DEVELOPMENT OF A FAULT DETECTION ALGORITHM FOR AN ALTERNATE AEROBIC/ANOXIC CYCLE NITROGEN REMOVAL PROCESS

Dissertation

submitted to and approved by the

Department of Architecture, Civil Engineering and Environmental Sciences
University of Braunschweig – Institute of Technology

and the

Department of Civil and Environmental Engineering
University of Florence

in candidacy for the degree of
Doktor-Ingenieur (Dr.-Ing.) / **Dottore di Ricerca** in
Civil and Environmental Engineering

by

Emanuele El Basri

born 16.03.1986
from Prato (PO), Italy

Submitted on:	27.02.2017
Oral examination on:	08.05.2017
Professorial advisors:	Prof. Stefano Marsili-Libelli Prof. Hermann G. Matthies Prof. Claudio Lubello Prof. Norbert Dichtl

2017

Abstract

An emerging need in the management of wastewater treatment plants (WWTPs) is the availability of efficient fault detection (FD) methods to maximize the performance of the process instrumentation.

This thesis presents a critical appraisal of several differing approaches to the design and testing of FD algorithms to monitor the instrumentation used in the alternate aerobic/anoxic cycles (AC) process for nitrogen removal. In this scheme, the nitrogen removal is achieved by intermittently switching the aeration on and off, thus enabling the ammonia oxidation in the "on" phase and reducing the oxidized nitrogen in the "off" phase, provided that enough organic carbon is available.

The first part of the thesis is devoted to the definition of faults and their parametrization based on simple process indicators. Several factors were selected at the basis of the diagnosis, involving the slope and the timing of the process measurements of the various nitrogen species. Then the faults were classified according to their impact on the process. Gross faults, involving total sensor failure, were then separated from more subtle malfunctions, like drift and process anomalies. Two distinct detection algorithms were developed, separating the anoxic and the aerobic phