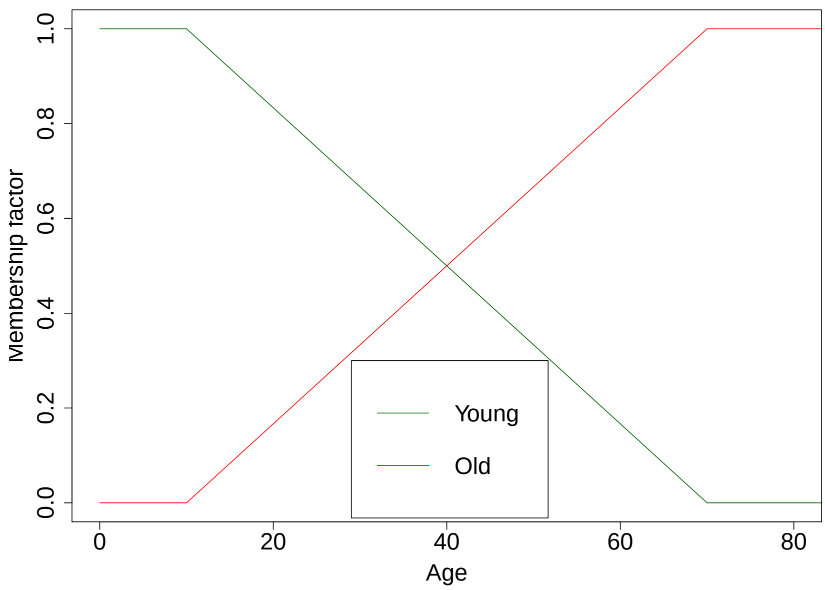


FIGURE 2.11: Membership function: example. There are 2 classes(Young and Old) and the age (on the x-axes) enables , for each class, the calculation of the corresponding value of the membership factor (on the y-axis).

by support (A), is the ordinary subset of X whose all elements have positive membership grades in A" (Starczewski, (2013), pag.2).

$$support(A) = \{x \in X | \mu_A(x) > 0\} \tag{2.9}$$

In figure 2.11, the support of the class 'Young' corresponds to a range of age between 0 and 70 years. In this range, the class 'Young' is still represented even if, when approaching the age of 70 years, the values of the membership factor are close to 0.

An input can be transformed in pairs of linguistic variables and membership factors. So for example, the age 35 could be read in fuzzy logic as ('Young',$\mu = 0.65$) or ('Old',$\mu = 0.35$). This transformation process of the input (in the example the age) in fuzzy language is called *fuzzification*.

In conclusion, the use of membership degree enables the representation of sets affected by vagueness and uncertainty with continuous functions, without increasing the number of classes. Zadeh, 1988 proposed a mathematical approach to process fuzzy information using the membership degree. This approach will be presented in the following subsections.

Rule base

The fuzzy algorithms require human-language-like rules to process the information at the input. A rule is composed by an IF/THEN structure with one or more input variables , and one variable at the output. For example, a set of fuzzy rules [1] could assume this form:

$$\begin{cases} \text{IF car\_age IS new AND motor IS high\_power THEN price IS high} \\ \text{IF car\_age IS new AND motor IS low\_power THEN price IS medium} \\ \text{IF car\_age IS old AND motor IS high\_power THEN price IS medium} \\ \text{IF car\_age IS old AND motor IS low\_power THEN price IS low} \end{cases} \tag{2.10}$$

In this case, there are two inputs (car age and motor typology) and one output (car price). In order to use the fuzzy rules, the input needs to be fuzzified as explained in the section *Memebership Degree and fuzzification*.

Rule implication: calculation of rule truth degree

The rules are based on a set of logic operators that connect the fuzzified input: AND, OR,NOT. Table 2.8 reports some common alternatives to define AND and OR logical operators with the probabilistic and the Zahed' approach. In fuzzy algorithms, this choice is customizable. The operator NOT, not reported in table 2.8 is always defined as:

$$\overline{(\mu_A(x))} = 1 - \mu_A(x)$$

The use of fuzzy operators to analyse a rule is generally called *implication*. With this operation, a truth degree is conferred to each rule. This truth degree expresses how much, according to the given inputs, the rule represents the condition of the system under analysis.

[1] The set of rules (2.10) is invented by the author of this thesis for the numerical example in Appendix A

TABLE 2.8: Operator in fuzzy sets -extract from (Dernoncourt, 2013)

Name	AND-Intersection	OR-Union
Zahed Operator	$\mu_{A\cap B}(x) = min(\mu_A(x), \mu_B(x))$	$\mu_{A(x)\cup B(x)} = max(\mu_A(x), \mu_B(x))$
Probabilistic	$\mu_{A\cap B}(x) = \mu_A(x) * \mu_B(x)$	$\mu_{A(x)\cup B(x)} = \mu_A(x) + \mu_B(x) - (\mu_A(x) * \mu_B(x))$

Identification of rule consequence

The rule consequence is the part of the rule after the 'THEN' logical operator. For example, in the system of fuzzy rules (2.10), 'price IS high' is the consequence of the first rule. The fuzzy logic algorithm needs a mathematical definition of the consequence. This is generally performed with two alternative approaches (Sivanandam, Sumathi, and Deepa, 2006, pag.119):

- Mamdani's fuzzy inference method
- Takagi–Sugeno–Kang inference method (often referred as Sugeno)

In the Mamdani method, as for the input, the rule consequence is a fuzzy set. Figure 2.12 shows an example of the fuzzy output for the car price of the set of rules in equation 2.10).

In the Mamdani method, the output result is the result of a linear equation depending on the input variables.

In the case of Sugeno approach, the rules are a bit different. For example, the first rule of the example 2.10 could assume this form:

IF car_age IS new AND motor IS high_power THEN price IS f(car_age,motor)

The element f(car_age,motor) is a function depending on car age and motor type, for example, like in equation 2.11 [2], in which 'motor' is the motor engine power expressed in hp and the car age is expressed in years.

$$Price = f(car\_age, motor) = 100 * motor - car\_age * 2000 \quad (2.11)$$

[2] Invented by the author of this thesis for demonstration purpouses.

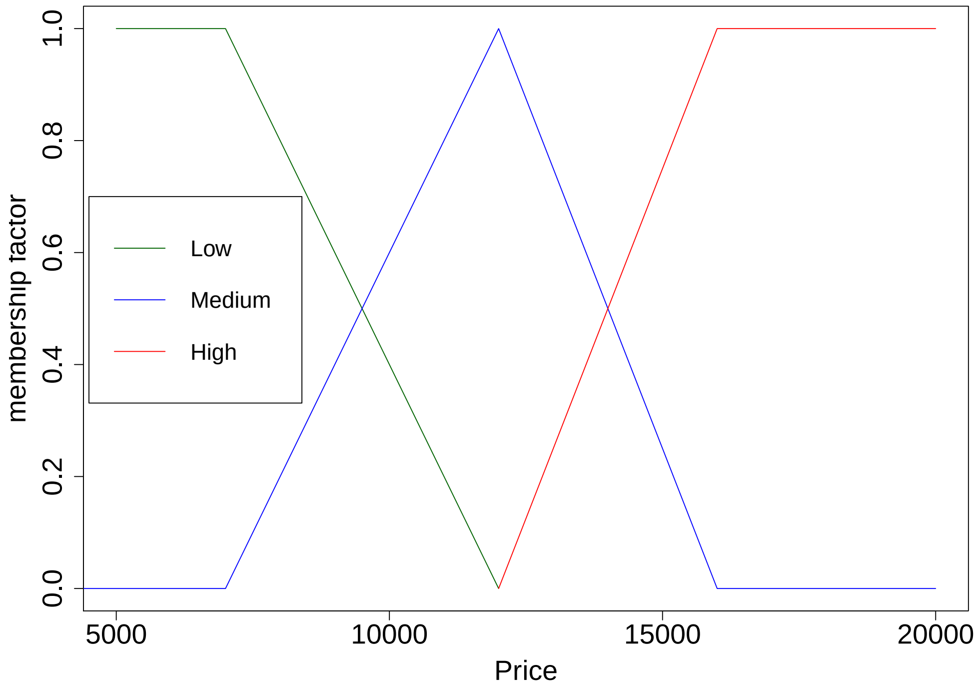


FIGURE 2.12: Example: Mamdani method for the definition of rule consequences

Defuzzification

When a degree of truth is defined for each rule, it is necessary to aggregate these results to have an overall representative value of the output. This process, called defuzzification, can be performed through two algorithms:

- mean of maxima (MeOM), generally associated only with the Mamdani approach;
- method of centre of gravity (COG).

The output, calculated as mean of maximum MeOM, is the result of equation 2.12, in which z values are the values of the output for which the truth degree of the rules is maximum and 'l' is the number of the rule(s) with higher membership factor. In the simplest case, there is just one dominant rule and l=1.

$$MeOM = \frac{\sum_{j=1}^{l} z_j}{l} \tag{2.12}$$

The COG is calculated as weighted average between the outputs of the rules weighted on their truth degrees. In equation 2.13, z_j is the j-value of the output and $t_j(z_j)$ is the truth degree of the rules associated to z_j. In the equation 2.13, 'j' represent the j-equation.

$$COG = \frac{\sum_{j=1}^{N} z_j * t_j(z_j)}{\sum_{j=1}^{N} t_j(z_j)} \tag{2.13}$$

2.5.2 Short considerations about fuzzy logic

Section 2.5 has shown how fuzzy logic can be used to process variables affected by uncertainty and vagueness. Moreover, fuzzy logic is able to store and process information with a human-like-language structure. Because of these characteristics, fuzzy logic methodology has been used in many domains and it is proposed as core methodology of the SK-DSS. A step-by-step numerical example for the reader interested in better understanding the methodology is available in Appendix A.

2.6 Random Forest

In this thesis, Random forest (RF) was used for regression purposes, in particular to estimate the missing values of COD after intensive comparison with other approaches (Torregrossa et al., 2016). RF is a popular technique for machine learning proposed by Breiman, 2001. This algorithm belongs to the class of 'supervised learning', i.e. the algorithm outputs are compared with references values. Random Forest can be used for classification and regression problems. The input of the algorithm is called 'training data' composed of the target variable (Y) and the independent-variables (X). The objective is to create a model that using X is able to predict or classify Y. If Y is composed by qualitative categories, it is a classification problem. If Y is composed by real numbers, it is a regression problem. In this section, an introduction to RF is provided, alongside some mathematical aspects. A more detailed explanation of this algorithm is provided by (Hastie, Tibshirani, and Friedman, 2009), which is also used as reference for this part of the thesis.

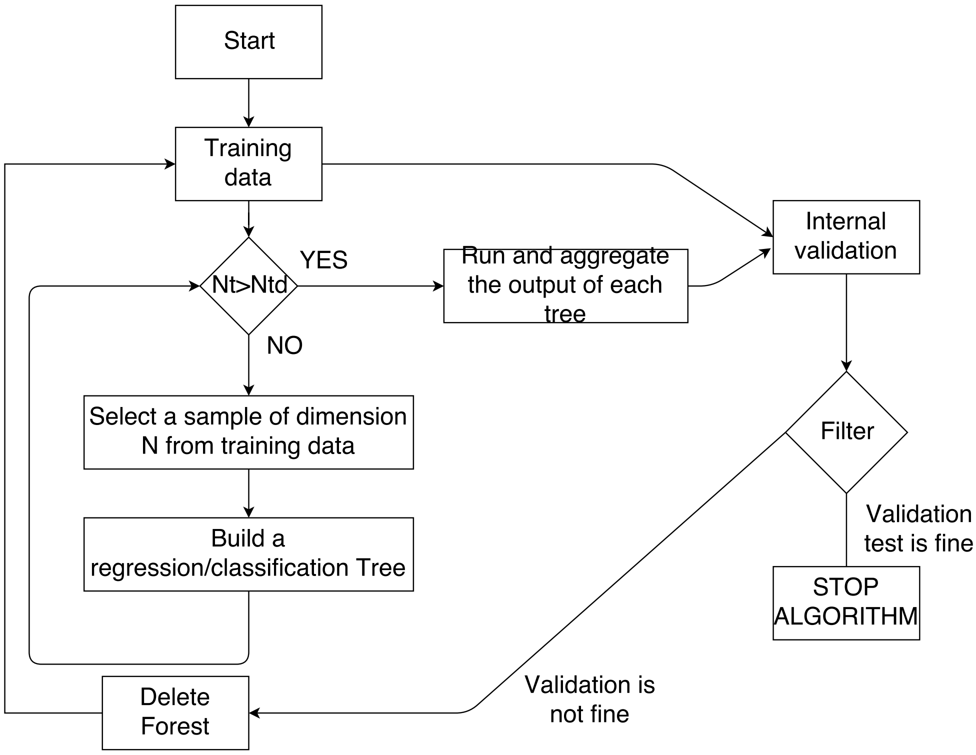


FIGURE 2.13: Procedure to train a Random Forest

Figure 2.13 shows the procedure to train a RF. The algorithm starts by uploading the dataset. After that, the following procedure is performed:

1. the first operation consists of checking if the number of decision trees (Nt) is higher than the desired one (Ntd). Ntd is customizable and the value has to be chosen for each specific problem (refer to section 2.6.2 for over-fitting issues);

2. until Nt>Ntd, the algorithm continues to build trees using subsamples of the training data. At the end of this procedure we have a set of trees, called 'forest'. The forest is random because the trees are generated with a random selection of subsets from the training data. Hence the name "random forest";

3. each tree is able to produce an independent output. The next step of the procedure consists of aggregating these output; generally, the aggregation is done by voting for classification and by mean for regression.

4. in order to check the quality of the response an internal validation is performed using training data. Generally, the evaluation is performed using the average error or the Root Mean Squared (RMS) Value. If the validation test gives a positive result, the algorithm stops, otherwise the forest is deleted and the process start again from point 1.

2.6.1 Variable importance

Another important use of RF is the estimation of input variable importance: the algorithm identifies which of the variables of the X dataset contributes more to the estimation of Y. Liaw and Wiener, 2002 proposed 2 methods to calculate the variable importance:

- testing the changes in prediction quality after permutation of variables in the decision trees of the forest;
- calculating the total decrease in node impurity produced by each parameter; the node impurity is the Gini index for classification and the sum of squared residuals for the regression.

These approaches are specific to Random Forests. As an alternative, analogously to other regression algorithms, a classic sensitivity analysis can be performed; the main difference is the following: the first two approaches use the available database and the effect of changes in RF structure to identify the variable importance (fixed input data, variable RF structure); in contrast, the sensitivity analysis consists of the generation of response curves by modifying the input data of a fixed regression model (variable input data, fixed structure).

This feature is important to generate new knowledge from the dataset. For example, in (Torregrossa et al., 2016), it was used to identify which parameter has a larger influence on the COD load concentration at the inlet of WWTPs to have a better comprehension of the interdependencies.

2.6.2 Noise, over-fitting and warnings for operators

As explained before, the RF procedure is based on a random selection of a data subset from training data. With large databases, if the largest part of the parameters are not relevant for the analysed problem, the probability to have inefficient trees in the forest is high. In simplest words, a big portion of the forest trees could be built relying on unimportant information. When this happens, the random forest is not efficient and the quality of the results is poor. In this case the simple use of a machine learning algorithm does not per se guarantee any meaningful result and an adequate knowledge of the physical problem is required to select the parameters and understand the results (Hastie, Tibshirani, and Friedman, 2009). Another common problem in machine learning is over-fitting. This phenomenon occurs when the generated model is dataset-specific and it is unable to efficiently work with new data. In the case of random forest, model over-fitting seems to be a limited problem, even if it cannot be ignored (Hastie, Tibshirani, and Friedman, 2009). In the case of RF, over-fittings depends on the number of trees chosen for the model; an extremely high number of trees can produce over-fitting. The adequate number of trees must be carefully selected and a post-evaluation of the model has to be done by experts in the field.

In conclusion, to reduce or avoid the risks of noise and over-fitting, an adequate knowledge of the field is always required. It is not possible to apply this technique (such as other machine learning algorithms) with scientific rigour, if an adequate comprehension of the phenomena under analysis is missing.

2.7 Final considerations

In this chapter, the main areas of interest for this thesis were introduced and discussed:

- technical aspect of WWTP technology;
- energy aspect in WWTPs;
- decision support systems;
- fuzzy logic methodology;
- random forest algorithm.

The wastewater treatment domain is characterised by relevant and complex problems. Moreover, decision makers have to deal with conflicting objectives and with a wide set of parameters affected by uncertainty. The WWTP energy efficiency is one sub-domain that inherits all these features. These characteristics make this domain interesting for the application of decision support systems. In fact, DSSs (and in particular the environmental DSS) perform well with semi-structured problems, characterised by multiple potential solutions and a large number of uncertain parameters. Decision support systems rely on multiple methodologies. Subsection 2.5 illustrates how fuzzy logic can be suitable to be coupled with decision support systems because of converging characteristics: the abilities to deal with multi-parameters scenarios, process uncertainty, store and use expert knowledge.

In conclusion, this chapter provided the basis to understand the methodology used in this thesis, the motivation to invest time and resources in WWTP energy optimisation, and the justification for the selection of the technologies adopted in this project. Chapter 3 will focus on SK-DSS methodology, in particular presenting the SK-DSS as a block diagram in which the global view is explained.

Chapter 3

A plant generic cooperative decision support system

LEGAL DISCLAIMER: The present chapter partially reproduces research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of the EdWARDS project.

The proposed decision support system aims to provide decision-makers with a manageable set of information obtained by analysing environmental, energetic, technical and economic data. This will be explained in detail in sections 3.1-3.1.9 :

- the fuzzy logic engine (Zadeh, 1965) is the core of this DSS;
- the fuzzy logic engine applied, not to raw process data, but key performance indicators;
- benchmarks are calculated for each plant and for given operational condition through the use of benchmark equations;
- a set of fuzzy rules is generated according to the availability of sensors installed in the plants.

Consequently, this DSS is able to treat data from various WWTPs regardless the layout, the plant size, the technology and the data availability. The fuzzy logic engine is able to produce a case-specific plant

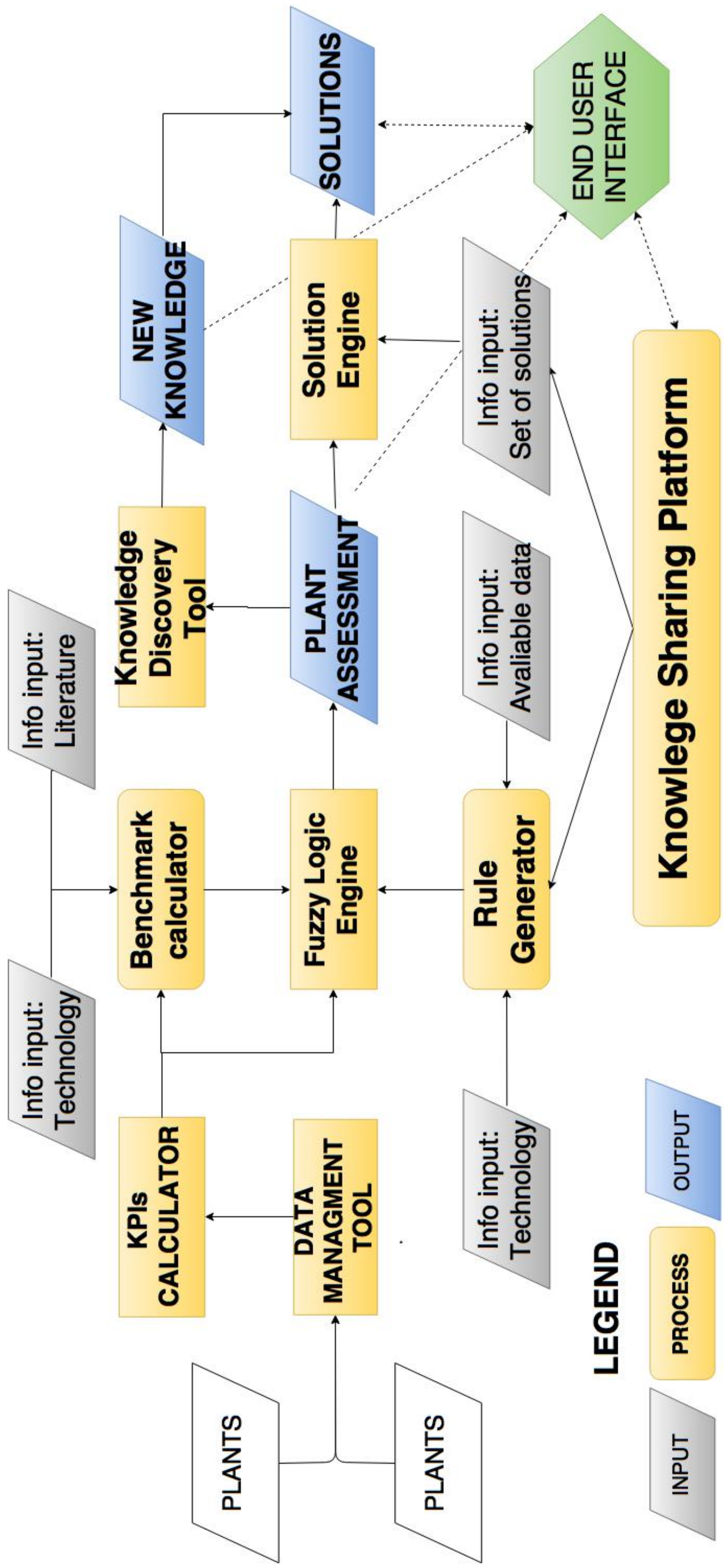


FIGURE 3.1: SK-DSS architecture

assessment, that is used to provide solutions. The reader will find a detailed explanation with numerical examples in chapter 5. Another important characteristic of this DSS is that it is based on a cooperative platform in which the end-user can update and share the solutions for a given problem. This characteristic is important, because it fills a gap in literature: in fact,to the best of the knowledge of the author of this thesis, before the publication of (Torregrossa et al., 2017a), a cooperative decision support system for energy saving in WWTPs did not exist. This is also the reason of the name given to this tool: Shared Knowledge Decision Support System (SK-DSS).

3.1 The proposed model

NOTE: The section 3.1 is mainly extracted from (Torregrossa et al., 2017a).

The SK-DSS architecture is shown in fig. 3.1. In the SK-DSS, various processes are interconnected to sequentially perform all the required operations: from rough data processing to the plant performance assessment and case-base solutions [1]. The system is composed of multiple tools: a data management tool, a KPI calculator, a benchmark calculator, a rule generator, a fuzzy logic engine, a solution engine and a knowledge discovery tool.

Table 3.1 shows the flow of information between the SK-DSS processes. This table distinguishes the processes in: set-up, preliminary, core and tertiary processes. Set-up processes are concerned with the connection of the plant to the system, the data processing, the nomenclature normalisation (data management tool) and the expert knowledge acquisition. The SK-DSS preliminary processes prepare the information required by the fuzzy logic engine: in particular the KPI calculator produces the KPIs to be evaluated, the benchmark calculator prepares the benchmark used to define the fuzzy model and the rules generator organises the information retrieved from the shared knowledge platform as fuzzy rules.

The fuzzy logic engine combines the KPIs, the benchmark values and the fuzzy rules to identify the plant operational conditions and to

[1] In the SK-DSS, the term 'solution' expresses the same concept of 'suggestions'. The two terms are interchangeable.

TABLE 3.1: Flow of information in the SK-DSS model

Process name	Input	Output
	Set-up	
Data management tool	Real-time data	Pre-processed dataset
Knowledge sharing platform	Human Knowledge	Case-base scenarios, solutions, Technological info
	Preliminary Processes	
KPI calculator	Pre-processed dataset	key performance indicator (KPI)
Benchmark calculator	Technological info	Benchmarks
	KPI	
Rule generator	Technological info Data availability info Case-base scenarios	Fuzzy rules
	Core Process	
Fuzzy Logic engine	KPI Benchmarks Fuzzy rules	Global assessment index; Rule truth degrees
	Tertiary Processes	
Knowledge discovery tool	Rule truth degrees	Most frequent scenarios
Solution engine	Most frequent scenarios, solutions	

produce an overall performance index. The fuzzy logic engine is the heart of the SK-DSS. A fuzzy methodology is proposed because:

- it enables plant managers to share knowledge in a human-like language;
- it is able to process and evaluate data by processing uncertainty;
- it can be combined with KPIs to produce a plant generic analysis;
- the analysis of the truth degree of the rules can be used as input for a case-based reasoning.

Other possible approaches, such as decision trees, cannot provide all of these desired features. The tertiary processes use the output of the fuzzy logic to provide a list of potential solutions and increase the comprehension of the plant operational conditions.

The data management tool and the calculation of KPIs are described in detail in Chapter 4. The benchmark calculator, the rule generator, the knowledge discovery tool, the fuzzy logic engine and the solution engine are explained in detail in chapter 5.

3.1.1 The data management tool

The data management tool receives data files on a daily basis from each WWTP (generally in .csv or .xml format) with the values generated by the plant sensors during the previous day. The files received have a plant specific nomenclature, a specific unit of measurement and a specific measurement time interval (for example some measurements are executed each 30s and others each 2h). The data management tool processes these files by extracting and validating the data, normalising to a standard nomenclature, converting the units of measurement and finally producing for each sensor a representative daily value. The data management tool is fully described in Chapter 4.

3.1.2 Shared Knowledge Platform

The Shared Knowledge Platform is an on-line database, that stores the expert knowledge as fuzzy rules and, in parallel, stores the set of solutions used by the solution engine. It is a very flexible tool, that enables plant managers to upload, share and access information. Consequently, each plant manager both ’produces’ and ’uses’ the knowledge which is shared. A first version of this shared knowledge platform relies on PostgreSQL database, while a new version of this tool relies on the YouTube platform (Torregrossa and Hansen, 2018). The use of the shared knowledge platform is described in Chapter 5.

3.1.3 KPI calculator

The KPI calculator processes the dataset prepared by the data management tool and produces KPIs that enable the comparison between different plants. For example, in the SK-DSS the use of specific energy (kWh/pe) allows the comparison of plant energy consumption regardless of the size of the plant. The KPIs calculation is performed according to the methodologies defined in the INNERS project (cf. INNERS report (INNERS, 2015) and in (Torregrossa et al., 2016)). In addition, SK-DSS can complete the plant data by estimating missing data. After this stage of the process, the information is harmonised and conformed with SK-DSS standards. Detailed information can be found in Chapter 4, in which the full process from data gathering to KPI calculation is explained in detail.

3.1.4 Benchmark calculator

In the benchmark analysis, each KPI is compared with the respective reference value (the benchmarks). The process for calculating the benchmark values is dynamic and takes into consideration human knowledge, literature values, plant technology, uncertainty and data availability. This is achieved with a set of benchmark equations for each KPI. The SK-DSS processes the available information for each plant and automatically selects the correct equations. These equations depend on operational values (such as: pollutant load, temperature or sludge age) and on the uncertainty inherent in the estimated values (trough the coefficient of variation, (Torregrossa et al., 2016)). Subsequently, the resulting benchmark values are used in the fuzzy logic engine. In chapter 5, the set of equations for blowers, pump energy consumption and the biogas production are presented and the entire procedure is explained in detail.

3.1.5 The rule generator

The fuzzy logic engine requires a set of IF/THEN rules to perform its analysis (Williams, 2009). Each rule describes a condition of the observed unit[2], which is specific for a given technology and a given set of available KPIs.

In order to maintain a plant-generic approach even with different sets of available information, the SK-DSS enables the selection of a set of rules with many options. The plant managers can select the set of rules, for example, according to sensor availability. At the current stage, the plant managers can also propose new rules, that are evaluated and validated with the platform and finally included in the selection option.

3.1.6 The fuzzy logic engine

Fuzzy logic methodology (Zadeh, 1965; Sivanandam, Sumathi, and Deepa, 2006; Starczewski, 2013) is currently used in different environmental applications for its flexibility in dealing with uncertainty and complex

[2]For the candidate "observed unit" is the object of the analysis. The observed unit can be the overall plant, a single stage or an aggregation of devices (like the blowers or the pumps).

phenomena (for example in (Wang et al., 2016; Chen and Lee, 2003; Castro, Paulo Carvalho, and Ribeiro, 2011)).
The SK-DSS fuzzy logic engine takes as input: set of rules, benchmarks and the KPI values to assess the observed units. The output of the engine is a performance score, inferred by the fuzzy logic algorithm for each unit. This performance score ranges between 0 (the unit performance is low) and 10 (the unit performance is good). Furthermore, the fuzzy logic engine produces, for each rule, a 'degree of truth', which, depending on the definition used, may also be called 'accuracy'. Each rule is a statement on the condition of the plant, the respective result corresponding to their degree of truth. The rule degree of truth ranges between 0 (the statement is false) and 1 (the statement is true). The set of rules describes the possible conditions of the observed unit and using the degree of truth, it is possible to identify the most relevant rule for a given plant at a given day. The rules and their degrees of truth are aggregated to produce the performance score. An extensive introduction to fuzzy logic is available in section 2.5.

3.1.7 Knowledge Discovery Tool

The Knowledge Discovery tool aims to retrieve useful information from WWTPs in order to enhance the decision-making process. Basic statistical analysis can be performed on the rough data to have a synthetic representation of the WWTPs and to produce reports. Such a statistical analysis is commonly performed on WWTPs data. Here, in contrast, a statistical analysis is performed on the output of the fuzzy logic engine. In the SK-DSS, this analysis is focused on the degrees of truth of each rule, which are stored in a database. Each rule, inherently, represents a condition of the plant. The calculation of statistical indexes (like the average) can provide useful information concerning the importance of each rule for a given plant. For example, a rule with a high average degree of truth describes a condition that is often verifiable in the plant. Chapter 5 presents numerical examples.

3.1.8 Solution Engine

The inputs of the solution engine are the results of the fuzzy logic engine and the set of solutions defined for each rule. The Knowledge

Discovery Tool produces for each rule a degree of truth, i.e. the likelihood that the operation condition described in the rule is true. The plant manager can visualise the most probable operational condition(s) of the plant under assessment and access the specific set of solutions.

3.1.9 End User Interface

As shown in fig. 3.2, multiple WWTPs can be connected simultaneously to the decision support system. The information flow is bi-directional: data and shared knowledge flow from the plants to the SK-DSS while plant assessments and lists of suggestions travel in the opposite direction. The web interface enables the end user to monitor this flow of information, visualize the outputs of the decision support system and access the shared knowledge platform. The reader can test the current version of the end user interface at this website https://dario-torregrossa.shinyapps.io/Ver2/.

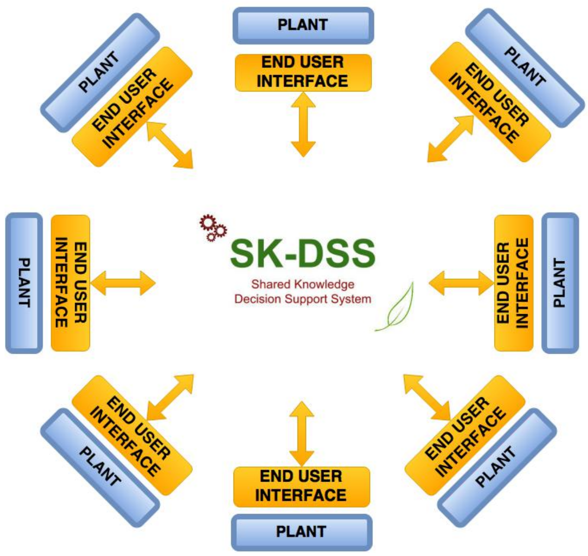


FIGURE 3.2: Network Structure

3.2 Conclusion

In this chapter the model of an innovative DSS was proposed and described in subsections 3.1-3.1.9. In particular, this chapter shows how the SK-DSS is a flexible tool, able to process key performance indicators, suitable for information sharing and able to deal with emerging challenges. The specific aspects of this tool will be discussed in the following chapters. In particular, chapter 4 will discuss the process from data gathering to KPI calculation. Chapter 5 will show the application of SK-DSS for pumps, blowers and biogas production. Moreover, Chapter 5 will present the implementation of the solution engine as a YouTube based cooperative platform. Chapter 6 describes in detail the use of the web-interface.

Chapter 4

SK-DSS in details: from plant data gathering to KPI calculation

LEGAL DISCLAIMER: The present chapter partially reproduces research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of the EdWARDS project.

WWTPs are complex facilities: the interaction of several processes driven by multiple parameters is determinant for the plant performance. This requires to monitor and control a large set of information.

This thesis adopts a classification by distinguishing environmental, process, design, and device parameters. *Environmental parameters* are those not depending on plant operation, for example: inflow rate, external temperature, pollution load. Environmental parameters are given by external conditions and they cannot be influenced by decisions. *Process parameters* are those related to the plant performance: for example the removal rate of pollutant concentration or the sludge production. Generally, process parameters are those to be optimized. *Design parameters* include plant 'static' information such as the layout of the plant, the technologies and the size of the tanks. Design parameters are considered 'static' from the operational point of view because they do not change on daily basis; obviously, plant design parameters can change because of plant updates. *Device parameters* are those connected with the operation of the facilities installed in WWTPs. For example, in a pump system, it is possible to measure the rotation per minute, the energy consumption, the pump temperature or even the vibration.

This classification has the objective to drive the reader into the complexity of the WWTP information set. Moreover, some parameters can fit with many classes: for example, the biogas-energy can be classified either as process parameter or device parameter (being an output of the Combined Heat and Power engines).

Another interesting and critical aspect of data management in WWTPs is the amount of information. This can varies according to the size of the plant and to the number of sensors installed. A well-equipped WWTP can produce an amount of values not manageable by plant operators without an IT support; for example, in the WWTP of Solingen-Burg, 200.000 measures per day are registered and stored in the database (Torregrossa et al., 2016). Consequently, it is necessary to process these information sets and produce meaningful synthetic parameters that can be easily used to support the decision-making process.

The aim of this thesis is to produce a plant-generic decision support system, i.e. able to simultaneously treat different WWTPs regardless of their size and technology. This requires that the database is plant generic, i.e. the database should contain comparable information. This chapter explains how this is achieved in the framework of Shared-Knowledge Decision Support System. Section 4.1 will explain how the dataset is produced, section 4.2 explains how to build a plant-generic database, section 4.3 explains how to estimate missing data, section 4.4 will introduce the calculation of KPIs, and section 4.5 will explain how the algorithm is automated.

4.1 Data gathering

The first step of the data processing is the data gathering that corresponds to the operations required to measure the parameters and store these values as a database record. In SK-DSS, there are 3 typologies of data gathering:

- remote sensing, that provides information measured on line (for example energy values);

- laboratory analysis, that provide information about waste water characteristics (such as the COD);

- manual filling of static data (such as the size of the tanks).

Subsection 4.1.1 explains how values are gathered from SCADA systems, subsection 4.1.2 describes the use of laboratory data and subsection 4.1.3 concerns the static data process.

4.1.1 Remote sensors and SCADA Systems

According to Bailey and Wright, 2003, Supervisory Control and Data Acquisition (SCADA) systems "refers to the combination of telemetry and data acquisition. SCADA encompasses the collecting of the information, transferring it back to the central site, carrying out any necessary analysis and control and then displaying that information on a number of operator screens or displays."

In the last years, SCADA systems became really popular in WWTP domain, but the stored datasets are still largely under-used (Torregrossa et al., 2017a).

In WWTP domain, a first attempt to produce a plant-generic database focussed on energy measurement has been done in the framework of INNERS project (INNERS, 2015). This INNERS database was called EOS (Energy Online System) and can be considered the 'father' of SK-DSS database. EOS was able to get, store, process and normalize data from WWTPs and create a daily plant-generic database. This was the starting point of this PhD project.

EOS suffered from data management issues because it was programmed to import all the information produced in the WWTPs, regardless their final use. This had negative consequences because:

- useless information was stored;

- the storage of the result of high-frequency sensors (1 value each 15 seconds) increased the size of the database;

- the extraction of the information from the database was inefficient and slow.

As consequence, the system collapsed when an additional plant has been added. SK-DSS database is the evolution of EOS database. It inherits the concept, but it stores only the information required for decision support: the daily data aggregation is performed on-the-fly and only the daily value is saved for each sensor.

This means, for example, that if a sensor produces 1440 values for a day (1 observation each minute), SK-DSS aggregates on-the-fly this value and stores only the daily aggregation: the computational advantage is evident.

Moreover, the high-frequency information is stored in the files that SK-DSS receives each day from WWTPs and, if necessary, it could be retrieved. At the moment, this aggregated information is sufficient because SK-DSS works at a daily time-resolution. Increasing the time-resolution is out of the scope of this thesis because:

1. high-time resolution benchmark and KPIs are not available;
2. a 1-day decision time is a good time-frame for decision support systems (a higher time resolution would suggest moving to automatic systems).

The diagram flow of SK-DSS database is illustrated in figure 4.1. The WWTP is composed of several stages; each stage can be equipped with sensors that observe the operational conditions (such as devices energy consumption, pH, and temperature). Each sensor is programmed to work with a specific frequency and with a specific unit of measurement. The SK-DSS requires a dataset normalization. In section 4.2, a detailed explanation of aggregation process is described. For the moment, it is important to consider that WWTP sensors produce datasets that can be different for time resolutions, units of measurement and nomenclature.

4.1.2 Laboratory analyses

The laboratory analyses are executed, generally at a regular time, in order to assess the water quality at the specific sampling point of WWTPs (for example inlet, outlet, digester). After the analyses, generally, the lab operator has to insert manually the results in the database. The main limitation of these analyses are:

- the cost and consequently the low sampling frequency;

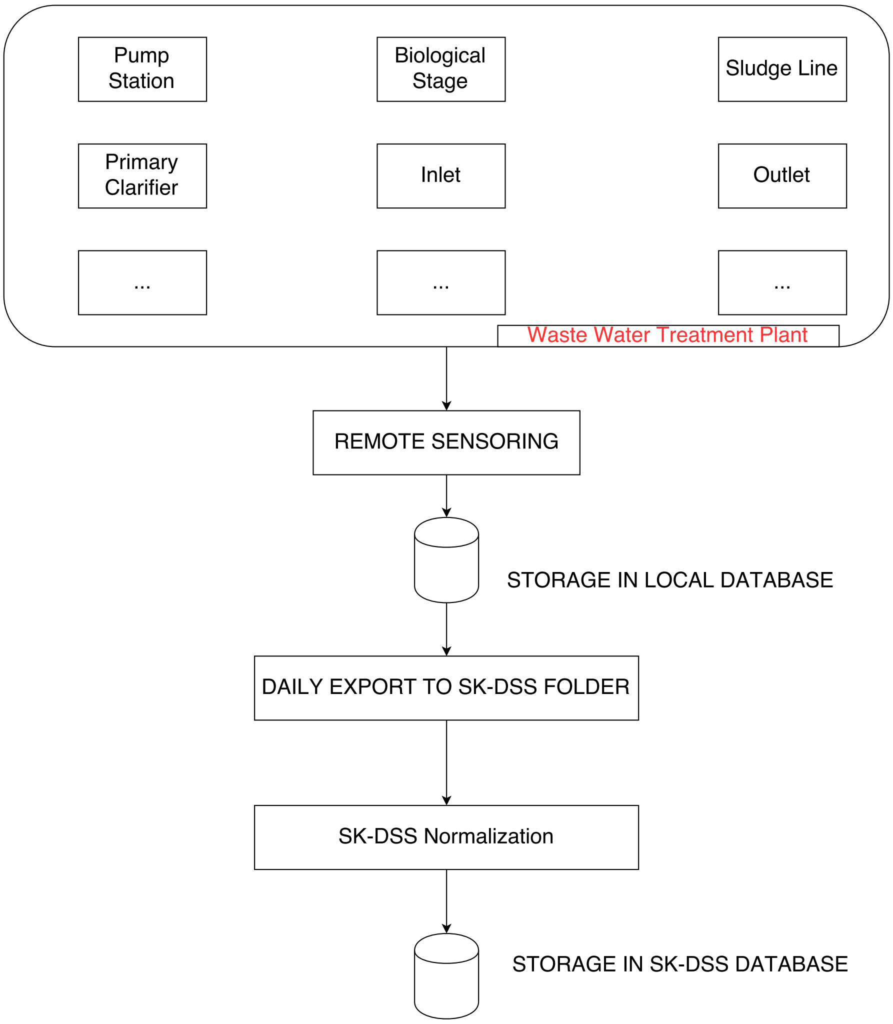


FIGURE 4.1: Diagram of SK-DSS database

- the time delay from the sampling and the availability of the value on the system.

These limitations lead to the unavailability of laboratory analysis for most of the days of the year. For example, in the WWTP of Solingen-Burg (GER) there is a set of laboratory measurements for once each fifteen days, while for the WWTP of Hidden[1]-City (NL) a set of laboratory values is available for each week. In (Torregrossa et al., 2016) and in section 4.3, the author of this thesis demonstrated how to use the artificial intelligence techniques to estimate the daily missing values of COD concentration.

[1]Real name omitted for confidential standards

4.1.3 Static data

The static data does not change daily (for example the size of the tanks). For example, static data are the installation power of pumps and blowers, the volume of the tanks or the designed capacity. These values need to be inserted in the SK-DSS database only once before the first usage of the system and updated if the values change.

4.2 A plant-generic database

As mentioned before, the main characteristic of SK-DSS is its flexibility, i.e. the capacity to efficiently manage datasets with different characteristic. WWTPs connected to SK-DSS produce data in 'CSV' and 'XML' formats, they use different time-slots for observations, different nomenclature and different units of measurement. For example, figure 4.2 shows the structure of a XML file and figure 4.3 shows the structure of a CSV file for the same information. In order to process data coming from different databases, SK-DSS has to normalize the information. The normalization process is illustrated in figure 4.1:

- data is observed in WWTP and stored in a local database managed by plant managers;
- each day, an automatic export transfers the measurement of the day before to a folder hosted in the SK-DSS server; these files currently are in CSV or XLM formats, but there is not a specific requirement;
- each dataset contained in these files has to be processed in order to normalize the nomenclature, the time aggregation and the units of measurement. This is done with a specific script for each plant;
- the normalized information is stored to SK-DSS database.

In the subsection 4.2.1, the structure of the SK-DSS plant table is described.

4.2.1 Data processing and structure of the table

Figure 4.4 shows the content and the structure of the SK-DSS plant data table. This table contains:

- the oid, i.e. an unique reference number for the record;
- the date;
- the name formatted according to the SK-DSS nomenclature;
- the value associated to the record;
- the plant name;
- the unit of measurement.

The data format reports only the year, the month and the day. Compared to the raw data each hour reference is lost because each record represents the aggregated value for given day. The decision to have daily aggregation has been taken because some parameters (such as the sludge age or the COD concentration) have not an higher time resolution and SK-DSS requires a normalized time aggregation: the daily aggregation is the most detailed that can be satisfied by all the parameters, because it is necessary to adopt the lower value of the parameter sampling frequency.

Nomenclature explanation

The name of the sensors in SK-DSS respects this nomenclature:
$[stage]\_[sub{-}stage]\_[device]\_[mean]\_[aggregation]\_[(observedparameter)]$.

The first part of the name reports the stage; for example WWTP if the sensor refers to the full plant, BIO if the sensor refers to the biological stage, PCL for the primary clarifier.

The sub-stage sometimes is added to refer to a particular position of the stage. For example, it is possible to use BIO_INLET, BIO_OUTLET, BIO_TANK to identify the inlet, the outlet or the tank of the biological process.

After the sub-stage, the nomenclature convention sets the device name, for example: PUMP for pumps, BLO for blowers.

In the 4rd position, the nomenclature places the mean: WW for wastewater, AIR for air, SLU for sludge.

The aggregation consists of 2 option: AVG if the daily parameter values come from aggregation by mean, SUM if the daily parameter values are calculated by sum.

LISTING 4.1: Figure

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<plant operator="Demo" description="Klaranlage DEMO">
  <report aggregation="2h" version="1.0" date="15.08.2025">

    <pv type="R" id="165" name="Gaserzeugung Faulbehalter 1 " unit="Nm3/h" aggregation-type="S">
      <tw unit="Nm3/d" value="1597" />
      <min time="02:00:00" value="59" />
      <max time="22:00:00" value="89" />
      <mw time_slot="00:00-02:00" value="59" />
      <mw time_slot="02:00-04:00" value="62" />
      <mw time_slot="04:00-06:00" value="60" />
      <mw time_slot="06:00-08:00" value="59" />
      <mw time_slot="08:00-10:00" value="59" />
      <mw time_slot="10:00-12:00" value="61" />
      <mw time_slot="12:00-14:00" value="68" />
      <mw time_slot="14:00-16:00" value="65" />
      <mw time_slot="16:00-18:00" value="86" />
      <mw time_slot="18:00-20:00" value="64" />
      <mw time_slot="20:00-22:00" value="89" />
      <mw time_slot="22:00-00:00" value="67" />
    </pv>
  </report>
</plant>
```

FIGURE 4.2: Example of plant export in xml format

Sensor_id,	Plant,	Time,	Value,	Unit
Biogas,	DEMO,	04:00,	62,	Nm3.h
Biogas,	DEMO,	08:00,	59,	Nm3.h
Biogas,	DEMO,	12:00,	61,	Nm3.h
Biogas,	DEMO,	16:00,	65,	Nm3.h
Biogas,	DEMO,	20:00,	64,	Nm3.h
Biogas,	DEMO,	24:00,	67,	Nm3.h

FIGURE 4.3: Example of plant export in CSV format

	oid	date date	sensor_id text	value double precision	plant text	unit_of_measurement text
1	58894	2016-08-25	WWTP_SUM_VOL	272105.854249272	DEMO1	[m3]
2	58895	2016-08-26	WWTP_SUM_VOL	272106.019770432	DEMO1	[m3]
3	58896	2016-08-27	WWTP_SUM_VOL	273108.530421633	DEMO1	[m3]
4	58897	2016-08-28	WWTP_SUM_VOL	274107.873693738	DEMO1	[m3]
5	58898	2016-08-29	WWTP_SUM_VOL	275105.273765347	DEMO1	[m3]
6	58899	2016-08-30	WWTP_SUM_VOL	276099.615697698	DEMO1	[m3]
7	58900	2016-08-31	WWTP_SUM_VOL	277091.605977738	DEMO1	[m3]
8	58901	2016-09-01	WWTP_SUM_VOL	278080.751730479	DEMO1	[m3]
9	58902	2016-09-02	WWTP_SUM_VOL	279066.236596328	DEMO1	[m3]
10	58903	2016-09-03	WWTP_SUM_VOL	280049.489181286	DEMO1	[m3]
11	58904	2016-09-04	WWTP_SUM_VOL	281029.937376283	DEMO1	[m3]
12	58905	2016-09-05	WWTP_SUM_VOL	282006.685722833	DEMO1	[m3]
13	58906	2016-09-06	WWTP_SUM_VOL	282980.889760415	DEMO1	[m3]
14	58907	2016-09-07	WWTP_SUM_VOL	283951.954320925	DEMO1	[m3]
15	58908	2016-09-08	WWTP_SUM_VOL	284919.145939317	DEMO1	[m3]
16	58909	2016-09-09	WWTP_SUM_VOL	285882.540558848	DEMO1	[m3]
17	58910	2016-09-10	WWTP_SUM_VOL	286843.772787588	DEMO1	[m3]
18	58911	2016-09-11	WWTP_SUM_VOL	287800.14191226	DEMO1	[m3]
19	58912	2016-09-12	WWTP_SUM_VOL	288753.856966472	DEMO1	[m3]
20	58913	2016-09-13	WWTP_SUM_VOL	289703.733118952	DEMO1	[m3]

FIGURE 4.4: SK-DSS Plant data table structure

In the last position, the observed parameter indicates the object of the measurement. For example, COD for chemical oxygen demand, EA for the energy, DO for dissolved oxygen.

Not all the field are necessary to be completed. For example, the global energy consumption takes the following name: WWTP_SUM_EA. In this case, the sub stage definition, the device definition and the mean are not applied. A first nomenclature convention imposed to fill the not relevant information with NA. In this case the global energy consumption should be WWTP_NA_NA_NA_SUM_EA. For sake of simplicity and after testing that the compact nomenclature preserves the meaning, when possible, the compact nomenclature was adopted. As additional support, a sensor name dictionary can be easily generated. Table 4.1 propose an sub-sample of such a dictionary.

TABLE 4.1: Sub-sample of the SK-DSS dictionary

Name	Explanation
"BIO_INLET_NH4N"	NH4H measured at inlet of biological stage
"WWTP_INLET_WW_BOD"	BOD measured at the inlet of WWTP
"WWTP_OUTPUT_PO4"	PO4 measured at the output of WWTP
"WWTP_SUM_EA"	total energy consumption of WWTP
"BIO_BLO_EA"	Energy consumption of blowers
"WWTP_PE"	Population equivalent
"WWTP_PUMP_EA"	Energy consumption of the pump
"BIO_BLO_AIR_VOL"	Air volume provided by the blowers of the biological stage

TABLE 4.2: List of relevant KPIs

	Name	Explanation
1	KPI_BLO_EAperPE	Specific energy consumption of the blowers: energy per population equivalent
2	KPI_PUMP_EAperPE	Energy consumption of the pumps per population equivalent
3	KPI_PUMP_EAperVOL	Energy consumption of the pumps per volume of wastewater
4	KPI_SUM_EAperPE	Specific energy consumption of the WWTPs: energy on connected population
5	BENCH_BLO_EA_v1	Benchmark of blower energy consumption. V1: depending on the connected population
6	BENCH_BLO_EA_v2	Benchmark of blower energy consumption. V1: depending on connected population and sludge age
7	BENCH_BLO_EA_v0	14 kWh/y/pe. Constant value
8	INDEX_BIO_BLO_EA_v1	Index of energy consumption: KPI_BLO_EAperPE/BENCH_BLO_EA_v1
9	INDEX_BIO_BLO_EA_v2	Index of energy consumption: KPI_BLO_EAperPE/BENCH_BLO_EA_v2
10	INDEX_BIO_BLO_EA_v0	Index of energy consumption: KPI_BLO_EAperPE/BENCH_BLO_EA_v0
11	INDEX_BLO_AIR	Air index, measured air flow divided per theoretical air flow

4.3 Estimation of missing data

NOTE: The methodology presented in this section has been already published by the autor of this thesis in (Torregrossa et al., 2016) during the second year of this PhD project.

In order to have a plant-generic assessment, SK-DSS works with key performance indicators instead of unprocessed parameters: for example, it uses the specific energy consumption per PE [kWh/p.e.], instead of energy consumption [kWh]. Consequently, SK-DSS can be programmed to use this information with multiple plants, regardless of different size. In order to perform the daily KPI calculation, it is necessary to have a set of information without missing values. For example EN_{PE} (equation 4.2) is the KPI used in (Torregrossa et al., 2016) for the specific electrical energy consumption. SK-DSS uses the equations 4.1 and 4.2 where $FLOW_w$, COD_{conc}, EN are respectively the daily wastewater flow [$m3$/day], the average daily COD concentration [mg/l] and the electrical energy consumption [KWh]. The Population Equivalent (PE) is related to the pollutant load and approximates the number of connected people [pop]. For ease of analysis a load coefficient to convert COD concentration into PE is often used. This load coefficient is country specific and the coefficient used here is f=120 gCOD/pop/day. In alternative it is possible to calculate the population equivalent based on BOD concentration.

$$PE = \frac{FLOW_w * COD_{conc}}{f}; [pop] \tag{4.1}$$

$$EN_{PE} = \frac{365 * EN}{PE}; [\frac{kWh}{pop * year}] \tag{4.2}$$

In this case, the limiting information is connected with the frequency of laboratory analysis. In fact, generally, the analysis of wastewater characteristic is not performed for each day and, for the most of the days, the concentration values for BOD or COD are not available. Consequently, the KPI calculation is not possible. For example, in the WWTP of Burg-Solingen (GER) the samplig and analysis of wastewater at inlet is execute once each 15 days. In the WWTP of Hidden-City (NL), this operation is performed once per week. In both case, the daily KPI calculation is not possible. The source of this incompleteness of information derives from the cost of laboratory analysis. In alternative, this information can theoretically be retrieved by on-line measurements with Total Organic Carbon (TOC) or Spectral Absorption Coefficient (SAC) analysers which are good proxies for COD. This is however rarely successfully achieved, because of operational and analytical difficulties (signal drift, fouling, blockages), high investment and operational costs and little perceived value by operational staff (Kern et al., 2014; Martin and Vanrolleghem, 2014).

4.3.1 Algorithms for estimation of concentration load

The estimation of missing concentration values of pollutants at inlet of WWTPs can be performed by using non-linear regression models. Let's be 'Y-set', the parameter (or the parameters) to be estimated and let's define 'X-set' the matrix with independent parameters. The X-set should include all the parameters linked to the Y-set. The selection of these parameters is not obvious, relevant parameters could not be available and consequently the models are expected to be affected by uncertainty.

In fact, as explained in (Torregrossa et al., 2016):

“The phenomena that determine the COD at the WWTP inlet are highly complex and they concern the COD production as well as physical and biological aspects in the sewer system. COD in wastewater originates mainly from domestic, industrial or commercial sources. The domestic production is relatively regular, while the industrial and the commercial COD is more variable. Moreover, infiltration and ex-filtration are unpredictable but related to holes or cracks in pipes or illegal connections. In addition, according to Rauch et al., (2002), four phenomena influence the wastewater dynamics: pollutant accumulation, pollutant wash-off, pollutant transport and pollutant processes. Essentially, during dry periods, sedimentation occurs in catchment surfaces and in pipes, so that only part of the produced COD is transported to the WWTP. During storm events, these partially degraded sediments are washed off and they can reach the WWTP. In order to model these processes, an ideal analytical or regressive model should properly take into account all parameters that influence the phenomena related to COD concentration."

The same reasoning can be extended to all the pollutant concentration (such as the BOD). In (Torregrossa et al., 2016), it has been explained that exponential or linear regressions don't not provide satisfactory results while artificial intelligence regression algorithms seem to be more suitable to model the complexity of the pollution load generation.

4.3.2 An algorithm based on random forest

Fig. 4.5 shows the diagram flow of the algorithm.

The starting point is the SK-DSS database that stores the X-set parameters and the Y-set. The X-set parameters has values for each day, while the Y-set present missing values. The idea behind this algorithm is the following: identify the relation between X-set and Y-set, then, because X-set is available for each day, Y-set can be estimated for each day. The X-set depends on the availability in the WWTPs. In Torregrossa et al., 2016, for the WWTP of Solingen-Burg, the X-set was composed by the following parameters:

- the wastewater volume for the target day and for the days before (at day-1, day-2,day-3...day-7) ;
- the temperature of the wastewater and of the outside air;
- the day of the week and the number of the month;
- the NH4 concentration.

Consequently, the following operation is identifying the parameters suitable to be included in X-set and generate a database called Dataset A. At this point, it is possible to start the model training. In figure 4.5, the starting and the end point of the regression model are marked with red points. This thesis will present the example of the random forest. Nevertheless, as demonstrated in (Torregrossa et al., 2016), the regression modelling can be based on other algorithm such as random forest or neural networks. In order to train a model, a subset without missing values is required. This is done by removing the NaN (not a number values) and producing the dataset B. The dataset B is composed by the

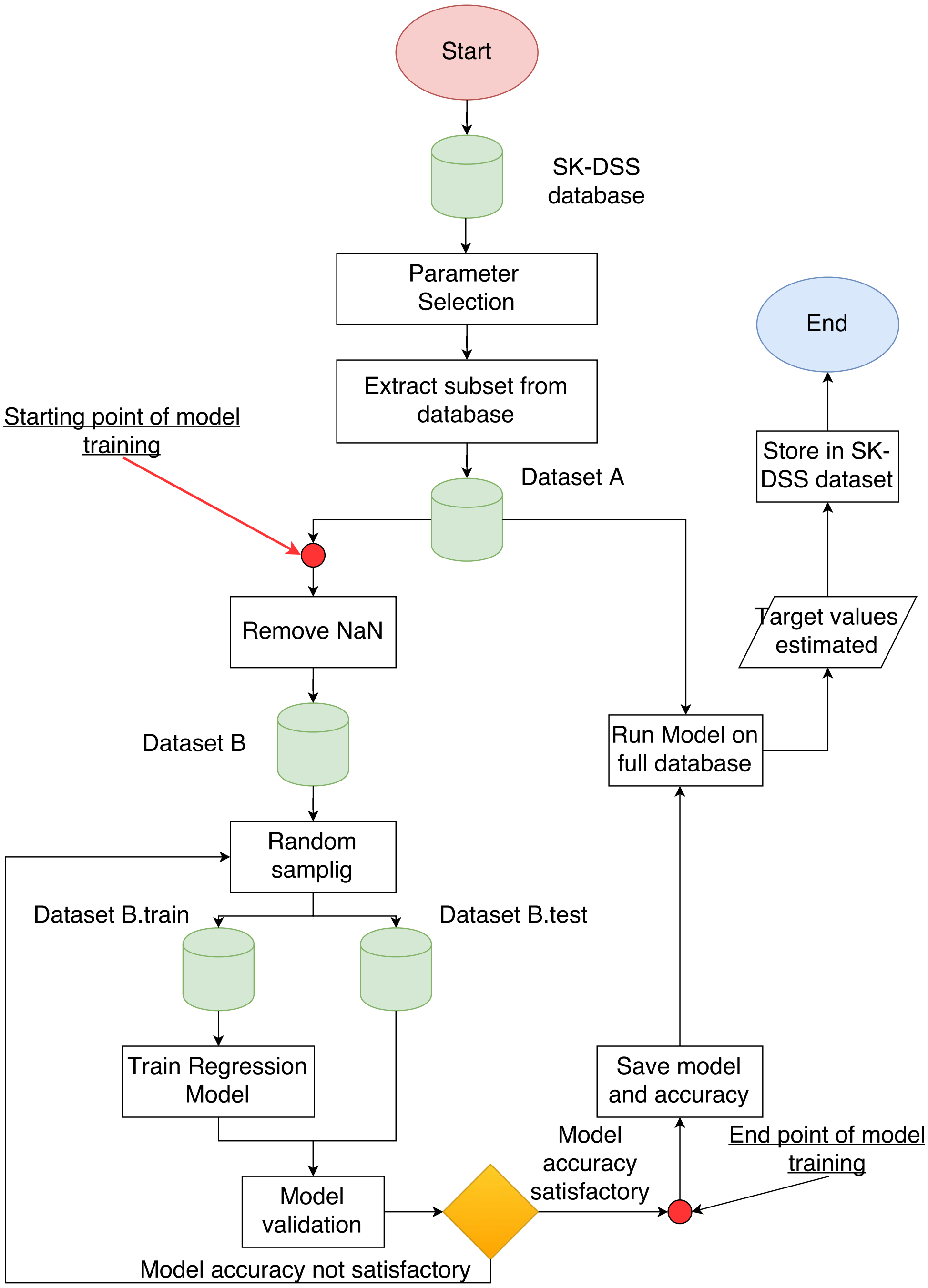


FIGURE 4.5: Diagram flow: algorithm for estimation of missing parameters

days in which both X-set and Y-set are complete; the size of dataset B is important to produce good model. In (Torregrossa et al., 2016), dataset B was composed by 130 rows; the experience acquired in these years of PhD suggest a minimum size of 100 rows in order to have a robust procedure. In fact, the dataset B needs to be randomly split in 2 subset for the train (B.train) and for the model validation (B.test). B.train is used to train the model, while B.test is used to obtain an independent validation the model accuracy.

The training of the model is performed with Random Forest. As explained in (Torregrossa et al., 2016), "Random Forests is a technique based on a combination of tree predictors already used to solve regression problems in WWTPs (Dürrenmatt and Gujer, 2012). The authors implemented a RF regression with 500 trees. Each tree is able to predict a value for COD and the output of the algorithm is their average value."

The model validation is performed by using the algorithm on dataset B (fig.4.5). Please note that the internal validation discussed in fig. 2.13 is different from the model validation of figure 4.5. The main difference is that the internal validation is done using the same dataset used for training. In the algorithm proposed in figure 4.5, the model validation is based on new data. If a portion of training data is used only for internal validation is called out-of-bag data (OBB).

B is then divided in his B.train (60% of values) and B.test (40% of values). The training of the model is performed with B.train, while B.test is used to validate it by comparing the estimated values with the real ones. In particular the coefficient of determination (R^2), the mean absolute error and the coefficient of variation of root mean square error are the most used Torregrossa et al., 2016. In SK-DSS, the mean absolute error is used: if this value is larger than 0.2 the model accuracy is considered not satisfactory and the algorithm restarts from the random data sampling. If the mean absolute error is less than 0.2, the model is accepted and it is saved to be available to the next step.

Now, the dataset A can be used as input of the model to predict the Y-set for each day. These values are calculated and updated on the SK-DSS dataset.

Figure 4.6, retrieved by (Torregrossa et al., 2016) shows that accurate results can be obtained, with values of $R^2 > 0.71$ for the independent validation executed on the portions of data not used for training.

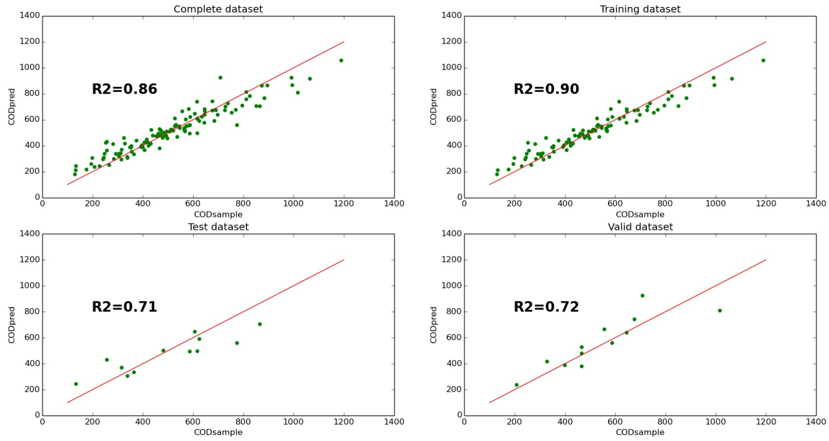


FIGURE 4.6: Results of random forest algorithm. Figure extracted from (Torregrossa et al., 2016)

4.3.3 Uncertainty management

The management of uncertainty is an aspect connected to the estimation of missing parameters. In fact, due to the complexity of the phenomena, it is necessary to accept that estimation values can diverge from real one. The sources of uncertainty can be various such as data quality, number of records in the database or selection of parameters. For decision making purposes, this uncertainty can be assessed and managed. In (Torregrossa et al., 2016), a first attempt to overcome this issue has been successfully performed by transforming the benchmark and the key performance indicators in ranges that take into consideration the uncertainty. In SK-DSS, the KPI are processed by fuzzy logic, the peculiarity of which is to manage efficiently value affected by uncertainty (Torregrossa et al., 2017a). Details on fuzzy logic process can be found in section 2.5.

4.4 KPI calculation

The KPI calculation step takes as input the table of daily data in which nomenclature, time aggregation and unit of measurement are normalised for each sensor and for each plant. Consequently, it is possible to set-up plant generic scripts to perform some calculations and obtain the KPI values. In SK-DSS, this operation is performed at the database level trough SQL queries. In order to explain in detail how this queries are structured, in fig. 4.7, the query code for the calculation of specific energy consumption (i.e. energy per population equivalent) is presented. The query of fig. 4.7 assumes this structure:

- the first 2 lines remove from the table called 'KPI', the old value of the KPI called KPI_SUM_EAperPE;
- the line 5 gives the instruction to re-fill the table with new values;
- the lines 8-12 format the information to be according to the requirements of KPI table; in particular it is necessary to provide a date, a KPI name a value and the plant name. The value of the KPI is calculated at the line 10; en.value is the value of the energy consumption, while pe.value is the value of the population equivalent;
- the block of lines 15-17 takes from the table called 'plant data' the information about energy consumption;
- the block of lines 21-23 gathers information about the population equivalent;
- the information of energy and connected population are connected trough line 19 and 25. Line 19 gives the instruction to take in consideration only the days in which there is the value of energy and pe. In case of missing values, no KPI is calculated for the specific day and the specific plant. The line 25 inserts the condition that the value calculation is done with values referring to the same date and the same plant.

The structure of this script enables to automatically calculate the same KPI for several plants and for several days and store the results in the KPI table. The other KPIs can be calculated with the same approach.

The resulting KPI set is normalized for further analyses. Table 4.2 shows a list of KPIs calculated for the global energy assessment, pumps and blowers.

SK-DSS has 3 typologies of indicators to be stored in 'KPI table': KPI, benchmark and indices.The KPIs are calculated with the combination of daily parameters (for example:energy and population equivalent). Benchmark are used as reference value for KPIs. The same benchmark

```
delete from  wwtp_data_gathering.KPI
where KPI_id='KPI_SUM_EAperPE';
--COMMENT: remove old values from the KPI table

insert into wwtp_data_gathering.KPI
--COMMENT: insert new values in the KPI table

select en.date,
'KPI_SUM_EAperPE' as "KPI_id",
(en.value/pe.value)*365 as "value",
en.plant from
--COMMENT: specify values to be inserted


(select * from wwtp_data_gathering.plant_data
where sensor_id='WWTP_SUM_EA') as en
--COMMENT: call values of total energy consumption

full join

(select * from wwtp_data_gathering.plant_data
where sensor_id='WWTP_WW_PE') as pe
--COMMENT: call values of population equivalent

on en.date=pe.date and en.plant=pe.plant
--COMMENT: couple information with same date
and same plant
```

FIGURE 4.7: Query of KPI: energy consumption for population equivalent

is calculated with different formulas that takes into account a different parameter availability. For example, in table 4.2, the benchmark for blower energy consumption is available as the result of a three formulas that takes as input (i) the PE, (ii) the PE and the sludge age and (iii) a constant value (Torregrossa et al., 2017a). Consequently, SK-DSS guarantees that each plant has a complete set of benchmark values, dynamically calculated for each day. The third category of indicators is composed by indices that are used to compare the KPI with their reference values. An index is really easy to calculate (ratio between kpi and benchmark values) and intuitive to read: if the value of the index is higher than 1, then the KPI is higher than the reference. If the index is equal to 1, it means that the KPI corresponds to the reference value, while if an index value is minor than 1 it means that the KPI value is smaller than the corresponding benchmark. SK-DSS dataset is an evolution of EOS dataset. In EOS, only the KPIs were calculated, while SK-DSS introduces some improvements:

- the benchmark are dynamically calculated for each day and they take into consideration the operational condition of the plants;
- the index values calculated in SK-DSS are more intuitive to be understood.

Indices, benchmarks and key performance indicators are now able to be visualized stand-alone or to be processed all together by the fuzzy-logic engine.

4.5 Algorithm automation

All the processes described in the previous sections can be automated in order to be more useful to the plant managers. The advantages of an automatic process are various:

- the automatic procedure is in theory free of computational errors; once the script set-up is correct, all the calculations should be correctly executed by the computers; the only source of error could be in the code; as alternative, a daily human-based calculation is generally considered more exposed to errors;
- it is possible to save time in calculation and set-up the system to prepare the results at given time in order to be ready for the operators (for example the calculations can be done during the night to be ready each day at 07.00 am);
- it is possible to increase the number of connected plants without any additional effort for the end-users.

This is realized with the algorithm explained in figure 4.8. The sensors produce the plant values at the plant level. These datasets are then transferred to a server in which SK-DSS stores the information sent by WWTPs. Each WWTP has its own folder that is observed with an Incrontab application (*http://inotify.aiken.cz/?section=incron&page=about&lang=en*).

Using this tool, it is possible to activate a script when a new file is created in the observed folder. In SK-DSS, plant-specific python scripts, activated by incrontab, process the new files to import raw plant data in SK-DSS daily table. At this stage, data normalisation is executed; this is the part of the decision support system in which plant data lose their specificities. The KPI calculation is performed at regular time intervals; currently the script is set-up to automatically start at 9:00 am for each morning. In Ubuntu server, this automatic routine is activated by Cron: "a system daemon used to execute desired tasks (in the background) at designated times" (`https://help.ubuntu.com/community/CronHowto`). After the KPI calculation, Cron is used to automatically perform the advanced analysis that will be explained in the following chapters. For advanced analysis, the Cron job is set to be activated each day at 11.00 am.

Cron and Incrontab are really flexible applications that can be activated at regular interval (each hour for example), at given day of the week or given month, or it is possible to perform specific routine for given plants. The flexibility of these tools enable SK-DSS to work with a great quantity of connected WWTPs, because a large (potentially critical) amount of data can be split in subsets and processed at different time interval. The current time scheduling currently adopted is illustrated in figure 4.8. A detailed discussion about these tools is out of the topic of this thesis.

4.6 Conclusions

In this chapter, it has been shown how data coming from different WWTPs are processed and normalized. In particular, the sections about KPI calculation and the estimation of missing data explained how SK-DSS produces a dataset standardized and complete to be further analysed. An interesting aspect of this process is the automatic activation of routine works. In few words, SK-DSS connects different WWTPs and automatically produces, for all of them, performance indices on daily basis.

Next chapter, chapter 5 will show the use of this information to produce the assessment of pumps, blowers and biogas. Chapter 6 describes in detail the representation of the produced information by mean of the web-interface.

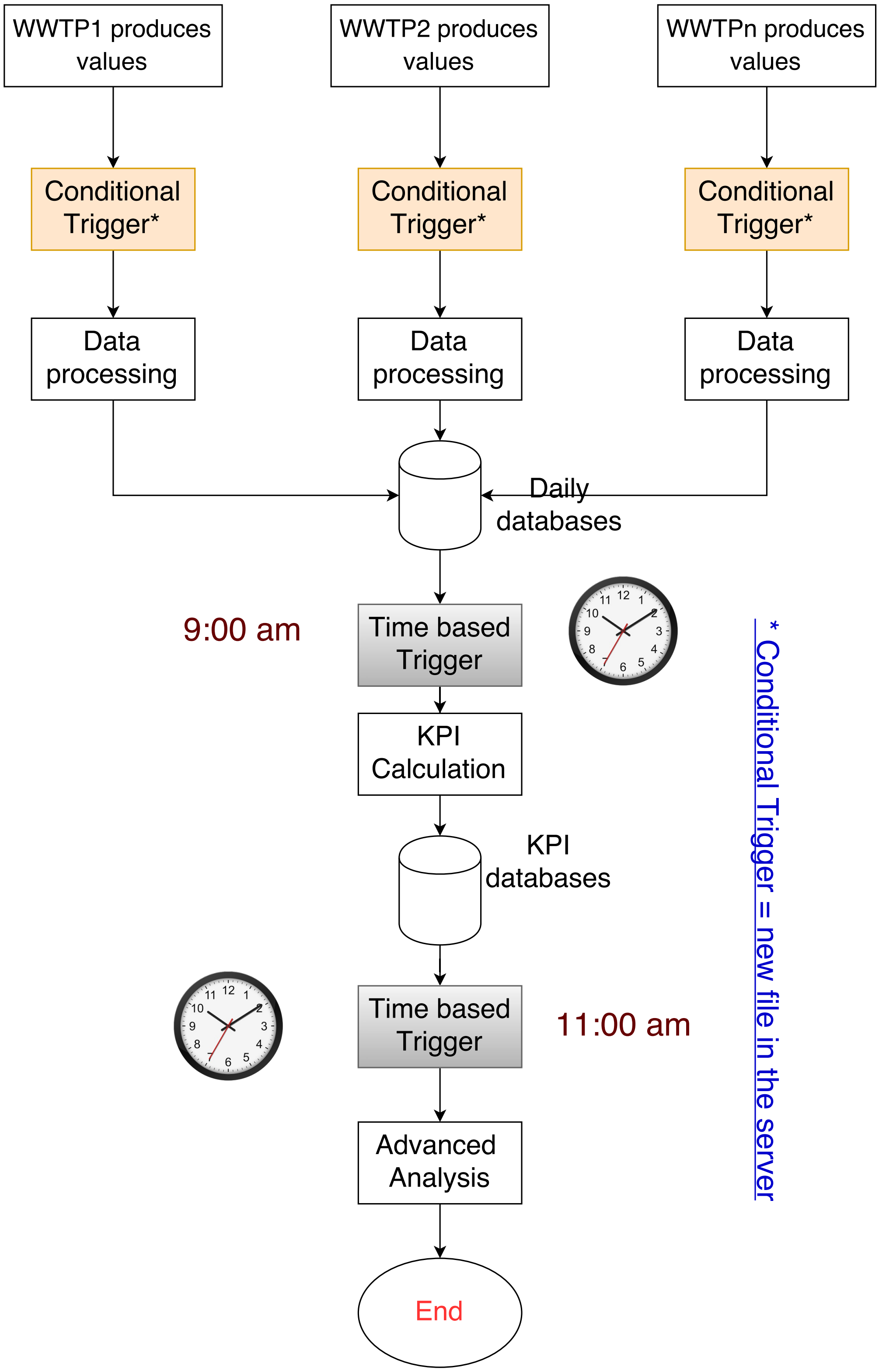


FIGURE 4.8: Algorithm automation

Chapter 5

SK-DSS in details: from KPIs to solution analysis

LEGAL DISCLAIMER: The present chapter partially reproduces research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of the EdWARDS project.

5.1 Introduction

In this chapter, some device-based applications of the SK-DSS methodology are presented. In particular, this PhD thesis is focussed on blowers, pumps and biogas production assessment. This choice has been done because blowers and pumps cover almost the 80% of the global WWTP energy consumption and an optimal biogas production can cover a relevant part of energy requirements (INNERS, 2015; Hansen, 2018).

The remaining energy consumers (such as those identified in fig. 2.9) were not taken into consideration because the duration of this PhD project was limited to 4 years. Moreover, it is important to mention that all the methodological aspects presented in this chapter have been peer-reviewed and accepted for publication (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b). In particular, the content of the section 5.2 is largely taken by (Torregrossa et al., 2017a) while the section 5.3 reflects the content of the paper (Torregrossa et al., 2017b).

5.2 Blowers analysis

The blower energy consumption is the most relevant part of the total energy consumption of WWTPs (fig. 2.9).

In the paper (Torregrossa et al., 2017a), the main novelties of this approach were discussed. "In particular, this approach

- uses the dynamic benchmark calculation;
- is plant generic, or, in other words, it is able to simultaneously analyse multiple WWTPs;
- is able to produce daily analysis reports;
- enables the data-mining over WWTP database;
- provides plant managers with case-based suggestions for energy efficiency."

The sections from 5.2.1 to 5.2.6 are largely extracted from (Torregrossa et al., 2017a).

5.2.1 Case study: blowers assessment

In the following subsections each process, shown in Figure 3.1 is explained in detail using a practical example carried out for the WWTPs of Burg-Solingen (GER) and Hidden-City (NL), called respectively BUR and NL1. BUR and NL1 provide different inputs for the SK-DSS. Consequently, SK-DSS has to work with two different sets of KPIs (Table 5.1).

The methodology can be divided in several sub-steps:

- rule set selection. In this step, a set of rules is selected for each plant according to the plant specificities (such as the data availability);
- benchmark calculator. For each day and for each plant, the system calculates a daily benchmark that take into account the daily operational conditions;

TABLE 5.1: Set of KPIs

KPI_name	Explanation	Units of measurement	BUR value	NL1 value
BLOair_VOL	Volume of air in the biological reactor	[m^3/day]	85647	N/A
O_{2-req}	Theoretical oxygen request	[m^3/day]	9272	19787
BLO_EA	Energy consumption of blowers	[kWh/day]	4587	10059
PEpred	Estimated population equivalent	number of people	109566	231437
$EN_{blo-spec}$	Consumption per population equivalent	[kWh/PE/y]	15.3	15.9
AIRindex	Theoretical Air Volume/Measured Air Volume	-	0.39	N/A
KPI_BIO_SLA	Sludge Age	[days]	15	N/A

- fuzzy logic analysis. KPI and benchmarks are analysed with a fuzzy logic engine that provides an overall evaluation of the plants and the analysis of several scenarios;
- case-based solution generation. The analysis of the scenarios performed by the fuzzy engine is used to provide solutions.

5.2.2 Definition of the set of rules

The selection of the set of rules for fuzzification has to take into account the available KPIs. In this case, the system requires two different sets of rules as a consequence resulting from the differences in the input data (KPI sets) for each plant. The set of rules for BUR includes the AIRindex and the $EN_{blo-spec}$. The set of rules for NL1 includes just the $EN_{blo-spec}$ due to insufficient data for calculating the AIRindex. These different sets of rules are shown in Table 5.2. The content of table 5.2 is written in a fuzzy logic language to be processed by SK-DSS. For example, the rule BIO.24 takes the following format:

- RULE BIO.24 : IF AIRindex IS Low AND $EN_{blo-spec}$ IS Medium THEN BIO_BLO_Score IS Low ;

5.2.3 Benchmark calculator

The SK-DSS requires benchmark values to decode the expression (High, Medium, Low) of the KPIs. For example, in order to use the RULE BIO.24, the system needs to know when 'AIRindex IS Low' and when '$EN_{blo-spec}$ IS Medium'. The benchmark calculation for $EN_{blo-spec}$ is dynamic i.e. it is calculated on a daily basis. The input in this step is the technology code and a set of equations derived from literature

TABLE 5.2: Set of rules

Plant	Rule	AIRindex	$EN_{blo-spec}$	BIO_BLO_Score
		Condition		Consequences
BUR	BIO.21	Low	High	Low
	BIO.22	High	High	Low
	BIO.23	Medium	High	Low
	BIO.24	Low	Medium	Low
	BIO.25	Medium	Medium	Medium
	BIO.26	High	Medium	Medium
	BIO.27	Low	Low	Medium
	BIO.28	Medium	Low	High
	BIO.29	High	Low	High
NL1	BIO.40		High	Low
	BIO.41		Medium	Medium
	BIO.42		Low	High

(Shi, 2011). The output is the set of benchmark values, which are used to fuzzify the inputs (fig. 5.1). Piecewise linear functions (triangular and trapezoidal) were chosen to represent the input universe, because, without decreasing the performance of the fuzzy engine, the construction per points is more intuitive (and consequently easier to share in the platform) than smoother functions (like the Gaussian or Bell Curve). Moreover the management of overlapping regions of the various curves is easier with piecewise linear functions (Yager and Filev, 1995; Mendel, 1995).

The system has generally three ways to calculate the benchmarks: an equation with all the correlated KPIs, an equation based on the connected population, or a static value. In the case of the specific energy consumption for blowers ($EN_{blo-spec}$) in a plant with anaerobic sludge digestion, the equations [1] have the following form:

- by using a complete equation with 2 depending variables:

$$\text{Benchmark for } EN_{blo-spec} = \begin{cases} \text{Low} & = 28.84 * SA^{0.26} * PE^{-0.15} \\ \text{Medium} & = 38.51 * SA^{0.26} * PE^{-0.15}\ [KWh/pe/y] \\ \text{High} & = 48.18 * SA^{0.26} * PE^{-0.15} \end{cases} \tag{5.1}$$

[1] These equations were retrieved by interpolation from the benchmark values provided by (Shi, 2011).

- by using a set of equations based on population equivalent:

$$\text{Benchmark for } EN_{blo-spec} = \begin{cases} \text{Low} & = 63.34 * PE^{-0.15} \\ \text{Medium} & = 87.01 * PE^{-0.15} \quad [KWh/pe/y] \\ \text{High} & = 110.85 * PE^{-0.15} \end{cases} \tag{5.2}$$

- by using static values:

$$\text{Benchmark for } EN_{blo-spec} = \begin{cases} \text{Low} & = 10.5 \\ \text{Medium} & = 14.5 \; [KWh/pe/y] \\ \text{High} & = 18 \end{cases} \tag{5.3}$$

In the set of equations 5.1 and 5.2 SA is the Sludge Age [day] and PE is the Population Equivalent [pe].

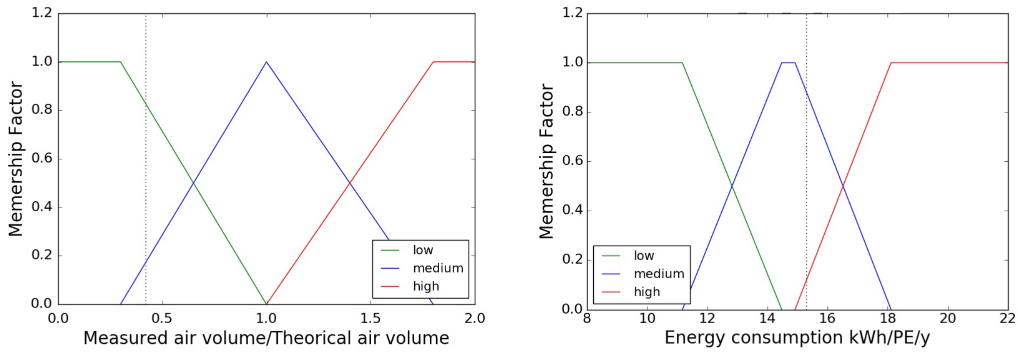


FIGURE 5.1: Fuzzification of the inputs in BUR

The system tries to use as first choice the complete equation (5.1). If some variables are missing, it attempts the calculation using the second equation (5.2). If data is still insufficient, the system uses the static values (5.3). These values are used for the fuzzification. Table 5.3 shows the benchmark values for the $EN_{blo-spec}$ with the 3 different equations calculated for BUR and NL1.

5.2.4 Fuzzy Logic Engine & Knowledge discovery

For a detailed explanation of fuzzy logic methodology, the reader can refer to these sources: (Zadeh, 1965; Starczewski, 2013) or to this thesis at the section 2.5.1.

The inputs for this process are benchmark values for fuzzification, KPI values and rules.

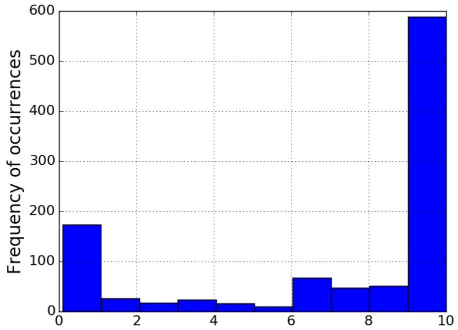
TABLE 5.3: Calculated benchmark values for $EN_{blo-spec}$

Plant		Benchmark values: $EN_{blo-spec}$		
		Complete equation	Partial equation	Static values
BUR	Low	10.99	11.94	10.5
	Medium	14.68	16.41	14.5
	High	18.37	20.9	18
NL1	Low	N/A	10.15	10.5
	Medium	N/A	13.94	14.5
	High	N/A	17.77	18

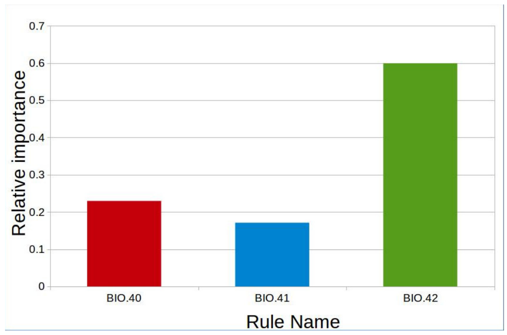
By coupling the benchmark definitions (fig. 5.1 and table 5.3) with the values of KPIs of table 5.1, for BUR, SK-DSS detects the $EN_{blo-spec}$ as 'medium' and the AIRindex between 'low' and 'medium'. These fuzzified KPIs are finally compared with the rules and consequently the degree of truth is calculated. For BUR, the rule with highest degree of truth is the rule BIO.24 that describes the condition of medium $EN_{blo-spec}$ and low AIRindex. The degree of truth for this rule corresponds to 0.78. By aggregating the degree of truth of each of the rules, SK-DSS derives, for BUR, a score equal to 3.4. The identical procedure applied to NL1, produces a score equal to 3.2.

The Score value is an index between 0 (worst condition) and 10 (optimum condition) and it represents the multi-perspective performance of the observed unit. In this example the score value is influenced by the air index and by the energy consumption per population equivalent. For the day under evaluation, the score of 3.4 for BUR shows a potential for efficiency gains because the energy consumption is somewhat high for the given air flow. The same reasoning can be reproduced to comment the score value of NL1.

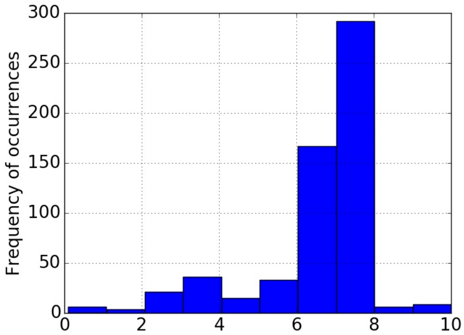
The Score values and the degrees of truths associated to each rule can have various usages to understand the plant behaviour. They can be aggregated to calculate the average performance of the system; for example the average score for the aeration system in December 2014 is 7.8 for NL1 and 7.15 for BUR. In fig. 4, on the left side, the histograms of the scores are shown, while on the right, the relative importance of the rules is presented. In NL1, the most important rule is BIO.42, that



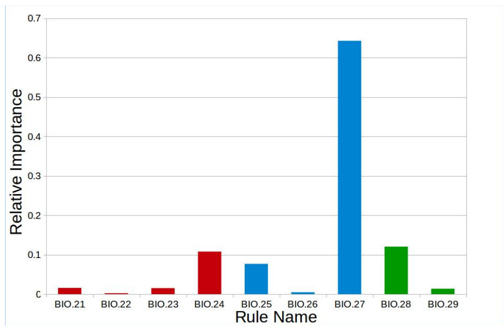
Histogram: NL1 Scores of biological stage



Bar plot: Relative importance of the rules in NL1



Histogram: BUR Scores of biological stage



Bar plot: Relative importance of the rules in BUR

Note: The rules are ordered according to the respective plant performances. In green the block of rules corresponding to 'good output' on the left. In blue the block of rules corresponding to an 'average output' in the middle. In red the block of rules corresponding to a 'bad output' on the right.

FIGURE 5.2: Statistical outputs from fuzzification

describes the condition in which the energy consumption of the blowers is low and the score is high. Using the same methodology, the most recurring situation in BUR is described by rule BIO.27 (low energy consumption and low air flow).

5.2.5 Engine for solutions

SK-DSS has a set of solutions associated to respective rules. These solutions are generally derived from literature analysis and experts' interviews. For each solution, the description of specific actions is provided alongside the expected energy savings, the response time and expected costs. In this step, SK-DSS takes into consideration the rule with degree of truth$\geq$0.5 and it shows the respective solutions. In this example the BUR analysis shows just the rules number BIO.24 and its respective solutions as in table 5.4. The same process is performed for NL1 (also table 5.4).

Having the information about the analysis of the WWTP and the list of possible solutions, the plant manager, using the plant specific information and existing experience, can select the solutions that are appropriate. One of the most recent version of SK-DSS uses the popular YouTube platform to connect the fuzzy logic rules to their respective solutions. This new method to propose solutions is discussed in section 5.7 and in (Torregrossa and Hansen, 2018) in which a case-study specifically focussed on blower assessment is provided.

5.2.6 Validation of the methodology and comparison with other approaches

A first external validation of the methodology and results obtained was performed during a meeting in which experts were invited to compare the results with their knowledge. The results of this peer-review were promising because there was overall agreement that the system works properly, produces outputs that seem reasonable and the values obtained are of the right magnitude. One of the recommendations was that the benchmark equations and rules should be extended to take in consideration more elements (such as the operational temperature and the dissolved oxygen concentration).

Another validation on the efficiency of the blowers only has been performed by using a comparison with other analysis: Panepinto et al., 2016 uses the formula 5.4 to evaluate the efficiency of blowers.

$$\eta = 3.28 \cdot T \cdot 10^{-4} \cdot \frac{Q_a}{P_a} \cdot [(\frac{p_2}{p_1})^{\frac{(k-1)}{k}} - 1] \tag{5.4}$$

in which T,Q_a,P_a,p_1,p_2 and k are respectively the temperature of the wastewater in the biological stage, the air flow rate, the consumed power, the input and output pressure of the blowers and k=1.395. The average efficiency of the blowers (η) obtained for BUR was 0.30 with a good efficiency reference value around 0.7 and a lower range of 0 (the minimum value of η for the extreme condition in which the air flow is equal to zero) . The equivalent value obtained using SK-DSS was 12.8 kWh/p.e./year, with 10 kWh/p.e./year considered good and 18 kWh/p.e./year considered bad. Therefore, the results of this comparison seems reasonable, because both analyses show a margin for energy saving while considering different parameters.

	Suggested Solution	E savings	Cost	Response time
	BUR Scenario: Energy consumption Medium with AIRindex Low			
Truth degree = 0.78	Is blower pressure nominal? Yes → System improvement: Blower capacity insufficient. Consider adding additional blower capacity.	Depends	Depends	Medium
	Check for potential pipe obstructions.	Low	Low	Low
	Fouled Diffuser: Bio-fouling in the pores or Calcium hardness fouling possible. A diffusers check and clean up required in order to lower the pressure in the blower system.	Low	Low	Low
	Is there brown foam on surface? Yes → Evaluate in fluent or internal side stream for septic conditions. Increase air supply to match organic loading.	Low	Low	Low
	Improve system design by installing additional DO sensors to perform a more detailed analysis of the oxygen concentration.	Medium	Medium	Low
	Improve system operation design: set-up the valves to reduce the blower pressure.	Low	Low	Low
	Control the Solid Retention Time (SRT). If SRT is too high, reduce it! Remember: the SRT is linked to temperature. Automatic and/or seasonal adjustments could be beneficial.	Low	Low	Low
	Control the Dissolved Oxygen concentration (DO). If DO is too high ($\geq 4mg/l$), the aeration system is providing more air than necessary. A correct set-up of DO can reduce the energy consumption.	Low	Low	Low
	NL1 Scenario: Energy consumption High			
Truth degree = 0.54	Improve system operation: plan a maintenance	Low	Low	Low
	Improve system design: Upgrade to fine bubble diffusion	Medium	High	Medium
	Is blower pressure nominal? No → Maintenance Scheduling: Check Blowers conditions. Blowers malfunction possible.	Depends	Depends	Depends
	Is blower pressure nominal? Yes → System improvement: Blower capacity insufficient. Consider adding additional blower capacity.	Depends	Depends	Medium
	Improve system control design: Install automated DO controls if not installed yet.	High	Medium	Medium
	Improve system operation: Reduce blower power by reducing its speed.	Low	Low	Low
	Improve system operation: Switch to a lower air capacity blower.	Low	Low	Low
	Improve system design: install variable frequency drivers (VFDs) on blowers.	Medium	High	Low
	Control the Solid Retention Time (SRT). If SRT is too high, reduce it! Remember: the SRT is linked to temperature. Automatic and/or seasonal adjustments could be beneficial/	Low	Low	Low
	Control the Dissolved Oxygen concentration (DO). If DO is too high ($\geq 4mg/l$), the aeration system is providing more air than necessary. A correct set-up of DO can reduce the energy consumption.	Low	Low	Low

TABLE 5.4: Extract of solutions based on scenario analysis for BUR and NL1

However, it is important to consider that SK-DSS takes into account a larger set of information than equation 5.4. In fact, SK-DSS also includes into the analysis operational condition parameters (such as the population equivalent and the oxygen demand in biological tank) and blowers parameters (such as the consumed energy and the volume of air generated). In contrast, equation 5.4 takes account of the blower parameters only. In other words, SK-DSS assesses the complete observed unit. In the data, it can be observed that when the blower parameters divert from their nominal values a similar effect in the SK-DSS output is evident. However, when the blower parameters are in a normal range, the SK-DSS score could be still low because of other factors.

5.2.7 Summary of blower analysis

In this section, a novel methodology for on-line blower assessment was presented. This methodology has a high added value because it provides a plant-generic, multi-perspective and high-frequency blower monitoring and provides case-based suggestions. In particular, the analysis of 2 WWTPs with a different set of input parameters showed that:

- the plant in The Netherlands has generally a low energy consumption. More detailed information could not be provided because no information are available about air inflow;
- the plant in Germany show a low energy consumption with a reduced amount of provided air.

Starting from the scenario analysis, the methodology presented in this section is able to provide case-based suggestions. This methodology offers a tool that can effectively compete with the classical energy benchmark approaches for many reasons. In fact, first of all, the period between two analyses is reduced to 1 day, and the analyses take into consideration new elements such as:

- benchmarks based on operational conditions;
- daily estimated values for pollution load;
- a larger set of information, here including the quantity of air effectively provided by the blowers;
- the set of available information in the WWTP.

Moreover, this set of information can be customized and adapted to the circumstances. For example, in (Torregrossa and Hansen, 2018), the same methodology took into account the oxygen concentration in the biological tank.

5.3 Pump analysis

In WWTPs, pump energy consumption represents a relevant portion of the overall energy consumption: Shi, (2011) and Metcalf and Eddy, (2014) reports that they account for the 12% of the WWTP total energy consumption, while INNERS final report shows a pump consumption corresponding to the 26.22% of the total (INNERS, 2015). The analysis of pump systems with the SK-DSS methodology has been published in (Torregrossa et al., 2017b) and the section 5.3 is largely extracted from this paper.

In this thesis and in the paper, the data used belongs to the Solingen-Burg WWTP (BUR) in Germany, which was processed by the EOS system (Torregrossa et al., 2016). BUR is equipped with a SCADA system which ultimately provides daily data on the energy consumption of significant devices and operational conditions relating to wastewater inflow, pollution load and wastewater composition. Within BUR, an intermediate pumping station (IPS), comprising 6 pumps (80 kW-power each one) equipped with variable speed drives, lifts the wastewater after primary treatment 10.33 m into the aeration basins. Figure 5.3 shows the histograms for the daily flow and the daily energy consumption.

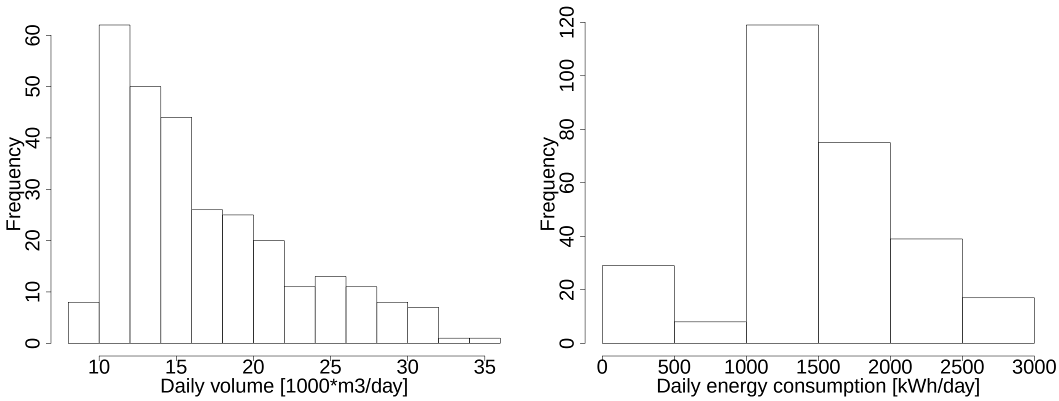


FIGURE 5.3: Histograms for the daily wastewater inflow (left) and daily energy consumption (right) of the intermediate pumping station

5.3.1 Literature review:methodologies for pump optimization

Various authors have proposed efficiency strategies for pumping systems, including design and calibration, assessment and reporting, maintenance and on-line monitoring (Hydraulic Institute and Department of Energy (US), 2006; da Costa Bortoni, Almeida, and Viana, 2008; Zhang, Zeng, and Kusiak, 2012; Chang et al., 2012; DeBenedictis et al., 2013; Zhuan and Xia, 2013; Berge, Lund, and Ugarelli, 2014; Olszewski, 2016; Zhang et al., 2016; Torregrossa et al., 2017d; Wang et al., 2017). In ((Hydraulic Institute and Department of Energy (US), 2006), p. 33), the authors elaborated the most common maintenance strategies for pump efficiency: the simplest approach is a fixed interval scheduled maintenance; a more advanced approach is predictive or conditional maintenance. The latter can be performed using vibration sensors, thermography, or infra-red (IR) scanning (Hydraulic Institute and Department of Energy (US), 2006); these sensors do not directly provide measures for energy performance but the detection of anomalies can indirectly produce energy savings. In (Torregrossa et al., 2017d), it is proposed a decision-making tool for the efficiency analysis and decision-support based on an economic consideration of WWTP pumps. In (da Costa Bortoni, Almeida, and Viana, 2008), a methodology is presented based on a mathematical optimization to identify the best flow repartition in a parallel-multi-pump system. Zhang, Zeng, and Kusiak, (2012) obtained positive results developing a methodology, based on a neural network, for the optimal scheduling of operations in multi-pump systems. Chang et al., (2012) used fuzzy logic algorithms to find the optimal operating point of multi-pump systems. Olszewski, (2016) focused his effort on the optimization of multi-pump systems with a genetic algorithm optimization.

DeBenedictis et al., (2013) proposed a methodology to evaluate the impact of variable speed drives and program-logic controllers (PLCs) on pump system efficiency. Berge, Lund, and Ugarelli, (2014) focused on the condition monitoring of pump systems using an array of sensors including: pump vibration, motor winding temperature, motor current, motor bearing temperature and pump inflow. Wang et al., (2017) proposed an optimization method for the design of a multi-pump system, which analyses the various energy losses in a pump system such as disk friction loss or hydraulic loss. Zhang et al., (2016) show that it is possible to save between 6% and 14% on pump energy with data-driven models and optimization approaches. Zhuan and Xia, (2013) demonstrate the feasibility of reducing maintenance and energy costs of multi-pump systems by using optimization algorithms.

5.3.2 Pumps efficiency in WWTPs

Pump performance can be dramatically decreased by several issues, such as cavitation, over-sizing, wear, leakages and, in extreme cases: blocking. In the WWTP domain pump design is often inaccurate because of significant differences between design and operational flows during the lifetime of the pump (Metcalf and Eddy, 2014; Hydraulic Institute and Department of Energy (US), 2006; Torregrossa et al., 2017c). The costs of inefficiencies can be reduced with proper design and management. This includes the detection of problems related to pump efficiency and an understanding of its implications (Hydraulic Institute and Department of Energy (US), 2006).

Despite the theoretical availability of approaches for enhancing pump energy performance (subsection 5.3.1), in the WWTP domain, pump energy management is generally sub-optimal for a variety of reasons, such as a lack of useful information (for example: vibration measurements) or infrequent assessments (once or twice per year). One of the most widely used analyses relies on the calculation of efficiency indices which are periodically benchmarked (Hydraulic Institute, Energy, and Energy, 2005). In (Spellman, 2003, p. 64), the author proposes reference values to evaluate the efficiency of a given pump system and some of its most important components: motor, pump and flow controller. In the WWTP domain, the assessment of energy performance of the pumps is mostly performed just a few times a year using the average values of flow and energy for the period being analysed. This approach is not ideal because it does not take into account seasonal phenomena and pump degradation and does not allow the early detection of failures.

Nowadays, in the WWTP domain, the availability of Supervisory Control And Data Acquisition (SCADA) systems theoretically allows an increased frequency of energy analysis and an improved plant management with a faster response time to inefficiencies found (Torregrossa et al., 2016). As for the global energy consumption (Torregrossa et al., 2016) , it is postulated here that a daily pump energy benchmark exercise can improve energy efficiency and reduce pump management costs.

However, the simple scaling of a yearly index to (for instance) a daily resolution is not necessarily beneficial for pump management and can even produce misleading results. The calculation of daily indices can produce time series with a high amplitude and a simple daily efficiency index is therefore difficult to interpret. In practice, pump systems are complex to analyse; some pump system parameters are constant, or change very slowly, over a long period of time (number of pumps, set-up of controller, lift and pump wear) while other parameters change continuously (inflow, temperature, particulates) (Gülich, 1998; Hydraulic Institute and Department of Energy (US), 2006). Many efficiency measures are interdependent and highly dynamic. For example, cavitation depends on temperature, inflow and hydraulic losses (WEF, 2008, chapter 8, p. 14) while the overall efficiency of the pump system is influenced by the best efficiency point (BEP) and the inflow ((Hydraulic Institute and Department of Energy (US), 2006) , p. 37).

Using a daily efficiency index, without thoroughly assessing the effects of operational conditions, could therefore lead to the debatable conclusion that the pump system performance changes day by day. A more precise statement should assert that the pump system performance is, within a sufficiently short interval (typically one day), highly dependant on the operational conditions. Consequently, a detailed daily analysis of pump system performance should be able to effectively separate the effects of system set-up and operational conditions.

5.3.3 Gaps in literature, objectives and novelty of this paper

A recent review paper on pump efficient control strategy edited by Arun Shankar et al., (2016) found that, despite the large number of literature contributions, to date, there seems to be a lack of methodologies for the analysis of pump energy consumption, that I) can be automatically performed (at least) on a daily basis, II) are able to separately assess the effects of system set-up and operational conditions, and III) rely solely on the parameters commonly available in WWTPs (inflow and energy consumption). Section 5.3.1 leads to the same conclusion. This is the knowledge-gap that the present chapter seeks to address by presenting a data-driven methodology and a tool for pump system energy efficiency which is capable of:

- automatically carrying out a daily pump efficiency analysis to detect potential inefficiencies at an early stage;
- identifying and separately addressing long term and short term efficiency patterns;
- producing an overall index that takes into consideration the pump operational conditions;
- supporting the plant managers in the early identification of imminent failures;
- suggesting potential solutions;
- evaluating the solutions according to economic criteria;
- improving the comprehension of the pump system behaviour examined;
- relying on information generally available to plant managers (inflow, pump energy consumption, and energy cost).

Torregrossa et al., 2016 explained how the Energy Online System (EOS) automatically records, collects and processes on-line data from WWTPs and aggregates them in daily KPIs. The tool described in the current paper uses the data automatically aggregated in EOS for pump system analysis. EOS is plant-generic and the pump decision support tool inherited this feature: it functions for many, if not all, WWTPs, despite their differences. For this reason, it was considered essential that it relies solely on information provided by commonly available sensors.

The methodology proposed in this chapter is new and innovative for the following reasons :

- it combines signal decomposition, fuzzy logic and benchmarking to analyse pump operational conditions;

- it uses new KPIs as input of fuzzy logic: η_t,η_f, τ, Z (please refer to subsections: 5.3.4, 5.3.4 and 5.3.4)
- it applies the analysis of truth degree of fuzzy logic rules to provide plant managers with case-based solutions concerning pump efficiency;
- it estimates a potential energy saving, relying on the trend value of efficiency (η_t), obtained with signal decomposition.

Existing competing methodologies (such as the one proposed by Berge, Lund, and Ugarelli, (2014)) come to similar conclusions given some of the constraints already discussed (such as the time frequency of the analysis and the early detection of problems). However, they are difficult to reproduce in the WWTP domain, because they rely on information generally unavailable in this area (such as pump vibration, current and motor winding temperature).

5.3.4 Inside the methodology

Calculation of the efficiency index

The EOS supplies the daily pumped volume and the daily energy consumption per pump (Torregrossa et al., 2016). First, equation 5.5 is used to calculate the efficiency of the intermediate pumping station, η. In this formula, *m* is the mass of wastewater lifted [kg], *h=10.33m* is the static head, *g* is the standard gravity [9.81 m$/s^2$] and E_{obs} is the observed energy [J].

$$\eta = \frac{mgh}{E_{obs}}[dimentionless] \tag{5.5}$$

When $\eta = 0.32$, the pump system is operating at the minimal acceptable performance, while $\eta = 0.80$ shows highly-performing pump systems (Spellman, 2003). The analysis of the daily time series of η (fig. 5.4) shows high fluctuations; as already discussed, a simple comparison of the daily values of η with a reference value is not an optimal strategy.

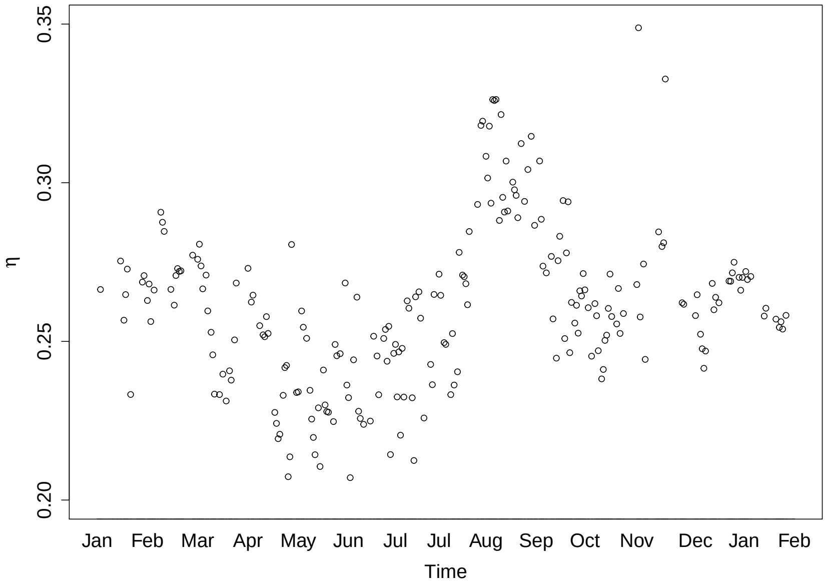


FIGURE 5.4: Subset of time series of η values. This figure shows daily efficiency values for the pump system from January 2015 to March 2016

Consequently, the methodology separates the trend value component of the efficiency index from the daily fluctuations in order to individually assess long-term phenomena (such as pump ageing) and short-term phenomena (mainly influenced by operational conditions)[2].

Trend calculation

The daily trend value is calculated using a rolling window median for the previous 90 days[3]. In a previous work and after extensive testing, it has been showed that this is a robust method for this type of data (Torregrossa et al., 2017d). The number of days included in the rolling median calculation is generally known as the rolling window (W_m). This

[2]For example, if in the period analysed, η has a normal distribution (let us suppose 0.30±0.1) it is more beneficial to consider the average value 0.30 as the representative value for the efficiency in this period and investigate the relation between the deviations from this representative value and the operational conditions. In any case, since it is not possible to establish a priori a normal distribution for η values, the median is used instead of the average.

[3]For example, let us consider the first 180 days of a year. The rolling median algorithm calculates the trend value for the 90^{th} day as the median value of days 1 to 90. Then, the algorithm attributes the median value of day 2 to 91 to the 91^{st} day. This operation is iterated over all the dataset. Please note that the calculation for the first 90 days is made with less data and must be used with caution

parameter can be customized as explained in detail in subsection (Torregrossa et al., 2017b).

If the value of W_m is large enough to be representative of the operational conditions, the trend values represent the performance of the pump system without the effect of short-term phenomena. This fluctuation is calculated as the difference between the daily trend and the daily value. Consequently:

$$\eta = \eta_t + \eta_f \tag{5.6}$$

In this equation, η_t is the trend value while η_f is the fluctuation.

In the following sections, the values of η_t and η_f are investigated separately and new operative parameters are defined.

Calculation of the efficiency slope

In order to use the values of η_t to identify changes in the trend, this paper proposes a parameter, τ, representing the average slope of η_t (Torregrossa et al., 2017d). τ is calculated daily with equation 5.7, in which η_{t_d} is the value of the trend for the reference day and $\eta_{t_{d-I_\tau}}$ is the value of the trend calculated for the I_τ days before the reference day. In our case, $I_\tau = 180$, note: this parameter is customizable as explained in (Torregrossa et al., 2017b). For the values of τ, which represents the pump system performance degradation, SK-DSS proposes two benchmark values: -0.4%/year for a good degradation rate and -1.6%/year as the threshold degradation rate requiring urgent maintenance (Torregrossa et al., 2017d). The values of τ are supposed to be negative if no maintenance is performed, otherwise a positive τ is expected (i.e. an increases in performance) after maintenance. Consequently, τ can be used to evaluate both pump ageing and the effect of maintenance.

$$\tau = 365 * \frac{\eta_{t_d} - \eta_{t_{d-I_\tau}}}{I_\tau * \eta_{t_{d-I_\tau}}} \tag{5.7}$$

Calculation of Z

The values of η_f can be used for the early identification of failures. Normally, the η_f values should fluctuate around the value $\eta_f = 0$. However, if a potential failure (for example a partial-obstruction) occurs in the pump system, the η_f time series will show a series of negative values. Before observing a change in the trend η_t, this effect has to be there long enough to change the 90 day median. Consequently, η_t and τ are not efficient for the early detection of failures. SK-DSS therefore proposes a binary parameter, Z, which is equal to 0 if the system registers 15 consecutive days with $\eta_f < 0$, otherwise Z=1 (cf. eq. 5.8 [4]).

$$\begin{cases} z = 0 \text{ if } max(\eta_{f;d=-1}; \eta_{f;d=-2}; ...; \eta_{f;d=-Wz}) < 0; \\ z = 1 \text{ if } max(\eta_{f;d=-1}; \eta_{f;d=-2}; ...; \eta_{f;d=-Wz}) > 0; \end{cases} \tag{5.8}$$

The length of the sequence of negative-fluctuation days (W_z) can be customized, as explained in (Torregrossa et al., 2017b).

Fuzzy logic and scenario analysis

Fuzzy logic has been shown to be able to store expert knowledge using a human-like language in a series of understandable statements, to deal with uncertainty and to efficiently process multiple parameters; a detailed explanation of fuzzy logic algorithms is provided by Zadeh, (1965) and Starczewski, (2013) or by this thesis at the section 2.5.1.

At this point, this methodology offers four parameters (η_t,η_f,τ,Z) which represent the efficiency trend, the fluctuation in the trend, the ageing of the pump and the existence of potential new failures respectively; consequently a detailed multi-perspective pump assessment can now be carried out. However, the information contained in the four parameters is still not obvious. In order to deliver clear performance information on the pump system under investigation, for the reasons above explained, SK-DSS analyses these parameters with a fuzzy logic engine.

This fuzzy logic approach is based on the set of rules reported in table 5.5. Each rule describes a condition of the pump system and the fuzzy logic produces a score for each rule. Table 5.5 reports the 9 rules of the fuzzy system implemented[5].

[4] In this equation, $\eta_{f;d=-n}$ corresponds to the value of η_f at the n-day before the reference day

[5] For example, the first rule, representing a condition in which the pump system has a high value for the trend and a positive fluctuation, can be read as:

IF η_t IS high AND η_f IS positive THEN Score IS High;

TABLE 5.5: Fuzzy logic rules used

id	η_t	η_f	τ	Z	**Score**
1	high	positive			High
2	high	negative			Medium
3	low	positive			Medium
4	low	negative			Low
5			low		Low
6			medium		Medium
7			high		High
8				low	Low
9				high	High

The first 5 columns report the rule Id, and the input parameters. η_t is the trend, η_f is the fluctuation, τ is the average slope of η_t and Z is the binary parameter as described in equation 5.8. The last column expresses the evaluation of the operational condition, depending on the inputs, that will be transformed to a fuzzy output.

The fuzzy logic algorithm requires the mathematical definition of the input variables, obtained by defining the membership functions which associate the input values to their membership degree[6] (Starczewski, 2013; Zadeh, 1988). The membership degrees associated with the functions are shown in figure 5.5.

For each day, the fuzzy logic engine calculates the result of each rule with the Mamdani implication method (Mamdani and Assilian, 1975), which produces a truth degree (TD) in the range 0-1 for each scenario[7]. In this thesis, the rule with the highest truth degree is defined as the 'winning rule'.

The winning rule is extremely useful for plant managers in a subsequent decision-making process because it effectively describes the current operational condition of the pump system. For example, if rule number 8 has the highest truth degree, there is a decreasing trend over

[6]For example, a trend $\eta_t > 0.6$ is considered 'high' with membership value of 1, a trend $\eta_t < 0.2$ is considered 'low' with membership value of 1, while in the region between 0.2 and 0.6 both definitions (high and low) are valid with a different membership degree (for example the trend $\eta_t = 0.4$ is at the same time 'low' and 'high' with a membership value of 0.5). Consequently, the membership values describe a state function.

[7]For example, the statement of the first rule is: 'η_t is high and η_f is positive'. If the output of the first rule is 0, this means that this statement is 'false'; if the output of the first rule is 1, this statement is 'true'. For values of truth degree between 0 and 1, the statement is partially 'true' and partially 'false'.

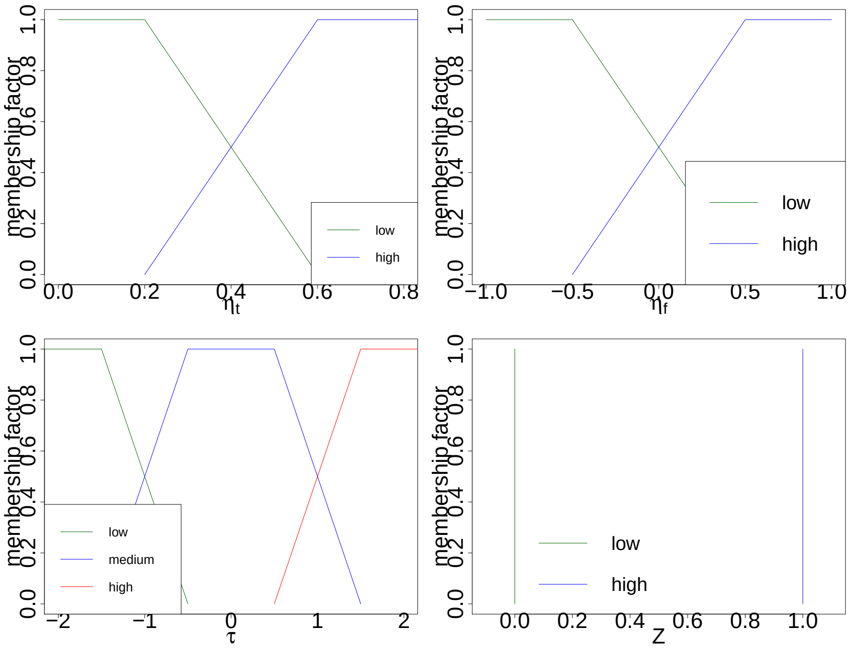


FIGURE 5.5: Membership function for each input

the last two weeks (η_f was negative) and an urgent investigation is required.

However, to fully understand the system behaviour, it is important to simultaneously observe the truth degree of each rule. It is also important to understand that the rules which use the same parameters are complementary and the sum of their membership values is always 1. In this case, there are 3 blocks of complementary rules: 1-4, 5-7, 8-9. The first block reports information about the pump conditions on the day of analysis, the second block analyses long term phenomena affecting pump degradation and the last block monitors potential failures. This rule structure was imposed to avoid illogical results: for example, a pump cannot be fully efficient and fully inefficient at the same time. With the current rule structure, the fuzzy engine will provide 3 fuzzy statements each on:

- pump condition (by analysing the block 1-4);
- pump degradation (by analysing the block 5-7);
- early detection of inefficiencies (by analysing the block 8-9).

The rules for each block are independent, which means that, for example, it is possible that a pump system with a high efficiency is experiencing a high degradation rate, or a normal degradation rate. The analysis of the winning rules provides the correct global view. Table 5.7 reports a set of remedial suggestions which are related to specific rules.

5.3.5 Economic consideration and flow-related issues

The methodology published in (Torregrossa et al., 2017b) presents also the opportunity to analyse the suggestions with an economic approach and a graphical approach to detect flow-related issues. In this thesis, this part is omitted because it was decided to focus on the fuzzy logic aspects. This part can be found in (Torregrossa et al., 2017b).

5.3.6 Results

Index calculation

After the calculation of the daily indices with equation 5.5, the signal is decomposed and the indicators calculated as described in subsections 5.3.4, 5.3.4 and 5.3.4. For each day, SK-DSS obtains daily values for η_t,η_f,τ and Z. Table 5.6 shows the summary of the decomposition analysis, i.e. representative statistical values for the key performance indicators calculated.

TABLE 5.6: Summary of decomposition analysis for the WWTP in Burg

	η_t	η_f	τ	Z
Min	0.23	-0.06	-0.18	0.00
1st Quartile	0.25	-0.02	0.10	1.00
Median	0.27	0.00	0.17	1.00
Mean	0.26	0.05	0.15	0.88
3rd Quartile	0.27	0.03	0.23	1.00
Max	0.29	0.68	0.40	1.00

Table 5.6 shows that the trend efficiency is below the threshold for the normal efficiency ($\eta_t < 0.32$, Spellman, 2003). The range of η_f reflects the impact of short term phenomena on pump efficiency. The τ values are in the range of normal pump performance deterioration. The first quartile value of Z shows that for at least 75% of the time, there is no 15-day sequence of negative fluctuations. In other words, table 5.6 shows a low-efficiency pump system, which is stable for long periods and with high fluctuations in efficiency in the short term.

5.3.7 Results of fuzzy logic and scenario analysis

The first output of the fuzzy logic system is an overall performance index in the range 0-100. For BUR, the fuzzy score for each day is below the optimal performance value. This score can provide a direct explanation of pump performance. For example, when figures 5.4 and 5.6 are compared in the period between July and September, the time series of η values varies greatly due to operational conditions; in the time series of the fuzzy score, the operational conditions have a minor impact on the overall evaluation and the score is stable in the region of 60%-70%, corresponding to a sub-optimal condition due to a low η_t. In the plot 5.6, there are points in which the score is under 40: these results occur because of Z (i.e., there was a sequence of negative values).

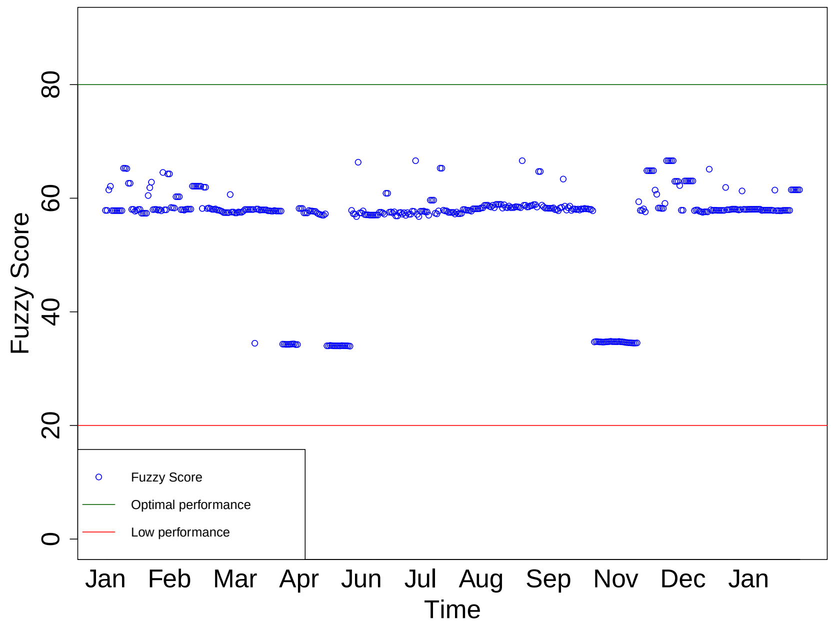


FIGURE 5.6: Fuzzy Score time series for BURG from 1 December 2015 to 1 March 2016

The analysis of the truth degree of each rule can explain the scenario which mostly influenced the fuzzy score. For example, the fuzzy score of 29 April 2015 was equal to 34. For that day, the rules with the highest truth degree were rule 4, rule 6 and rule 8. This means that, the η_t was low, the fluctuation was negative, the deterioration speed was low and there was no sequence of 15 negative fluctuations. This example highlights the large amount of information obtainable relative to the

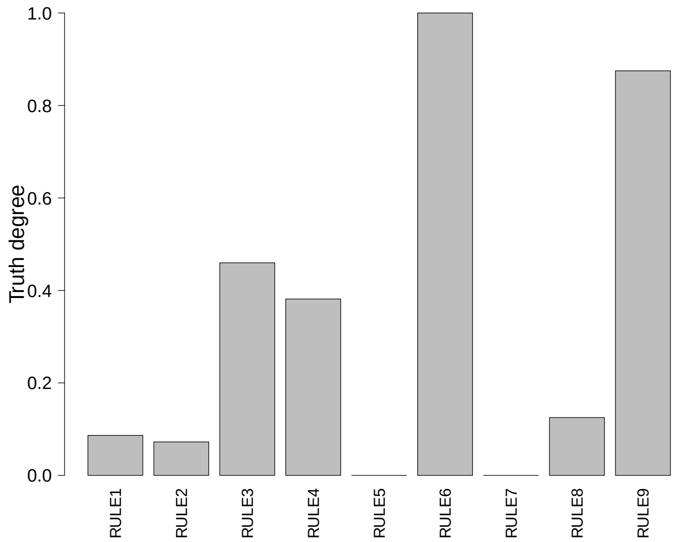


FIGURE 5.7: Strength of the rules

classical η calculation. With the classical approach, the only information for that day is that $\eta = 0.20$.

Fuzzy logic can also support the analysis of patterns in a period. For example, figure 5.7 reports the average truth degree for each rule in the period between 1 January 2015 and 1 March 2016. The plot shows that rules 3, 6 and 9 have a truth degree higher than 0.5 and consequently they describe the operational conditions dominant in the system.

When these results are related to the set of solutions shown in table 5.7, it can be observed that a potential solution for the operational condition described by rule 3 corresponds to a non-urgent maintenance action, while rules 6 and 8 do not require any action. Consequently, the plant manager should schedule a non-urgent maintenance. The analysis of potential cost savings estimates the maximum investment that the plant managers should accept to pay: €60022 (details in Torregrossa et al., 2017b). A numerical example of the calculation of the potential cost savings is available in Appendix B.

Availability of the current methodology for field applications

The methodology is currently suitable for being applied directly to centrifugal pump systems, regardless of their size or the pump system configuration. The methodology relies on data measured on-line and aggregated at a daily resolution, as proposed in the EOS system (Torregrossa et al., 2016). This requires an initial set-up of a system able to gather data from remote sensors, process, aggregate and make them available to the decision tool. In the WWTP domain, the required sensors are generally available. After the connection of the WWTPs to the EOS dataset, two options are available to apply this methodology:

- on-line analysis; an automated script processes the data at fixed-time interval and submits the results to the SCADA system;
- on-the-fly analysis; a script, with a graphical interface using customizable parameters, analyses the data and produces the results on-screen.

In the first case, the results are immediately accessible for plant managers, no interaction with the software is required and the analysis is performed over all the records. In the second case, the software requires a parameter set-up but the plant manager can carry out specific tests, for example analysing a specific period. These two approaches can coexist.

TABLE 5.7: Example (subset) of suggestions related to the rules

Rule	**Suggestion**
1	Do nothing
2	If rule 9 has TD=1, do nothing. If rule 8 has TD=1, plan maintenance urgently.
3	The efficiency is not sufficient. Plan maintenance or evaluate pumps replacement. Are there flow-related patterns? If yes, install a pony pump
4	The efficiency is not sufficient. Plan maintenance, urgently if, for rule 8, the TD=1. Evaluate pump replacement. Are there flow-related patterns? If yes, install a pony pump If the pump is more than 15 years-old, consider replacing it
5	Pump performance is decreasing fast. Plan maintenance.
6	Pump performance slope has a normal value. Do nothing.
7	Pump performance is increasing. Is this the effect of maintenance? If not, check the data consistency.
8	In the last 15 days, the fluctuations have been negative. Are there flow-related patterns? If yes, install a pony pump.
9	The fluctuations are regular.

5.4 Considerations about the assessment of energy consumer devices

In this section, it has been explained how SK-DSS can deal with WWTP pump monitoring. This is obtained through a plant generic approach able to perform a high-frequency analysis of the pump systems and provide case-based suggestions. The pump monitoring tool jointly with the blower monitoring tool can contribute to the energy optimization of the most energy-intensive devices. This approach is so flexible that it is immediate to imagine the application to other devices. The analysis of more devices has not be done because of time limitations and it was better evaluated to focus on the biogas production optimization (ref. to section 5.5).

5.5 Biogas analysis

The energy production by biogas plays a key role in the WWTP energy balance. In WWTPs, biogas is the product of sludge anaerobic digestion (AD) and it is mainly composed by methane (CH_4,50-70%) and carbon dioxide (CO_2,30-50%) (Shen et al., 2015); the energy is extracted by the methane portion through combustion. For USA, Metcalf and Eddy, 2014 (pag. 1521) reports a maximum value of biogas production equal to 28 l/persons. In (Hansen, 2018) and in figure 5.8, for Europe, a normal production of biogas is estimated to be between 7 $.3m^3/PE/year$, corresponding to 20 l/persons/day. The biogas can then be converted in electric and thermal energy. The efficient conversion of this quantity of biogas in energy has a relevant role in energy management. Gude, 2015 reports that on average the biogas production can satisfy between the 25-50% of the WWTP electric energy demand. According to Shi, 2011, this figure is around 33%. In (INNERS, 2015), the power self-sufficiency rate is estimated in the range 55-70 %, with the possibility to increase this threshold by co-digestion with organic material from other sources. On the other hands, a non optimal management of biogas production can become an intensive source of green house gases (GHG) and increase the carbon footprints of WWTPs (Shen et al., 2015). All these considerations increase the environmental and economic interest for the optimal management of biogas production.

Therefore, in the recent years, many authors focussed their efforts to increase the biogas production and the energy generation. In particular, literature shows two main approaches: development of new technologies and optimal management of existing ones. For example, for the first category, Budych-Gorzna, Smoczynski, and Oleskowicz-Popiel, 2016 obtained an increased biogas production (+80%) by co-digestion of sewage sludge and industrial waste. MosayebNezhad et al., 2017 coupled solid oxide fuel cell (SOFC) systems and micro gas turbines to increase the energy self-coverage of WWTPs by 15%. Maragkaki et al., 2017 co-digested the sewage sludge with an organic dried mixture and increased the biogas production by 2.7 times.

The second category (optimal management) accounts also an interesting amount of applications. Björnsson, Murto, and Mattiasson, 2000

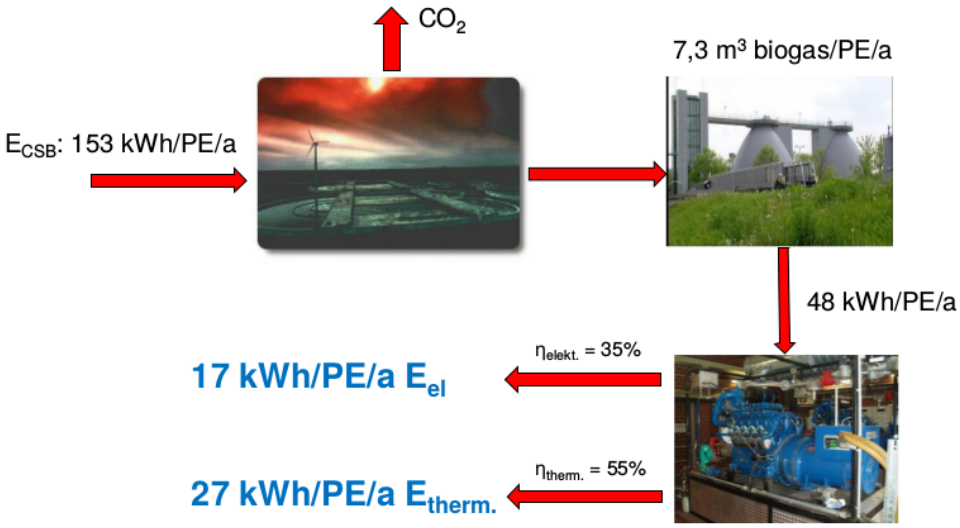


FIGURE 5.8: Energy balance biogas, extracted from (Hansen, 2018)

monitored on-line the AD of a wastewater treatment plant using the values of pH, alkalinity and Volatile Fats Acids. Bernard et al., 2005 presented the TELEMAC system, able to remotely monitor the anaerobic digester and identify potential faults. Boe et al., 2010 proposed an AD monitoring based on on-line and off-line parameters and tested the response of the system to increasing of organic load. Madsen, Holm-Nielsen, and Esbensen, 2011 proposed a review of anaerobic digestion monitoring approaches. Kusiak and Wei, 2014 applied a data mining approach to model and predict the methane production from a WWTP; using an Adaptive Neuro-Fuzzy Inference System algorithm, Kusiak and Wei, 2014 identified the flow rate, the volatile fatty acids loads and the detection time as the most influencing parameters. Akbaş, Bilgen, and Turhan, 2015 used neural networks to predict, model and optimize the biogas production (+71%) by monitoring and control many process parameters (such as sludge retention time, alkalinity, pH and temperature). The biogas production was one of the parameters taken into consideration by Panepinto et al., 2016 for the evaluation of energy efficiency of a large WWTP. The energy on-line system (EOS,Torregrossa et al., 2016) perform the on-line monitoring of the energy production from biogas. An interesting application of on-line monitoring of digestion processes has been performed by Li et al., 2017 on piggery wastewater. Recently, Robles et al., 2017 proposed a fuzzy-logic controller based on the effluent total volatile fatty acids concentration to control the methane production of industrial winery wastewater. Moreover, Garrido-Baserba et al., 2015 and Turunen, Sorvari, and Mikola, 2018 built a decision support system for the optimal selection of sludge treatment.

As already done for pump and blower, in this section, a decision support tool is developed to be included in SK-DSS.

The demonstration of biogas tool is done with the YouTube based version of DSS described in (Torregrossa and Hansen, 2018) and in the section 5.7. Being part of SK-DSS, this tool inherits these characteristics: 1) it is plant generic, i.e. able to simultaneously work with many WWTPs regardless their specificities (such as the size or the connected population); 2) it is fully based on on-line measurements; 3) it is based on the daily assessment of key performance indicators; 4) it is coupled with artificial intelligence algorithm for the estimation of missing values; 5) it provides a fuzzy-logic based scenario evaluation; 6) it provides case-based suggestions; 7) it enables the cooperation between experts and end-users. This decision support tool has an **added value** because:

- it is flexible to be applied to many WWTPs;
- it takes into consideration the parameters with time-lag in order to account the events in the past;
- it provides a user-friendly interface based on YouTube to enable the cooperation between end-users.

Despite the large interest of researchers, literature review did not show any decision support tools with the above mentioned characteristic for the biogas optimization.

5.5.1 Material and Method

The wastewater treatment plant used to test the methodology is a conventional activated sludge plant equipped with a biogas digester and a CHP engine. The design inflow is around 5000 m^3/d, the average pollution load is around 2800 kg BOD_5/d. The plant is equipped with 2 aeration basins with a volume 3000 m^3. The sludge digester has a volume of 1150 m^3 and the gas storage tank consists of 450 m^3.

The available information useful for this investigation are: daily inflow, 1 value/week of BOD5 concentration at inlet, the amount of produced biogas, the sludge flow to digester, the pH of the digester and the temperature of the digester.

The methodology proposed is explained in fig. 5.9. The first 4 blocks (from data gathering to KPIs calculation) are explained in detail in (Torregrossa et al., 2016). The data are gathered from on-line sensors, cleaned and aggregated in daily data. The chemicals parameters such as BOD and COD are measured once each two weeks and the daily missing values are estimated with a random forest algorithm in order to enable the calculation of daily KPIs. The uncertainty of estimation model is calculated to be taken into consideration in the following steps. A detailed dissertation can be found in (Torregrossa et al., 2016). At the end of these steps the algorithm can calculate KPIs for each day.

The daily available information for biogas process monitoring are: date, pH of the digester, solid retention time, temperature in the digester, estimated population equivalent with uncertainty and biogas production. Other parameters such as the quality of the sludge or the biogas composition are available with low frequency (few information per months) and without regularity; therefore they are not included in the automatic algorithm but they can be used by plant managers as additional information to be integrated in the decision making process.

After these steps, the model shown in fig. 5.9 consists of two independent assessments:

- benchmarking of biogas production that takes as input the estimated population equivalent with uncertainty and biogas production;
- the fuzzy-logic based evaluation of the process that takes into account the pH and temperature of the digester, and solid retention time.

Each assessment analyses the anaerobic digestion process from different perspective; the first one looking at the output of the process, the second one looking at the process parameters. The results of these two assessments are independent and can be compared. If they converge the fuzzy logic analysis can be used to explain the operational scenario and suggest solutions as proposed in (Torregrossa et al., 2017a). If they diverge, it is necessary to look for additional information. The divergence can be caused by the uncertainty associated to the estimated population equivalent and/or to the effect of parameters not automatically taken into consideration (such as the concentration of volatile fat acids in the digester). In this case, it is strongly suggested to retrieve more information to be included in the decision making process. Anyway, as it will be explained in the results, the test divergence in this analysis is rare. The subsections 5.5.3 discusses in detail the first test, while the subsection 5.5.4 explains the evaluation of AD process quality through a fuzzy logic analyser.

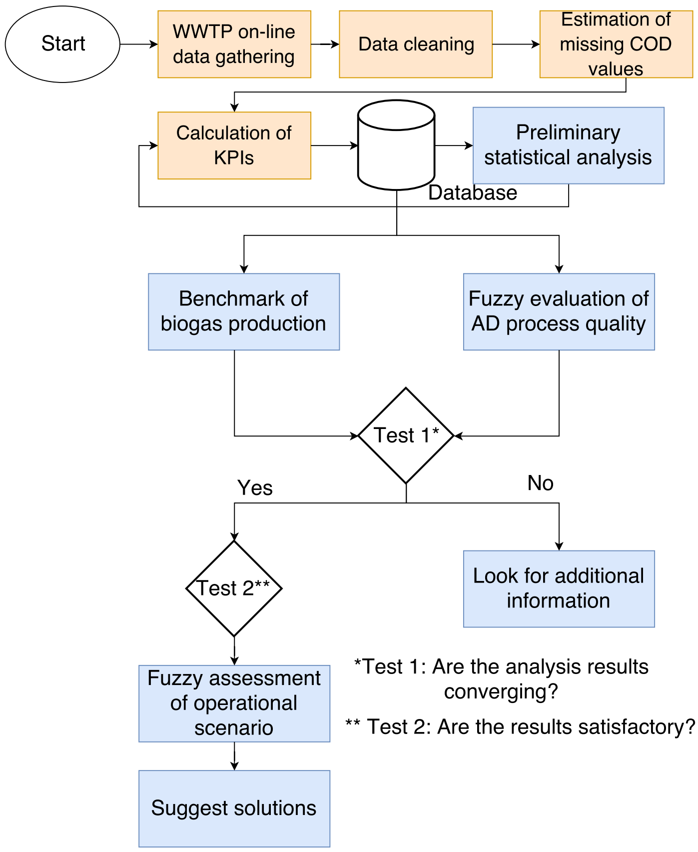


FIGURE 5.9: Diagram flow algorithm

5.5.2 Preliminary statistical analysis

Once obtained the historical data, it is worthy to perform some statistical test to better identify the relationship between variables. In particular, since the AD parameters could need many days to return in a normal range after a shock (Björnsson, Murto, and Mattiasson, 2000), it is worthy to identify the optimal time frame for the investigation. Therefore a multivariate VAR model has been generated and the analysis of Akaike's Information Criterion (AIC,Shumway and Stoffer, 2011) identifies the optimal time-frame in the range [day_t, day_{t-8}], in which day_t is the day under analysis and day_{t-8} is the 8^{th} day before day_t. Therefore, in the following analyses the tests will take into consideration the events in a time frame of 8 days. Moreover, it has been calculated the uncertainty in the estimation of population equivalent: 12.95%.

5.5.3 Benchmark of biogas production

The first test consists of benchmarking the biogas production against a reference value. Metcalf and Eddy, 2014 proposed a target value of 28 l/pe/day, while Hansen, 2018 propose a value of 20 l/pe/day. Haberkern, Maier, and Schneider, 2008 adopt 20 l/pe/day as guide value (i.e. a value normally obtained by well managed WWTPs) and 30 l/pe/day as target value to be obtained with an optimal management. In this thesis, the value of 25 l/pe/day is assumed as benchmark for an ideal biogas production.

The key performance indicators to be compared to this values have to take into consideration 2 elements: the optimal time frame and the uncertainty associated to the population equivalent. Therefore, instead of calculating a single KPI, the system calculates a KPI range standing between a minimum (equation 5.9) and a maximum value (equation 5.10).

$$Biog_{min} = \frac{biogas}{\overline{PE}_8(1 + 12.95\%)}[l/pe/day] \tag{5.9}$$

$$Biog_{max} = \frac{biogas}{\overline{PE}_8(1 - 12.95\%)}[l/pe/day] \tag{5.10}$$

In equations 5.9) and 5.10), the elements have the following meaning: $Biog_{min}$ and $Biog_{max}$ are the minimum and the maximum specific

TABLE 5.8: Example of KPI calculation

day	PE	Biogas [l]
1	99000	2697750
2	109000	2534250
3	100000	2475000
4	90000	2295000
5	107000	2755250
6	97000	2473500
7	99000	2549250
8	104000	2444000
Mean PE	100625	
	8th Day	
$Biog_{max}$	27.60	[l/pe/day]
$Biog_{min}$	21.69	[l/pe/day]

biogas production [l/pe/day]; the term $biogas$ corresponds to the daily amount of biogas production measured in the plant [l/day], $\overline{PE}_8$ is the average value of population equivalent in the period [day_t, day_{t-8}] and the value 12.95% corresponds to the correction factor due to uncertainty.

This way to perform the benchmarking of biogas production is innovative, because, currently, the benchmark practice does not take into consideration the time-lag. Table 5.8 shows an example of calculation of the KPIs referred to day 8 of a simulated time-series. The average value of population equivalent in the 8 days is 100625 PE. The value of PE is estimated with random forest and, in this case, there is an associated uncertainty of 12.95%. The value of biogas at the 8^{th} day is 2444000 l. According to equations 5.9 and 5.10, the minimum value of KPI is 21.69 l/pe/day and the maximum is 27.60 l/pe/day. In this example, the biogas benchmark is between maximum and minimum KPIs, i.e. it is necessary to perform additional investigations by taking into consideration other information (such as the process parameters).

The result can be represented as a time series (ref. upper plot , fig. 5.10).

5.5.4 Fuzzy evaluation of AD process quality

The second test processes the pH of the digester, the temperature and the solid retention time to provide a quality assessment of the process. In order to take into consideration the effect of time lag, the system calculates the following parameters:

- $pH_{min,8}$; the minimum value of pH in the period [day_t, day_{t-8}];
- $pH_{max,8}$; the maximum value of pH in the period [day_t, day_{t-8}];
- $T_{min,8}$; the minimum value of temperature in the period [day_t, day_{t-8}];
- $T_{max,8}$; the maximum value of temperature in the period [day_t, day_{t-8}];
- $SRT_{min,8}$; the minimum value of SRT in the period [day_t, day_{t-8}];
- $SRT_{max,8}$; the maximum value of SRT in the period [day_t, day_{t-8}];

These parameters are used as input of the rules reported in tab. 5.9. The rule in this table need to be translated by the system in fuzzy rules. For example, the first rule becomes:

RULE 1: IF $pH_{min,8}$ is good AND $pH_{max,8}$ is good THEN Score IS High;

In few words, the system of rules applies a high score to the rules in which all the parameters are in the correct range otherwise it applies a low score value. The final score is calculated with the combination of the rules performed with the Mamdani fuzzy inference method (Starczewski, 2013). The parameters of the fuzzification process are the following:

- $pH_{min,8}$ is 'good' with support in range 6.9-7.3 and kernel in range 7-7.2;
- $pH_{max,8}$ is 'good' with support in range 6.9-7.3 and kernel in range 7-7.2;
- $T_{min,8}$ is 'good' with support in range 35-40 °C and kernel in range 37-38 °C;

TABLE 5.9: Fuzzy rules

Rule	$pH_{min,8}$	$pH_{max,8}$	$T_{min,8}$	$T_{max,8}$	$SRT_{min,8}$	$SRT_{max,8}$	Connector	Score
RULE1	good	good					AND	High
RULE2	NOT good	NOT good					OR	Low
RULE3			good	good			AND	High
RULE4			NOT good	NOT good			OR	Low
RULE5					good	good	AND	High
RULE6					NOT good	NOT good	OR	Low

- $T_{max,8}$ is 'good' with support in range 35-40 °C and kernel in range 37-38 °C;
- $SRT_{min,8}$ is 'good' with support in range 10-48 days and kernel in range 15-45 days;
- $SRT_{max,8}$ is 'good' with support in range 10-48 days and kernel in range 15-45 days.

For a rigorous mathematical definition of support and kernel, please refer to (Starczewski, 2013).

The fuzzy logic analysis returns for each day a quality evaluation score depending on the above cited parameters. This score can be represented as a time series (ref. fig. 5.10, bottom plot) and compared with the result of the first evaluation.

Analysis of the results

The assessments performed with biogas production benchmarking and the fuzzy evaluation of the AD process need to be compared in order to produce a robust analysis. The comparison of the results with their relative reference values (25 l/pe/day for the first assessment and a fuzzy score equal to 80 for the second assessment) can generate four cases:

1. both assessments indicate a non optimal operational condition of AD;
2. both assessments indicate an optimal operational condition of AD;
3. the biogas production is satisfactory while the fuzzy analysis of process parameters produces a score not sufficient;
4. the biogas production is not-satisfactory even if the fuzzy analysis of process parameters indicates a good process operation.

For the decision process, the cases 1 and 2 are not ambiguous and the result of the assessments can be considered robust because based on two independent approaches. In this case, as already done in (Torregrossa et al., 2017a) for the blower assessment and in (Torregrossa et al., 2017b) for the pump assessment, the analysis of membership factors of the fuzzy rules can be used to explain the phenomena occurring in the plant (fig. 5.11).

In the cases 3 and 4, the two analyses diverge. The reason of this divergence can originate from a high uncertainty in the estimation of the connected population or from an inadequate set-up of parameters not in the fuzzy logic input (such as the alkalinity). In this case, the proposed solution consists of a more detailed investigation that takes into consideration new parameters.

5.5.5 YouTube based cooperative decision-making support

After the identification of the operational conditions of the AD trough the fuzzy rules, the algorithm is able to provide case-based suggestions. This is done using the popular video-sharing website YouTube, that in this case is forced to work as a platform for the cooperative decision support system. Each scenario is linked to a YouTube web page, in which a short video explains the occurring operational condition. In add, the YouTube platform can be used to share comments, videos and documents between operators in a high interactive environment. Moreover the YouTube platform has a 'like' system that can be used by end-users to evaluate and peer-review the suggestions of the network. For a detailed explanation of YouTube platform please refer to (Torregrossa and Hansen, 2018) and section 5.7.

5.5.6 Results

After the preliminary analysis of data, the KPIs and the input of fuzzy logic parameters were calculated over a period of 4 years. In this period the specific biogas production was comprised in a range between 6.61 and 30.67 l/pe/day (with a mean value of 16.52). The pH of the digester shows a great variability being in range between 2 and 12 (the average value is close to 10). The temperature is comprised between a minimum

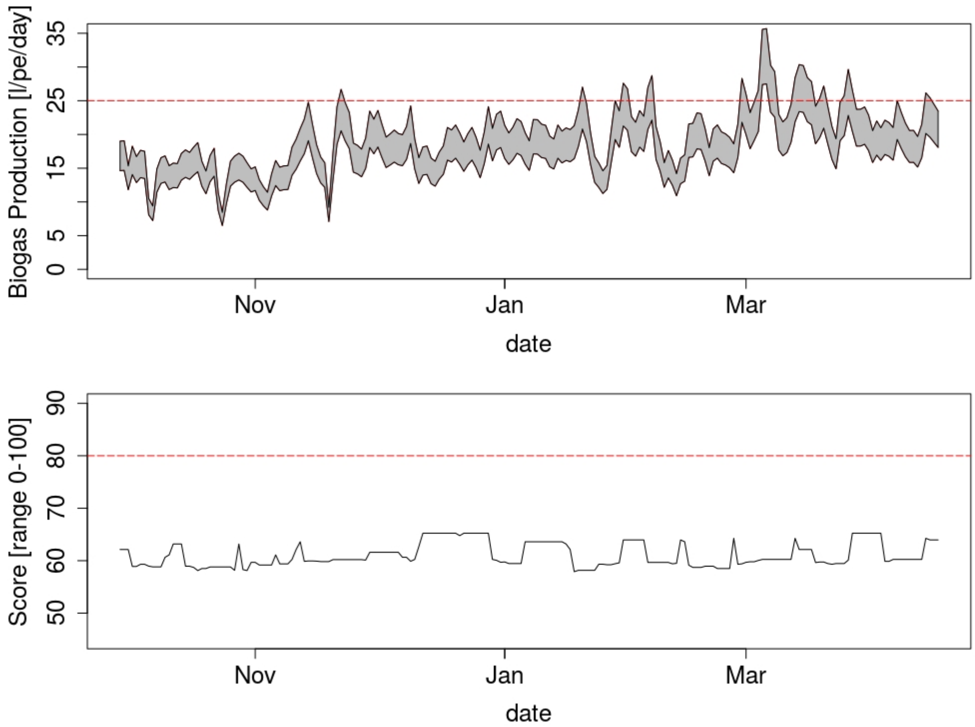


FIGURE 5.10: Result of tests between Oct-2014 and April 2015. First plot: benchmarking of biogas production . Second plot: time series of fuzzy logic classification

of 33-41 °C and the solid retention time between 15-49 days. This first summary of process parameters helps highlighting some aspects:

- the biogas production is generally inefficient with a great variability;
- potential issues are connected to the control of the pH.

Nevertheless, this kind of analysis of aggregated data is not accurate enough to improve our knowledge about the AD dynamics and to use the high-frequency data to automatically support the decision making-process in the framework of SK-DSS system (Torregrossa et al., 2017a). The main interest is in a methodology able to automatically process the data flowing day by day from the plant to the SCADA system and provide case-based solutions (Torregrossa et al., 2017a). In order to integrate the biogas assessment in such a decision support system, the algorithm shown in fig. 5.9 needs to be automatically run with daily frequency.

The results can be accessed though database queries or in a visual form as shown in fig. 5.10. This figure helps to compare the results of

TABLE 5.10: Link to YouTube cooperative platform

Rule id	Link to platform
Rule 1	https://youtu.be/T3llzU8KWOs
Rule 2	https://youtu.be/e0pqNxlwmjw
Rule 3	https://youtu.be/-i248rdkDLs
Rule 4	https://youtu.be/fmQgePW7_us
Rule 5	https://youtu.be/dLl01NZ0-hk
Rule 6	https://youtu.be/jjGB5aChZFk

the biogas benchmarking with the results of the fuzzy process; in particular it is observed that the biogas production range is always under the desired value of 25 l/pe/day, except for some days, in which the biogas production range crosses the benchmark horizontal line (upper part of fig. 5.10). Coherently, the fuzzy logic engine produces a sub-optimal score (bottom part of fig. 5.10). The convergence of these two independent analyses enables to use fuzzy logic to enlighten the AD dynamic. In particular, the plot 5.11 reports the importance of the rules expressed in table 5.9. The rule 1 and 2 describe the conditions related to pH conditions, rule 3 and 4 concern the temperature operational range and the rule 5 and 6 relate to the solid retention time. As explained before, these parameters are processed to take into account an optimal time-lag to take into consideration the events in the past. The analysis of the rules has to be done for rule blocks (fig. 5.11). The block 1-2 explains that there is a recurring wrong set-up of the pH range, that never is optimal. The block 3-4 shows that in more of 60% of the selected day, the temperature is out of range. The block 5-6 shows a recurring not optimal value of solid retention time. More detailed information can be obtained by reducing the time frame of block analysis (up to 1 day assessment). Therefore the priority for a plant manager should be to restore optimal pH conditions.

The case-based strategies for improvements can be discussed in detail in the YouTube based cooperative platform. The link associated to the platform web-pages are available in table 5.10 and reported in table 5.11.

TABLE 5.11: Example of solutions

Rule id	**Description**	**Solution**
Rule 1	pH in the range	No action
Rule 2	pH out the range	If $ph < 6.8$ inhibits the methane production. You can correct it using lime or sodium bicarbonate
Rule 3	Temperature range is optimal	No Action Check the fuzzification process [mesophilic/thermofilic set-up]
Rule 4	Temperature out the range	Adjust the temperature set-up Verify the efficiency of temperature controller
Rule 5	SRT in the range	No action Optimal range: 20-40 days
Rule 6	SRT out the range	Manage the digester flow Verify the reduction of effective volume caused by grit accumulation

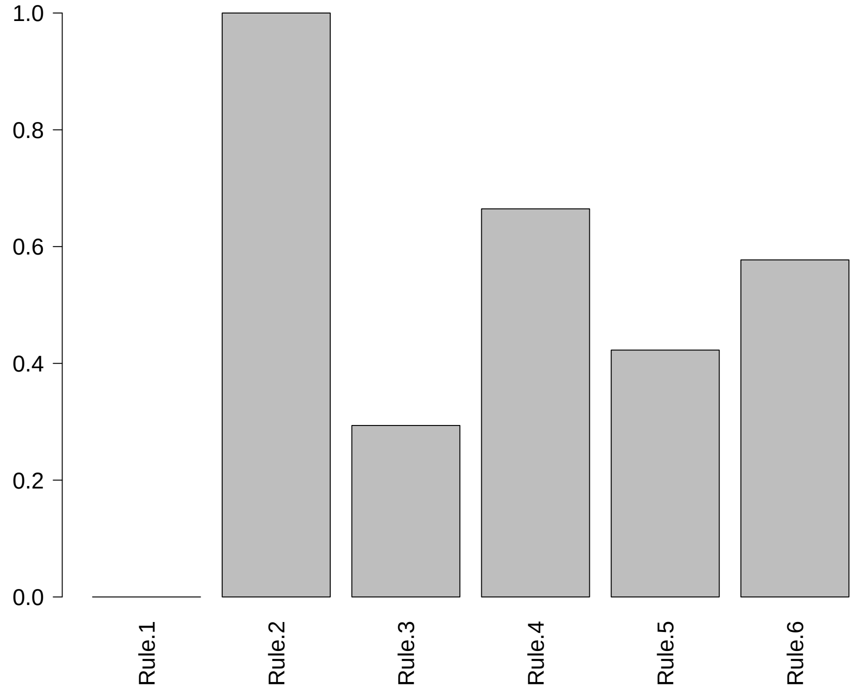


FIGURE 5.11: Result of tests between Oct-2014 and April 2015. The block 1-2 explains that there is a recurring wrong set-up of the pH range, that never is optimal. The block 3-4 shows that in more of 60% of the selected day, the temperature is out of range. The block 5-6 shows a recurring not optimal value of solid retention time.

5.6 A multi-level fuzzy logic

In the previous sections, this thesis has shown how fuzzy logic can be applied to different parts of the plants to produce specific analyses. With the application of SK-DSS to several devices, the amount of information to be analysed increases, and a synthetic plant performance index becomes necessary. This section of the thesis will show the concept of a system to produce a global index and evaluate the global performance of the plant devices.

Figure 5.12 shows the diagram flow: there are many specific fuzzy systems (for pumps, blowers, biogas, etc...) that produce a synthetic score for each device (or device groups); on the top, there is another layer of analysis, in which a global fuzzy system processes the synthetic scores of devices in order to produce a global evaluation score.

Table 5.12 shows the set of rules proposed. The scores of pump,

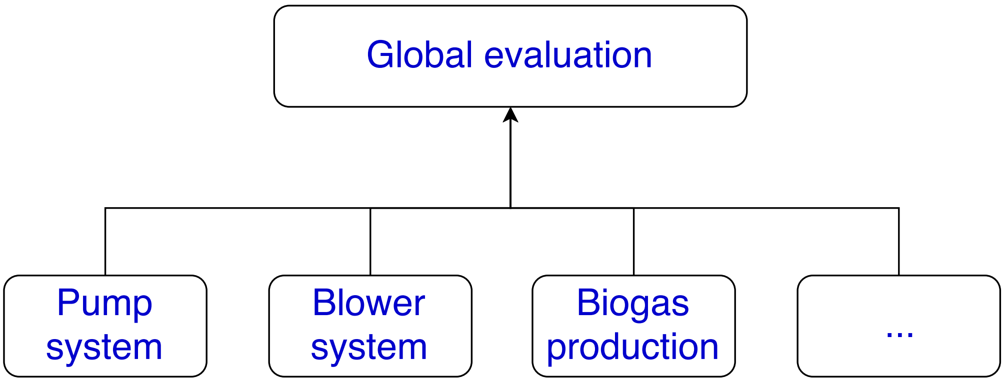


FIGURE 5.12: Flowchart of multilevel fuzzy logic

blower and biogas are in a range 0-100. For the fuzzification process, the system consider 'high' the device performances with $fuzzy - score > 80$ and 'low' the performance with $fuzzy - score < 20$. According to fuzzy logic algebra, all the intermediate stages can be defined at the same time 'low' and 'high' with different degree of truth (sect. 2.5).

This global evaluation layer is necessary in order to monitor many devices and have a synthetic information at a glance. In future developments, this global evaluation index will be linked to an alert system able to reach instantaneously the plant managers (for example by mails or sms) when a critical performance is detected.

TABLE 5.12: Rules of multilevel fuzzy system

Rule	Pump Score	Blower Score	Biogas Score	Global Score
1	Low	High	High	Medium-High
2	Low	High	Low	Medium-Low
3	Low	Low	High	Medium-Low
4	Low	Low	Low	Low
5	High	High	High	High
6	High	High	Low	Medium
7	High	Low	High	Medium
8	High	Low	Low	Low

Fig. 5.13 shows the results of an hypothetical WWTP in which pump and biogas performance are stable (on the long period) and mainly low, while the blower system performance starts from higher levels and drop down. Because of the weight attribution obtained with the rule set-up (Table 5.12), the global performance trend is strongly affected by the blower performance.

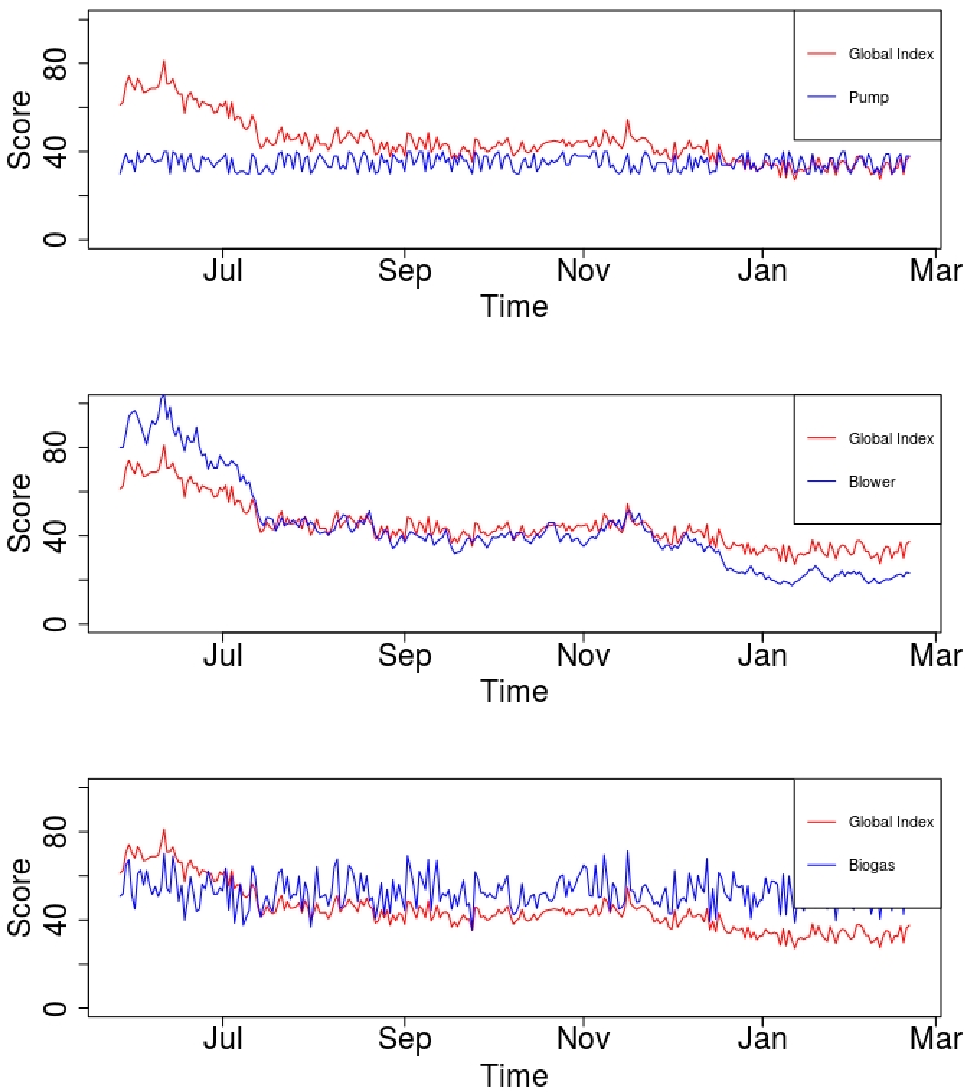


FIGURE 5.13: Results of multilevel fuzzy logic

This system is really flexible because it enables to customize the main parameters and attribute a different weights to the devices at the bottom layer. The results are easy to read and there is no limit to the number of devices to be analysed at the bottom layer. Moreover, in case the number of devices increases too much, it is possible to aggregate them with the present methodology by adding additional bottom-layers.

5.7 A YouTube Based Platform

SK-DSS needs to detect operational condition occurring in the plants, provide solutions and enable the cooperation between end-users.

In (Torregrossa and Hansen, 2018), a YouTube-based version of the cooperative platform has been presented. This is called SK-DSSy in which the 'y' is added to remark the interaction between SK-DSS and YouTube.

From (Torregrossa and Hansen, 2018):
" In SK-DSS (Torregrossa et al., 2017a), the cooperative platform was based on a PostgreSQL table that plant managers can use to visualize the results and upload new solutions. The main limitation of this approach was that plant managers should be able to use PostgreSQL queries. A not-addressed question concerned the evaluation of the solutions proposed by the common platform; according to the philosophy of this cooperative platform, each plant operator is authorized to upload solutions. This open-access approach does not guarantee the quality of proposed solutions. The quality of the end-user contribution could be decreased by several issues, for example: a limited comprehension of the variables, the upload of plant-specific solutions or the limited experience of the end users.

SK-DSSy proposes to overcome these issues by incorporating YouTube in the cooperative platform. In SK-DSSy, the fuzzy logic rules are connected to YouTube web-pages in which the solutions can be uploaded as video or as comment. SK-DSSy identifies the dominant scenario and lead the end-user to an associated video-page. For each rule, the main video is inserted by the page administrator; it explains the operational scenario connected to the rule, some basic solutions and, in order be time efficient, it lasts less than 1 minute. The end-users can visualize the videos, comment it with a text, add videos or link to external resources (such as papers or other web-pages). Another important YouTube function is the 'like' command. Each user can mark with a 'like' the useful suggestions and with a 'dislike' the comments considered not useful. The comments are consequently scored and it is possible to sort them by popularity.

The advantages of this approach are: i) the knowledge can be shared in many formats (such as video, text or links), ii) the use of YouTube

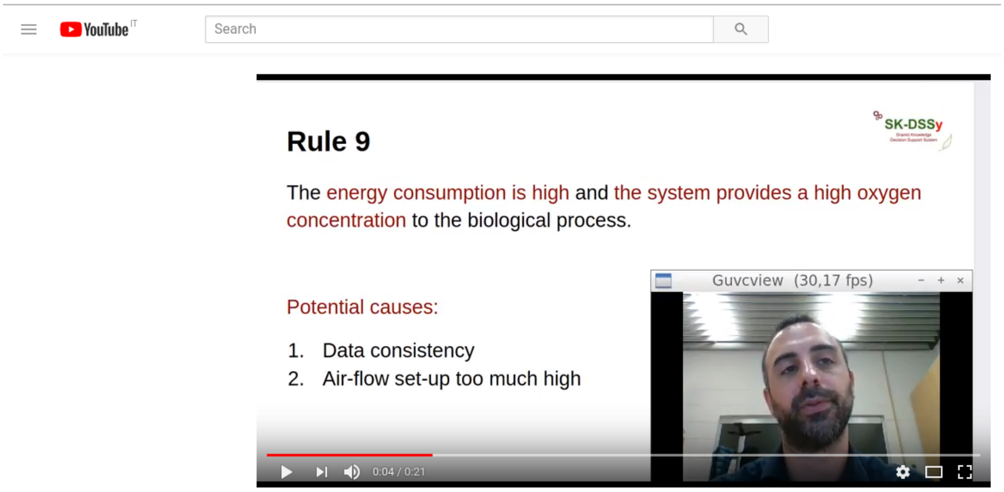


FIGURE 5.14: Screen-shot of intro-video

platform is extremely user-friendly, iii)the server and the maintenance of a part of SK-DSSy is externalized to Google-YouTube services without costs, iv) possible assessment of the proposed solutions.

....

The end-user is redirected to the YouTube page with:

- A video that explains the operational condition of the blower system;
- The discussion between plant operators;
- Additional resources (link to papers, repository, software);
- The evaluation of each suggestion by means of the 'like' system

Figure 5.14 shows a screen-shot of a video connected to a blower rule. In this videos, scenario explanations with potential causes are provided.

In the same web-page, the comment thread is activated. Figure 5.15 shows for example a piece of conversation in which a sensor failure question is addressed. The comment thread can be also used to share other documents; for example, in the answer shown by fig. 5.15, a link to an external book is provided. The 'like' system is used to score the answer and the contributions: the contributions with more 'like' can be sorted and visualized on the top of the page.

In summary, with SK-DSSy, the plant operators have a decision-support system able to perform the on-line analysis of WWTPs and to

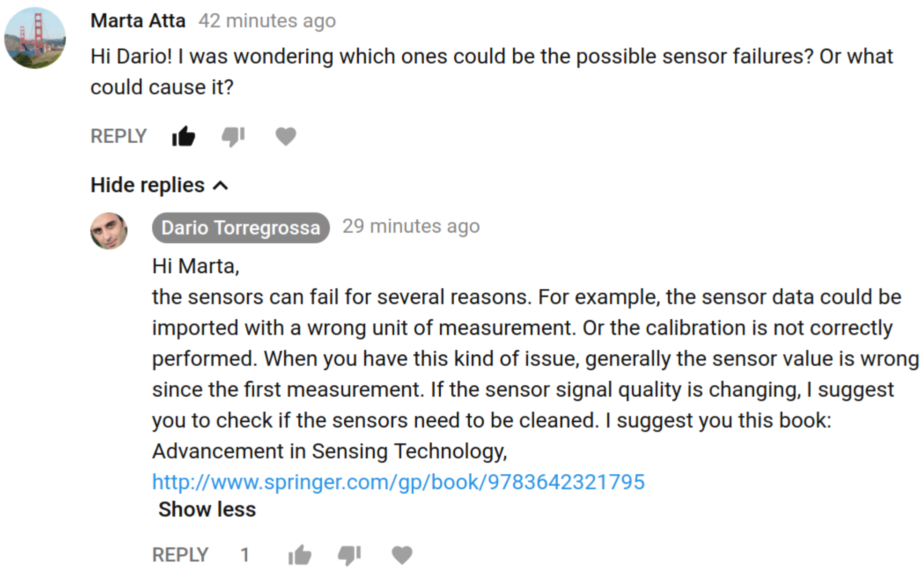


FIGURE 5.15: Screen-shot of cooperative use of YouTube platform for decision support

share knowledge. The contributions of the network can be voted and ranked. Moreover, the sharing-knowledge platform relies on YouTube platform that is well-known and generally considered user-friendly. "

5.8 Potential improvements and conclusions

The approach presented in this chapter is well defined and efficient in the WWTP monitoring. The management of expert knowledge can be still improved for what concerns the acquisition, the validation and the exchange of information. The first potential work can be performed on expert knowledge acquisition. Currently, this aspect is performed with expert interview and literature review. This operation is time expensive and it could be not efficient to feed the fuzzy rules because the input is not standardized (each paper is different, the expert interview is a dynamic talk). In alternative, it is possible to develop a tool that drives the expert to create fuzzy rules and attribute scores to the different operational scenarios. A follow-up of this project should include the development of such a tool.

In (Torregrossa and Hansen, 2018), it has been demonstrated that the exchange of solutions can be efficiently performed with not-conventional tools such as YouTube and with a large amount of formats (such text, document, comment thread, video, audio...). In order to expand the potential of cooperation between end-users, it is necessary to develop a platform that enables the multi-format knowledge sharing.

Moreover, it is necessary to develop a platform that efficiently evaluates the solutions. Different approaches can be imagined. For example, the 'like' system (Torregrossa and Hansen, 2018) can be used to evaluate the proposed solutions, but new features can be added in a new platform. A 'tag' system could associate a key-word to each solution. For example, a set of tag ('low budget', 'small WWTPs', 'small response time') could help the decision maker to easily select the optimal solution with more information. These aspects are discussed in detail in the last chapter.

In conclusion, at this stage, the SK-DSS was successfully tested for real applications, but a large scale test has still to be performed. The system has been successfully tested on few WWTPs, with few experts involved in the process and no external end-users a part from the PhD candidate. In order to make SK-DSS easy to use, an important element is the web-interface, discussed in detail in next chapter (chapt. 6). After that the last two chapters will report a synthesis of the obtained results (chapt. 7) and a discussion about potential developments (chapt.8).

Chapter 6

A web-application for the end-users

LEGAL DISCLAIMER: The present chapter partially reproduces research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of the EdWARDS project.

The first prototype of a web interface has been built with Shiny R application (https://www.shinyapps.io/) and posted online at the following address: https://dario-torregrossa.shinyapps.io/Ver2/. This interface includes the device monitoring tool discussed in the previous chapters of this thesis. The reader can test this demo by connecting to the website. This interface appears as a web-page, in which the end-user can customize the parameters and instantaneously visualize the results automatically produced in the background. While testing the application, the reader should be aware that the basic service of shinyapp has a limited number of hours of activation per month (25 hours), the power of the processor is limited and the application is therefore slow. Tests executed on a desktop version show that, with a normal processor [1] the application is fast, reactive and stable. An update of the Shinyapps service to the professional version would fix the problems of performance but for the purpose of this demonstration, this solution is considered too expensive.

[1]8 Gb RAM, 4 CPUs

6.1 Organization of the web-app

The web-app is organized in several web-pages with different functions:

- a first page, in which the end-user can analyse the raw sensor data;
- a page in which the end user gets information about KPIs;
- a page with the analysis of blowers;
- a page with the analysis of pumps;
- a page with the global evaluation;
- a page with the references.

On each page, the end user can customize the data range as well as the specific parameters required for the analysis. In order to respect data protection standards, the web-app runs with simulated data. At the current stage of developments, the biogas analysis is not included in the web-interface, in order to have a lighter environment for the show-case application considering that the Shiny basic service has limited performance. However, an additional web-page would have no relevant impact on the effectiveness and the purposes of this demo, in which the main functionalities are already fully exposed.

6.2 Analysis of raw data

The first page supports the end user in the analysis of raw data. Each sensor is listed in the selector, as well the analysis interval time and the plants connected to the system. The end-user can select plant, interval and sensor in order to visualize a time-series and a histogram (fig. 6.1). A part from a filter to remove outliers, no manipulation is performed on this dataset and this screen offers 'just' a visualization support tool.

6.3 Analysis of KPIs

The second page is dedicated to the analysis of the key performance indicators discussed in chapter 4. This web-page is strongly influenced

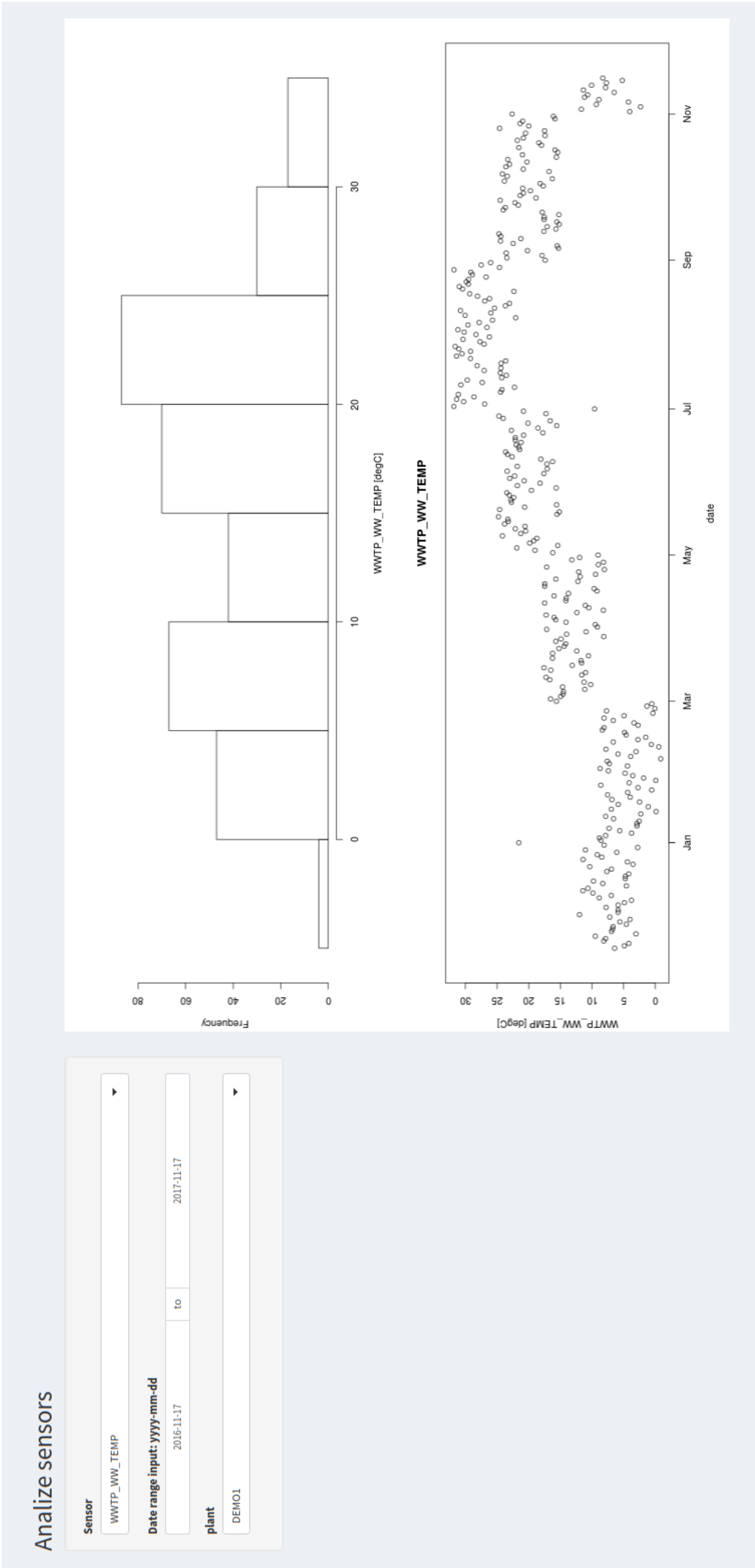


FIGURE 6.1: ScreenShot - raw data - online application

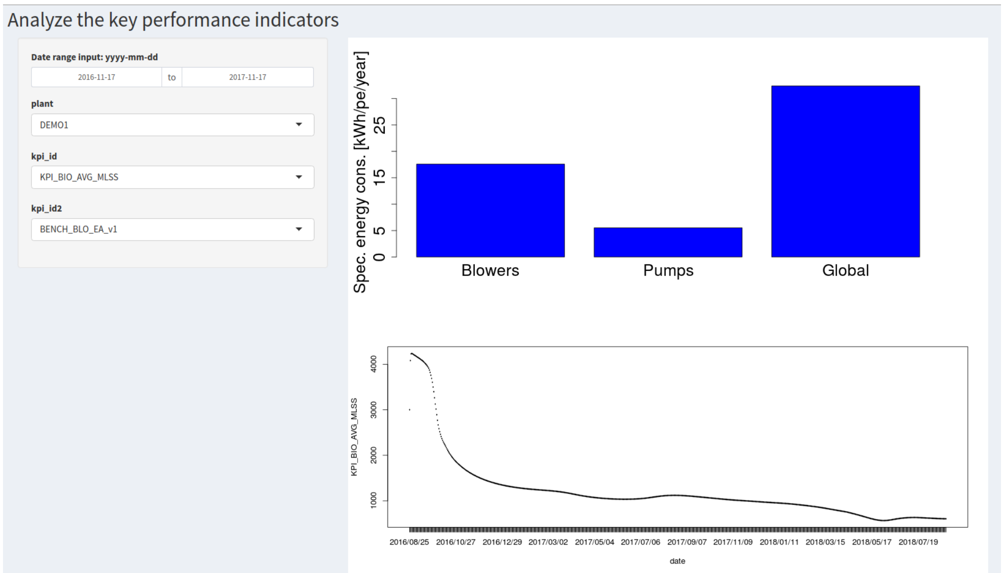


FIGURE 6.2: ScreenShot - KPI analyzer - online application

by the INNERS project (INNERS, 2015), with the difference in methodology discussed in subsection 4.1.1.

A first bar plot shows the main energy KPIs (global energy consumption, pumps, blowers) in order to give a first impression of plant performance. The axis variables of the bar plot are fixed (fig. 6.2), while the parameter values depend on the time frame selected.

Under the bar plot, the end user can visualise the time series of the KPIs chosen with the selection box on the left. These two plots used in combination can provide the following information:

- a global view of energy performance;
- a time series to identify changes in KPI values;

In summary, with this page an end-user can examine the following: is the plant performance satisfactory? Is the plant performance affected by a specific KPI? Is there a specific moment in which plant performance started to diverge from their expected behaviour?

On this page, the end-user can manipulate parameters in a very flexible way in order to test hypotheses and investigate the plants. The other web-pages offer more detailed and automatized analyses with a higher level of complexity that should augment and not replace a simple data visualization tool like this.

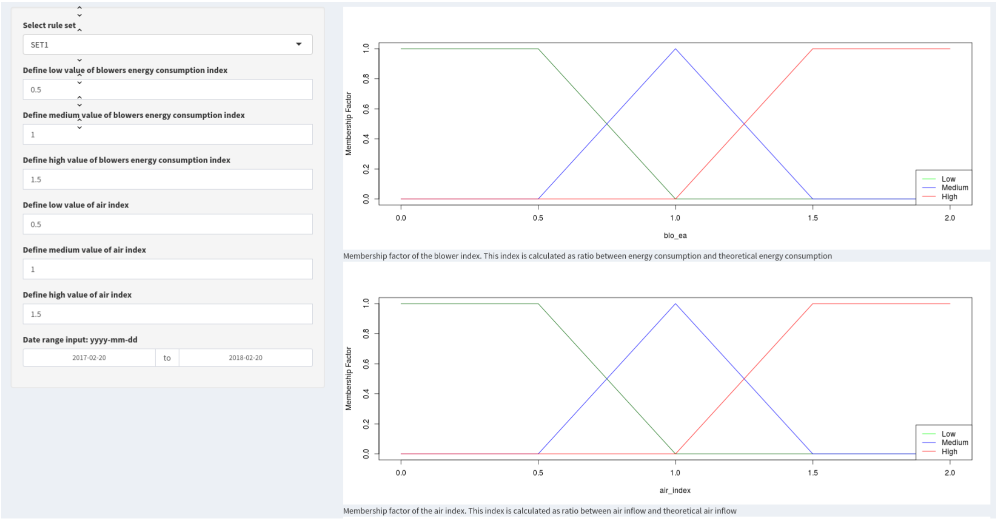


FIGURE 6.3: ScreenShot - blower analysis - Inputs

6.4 Analysis of blower consumption

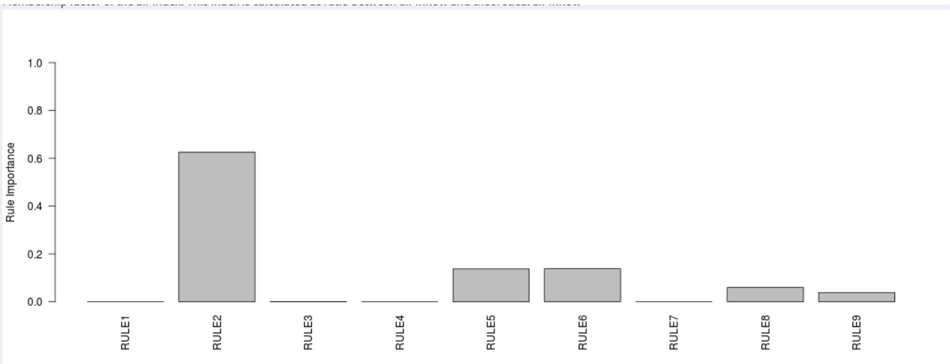
The analysis of blower consumption is performed with the methodology explained in section 5.2 and in (Torregrossa et al., 2017a). The web-application enables the selection of the necessary parameters for the analysis. In particular, it makes possible the selection of the KPI rules, the fuzzification parameters for blower consumption and air index, and the time of analysis.

The web-page returns the shape of fuzzification function for each parameter, the importance of the rules and a links to a YouTube web-interface for each rule. Here the solutions for given operational scenario are given (Torregrossa and Hansen, 2018).

Fig. 6.3 shows a screen-shot in which there are input fields on the left side and, on the right, plots return the fuzzification function of the input variables. Fig. 6.4 represents a screen-shot in which the importance of the rules and the links to the solutions are presented. These links open a YouTube web-page in which the end users can interact as discussed in section 5.7.

6.5 Analysis of pump consumption

The web-interface also provides a tool for the analysis of pump energy consumption, according to the methodology described in section 5.3



Importance of each rule: it corresponds to the truth degree of each rule. For further explanation refer to the literature

Show 25 entries

Search:

Ruleid	AIR.INDEX	BLO.INDEX	Score	Link.to.solution
Rule1	Low	High	Low	Visit our HTML tutoriLink>
Rule2	High	High	Low	Visit our HTML tutoriLink>
Rule3	Medium	High	Low	Visit our HTML tutoriLink>
Rule4	Low	Medium	Low	Visit our HTML tutoriLink>
Rule5	Medium	Medium	Medium	Visit our HTML tutoriLink>
Rule6	High	Medium	Medium	Visit our HTML tutoriLink>
Rule7	Low	Low	Medium	Visit our HTML tutoriLink>
Rule8	Medium	Low	Good	Visit our HTML tutoriLink>
Rule9	High	Low	Good	Visit our HTML tutoriLink>
Ruleid	AIR.INDEX	BLO.INDEX	Score	Link.to.solution

FIGURE 6.4: ScreenShot - blower analysis - rules and solutions

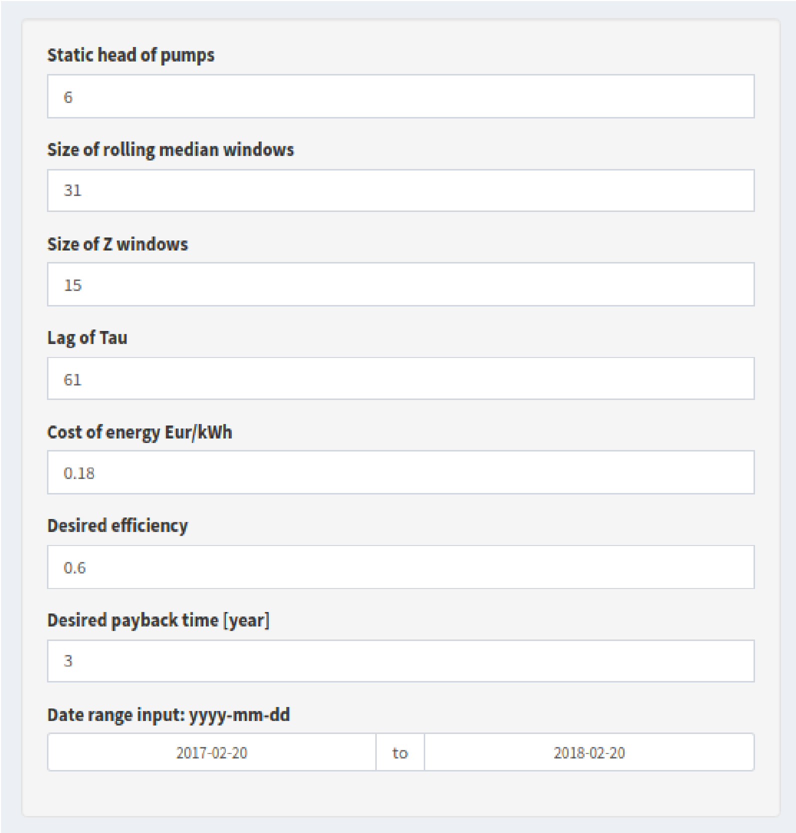


FIGURE 6.5: ScreenShot - pumps analysis - Inputs

and in (Torregrossa et al., 2017b). In particular, fig. 6.5 shows the grey-box to manipulate input parameters and fig. 6.6 shows the outputs and the connection to the solutions. The same schema is applied to pumps, blowers and biogas energy management; each web-application page differs from the others only for the required parameters. In the end-user interface, the format of the results is identical for each application while, the algorithms adopted in the background are different (ref. to chapter 5).

6.6 A global evaluation index

Figure 6.7 shows the global performance of the WWTPs. In particular, it is possible to visualize the indices of blowers and pumps and a global index calculated with the procedure explained in section 5.6. In the example shown in fig. 6.7, a drop of global efficiency occurred in mid-July, which mainly depends on blower efficiency. In this case, it is important to remark that the information presented in this figure is

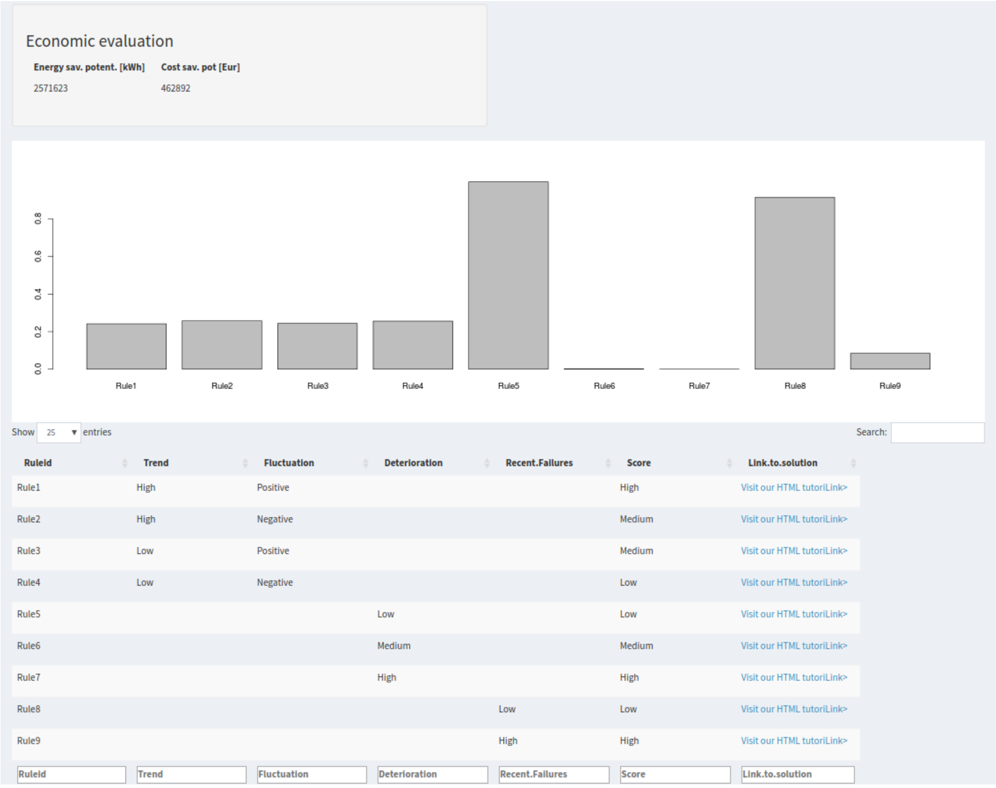


FIGURE 6.6: ScreenShot - pumps analysis - rules and solutions

different to those of fig. 6.2. In fact, fuzzy logic takes into consideration many elements (for example the air index of the blowers, the fluctuations of the pump or the energy consumption), while in the KPI pages just the energy consumption is considered.

6.7 From diagnosis to solutions

An important aspect concerns the suggestion and the evaluation of case-based solution. The knowledge sharing platform enables the end users of the network to cooperate and exchange information. Two options were investigated during this Ph.D. thesis: I) a SQL based knowledge based platform and II) a YouTube based cooperative platform. The first option was proposed in (Torregrossa et al., 2017a) and the second in (Torregrossa and Hansen, 2018). In this web-application, the YouTube based version is proposed and was chosen as the final option.

As explained in (Torregrossa and Hansen, 2018), the YouTube based platform “

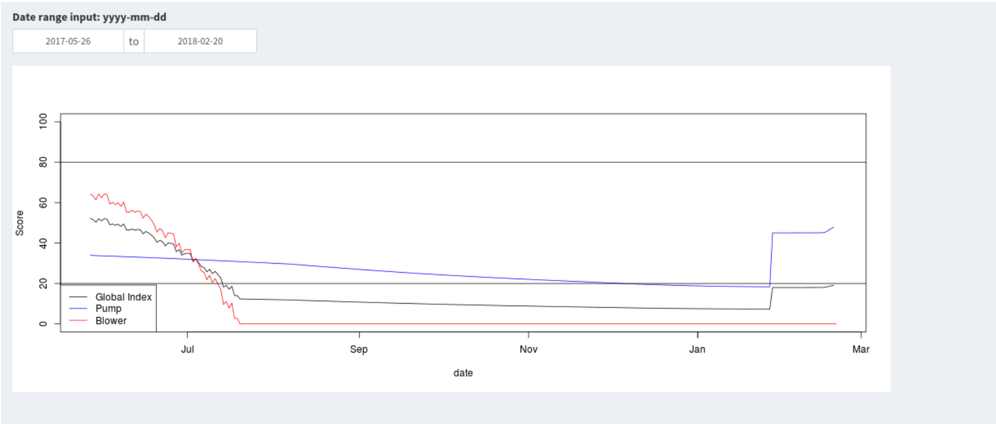


FIGURE 6.7: Screen Shot -global evaluation - Output

...presents some added values when compared to the former version:

- it is user-friendly, because the YouTube platform is one of the most used all over the world;
- plant managers do not have to deal with PostgreSQL to insert their suggestions;
- the knowledge is shared in a video-format; this is an advantage because the information provided in a video can be much more detailed and supported by images and animations;
- the YouTube 'like' system can be used to give a score to the suggestions;
- the YouTube video-suggestions can be commented;
- it is possible to access the suggestions from every device connected to the internet (smart-phone, tablet, pc).

"

The web-application links the rule analysis with the corresponding YouTube link. More details about the advantages, disadvantages, and opportunities can be found in (Torregrossa and Hansen, 2018) and in section 5.7. Moreover, it is important to mention that YouTube is only one of the social media that can be included in the SK-DSS and it can be replaced other web-services with similar features (i.e. Facebook or LinkedIn).

6.8 References

The methodologies applied in the web-application were published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b). A specific web-page is dedicated to references in order to provide the end-users with an explanation of the complete methodology. After the publication of this thesis, an additional link will be added to access this resource.

Moreover, the reference contact of the system administrator is provided.

6.9 Further improvements

This interface is a proof of concept of a SK-DSS provided as a web-service. There are still some improvements needed to make this interface ready for market:

- it must be exported in a highly powered private server; the current version that the reader can test is provided using a free service with limited power and usage time;
- the privacy has to be set up. A login system is necessary in order to prevent access to confidential data;
- visually, the SK-DSS needs to become more appealing;
- a demonstration environment should be made available to the end-users.

These aspects were to date not prioritised because they were less important than methodological aspects under the scientific profile.

Chapter 7

Synthesis and results

LEGAL DISCLAIMER: The present chapter partially reproduces research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of the EdWARDS project.

In this thesis, a cooperative decision support system for energy saving and production in WWTPs has been presented. The characteristics of this decision support system are aligned with the original research question and with the seven specific objectives presented in the introduction. Section 7.1 presents a comparison between the original objectives and the obtained results using the chapters of this thesis and the publications produced during the EdWARDS project.

7.1 Discussion of project results

The original research question stated in the submission to the National Research Fund of Luxembourg (FNR) was the following:

““Is it possible to develop a methodology that, based on benchmarking of on-line data from different WWTPs, enriched with expert knowledge, will be able to support a decision process aiming to increase the energy efficiency of WWTPs? Is it possible to apply this methodology to multiple plants simultaneously and at the same time to provide case-sensitive targeted advice?” (*FNR Application 7871388, 2014-03-20*).

This research question was enlarged during the project definition (section 1.3) and completed with a list of seven specific objectives (section 1.4).

The **first objective** consisted of the identification of a global model to process information from several WWTPs and provide decision making support. This model was discussed in chapter 3 and in (Torregrossa et al., 2017a). The proposed model presents a coherent structure able to integrate information on technology, data gathered online, static data, and expert knowledge to provide decision making support. In (Torregrossa et al., 2017a) the full SK-DSS was presented for the first time. This paper elaborates all the aspect concerning this cooperative decision support system, including the explanation of the information flow, the functional blocks and their interactions. SK-DSS was evaluated positively by the anonymous journal reviewers especially for its novelty, added value and solid structure. From the perspective of the candidate, the publication of (Torregrossa et al., 2017a) together with the case-study applications of blowers (Torregrossa et al., 2017a) and finally also pumps (Torregrossa et al., 2017b) are sufficient to positively answer the original research question.

The data normalisation **(second objective)** consists of WWTP information management, nomenclature normalisation, application of a uniform set of measurement units and the calculation of comparable key performance indicators. This objective, already addressed in INNERS project (INNERS, 2015), was improved with a new methodology that starting from the INNERS-EOS results provides a faster and more stable approach. INNERS-EOS calculated and stored all the available parameters, while the SK-DSS processes only the parameters useful for the decision support process. This improvement makes the calculations faster and more stable, ultimately enabling the connection of a larger number of WWTPs. These aspects were discussed in detail in chapter 4.

Chapter 4 and (Torregrossa et al., 2016) deal with the estimation of the missing data required for a daily KPI calculation **(third objective)**; this aspect is essential to enable a daily plant assessment which requires information about pollution load generally not available at a daily frequency. An original approach based on the random forest algorithm was developed; the estimation of missing parameters performed with this algorithm was shown to be of adequate accuracy to be included in the decision making process. Moreover, the core of this analysis is a fuzzy logic methodology, well-known for its efficiency in processing values affected by uncertainty (Starczewski, 2013). The selection of the random forest algorithm was done after a comparison with several alternatives presented in (Torregrossa et al., 2016).

The **fourth objective**, consisting of the application of the methodology to blowers and pumps was presented in sections 5.2 and 5.3 and in the papers (Torregrossa et al., 2017a; Torregrossa et al., 2017b). The presented tools are based on an innovative combination of KPIs and fuzzy logic. One of the main advantages of these methodologies is the plant-generic feature. The blower analysis can be done with 3 different information sets; from a more complete set (including for example air flow and energy sensors) to a basic set of information (relying only on energy consumption). The end-user can select the set to be analysed according to the sensor availability or personal preference. The pump analysis is performed using a basic set of information (pumped flow and energy consumption). With this information (generally available in WWTPs), complex analyses concerning the pump ageing, detection of early failures and identification of flow-related inefficiencies is performed. This methodology can be applied to all the centrifugal pump applications, even outside of the WWTP domain, and received a high consideration from the reviewers of (Torregrossa et al., 2017b).

Section 5.5 is dedicated to the **fifth objective**, i.e. the development of a methodology to monitor biogas production. An innovative dual-performance assessment is proposed. The first analysis is a benchmarking evaluation which compares energy production against expected values and the second one analyses process quality by monitoring the main operational parameters. The biogas assessment takes into account the time-lag effect of inefficient parameter set-up and the uncertainty PE estimation. This approach is widely applicable in WWTPs because it considers a set of sensors (considered minimal for WWTPs equipped with a digestion unit) that includes temperature, pH, solid retention time and biogas production.

The **sixth objective** was developed in section 5.7 and (Torregrossa and Hansen, 2018), in which a platform for cooperation was proposed. This innovative idea consists of connecting the fuzzy logic analyser to the popular YouTube platform, enabling a user-friendly interface to share multi-medial suggestions, links and comments. The advantages of such an approach are numerous: the possibility to share ideas in several formats (video, document, link, comment), the wide accessibility (computer, smart-phone, tablet), the externalisation of costs (maintenance, server space) and the user-friendly YouTube usage. The enthusiastic comments of anonymous reviewers of (Torregrossa and Hansen, 2018) lead to the publication of this manuscript without any requested improvement.

Chapter 6 exposes the user-friendly graphical interface (**seventh objective**). This on-line platform enables the easily customisation of the parameters, rapid visualisation of the results and access to the YouTube platform with suggestions for energy saving. This interface, still at its prototype stage, shows great potential for end users which can access on-line the assessment of their plants.

Given the original targets and the results obtained, it is possible to postulate here that the objectives were addressed. This statement is reinforced by this thesis and by the positive peer reviewed evaluations of the methodologies published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017; Torregrossa and Hansen, 2018).

The papers published by the candidate during the PhD received a positive evaluation from high-ranked journals [1] as well as from the highly qualified audience at the conferences attended. Moreover, the high values of FWCI [2] obtained by the published papers shows the appreciation of the scientific community for the methodology. For example, on 27-03-2018, the key-paper in which the SK-DSS was firstly presented (Torregrossa et al., 2017a) has a FWCI=13.15, i.e. it has been cited thirteen times more than the average of similar papers.

In summary, in this PhD project, the candidate presented a decision support system which addresses the original research question, as well

[1] Environmental Research, Journal of Cleaner Production, Applied Energy

[2] "Field-Weighted Citation ImpactField-Weighted Citation Impact shows how well cited this article is when compared to similar articles. A FWCI greater than 1.00 means the article is more cited than expected according to the average. It takes into account: The year of publication Document type, and Disciplines associated with its source. The FWCI is the ratio of the article's citations to the average number of citations received by all similar articles over a three-year window. Each discipline makes an equal contribution to the metric, which eliminates differences in researcher citation behavior." Definition from Scopus.

as the specific objectives.

The development of the SK-DSS stimulated new research questions and new ideas that were not developed because of time limitations. In chapter 8, the thesis will focus on these new directions, possible improvements and potential follow-ups.

Chapter 8

Outlook

LEGAL DISCLAIMER: The present chapter partially reproduces the research work already published in (Torregrossa et al., 2016; Torregrossa et al., 2017a; Torregrossa et al., 2017b; Torregrossa et al., 2017c; Torregrossa et al., 2017d; Torregrossa, Hansen, and Leopold, 2017). All the scientific content, the methodology, the scripts, and the results are the original production of the candidate in the framework of EdWARDS project.

In the water domain, the decision support science is an open sector.

A detailed explanation of not addressed challenges and research questions would deserve a full book. Nevertheless, in the framework of this thesis, it is useful to mention some of them with the aim to stimulate a discussion.

8.1 Application of the SK-DSS on large scale

The SK-DSS is designed to process information coming from different WWTPs. The normalization of the information format necessary in this method has been already discussed in section 4.2.

However, the normalization process, even if plant-specific and time-expensive, is not really an issue for the connection of the plant to the SK-DSS. From the experience gained in this project, the main challenge is the maintenance of the stability of the import process. Plant operators frequently update the database or perform operations that can cause failures. In this project, for example, some paradigmatic issues were experienced:

- changes of sensor names; in this case, some data is not imported;

- changes of the access permission of files to be processed;in this case, some data is not imported;
- changes in measurement units; in this case, the data is imported but the values are wrong. In the best case, the data is recognized as outliers and the problem is evident. In the worst case, the data is wrong but the error is not detected.

Currently, having only three WWTPs connected to the SK-DSS, these issues are easily detected and fixed. However, a large scale application of SK-DSS requires a more robust approach to avoid this kind of problems.

Asking plant operators to avoid performing maintenance to their SCADA system is not realistic. Given that, potential solutions are: i) a software that automatically controls the data inflow; ii) a strong communication protocol between SK-DSS operators and plant operators to efficiently manage potential SCADA maintenance issues, or iii) an interface to assist the operator in database updates.

The software should be able to control at least the data records over time and detect missing values and outliers. This tool should generate alerts in two cases:

- the number of records generated by the plant decreases;
- the number of outlier increases.

This tool should be able to distinguish one-time issues from permanent failures. For example, such a system should be able to distinguish if a sensor experienced an isolated or a permanent failure, or if the number of failures is increasing over the time.

The protocol between SK-DSS and plant operators should be based on some principles:

- it is necessary that, for each plant, a person responsible for SCADA system is identified. This operator should be able to understand the requirements of the SK-DSS, to communicate with SK-DSS operators and to plan changes compatible with the decision support system;
- the changes in SCADA system and in WWTPs need to be documented and this documentation need to be accessible to the SK-DSS operators;

- the changes in the SCADA system need to be reduced to a minimum;
- the changes should be communicated in advance to the SK-DSS operator.

In return, the SK-DSS provider should guarantee a staff with an adequate number of workers that work on the system maintenance. As an empirical rule, the candidate would suggest that the maintenance staff should include 1 full-time operator every 20 or 25 plants.

8.2 Scalability of the database

Another important issue concerns the scalability of the database. Currently the information is stored in a PostgreSQL database. This is a relational database, and as has already occurred (sect. 4.1.1), a large amount of data can generate failures. The large-scale application of this SK-DSS will require an update to another database technology.

DataJobs.com, 2017 suggests that the storage of "big data" is better achieved with non-relational databases:

" NoSQL (commonly referred to as "Not Only SQL") represents a completely different framework of databases that allows for high-performance, agile processing of information at massive scale. In other words, it is a database infrastructure that as been very well-adapted to the heavy demands of big data.

The efficiency of NoSQL can be achieved because unlike relational databases that are highly structured, NoSQL databases are unstructured in nature, trading off stringent consistency requirements for speed and agility. NoSQL centers around the concept of distributed databases, where unstructured data may be stored across multiple processing nodes, and often across multiple servers. This distributed architecture allows NoSQL databases to be horizontally scalable; as data continues to explode, just add more hardware to keep up, with no slowdown in performance. The NoSQL distributed database infrastructure has been the solution to handling some of the biggest data warehouses on the planet – i.e. the likes of Google, Amazon, and the CIA."

As shown in fig. 8.1, the performance of non-relational databases is not affected by the increasing volume of data. Consequently, this seems

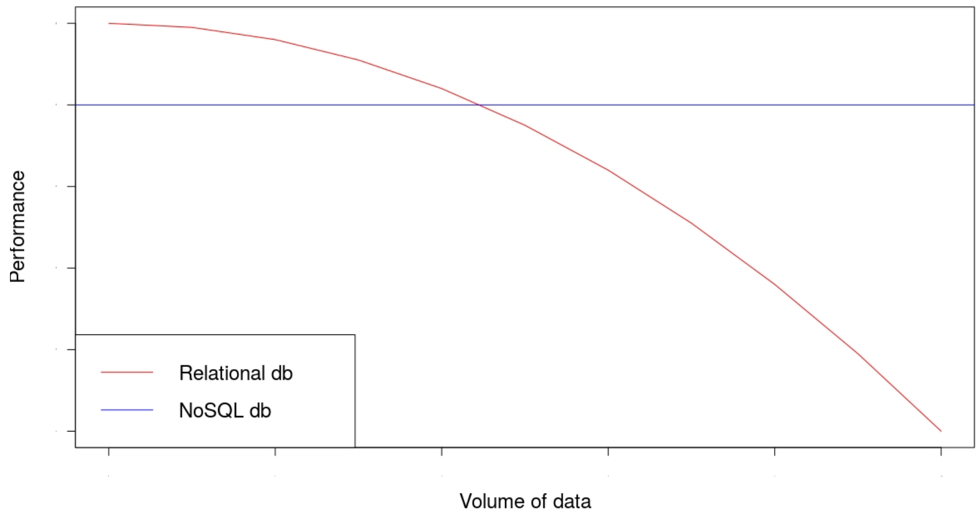


FIGURE 8.1: Scalability of NOSQL Database. Figure adapted from DataJobs.com, 2017. In x-axis the volume of data (increasing from left to right), in y-axis the performance (increasing from bottom to top)

to be a potentially efficient solution for a SK-DSS operating for several years with many WWTPs.

8.3 The knowledge challenge

The title of this section is 'The knowledge challenge' and concerns one of the most delicate aspects of SK-DSS: the knowledge management. In order to provide a high performance, the SK-DSS needs to rely on a robust, validated and updated set of information. This challenge can be divided into the sub-packages proposed in the following subsections.

8.3.1 Expert knowledge gathering and Knowledge sharing

First, the expert knowledge should be gathered and constantly updated in order to provide a high-level set of fuzzy-rules. A potential solution is the definition of a standard procedure to interview plant operators and obtain useful information for the SK-DSS. Another solution is the development of an application for expert knowledge gathering. The basic idea could be the development of a data simulator that presents some scenarios to experts for a standardized evaluation. The answers should be registered and processed to define the information set in a SK-DSS compatible format.

Another relevant aspect concerns the knowledge sharing trough a platform. In this case, it is necessary to provide an user-friendly interface and some incentives to cooperation. For example, in the scientific domain, the h-index of each author is a stimulus to share information through the publications of papers. A similar mechanism could be developed to promote cooperation between plant managers. This objective seems realistic if SK-DSS is adopted by professional networks, consortia and syndicates. For example, the PhD candidate was in contact with Croon (NL) and the Vereniging van Zuiveringsbeheerders (NL) and Wupperverband(GER); these organizations are active in collecting data and monitoring many WWTPs in the Netherlands and in Germany; this kind of organizations should be able to provide stimulus and incentives for the cooperation between plant managers.

8.3.2 Validate Knowledge

Another challenge is knowledge validation. The validation of solutions must be done in two stages:

- validation ex-ante. At this stage, it is necessary that wrong knowledge is not stored in the system and adequate strategies need to be put in place to prevent this;

- validation ex-post. This corresponds to the continuous assessment of the proposed solutions operated by the network of plant operators. It is necessary to periodically control the results of solution validation and remove those not considered adequate.

Moreover, it is necessary to set-up a long-term evaluation of solutions. While validations ex-post and ex-ante are based on expert knowledge, this-long term evaluation has to rely on effective results. Furthermore, the solutions should be updated with the development of new technologies.

8.3.3 A new set of benchmarks

The benchmarks available in literature are calculated using yearly averaged data and they generally do not take into account aspects such as the capacity utilization of the plant or the seasonality.

A new set of benchmarks aiming to assess the daily operational condition should be able to include all the variables relevant a daily timeframe. For example, Sala-Garrido, Molinos-Senante, and Hernández-Sancho, 2012 have shown that seasonal-plants are generally less efficient. The plant operator cannot obtain a performance comparable to the not-seasonal plants because of external factors out of his decision sphere (such as seasonality and design factors).

The daily assessment of the plant should take into consideration a set of benchmarks that accounts the effect of daily phenomena.

8.3.4 New kind of information

SK-DSS accounts only for the plant-related information. New development of this technology should also relate to the decision-making process and to the socio-economic environment around the facilities. For example, in the subsection 2.1.6, some opportunities such as the water reuse have been discussed. Such new challenges require an investigation of the complex relationships of WWTPs with the socio-economic context, the environment and the regulation. In order to deal with these, a new set of information need to be integrated to SK-DSS. For example, an enlarged set of sensors can include the use of satellite imaging, or the Geographic Information System (GIS) Maps, to classify the socio-economic environment around the plants.

8.4 Energy price

In future developments of the decision support system, energy tariffs should be taken into consideration, to minimize the total energy cost. The SK-DSS pump application (Torregrossa et al., 2017b) included a first economic evaluation of solutions based on the comparison of cost of solutions and cost of energy. This approach should be extended and also take into consideration the hourly variation in the cost of energy.

8.5 Smart Grid

In this thesis, WWTPs were mainly considered as isolated energy consumers/producers. In reality, WWTP energy balance requires the interaction with energy grids. From this perspective, WWTPs can be considered as 'agents' of the electric energy systems and SK-DSS could be expanded towards the optimal management of a larger system comprising producers, many consumers and the WWTPs. In particular, the cooperative platform (Torregrossa and Hansen, 2018) can be used to stimulate the exchange of information between different stakeholders.

8.6 Integration of sewer information

WWTP management could benefit also from the use of sewer information. At its current state, SK-DSS has a restricted set of information from the sewer, consisting of inflow rates and pollutant loads. The integration of a larger set of information (such as rain water volumes or weather forecasts) or the combination of SK-DSS with a detailed sewer model could increase the accuracy of analysis and suggestions. Another solution could be the integration of the decision support system with the dynamic management of the sewer systems (an application is provided by RTC4Water, `https://www.rtc4water.com/index.php?lang=en`)).

8.7 Micro-pollutant

The management of micro-pollutants is becoming more and more interesting for plant operators and the scientific community because of increasing concentrations in waste water (Hansen, 2018). In theory, SK-DSS can be applied to optimize micro-pollutant removal processes, albeit that this field is affected by uncertainty, conflicting objectives and restricted data availability.

8.8 The assessment of environmental impacts

The decision support system presented in this thesis considers the energy consumption of plants. However, energy is not the only indicator of environmental sustainability of the process. A more detailed analysis could take into account other elements such as the consumption of chemicals or the electricity production mix and many methodologies are suitable for this scope (such as life cycle assessment and carbon footprint).

8.9 Conclusions

This thesis is one of the outputs of 4-years work. During the development of this PhD project, the given tasks were accomplished and some potential future developments identified. This thesis shows that it is possible to build a plant generic decision support system specifically oriented to optimize the energy balance in WWTPs. This decision support system is characterised by several innovative features. This last chapter, with its dissertation about methodology limitations and new opportunities, is a bridge to future developments and follow-ups. Many potential follow-up aspects were identified. The two main lines consists of 'market-oriented' and 'research oriented' projects. The first line is the development of existing methodologies with the aim to deliver a product/service to the market. The second line consists of the development of new methodologies for decision making support. At the moment of writing of this thesis, it is not possible to make accurate predictions about these developments because they depend on many factors which cannot currently be carefully evaluated. Instead, at the moment, it is possible to report a strong interest from the candidate and the partners to continue exploring the topic with new projects and collaborations.

Appendix A

Fuzzy logic explained with a numerical example

A.1 Fuzzy Logic Example

In this section, an example of fuzzy logic is performed with the rules of equation 2.10. Let's assume that, using the rules of the system equation 2.10, the fuzzy system has to estimate the cost of cars according with 2 parameters: age and motor power. In the example, the input are:

- car_age=5 years
- motor_power=170 hp

A.1.1 Example: fuzzification of inputs

Having the rules, it is necessary to identify the membership factor associated to the input parameters. The systems of equations A.1 and A.2 model the membership factors for the age of the car and for the motor power. In equation A.1 : Car_age[New] is the membership factor of the car age associated to the adjective 'New', Car_age[Old] is the membership factor of the car age associated to the adjective 'Old', car_age is the age of the car [year]. In equation A.2: Motor_power[Low] is the membership factor of the motor power associated to the adjective 'Low-power', Motor_power[High] is the membership factor of the motor power associated to the adjective 'High-power', motor_power is the power of the car engine [hp].

$$\begin{cases} Car\_age[New] = min(1, \dfrac{car\_age - 3}{7 - 3}), car\_age \in [0,7] \\ Car\_age[Old] = min(1, 1 - \dfrac{car\_age - 3}{7 - 3}), car\_age \in [7,10] \end{cases} \tag{A.1}$$

$$\begin{cases} Motor\_power[Low] = min(1, 1 - \dfrac{motor\_power - 100}{200 - 100}), motor\_power \in [50{,}200] \\ Motor\_power[High] = min(1, \dfrac{motor\_power - 100}{200 - 100}), motor\_power \in [100{,}250] \end{cases} \tag{A.2}$$

Figure A.1 shows the graphical representations of the systems of equations A.1 and A.2 . With the given inputs, the results of these equations are:

- Car_age[New] =0.5
- Car_age[Old]=0.5
- Motor_power[Low]=0.3
- Motor_power[High]=0.7

In other words, the membership factors show a car neither old neither new and with a engine power quite high.

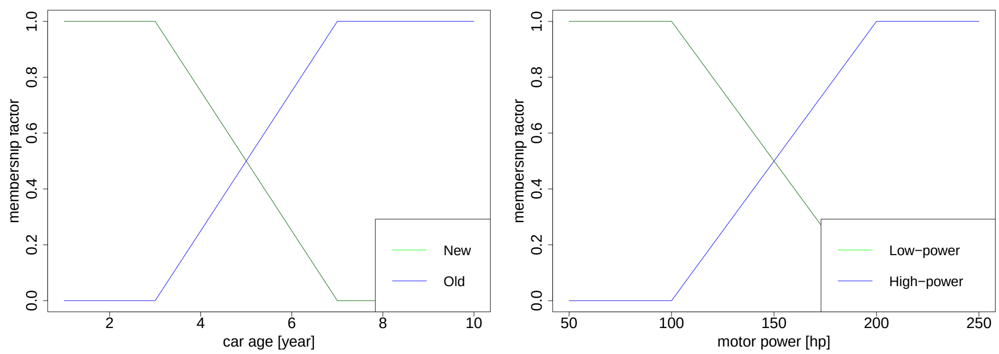


FIGURE A.1: Fuzzification of inputs

Example: infer rules and calculate truth degree

At this stage, it is necessary to calculate the truth degree of each rule. The mathematical operation corresponding to logic operator are reported in table 2.8. In this example, the probabilistic approach for AND operator is used, consequently the truth degree of each rules is calculated as multiplication between antecedent membership factors. In the system of equation A.3, the truth degrees of each rule are calculated.In this equation, Rule n is the truth degree of the rule n.

$$\begin{cases} Rule1 = Car\_age[New] * Motor\_power[High] = 0.5 * 0.7 = 0.35 \\ Rule2 = Car\_age[New] * Motor\_power[Low] = 0.5 * 0.3 = 0.15 \\ Rule3 = Car\_age[Old] * Motor\_power[High] = 0.5 * 0.7 = 0.35 \\ Rule4 = Car\_age[Old] * Motor\_power[Low] = 0.5 * 0.3 = 0.15 \end{cases} \tag{A.3}$$

Example: identification of rule output

At this stage, it is necessary to define the output, that in this case corresponds to the price of the car. For sake of simplicity, the price is defined in the system of equation A.4 as a Mamdani singleton (Sivanandam, Sumathi, and Deepa, (2006), pag. 120).

$$\begin{cases} Price[Low] = 5.000 \text{ €} \\ Price[Medium] = 12.000 \text{ €} \\ Price[High] = 20.000 \text{ €} \end{cases} \tag{A.4}$$

Example: defuzzification

The defuzzification is performed with the equation 2.13, that for the specific case becomes:

$$\begin{aligned} Price = & \frac{Price[High] * Rule1 + Price[Medium] * Rule2}{Rule1 + Rule2 + Rule3 + Rule4} \\ & + \frac{Price[medium] * Rule3 + Price[low] * Rule4}{Rule1 + Rule2 + Rule3 + Rule4} = \\ & = 13.750 \text{ €} \end{aligned}$$

For this example, the final result is 13750 €, that it is the price of a 5-years-old car of high engine power . Table A.1 reports other possible combinations between inputs and output calculated with this model. According to the common sense and the fuzzy set-up, older the car lower the price and, higher the power higher the price. Table A.1 shows how fuzzy logic is efficient in reproduce the human-reasoning.

TABLE A.1: Example of other inputs-output combination

Car age [year]	Motor power[hp]	Fuzzy_Price [€]
1	226	20000
1	77	12000
2	152	16160
3	128	14240
4	196	17690
4	123	12032
5	153	12475
7	209	12000
7	81	5000
8	227	12000
9	221	12000
9	143	8010
10	247	12000
10	224	12000
10	67	5000

Appendix B

Economic Calculation

B.1 Theoretical framework

The set of equations B.1 illustrates the mathematical procedure with the following nomenclature:

- η_t is the value of efficiency trend calculated with the rolling median (cf. section 5.3.4) ;
- η_{id} is the ideal value of efficiency;
- E_t is the value of energy consumption calculated with η_t;
- E_{id} is the ideal value of energy consumption;
- pes is the potential energy saving;
- C_{en} is the cost of energy [€/kWh];
- pcs is the potential cost saving calculated for each day;
- $\overline{pcs}$ is the average value of pcs for the period under study; in our case we calculate the average for the previous 180 days;
- $N_{day-3years}$ is the number of days in 3 years;
- mpeb is the maximum potential economic benefit resulting from maintenance over a period of 3 years. This value corresponds to the maximum investment that the plant manager should accept for extraordinary maintenance.

$$\begin{cases} \eta_{id} = 0.32 \\ E_t = \dfrac{mgh}{\eta_t}[J] \\ E_{id} = \dfrac{mgh}{\eta_{id}}[J] \\ pes = (E_t - E_{id}) * 2.78 * 10^{-7}[kWh] \\ pcs = pes \quad * \quad C_{en}[Euro] \\ mpeb = \overline{pcs} \quad * \quad N_{day-3years}[Euro] \end{cases} \tag{B.1}$$

B.2 Numerical example

Let's assume that for a given day, the pump system lift 1000 kg of water for 10 meters.

Let be:

- $\eta_t = 0.25$
- $\eta_{id} = 0.32$
- $m = 1000$ kg
- $g = 9.8 m/s^2$

Consequently, the value of energy consumption without the effects of short-term phenomena is $E_t = 392000$ J against an ideal value $E_{id} = 306250$ J.

The potential energy saving is equal to the difference between the energy consumption and the ideal energy consumption.

$$pes = (E_t - E_{id}) * 2.78 * 10^{-7}[kWh] = 217.379 \text{ kWh} \tag{B.2}$$

Now, it is necessary to transform the potential energy saving in potential cost saving. Given a cost of energy $C_{en} = 0.12$ €/kWh,

$$pcs = pes * C_{en}[Euro] = 26.085€ \tag{B.3}$$

If the average cost saving in 1 day is equal to 26.085 €, in 3 years the potential economic saving corresponds to 28563.6 €, equal to maximum potential economic benefit and the maximum investment with the desired pay-back time.

Bibliography

Abuşoğlu, Ayşegül et al. (2016). "Energy and economic analyses of models developed for sustainable hydrogen production from biogas-based electricity and sewage sludge". In: *International Journal of Hydrogen Energy* 41.31, pp. 13426–13435. ISSN: 03603199. DOI: 10.1016/j.ijhydene.2016.05.105. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0360319916315439.

Aiello, Giuseppe et al. (2017). "A decision support system based on multisensor data fusion for sustainable greenhouse management". In: *Journal of Cleaner Production*. ISSN: 09596526. DOI: 10.1016/j.jclepro.2017.02.197. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0959652617304237.

Akbaş, Halil, Bilge Bilgen, and Aykut Melih Turhan (2015). "An integrated prediction and optimization model of biogas production system at a wastewater treatment facility". In: *Bioresource Technology* 196, pp. 566–576. ISSN: 09608524. DOI: 10.1016/j.biortech.2015.08.017. URL: http://www.sciencedirect.com/science/article/pii/S0960852415011219.

Alex, J. et al. (2015). "Benchmark for evaluating control strategies in wastewater treatment plants". In: *European Control Conference, ECC 1999 - Conference Proceedings*, pp. 3746–3751. URL: http://www.engineeringvillage.com/blog/document.url?mid=cpx{_}M5c438e5a14e4fa33ee7M7b6a10178163171{\&}database=cpx.

Angulo, Cecilio et al. (2012). "Fuzzy expert system for the detection of episodes of poor water quality through continuous measurement". In: *Expert Systems with Applications* 39.1, pp. 1011–1020. ISSN: 09574174. DOI: 10.1016/j.eswa.2011.07.102. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0957417411010645.

Argyris, Nikolaos and Simon French (2017). "Nuclear emergency decision support: A behavioural OR perspective". In: *European Journal* *of Operational Research* 262.1, pp. 180–193. ISSN: 03772217. DOI: 10.1016/j.ejor.2017.03.059. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0377221717302813.

Arun Shankar, Vishnu Kalaiselvan et al. (2016). "A comprehensive review on energy efficiency enhancement initiatives in centrifugal pumping system". In: *Applied Energy* 181, pp. 495–513. ISSN: 03062619. DOI: 10.1016/j.apenergy.2016.08.070. URL: http://dx.doi.org/10.1016/j.apenergy.2016.08.070.

Baeza, J., D. Gabriel, and J. Lafuente (1999). "An expert supervisory system for a pilot WWTP". In: *Environmental Modelling & Software* 14.5, pp. 383–390. ISSN: 13648152. DOI: 10.1016/S1364-8152(98)00101-7. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1364815298001017.

Bailey, David (Telecommunications engineer) and Edwin (Engineer) Wright (2003). *Practical SCADA for industry*. Elsevier, p. 288. ISBN: 9780080473901.

Baumann, Peter, Manfred Roth, and Peter Maurer (2014). *Senkung des Stromverbrauchs auf Klaranlagen: Praxisleitfaden fur das Betriebspersonal – Systematisches Vorgehen zur Steigerung der Energieeffizienz durch Nutzung des Einsparpotenzials*. ISBN: 3456846665.

Becker, Markus and Joachim Hansen (2013). "Is the energy-independency already state-of-art at NW-European wastewater treatment plants". en. In: *Proceedings of the IWA-conference 'Asset Management for enhancing energy efficiency in water and wastewater systems'*, pp. 19–26. URL: http://orbilu.uni.lu/handle/10993/7857.

Berge, S P, B F Lund, and R Ugarelli (2014). "Condition monitoring for early failure detection. Frognerparken pumping station as case study". In: *Procedia Engineering* 70.1877, pp. 162–171. ISSN: 18777058. DOI: 10.1016/j.proeng.2014.02.019. URL: http://dx.doi.org/10.1016/j.proeng.2014.02.019.

Bernard, Olivier et al. (2005). "An integrated system to remote monitor and control anaerobic wastewater treatment plants through the internet". In: *Water Science and Technology* 52.1-2, pp. 457–464. ISSN: 02731223.

Bertanza, Giorgio et al. (2014). "How can sludge dewatering devices be assessed? Development of a new DSS and its application to real case studies". In: *Journal of Environmental Management* 137, pp. 86–92. ISSN: 03014797. DOI: 10.1016/j.jenvman.2014.02.002.

Björnsson, L, M Murto, and B Mattiasson (2000). "Evaluation of parameters for monitoring an anaerobic co-digestion process". In: *Applied Microbiology and Biotechnology* 54.6, pp. 844–849. ISSN: 01757598. DOI: 10.1007/s002530000471.

Boe, Kanokwan et al. (2010). "State indicators for monitoring the anaerobic digestion process". In: *Water Research* 44.20, pp. 5973–5980. ISSN: 00431354. DOI: 10.1016/j.watres.2010.07.043. URL: http://dx.doi.org/10.1016/j.watres.2010.07.043.

Bottero, Marta, Giulio Mondini, and Alessandra Oppio (2016). "Decision Support Systems for Evaluating Urban Regeneration". In: *Procedia - Social and Behavioral Sciences* 223, pp. 923–928. ISSN: 18770428. DOI: 10.1016/j.sbspro.2016.05.319. URL: http://linkinghub.elsevier.com/retrieve/pii/S1877042816303998.

Breiman, Leo (2001). "Random forests". In: *Machine Learning* 45.1, pp. 5–32. ISSN: 08856125. DOI: 10.1023/A:1010933404324. arXiv: /dx.doi.org/10.1023{\%}2FA{\%}3A1010933404324 [http:].

Budych-Gorzna, Magdalena, Marcin Smoczynski, and Piotr Oleskowicz-Popiel (2016). "Enhancement of biogas production at the municipal wastewater treatment plant by co-digestion with poultry industry waste". In: *Applied Energy* 161, pp. 387–394. ISSN: 03062619. DOI: 10.1016/j.apenergy.2015.10.007. URL: http://dx.doi.org/10.1016/j.apenergy.2015.10.007.

Caeiro, S. et al. (2017). "Environmental risk assessment in a contaminated estuary: An integrated weight of evidence approach as a decision support tool". In: *Ocean & Coastal Management* 143, pp. 51–62. ISSN: 09645691. DOI: 10.1016/j.ocecoaman.2016.09.026. URL: http://linkinghub.elsevier.com/retrieve/pii/S0964569116302216.

Campo, Giuseppe et al. (2017). "Enhancement of waste activated sludge (WAS) anaerobic digestion by means of pre- and intermediate treatments. Technical and economic analysis at a full-scale WWTP". In: *Journal of Environmental Management*. ISSN: 03014797. DOI: 10.1016/j.jenvman.2017.05.025. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S030147971730484X.

Caniani, Donatella et al. (2015). "Towards A New Decision Support System for Design, Management and Operation of Wastewater Treatment Plants for the Reduction of Greenhouse Gases Emission". In:

Water 7.10, pp. 5599–5616. ISSN: 2073-4441. DOI: 10.3390/w7105599. URL: http://www.mdpi.com/2073-4441/7/10/5599/.

Castellet, Lledo Lledó and Maria María Molinos-Senante (2016). “Efficiency assessment of wastewater treatment plants: A data envelopment analysis approach integrating technical, economic, and environmental issues”. In: *Journal of Environmental Management* 167, pp. 160–166. ISSN: 10958630. DOI: 10.1016/j.jenvman.2015.11.037. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0301479715303911http://ac.els-cdn.com/S0301479715303911/1-s2.0-S0301479715303911-main.pdf?{_}tid=0e4aac54-7b02-11e6-bfcd-00000aacb35e{\&}acdnat=1473916124{_}60e44e8ab4f5b1e2ad97f36609694bae.

Castillo, A. et al. (2016). “Validation of a decision support tool for wastewater treatment selection”. In: *Journal of Environmental Management* 184, pp. 409–418. ISSN: 03014797. DOI: 10.1016/j.jenvman.2016.09.087. URL: http://linkinghub.elsevier.com/retrieve/pii/S0301479716307599.

Castro, Ana C Meira, Joao Paulo Carvalho, and S. Ribeiro (2011). “Prescribed burning impact on forest soil properties-A Fuzzy Boolean Nets approach”. In: *Environmental Research* 111.2, pp. 199–204. ISSN: 00139351. DOI: 10.1016/j.envres.2010.03.004. URL: http://dx.doi.org/10.1016/j.envres.2010.03.004.

Chang, Wei et al. (2012). “Energy-saving algorithm for pumping systems based on fuzzy decision making”. In: *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, pp. 2395–2400. ISBN: 978-1-4673-1714-6. DOI: 10.1109/ICSMC.2012.6378101. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6378101.

Chen, Jim Q. and Sang M. Lee (2003). “An exploratory cognitive DSS for strategic decision making”. In: *Decision Support Systems* 36, pp. 147–160. ISSN: 01679236. DOI: 10.1016/S0167-9236(02)00139-2.

Comas, J et al. (2001). “Knowledge discovery by means of inductive methods in wastewater treatment plant data”. In: *Ai Communications* 14.1, pp. 45–62. ISSN: 0921-7126. URL: wos:000168269900006.

Comas, Joaquim et al. (2004a). "Development of a knowledge-based decision support system for identifying adequate wastewater treatment for small communities". In: *Water Science and Technology* 48.11-12, pp. 393–400. ISSN: 02731223.

Comas, Joaquim et al. (2004b). "Development of a knowledge-based decision support system for identifying adequate wastewater treatment for small communities". In: *Water Science and Technology* 48.11-12, pp. 393–400. ISSN: 02731223.

Corominas, Lluis et al. (2008). "Improving process performance in wastewater treatment plants using an intelligent environmental decision support system". In: *Proc. iEMSs 4th Biennial Meeting - Int. Congress on Environmental Modelling and Software: Integrating Sciences and Information Technology for Environmental Assessment and Decision Making, iEMSs 2008*. Vol. 1, pp. 295–302. ISBN: 9788476530740. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84858302311{\%}7B{\&}{\%}7DpartnerID=tZOtx3y1http://www.scopus.com/inward/record.url?eid=2-s2.0-84858302311{\&}partnerID=tZOtx3y1.

Cortés, U. and M. Sànchez-Marré (2001). "Environmental Decision Support Systems". In: *AI Communications* 14.1, pp. 1–2. ISSN: 09217126. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-0035073303{\&}partnerID=tZOtx3y1.

Council of the European Union (1991). "Council Directive of 21 May 1991 concerning Urban Waste Water Treatment (91/271/EEC)". In: *Official J. Eur. Commun. L* 34.May 1991, pp. 1–16. URL: http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:COUNCIL+DIRECTIVE+of+21+May+1991+concerning+urban+waste+water+treatment+(91/271/EEC){\#}1.

da Costa Bortoni, Edson, Roberto Alves Almeida, and Augusto Nelson Carvalho Viana (2008). "Optimization of parallel variable-speed-driven centrifugal pumps operation". In: *Energy Efficiency* 1.3, pp. 167–173. ISSN: 1570646X. DOI: 10.1007/s12053-008-9010-1.

DataJobs.com (2017). *What is Hadoop? What is NoSQL? What is MapReduce?* URL: https://datajobs.com/what-is-hadoop-and-nosql (visited on 11/15/2017).

De Gussem, K. et al. (2011). "Cost optimisation and minimisation of the environmental impact through life cycle analysis of the waste water treatment plant of Bree (Belgium)". In: *Water Science and Technology* 63.1, pp. 164–170. ISSN: 02731223. DOI: 10.2166/wst.2011.027.

DeBenedictis, A. et al. (2013). "Operational energy-efficiency improvement of municipal water pumping in California". In: *Energy* 53, pp. 237–243. ISSN: 03605442. DOI: 10.1016/j.energy.2013.02.012. URL: http://dx.doi.org/10.1016/j.energy.2013.02.012.

Dernoncourt, Franck (2013). "Introduction to fuzzy logic". In: January.

Djukic, Malisa et al. (2016). "Cost-benefit analysis of an infrastructure project and a cost-reflective tariff: A case study for investment in wastewater treatment plant in Serbia". In: *Renewable and Sustainable Energy Reviews* 59, pp. 1419–1425. ISSN: 13640321. DOI: 10.1016/j.rser.2016.01.050. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1364032116000800.

Doherty, Edelle (2017). "Development of new benchmarking systems for wastewater treatment facilities". PhD thesis, pp. 67–77. URL: http://hdl.handle.net/10379/6637.

Duah, D. and M. Syal (2016). "Intelligent decision support system for home energy retrofit adoption". In: *International Journal of Sustainable Built Environment* 5.2, pp. 620–634. ISSN: 22126090. DOI: 10.1016/j.ijsbe.2016.05.003. URL: http://linkinghub.elsevier.com/retrieve/pii/S2212609015300558.

Dürrenmatt, David JérÔme and Willi Gujer (2012). "Data-driven modeling approaches to support wastewater treatment plant operation". In: *Environmental Modelling & Software* 30, pp. 47–56. ISSN: 13648152. DOI: 10.1016/j.envsoft.2011.11.007. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815211002623.

DWA (2015). *Energiecheck und Energieanalyse - Instrumente zur Energieoptimierung von Abwasseranlagen*, p. 62. ISBN: 978-3-88721-276-6. URL: http://www.dwa.de/dwa/shop/shop.nsf/Produktanzeige?openform{\&}produktid=P-DWAA-A4W8YH.

EEC Council (1991). "91/271/EEC of 21 May 1991 concerning urban waste-water treatment". In: *EEC Council Directive*, p. 10.

Elmahdi, A. and D. McFarlane (2009). "A decision support system for a groundwater system case study: Gnangara sustainability strategy Western Australia". In: *18th World IMACS Congress and MODSIM09*

International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings July, pp. 3803–3809. ISSN: 17433541. DOI: ExportDate9July2012\nSourceScopus. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-80052994030{\&}partnerID=tZOtx3y1.

European Commission (2016). *Eighth Report on the Implementation Status and the Programmes for Implementation (as required by Article 17) of Council Directive 91/271/EEC concerning urban waste water treatment*. Tech. rep., pp. 1–17. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3.

European Environmental Agency. *Overview of electricity production and use in Europe — European Environment Agency*. URL: http://www.eea.europa.eu/data-and-maps/indicators/overview-of-the-electricity-production-1/assessment.

European Parliament (2000). "Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy". In: *Official Journal of the European Parliament* L327.October 2000, pp. 1–82. ISSN: 0144557X. DOI: 10.1039/ap9842100196. arXiv: 534.

— (2006). "Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC". In: *Official Journal of the European Union* L 064.1882, pp. 37–51. DOI: L102/15.

Eurostat. *WebPage: Gross Inland energy consumption (Eurostat)*. URL: http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Electricity{_}consumption{_}and{_}trade,{_}GWh,{_}2014{_}new.png (visited on 06/27/2017).

Fiter, M et al. (2005). "Energy saving in a wastewater treatment process: an application of fuzzy logic control." In: *Environmental technology* 26.11, pp. 1263–1270. ISSN: 0959-3330. DOI: 10.1080/09593332608618596. URL: http://www.ncbi.nlm.nih.gov/pubmed/16335601.

Foladori, Vaccari, and Vitali (2015). "Energy audit in small wastewater treatment plants: methodology, energy consumption indicators, and lessons learned". In: 1739, pp. 1–13. DOI: 10.2166/wst.2015.306.

Friedler, Eran and Ehud Pisanty (2006). "Effects of design flow and treatment level on construction and operation costs of municipal wastewater treatment plants and their implications on policy making". In: *Water Research* 40.20, pp. 3751–3758. ISSN: 00431354. DOI: 10.1016/j.watres.2006.08.015.

Gardoni, Davide, Arianna Catenacci, and Manuela Antonelli (2015). "Reuse of process water in a waste-to-energy plant: An Italian case of study". In: *Waste Management* 43, pp. 196–202. ISSN: 0956053X. DOI: 10.1016/j.wasman.2015.05.004. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0956053X15003554.

Garrido-Baserba, M et al. (2015). "Selecting sewage sludge treatment alternatives in modern wastewater treatment plants using environmental decision support systems". In: *Journal of Cleaner Production* 107, pp. 410–419. ISSN: 09596526. DOI: 10.1016/j.jclepro.2014.11.021. URL: http://linkinghub.elsevier.com/retrieve/pii/S0959652614012013.

Garrido-Baserba, M. et al. (2016). "Application of a multi-criteria decision model to select of design choices for WWTPs". In: *Clean Technologies and Environmental Policy* 18.4, pp. 1097–1109. ISSN: 16189558. DOI: 10.1007/s10098-016-1099-x.

Gibert, Karina, Dante Conti, and Darko Vrecko (2012). "Assisting the end-user in the interpretation of profiles for decision support. An application to wastewater treatment plants". In: *Environmental Engineering and Management Journal* 11.5, pp. 931–944. ISSN: 15829596.

Gisi, Sabino De et al. (2015). "An integrated approach for monitoring efficiency and investments of activated sludge-based wastewater treatment plants at large spatial scale". In: *Science of The Total Environment* 523, pp. 201–218. ISSN: 00489697. DOI: 10.1016/j.scitotenv.2015.03.106. URL: http://www.sciencedirect.com/science/article/pii/S0048969715003903http://dx.doi.org/10.1016/j.scitotenv.2015.03.106.

Goldstein, R. and W. Smith (2002). "Water & Sustainability (Volume 4): U.S. Electricity Consumption for Water Supply & Treatment - The Next Half Century". In: *Water Supply* 4.Volume 4, p. 93. ISSN: 1098-6596. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3.

Gómez-López, M.D. D. et al. (2009). "Decision support in disinfection technologies for treated wastewater reuse". In: *Journal of Cleaner Production* 17.16, pp. 1504–1511. ISSN: 09596526. DOI: 10.1016/j.jclepro.2009.06.008. URL: http://linkinghub.elsevier.com/retrieve/pii/S0959652609002078.

Grant, S. B. et al. (2012). "Taking the "Waste" Out of "Wastewater" for Human Water Security and Ecosystem Sustainability". In: *Science* 337.6095, pp. 681–686. ISSN: 0036-8075. DOI: 10.1126/science.1216852. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1216852.

Gu, Yifan et al. (2017). "The feasibility and challenges of energy self-sufficient wastewater treatment plants". In: *Applied Energy*. ISSN: 03062619. DOI: 10.1016/j.apenergy.2017.02.069. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0306261917302179.

Gude, Veera Gnaneswar (2015). "Energy and water autarky of wastewater treatment and power generation systems". In: *Renewable and Sustainable Energy Reviews* 45, pp. 52–68. ISSN: 13640321. DOI: 10.1016/j.rser.2015.01.055. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364032115000659.

Guerrero, J. et al. (2012). "Multi-criteria selection of optimum WWTP control setpoints based on microbiology-related failures, effluent quality and operating costs". In: *Chemical Engineering Journal* 188, pp. 23–29. ISSN: 13858947. DOI: 10.1016/j.cej.2012.01.115. URL: http://dx.doi.org/10.1016/j.cej.2012.01.115.

Guerrero, Javier et al. (2011). "Improving the performance of a WWTP control system by model-based setpoint optimisation". In: *Environmental Modelling & Software* 26.4, pp. 492–497. ISSN: 13648152. DOI: 10.1016/j.envsoft.2010.10.012. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815210002847.

Gülich, Johann Friedrich (1998). *Centrifugal Pumps*, p. 780. ISBN: 978-3-540-73694-3.

Haberkern, Bernd, Werner Maier, and Ursula Schneider (2008). *Steigerung der Effizienz auf kommunalen Klaranlagen*. Tech. rep.

Hakanen, J., K. Sahlstedt, and K. Miettinen (2013). "Wastewater treatment plant design and operation under multiple conflicting objective functions". In: *Environmental Modelling & Software* 46, pp. 240–

249. ISSN: 13648152. DOI: 10.1016/j.envsoft.2013.03.016. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815213000790.

Hamouda, M. a., W. B. Anderson, and P. M. Huck (2009). "Decision support systems in water and wastewater treatment process selection and design: A review". In: *Water Science and Technology* 60.7, pp. 1767–1770. ISSN: 02731223. DOI: 10.2166/wst.2009.538.

Hansen, Joachim (2018). *Lecture Notes in Sustainable Water Resources Management*.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). "The Elements of Statistical Learning". In: *Elements* 1, pp. 337–387. ISSN: 03436993. DOI: 10.1007/b94608. URL: http://www.springerlink.com/index/10.1007/b94608.

Hernández-Sancho, F, M Molinos-Senante, and R Sala-Garrido (2011). "Energy efficiency in Spanish wastewater treatment plants: a non-radial DEA approach." In: *The Science of the total environment* 409.14, pp. 2693–2699. ISSN: 1879-1026. DOI: 10.1016/j.scitotenv.2011.04.018. URL: http://www.ncbi.nlm.nih.gov/pubmed/21549411.

Hernandez-Sancho, F. et al. (2009). "Environmental benefits of wastewater treatment: an economic valuation". In: *Risk Management of Water Supply and Sanitation Systems*, pp. 251–260.

Hernández-Sancho, Francesc and Ramón Sala-Garrido (2009). "Technical efficiency and cost analysis in wastewater treatment processes: A DEA approach". In: *Desalination* 249.1, pp. 230–234. ISSN: 00119164. DOI: 10.1016/j.desal.2009.01.029. URL: http://linkinghub.elsevier.com/retrieve/pii/S0011916409008236.

Hudson, Donna L and Maurice E Cohen (2012). "Overcoming barriers to development of cooperative medical decision support models." In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* 2012, pp. 2194–7. ISSN: 1557-170X. DOI: 10.1109/EMBC.2012.6346397. URL: http://www.ncbi.nlm.nih.gov/pubmed/23366358.

Hydraulic Institute and Department of Energy (US) (2006). *Improving Pumping System Performance: A Sourcebook for Industry*. Tech. rep.

Hydraulic Institute, U S Department of Energy, and U S Department of Energy (2005). *Energy Tips – Pumping Systems*. Tech. rep. September. URL: www.eere.energy.gov.

INNERS (2015). *Innovative Energy Recovery Strategies in the urban water cycle*. Tech. rep., p. 105.

Inotify. *http://inotify.aiken.cz/?section=incron&page=about&lang=en*. URL: http://inotify.aiken.cz/?section=incron{\&}page=about{\&}lang=en (visited on 09/18/2017).

IPCC, Intergovernmental Panel On Climate Change (2007). "Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC". In: *Science* October 2009, p. 1009. ISSN: 0958-305X. DOI: volume. URL: http://www.amazon.com/Climate-Change-2007-Contribution-Assessment/dp/0521880092.

Kalbar, Pradip P., Subhankar Karmakar, and Shyam R. Asolekar (2016). "Life cycle-based decision support tool for selection of wastewater treatment alternatives". In: *Journal of Cleaner Production* 117, pp. 64–72. ISSN: 09596526. DOI: 10.1016/j.jclepro.2016.01.036. URL: http://linkinghub.elsevier.com/retrieve/pii/S0959652616000718.

Kallis, Giorgos and David Butler (2001). "The EU water framework directive: Measures and implications". In: *Water Policy* 3.2, pp. 125–142. ISSN: 13667017. DOI: 10.1016/S1366-7017(01)00007-1.

Kern, Peter et al. (2014). "COD and NH 4 -N Estimation in the Inflow of Wastewater Treatment Plants using Machine Learning Techniques". In: *Ieee*.

Kim, Minsoo et al. (2016). "Operator decision support system for integrated wastewater management including wastewater treatment plants and receiving water bodies". In: *Environmental Science and Pollution Research* 23.11, pp. 10785–10798. ISSN: 16147499. DOI: 10.1007/s11356-016-6272-6.

Ko, Kang-Young Ko Kang-Young et al. (2003). "Neural network model for wastewater treatment plant control". In: *IEEE International Workshop on Soft Computing Techniques in Instrumentation, Measurement and Related Applications, 2003. SCIMA 2003.* DOI: 10.1109/SCIMA.2003.1215928.

Krampe, J. (2013). "Energy benchmarking of South Australian WWTPs". In: *Water Science and Technology* 67.9, pp. 2059–2066. ISSN: 02731223. DOI: 10.2166/wst.2013.090. URL: http://www.ncbi.nlm.nih.gov/pubmed/23656950.

Kusiak, Andrew and Xiupeng Wei (2014). "Prediction of methane production in wastewater treatment facility: A data-mining approach". In: *Annals of Operations Research* 216.1, pp. 71–81. ISSN: 15729338. DOI: 10.1007/s10479-011-1037-6.

Leverenz, Harold L., George Tchobanoglous, and Takashi Asano (2011). "Direct potable reuse: a future imperative". In: *Journal of Water Reuse and Desalination* 1.1, p. 2. ISSN: 2220-1319. DOI: 10.2166/wrd.2011.000. URL: http://jwrd.iwaponline.com/cgi/doi/10.2166/wrd.2011.000.

Li, Yi et al. (2017). "Study on indicators for on-line monitoring and diagnosis of anaerobic digestion process of piggery wastewater". In: *Environmental Technology and Innovation* 8, pp. 423–430. ISSN: 23521864. DOI: 10.1016/j.eti.2017.09.008.

Liaw, Andy and Matthew Wiener (2002). "Classification and Regression by randomForest". In: *R News* 2.3, pp. 18–22. URL: http://cran.r-project.org/doc/Rnews/.

Little, Keith W. et al. (2017). "Decision support for environmental management of industrial non-hazardous secondary materials: New analytical methods combined with simulation and optimization modeling". In: *Journal of Environmental Management* 196, pp. 137–147. ISSN: 03014797. DOI: 10.1016/j.jenvman.2017.02.075. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0301479717301937.

Longo, Stefano et al. (2016). "Monitoring and diagnosis of energy consumption in wastewater treatment plants. A state of the art and proposals for improvement". In: *Applied Energy* 179, pp. 1251–1268. ISSN: 03062619. DOI: 10.1016/j.apenergy.2016.07.043. URL: http://linkinghub.elsevier.com/retrieve/pii/S0306261916309850.

Lorenzo-Toja, Yago et al. (2016). "Benchmarking wastewater treatment plants under an eco-efficiency perspective". In: *Science of the Total Environment* 566-567, pp. 468–479. ISSN: 18791026. DOI: 10.1016/

j.scitotenv.2016.05.110. URL: http://dx.doi.org/10.1016/j.scitotenv.2016.05.110.

Madsen, Michael, Jens Bo Holm-Nielsen, and Kim H. Esbensen (2011). "Monitoring of anaerobic digestion processes: A review perspective". In: *Renewable and Sustainable Energy Reviews* 15.6, pp. 3141–3155. ISSN: 13640321. DOI: 10.1016/j.rser.2011.04.026. URL: http://dx.doi.org/10.1016/j.rser.2011.04.026.

Mamdani, E.H. and S. Assilian (1975). "An experiment in linguistic synthesis with a fuzzy logic controller". In: *International Journal of Man-Machine Studies* 7.1, pp. 1–13. ISSN: 00207373. DOI: 10.1016/S0020-7373(75)80002-2.

Management Study Guide (2017). *Gaining Competitive Advantage with Decision Support Systems*. URL: https://www.managementstudyguide.com/gaining-competitive-advantage-with-decision-support-systems.htm (visited on 06/14/2017).

Maragkaki, A. E. et al. (2017). "Improving biogas production from anaerobic co-digestion of sewage sludge with a thermal dried mixture of food waste, cheese whey and olive mill wastewater". In: *Waste Management* 71, pp. 644–651. ISSN: 18792456. DOI: 10.1016/j.wasman.2017.08.016. URL: https://doi.org/10.1016/j.wasman.2017.08.016.

Martin, Cristina and Peter A. Vanrolleghem (2014). "Analysing, completing, and generating influent data for WWTP modelling: A critical review". In: *Environmental Modelling and Software* 60, pp. 188–201. ISSN: 13648152. DOI: 10.1016/j.envsoft.2014.05.008. URL: http://linkinghub.elsevier.com/retrieve/pii/S136481521400139X.

McIntosh, B.S. S et al. (2011). "Environmental decision support systems (EDSS) development – Challenges and best practices". In: *Environmental Modelling & Software* 26.12, pp. 1389–1402. ISSN: 13648152. DOI: 10.1016/j.envsoft.2011.09.009. URL: http://www.sciencedirect.com/science/article/pii/S1364815211002076http://linkinghub.elsevier.com/retrieve/pii/S1364815211002076.

Melorose, J., R. Perroy, and S. Careas (2015). *World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP.241*. Tech. rep., pp. 1–59. DOI: 10.1017/CBO9781107415324.

004. arXiv: arXiv:1011.1669v3. URL: http://www.ncbi.nlm.nih.gov/pubmed/21798940.

Mendel, Jerry M. JM (1995). "Fuzzy logic systems for engineering: a tutorial". In: *Proceedings of the IEEE* 83.9408047, pp. 345–377. ISSN: 00189219. DOI: 10.1109/5.364485. URL: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=364485.

Metcalf, E. and H. Eddy (2014). *Wastewater engineering: treatment and reuse*, p. 1819. ISBN: 0-07-041878-0. DOI: 10.1016/0309-1708(80)90067-6. URL: http://www.lavoisier.fr/notice/fr097556.html.

Mizuta, Kentaro and Masao Shimada (2010). "Benchmarking energy consumption in municipal wastewater treatment plants in Japan". In: *Water Science and Technology* 62.10, pp. 2256–2262. ISSN: 02731223. DOI: 10.2166/wst.2010.510. URL: http://www.ncbi.nlm.nih.gov/pubmed/21076210.

Molinos-Senante, M., F. Hernandez-Sancho, and R. Sala-Garrido (2014). "Benchmarking in wastewater treatment plants: A tool to save operational costs". In: *Clean Technologies and Environmental Policy* 16.1, pp. 149–161. ISSN: 16189558. DOI: 10.1007/s10098-013-0612-8. URL: http://link.springer.com/10.1007/s10098-013-0612-8.

Molinos-Senante, María, Francesc Hernández-Sancho, and Ramón Sala-Garrido (2010). "Economic feasibility study for wastewater treatment: A cost-benefit analysis". In: *Science of the Total Environment* 408.20, pp. 4396–4402. ISSN: 00489697. DOI: 10.1016/j.scitotenv.2010.07.014. URL: http://www.ncbi.nlm.nih.gov/pubmed/20667582.

MosayebNezhad, M et al. (2017). "Techno-economic assessment of biogas-fed CHP hybrid systems in a real wastewater treatment plant". In: *Applied Thermal Engineering*. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2017.10.115. URL: http://linkinghub.elsevier.com/retrieve/pii/S1359431117324213.

Muller, Kobel, and Schmid (2010). *Energie in ARA: Leitfaden zur Energieoptimierung auf Abwasserreinigungsanlagen; Handbuch*. Bundesamt für Energie. URL: https://books.google.lu/books?id=pyGImwEACAAJ.

Muller, E A, R Thommen, and P Stahli (1994). *Energie in ARA: Energiesparmassnahmen in Abwasserreinigungsanlagen : Handbuch*. Eidg. Dr.-Sachen- und Materialzentrale. ISBN: 9783905232493. URL: https://books.google.lu/books?id=Eb{_}5GwAACAAJ.

Mustajoki, Jyri and Mika Marttunen (2017). "Comparison of multi-criteria decision analytical software for supporting environmental planning processes". In: *Environmental Modelling & Software* 93, pp. 78–91. ISSN: 13648152. DOI: 10.1016/j.envsoft.2017.02.026. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1364815216302584.

Naidu, Ravi et al. (2016). "Emerging contaminants in the environment: Risk-based analysis for better management". In: *Chemosphere* 154, pp. 350–357. ISSN: 18791298. DOI: 10.1016/j.chemosphere.2016.03.068.

Nakagawa, Naoko et al. (2006). "Field survey of a sustainable sanitation system in a residential house". In: *Journal of Environmental Sciences (China)* 18.6, pp. 1088–1093. ISSN: 10010742. DOI: 10.1016/S1001-0742(06)60044-2.

Navarro-Hell?n, H. et al. (2016). "A decision support system for managing irrigation in agriculture". In: *Computers and Electronics in Agriculture* 124, pp. 121–131. ISSN: 01681699. DOI: 10.1016/j.compag.2016.04.003. URL: http://linkinghub.elsevier.com/retrieve/pii/S016816991630117X.

Ng, Bernard J H et al. (2014). "Environmental life cycle assessment of different domestic wastewater streams: Policy effectiveness in a tropical urban environment". In: *Journal of Environmental Management* 140, pp. 60–68. ISSN: 10958630. DOI: 10.1016/j.jenvman.2014.01.052. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0301479714001431.

Olszewski, Pawel (2016). "Genetic optimization and experimental verification of complex parallel pumping station with centrifugal pumps". In: *Applied Energy* 178, pp. 527–539. ISSN: 03062619. DOI: 10.1016/j.apenergy.2016.06.084. URL: http://dx.doi.org/10.1016/j.apenergy.2016.06.084http://linkinghub.elsevier.com/retrieve/pii/S0306261916308601.

Pablo-Romero, María del P. et al. (2017). "An overview of feed-in tariffs, premiums and tenders to promote electricity from biogas in the

EU-28”. In: *Renewable and Sustainable Energy Reviews* 73, pp. 1366–1379. ISSN: 13640321. DOI: 10.1016/j.rser.2017.01.132. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1364032117301429.

Panepinto, Deborah et al. (2016). “Evaluation of the energy efficiency of a large wastewater treatment plant in Italy”. In: *Applied Energy* 161, pp. 404–411. ISSN: 03062619. DOI: 10.1016/j.apenergy.2015.10.027. URL: http://dx.doi.org/10.1016/j.apenergy.2015.10.027.

Papa, M., G. Bertanza, and A. Abbà (2016). “Reuse of wastewater: a feasible option, or not? A decision support system can solve the doubt”. In: *Desalination and Water Treatment* 57.19, pp. 8670–8682. ISSN: 1944-3994. DOI: 10.1080/19443994.2015.1029532. URL: http://www.tandfonline.com/doi/full/10.1080/19443994.2015.1029532.

Papa, Matteo et al. (2017). “How far are we from closing the loop of sewage resource recovery? A real picture of municipal wastewater treatment plants in Italy”. In: *Journal of Environmental Management* 198, pp. 9–15. ISSN: 03014797. DOI: 10.1016/j.jenvman.2017.04.061. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0301479717304097.

Paraskevas, P A, I S Pantelakis, and T D Lekkas (1999). “Advanced integrated expert system for wastewater treatment plants control”. In: *Knowledge-Based Systems* 12.7, pp. 355–361. ISSN: 09507051. DOI: 10.1016/S0950-7051(99)00040-4. URL: http://linkinghub.elsevier.com/retrieve/pii/S0950705199000404.

Petrie, Bruce, Ruth Barden, and Barbara Kasprzyk-Hordern (2014). “A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring”. In: *Water Research* 72, pp. 3–27. ISSN: 18792448. DOI: 10.1016/j.watres.2014.08.053.

Poch, Manel et al. (2004). “Designing and building real environmental decision support systems”. In: *Environmental Modelling & Software* 19.9, pp. 857–873. ISSN: 13648152. DOI: 10.1016/j.envsoft.2003.03.007. URL: http://www.sciencedirect.com/science/article/pii/S1364815203002068.

Poch, Manel et al. (2014). "Where are we in wastewater treatment plants data management? A review and a proposal". In: *Proceedings - 7th International Congress on Environmental Modelling and Software: Bold Visions for Environmental Modeling, iEMSs 2014*. Vol. 3. International Environmental Modelling and Software Society, pp. 1450–1455. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84911903433{\&}partnerID=tZOtx3y1http://www.scopus.com/inward/record.url?eid=2-s2.0-84911903433{\%}7B{\&}{\%}7DpartnerID=tZOtx3y1.

Power, Dan. *DSSresource.com*. URL: http://dssresources.com/.

— (2000). "Decision Support Systems Hyperbook". In: *Cedar Falls, IA: DSSResources.COM, HTML version, Fall 2000*, pp. 1–22. URL: http://dssresources.com/dssbook/ch1sbdm.pdf.

Pretel, R. et al. (2015). "Designing an AnMBR-based WWTP for energy recovery from urban wastewater: The role of primary settling and anaerobic digestion". In: *Separation and Purification Technology* 156, pp. 132–139. ISSN: 13835866. DOI: 10.1016/j.seppur.2015.09.047. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1383586615302227.

Rahmanpour, Mehdi and Morteza Osanloo (2017). "A decision support system for determination of a sustainable pit limit". In: *Journal of Cleaner Production* 141, pp. 1249–1258. ISSN: 09596526. DOI: 10.1016/j.jclepro.2016.09.205. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0959652616315530.

Rauch, W et al. (2002). "Deterministic modelling of integrated urban drainage systems." In: *Water science and technology : a journal of the International Association on Water Pollution Research* 45.3, pp. 81–94. ISSN: 0273-1223.

Rawal, Nekram and S K Duggal (2016). "Life Cycle Costing Assessment-Based Approach for Selection of Wastewater Treatment Units". In: *National Academy Science Letters* 39.2, pp. 103–107. ISSN: 0250541X. DOI: 10.1007/s40009-016-0429-1.

Rehman, Fahad et al. (2015). "Fluidic oscillator-mediated microbubble generation to provide cost effective mass transfer and mixing efficiency to the wastewater treatment plants". In: *Environmental Research* 137, pp. 32–39. ISSN: 10960953. DOI: 10.1016/j.envres.

2014.11.017. URL: http://www.ncbi.nlm.nih.gov/pubmed/25483415.

Reinders, Michael et al. (2012). "Solution approaches for energy optimization in the water sector". In: *IWA World Congress on Water, Climate and Energy*, pp. 1–8. URL: https://keynote.conference-services.net/resources/444/2653/pdf/IWAWCE2012{_}0381.pdf.

Rizzoli, A.E. and W.J. Young (1997). "Delivering environmental decision support systems: software tools and techniques". In: *Environmental Modelling & Software* 12.2-3, pp. 237–249. ISSN: 13648152. DOI: 10.1016/S1364-8152(97)00016-9. URL: http://linkinghub.elsevier.com/retrieve/pii/S1364815297000169.

Robles, A. et al. (2017). "A fuzzy-logic-based controller for methane production in anaerobic fixed-film reactors". In: *Environmental Technology (United Kingdom)* 38.1, pp. 42–52. ISSN: 1479487X. DOI: 10.1080/09593330.2016.1184321.

Rose, David C. et al. (2016). "Decision support tools for agriculture: Towards effective design and delivery". In: *Agricultural Systems* 149, pp. 165–174. ISSN: 0308521X. DOI: 10.1016/j.agsy.2016.09.009. URL: http://linkinghub.elsevier.com/retrieve/pii/S0308521X16305418.

Ruano, M.V. V et al. (2010). "A systematic approach for fine-tuning of fuzzy controllers applied to WWTPs". In: *Environmental Modelling {&} Software* 25.5, pp. 670–676. ISSN: 13648152. DOI: 10.1016/j.envsoft.2009.05.008. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1364815209001212http://linkinghub.elsevier.com/retrieve/pii/S1364815209001212.

Sala-Garrido, Ramón, Francesc Hernández-Sancho, and María Molinos-Senante (2012). "Assessing the efficiency of wastewater treatment plants in an uncertain context: A DEA with tolerances approach". In: *Environmental Science and Policy* 18, pp. 34–44. ISSN: 14629011. DOI: 10.1016/j.envsci.2011.12.012.

Sala-Garrido, Ramón, María Molinos-Senante, and Francesc Hernández-Sancho (2012). "How does seasonality affect water reuse possibilities? An efficiency and cost analysis". In: *Resources, Conservation and Recycling* 58, pp. 125–131. ISSN: 09213449. DOI: 10.1016/j.

resconrec.2011.11.002. URL: http://dx.doi.org/10.1016/j.resconrec.2011.11.002.

Shen, Yanwen et al. (2015). "An overview of biogas production and utilization at full-scale wastewater treatment plants (WWTPs) in the United States: Challenges and opportunities towards energy-neutral WWTPs". In: *Renewable and Sustainable Energy Reviews* 50, pp. 346–362. ISSN: 13640321. DOI: 10.1016/j.rser.2015.04.129. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1364032115003998.

Shi, Cao Ye (2011). *Mass Flow and Energy Efficiency of Municipal Wastewater Treatment Plants*. Ed. by I W A Publishing. IWA Publishing, p. 55. ISBN: 9781843393825.

Shumway, R H and D S Stoffer (2011). *Time Series Analysis and Its Applications*. Vol. 97, pp. 417–437. ISBN: 9781441978646. DOI: 10.1007/978-1-4419-7865-3. arXiv: arXiv:1011.1669v3. URL: http://www.springerlink.com/index/10.1007/978-1-4419-7865-3.

Sivanandam, S.N. N N, S. Sumathi, and S. N. Deepa (2006). *Introduction to Fuzzy Logic using MATLAB*, p. 430. ISBN: 9783540357803. URL: http://www.amazon.com/Introduction-Fuzzy-Logic-using-MATLAB/dp/3540357807.

Spellman, Frank (2003). *Handbook of water and wastewater treatment plant operations*. CRC Press, p. 653. ISBN: 9781466553385. DOI: 10.1016/B978-0-12-369520-8.50043-7. URL: http://books.google.com/books?hl=en{\&}lr={\&}id=dXxcAgAAQBAJ{\&}oi=fnd{\&}pg=PP1{\&}dq=Handbook+of+Water+and+Wastewater+Treatment+Plant+Operations{\&}ots=I2ToGvx3v8{\&}sig=VYDsJINek6Nl4FwUFEE3yO{_}tf3s{\%}5Cnhttp://books.google.com/books?hl=en{\&}lr={\&}id=dXxcAgAAQBAJ{\&}oi=fnd{\&}pg=PP1{\&}dq.

Starczewski, Janusz T. (2013). *Advanced concepts in fuzzy logic and systems with membership uncertainty*. Vol. 284, pp. 1–319. ISBN: 9783642295195. DOI: 10.1007/978-3-642-29520-1.

Thürlimann, Christian M., David J. Dürrenmatt, and Kris Villez (2015). "Energy and process data processing and visualisation for optimising wastewater treatment plants". In: *Water Practice and Technology* 10.1, pp. 10–18. ISSN: 1751231X. DOI: 10.2166/wpt.2015.002.

URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84925456557{\&}partnerID=tZOtx3y1.

Tomei, Maria Concetta et al. (2016). "Techno-economic and environmental assessment of upgrading alternatives for sludge stabilization in municipal wastewater treatment plants". In: *Journal of Cleaner Production* 112, pp. 3106–3115. ISSN: 09596526. DOI: 10.1016/j.jclepro.2015.10.017. URL: http://linkinghub.elsevier.com/retrieve/pii/S0959652615013980.

Torregrossa, D et al. (2016). "Energy saving in WWTP: Daily benchmarking under uncertainty and data availability limitations". In: *Environmental Research* 148, pp. 330–337. ISSN: 00139351. DOI: 10.1016/j.envres.2016.04.010. URL: http://www.sciencedirect.com/science/article/pii/S0013935116301360.

Torregrossa, D. et al. (2017a). "Energy saving in wastewater treatment plants: A plant-generic cooperative decision support system". In: *Journal of Cleaner Production* 167. ISSN: 09596526. DOI: 10.1016/j.jclepro.2017.08.181. URL: http://www.sciencedirect.com/science/article/pii/S0959652617319145.

Torregrossa, Dario and Joachim Hansen (2018). "SK-DSSy: how to integrate the YouTube platform in a cooperative decision support?" In: *Springer book Lecture Notes In Business Information Processing - IN PRINTING*, pp. 1–12.

Torregrossa, Dario, Joachim Hansen, and Ulrich Leopold (2017). "A hybrid data-model decision tool for the assessment of the pump cavitation risk in wastewater treatment plants". In: *From Science to Society: The Bridge provided by Environmental Informatics*. Vol. 00. June 2017. Shaker Verlag, pp. 1–7. ISBN: 978-3-8440-5495-8. URL: https://www.shaker.de/de/content/catalogue/index.asp?lang=de{\&}ID=8{\&}ISBN=978-3-8440-5495-8{\&}search=yes.

Torregrossa, Dario et al. (2017b). "A data-driven methodology to support pump performance analysis and energy efficiency optimization in Waste Water Treatment Plants". In: *Applied Energy* 208.June, pp. 1430–1440. ISSN: 03062619. DOI: 10.1016/j.apenergy.2017.09.012. URL: http://dx.doi.org/10.1016/j.apenergy.2017.09.012http://linkinghub.elsevier.com/retrieve/pii/S0306261917312928.

Torregrossa, Dario et al. (2017c). "A Tool for Energy Management and Cost Assessment of Pumps in Waste Water Treatment Plants". In: *Decision Support Systems VII. Data, Information and Knowledge Visualization in Decision Support Systems*. Vol. 282. Springer International Publishing AG, pp. 148–161. ISBN: 978-3-319-57486-8. DOI: 10.1007/978-3-319-57487-5_11. URL: http://doi.org/10.1007/978-3-319-57487-5{_}11.

Torregrossa, Dario et al. (2017d). "Pump Efficiency Analysis of Waste Water Treatment Plants: A Data Mining Approach Using Signal Decomposition for Decision Making". In: *Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2017. ICCSA 2017. Lecture Notes in Computer Science*. Springer, Cham, pp. 744–752. DOI: 10.1007/978-3-319-62407-5_56. URL: http://link.springer.com/10.1007/978-3-319-62407-5{_}56.

Tran, Quynh K, Kurt A Schwabe, and David Jassby (2016). "Wastewater reuse for agriculture: Development of a regional water reuse decision-support model (RWRM) for cost-effective irrigation sources". In: *Environmental Science and Technology* 50.17, pp. 9390–9399. ISSN: 15205851. DOI: 10.1021/acs.est.6b02073.

Turunen, Ville, Jaana Sorvari, and Anna Mikola (2018). "A decision support tool for selecting the optimal sewage sludge treatment". In: *Chemosphere* 193, pp. 521–529. ISSN: 18791298. DOI: 10.1016/j.chemosphere.2017.11.052. URL: https://doi.org/10.1016/j.chemosphere.2017.11.052.

United States Environmental Protection Agency (USEPA) (2012). *The Clean Water Act: Protecting and Restoring our Nation's Waters*. URL: http://water.epa.gov/action/cleanwater40/cwa101.cfm{\#}nonpoint.

Venkatesh, G. and Rashid Abdi Elmi (2013). "Economic–environmental analysis of handling biogas from sewage sludge digesters in WWTPs (wastewater treatment plants) for energy recovery: Case study of Bekkelaget WWTP in Oslo (Norway)". In: *Energy* 58, pp. 220–235. ISSN: 03605442. DOI: 10.1016/j.energy.2013.05.025. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0360544213004325.

Verstraete, Willy, Peter Clauwaert, and Siegfried E. Vlaeminck (2016). "Used water and nutrients: Recovery perspectives in a 'panta rhei' context". In: *Bioresource Technology* 215, pp. 199–208. ISSN: 09608524. DOI: 10.1016/j.biortech.2016.04.094. URL: http://linkinghub.elsevier.com/retrieve/pii/S0960852416305909.

Wang, Chuan et al. (2017). "Optimal design of multistage centrifugal pump based on the combined energy loss model and computational fluid dynamics". In: *Applied Energy* 187, pp. 10–26. ISSN: 03062619. DOI: 10.1016/j.apenergy.2016.11.046. URL: http://dx.doi.org/10.1016/j.apenergy.2016.11.046.

Wang, Dong et al. (2016). "A cloud model-based approach for water quality assessment." In: *Environmental research* 148, pp. 24–35. ISSN: 1096-0953. DOI: 10.1016/j.envres.2016.03.005. URL: http://www.sciencedirect.com/science/article/pii/S0013935116300858.

WEF (2008). *Operation of Municipal Wastewater Treatment Plants*. Vol. 1542. 9. WEF Press, pp. 33–36. ISBN: 9788578110796. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3.

Wett, B, K Buchauer, and C Fimml (2007). "Energy self-sufficiency as a feasible concept for wastewater treatment systems". In: *Proceedings of the IWA Leading Edge Technology Conference (Singapore)*. September, pp. 21–24. ISBN: 3905232499. URL: http://www.araconsult.at/download/literature/let07{_}wett{_}energy{_}selfsufficiency3.pdf.

Wikiversity. *Introduction to Ill-Structured Problems - Wikiversity*. URL: https://en.wikiversity.org/wiki/Introduction{_}to{_}Ill-Structured{_}Problems (visited on 06/13/2017).

Williams, John K. (2009). "Introduction to fuzzy logic". In: *Artificial Intelligence Methods in the Environmental Sciences*, pp. 127–151. ISBN: 9781402091179. DOI: 10.1007/978-1-4020-9119-3_6.

Yager, Ronald R. and Dimitar P. Filev (1995). "Essentials of fuzzy modeling and control". In: *Control Engineering Practice* 3.8, pp. 1200–1201. ISSN: 09670661. DOI: 10.1016/0967-0661(95)90104-3. URL: http://linkinghub.elsevier.com/retrieve/pii/0967066195901043.

Yang, Gaiqiang et al. (2017). "A flexible decision support system for irrigation scheduling in an irrigation district in China". In: *Agricultural Water Management* 179, pp. 378–389. ISSN: 03783774. DOI: 10.1016/j.agwat.2016.07.019. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S0378377416302657.

Zadeh, L.a. (1965). "Fuzzy sets". In: *Information and Control* 8.3, pp. 338–353. ISSN: 00199958. DOI: 10.1016/S0019-9958(65)90241-X.

Zadeh, Lofti A. (1988). "Fuzzy Logic". In: *Computer* 21.4, pp. 83–93. ISSN: 00189162. DOI: 10.1109/2.53. arXiv: arXiv:1011.1669v3.

Zhang, Zijun, Yaohui Zeng, and Andrew Kusiak (2012). "Minimizing pump energy in a wastewater processing plant". In: *Energy* 47.1, pp. 505–514. ISSN: 03605442. DOI: 10.1016/j.energy.2012.08.048. URL: http://dx.doi.org/10.1016/j.energy.2012.08.048.

Zhang, Zijun et al. (2016). "Modeling and optimization of a wastewater pumping system with data-mining methods". In: *Applied Energy* 164, pp. 303–311. ISSN: 03062619. DOI: 10.1016/j.apenergy.2015.11.061. URL: http://dx.doi.org/10.1016/j.apenergy.2015.11.061.

Zhuan, Xiangtao and Xiaohua Xia (2013). "Optimal operation scheduling of a pumping station with multiple pumps". In: *Applied Energy* 104, pp. 250–257. ISSN: 03062619. DOI: 10.1016/j.apenergy.2012.10.028. URL: http://dx.doi.org/10.1016/j.apenergy.2012.10.028.

Zodiatis, G. et al. (2016). "The Mediterranean Decision Support System for Marine Safety dedicated to oil slicks predictions". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 133, pp. 4–20. ISSN: 09670645. DOI: 10.1016/j.dsr2.2016.07.014. URL: http://linkinghub.elsevier.com/retrieve/pii/S0967064516302144.

Zulkafli, Zed et al. (2017). "User-driven design of decision support systems for polycentric environmental resources management". In: *Environmental Modelling & Software* 88, pp. 58–73. ISSN: 13648152. DOI: 10.1016/j.envsoft.2016.10.012. URL: http://www.sciencedirect.com.proxy.bnl.lu/science/article/pii/S1364815216308799.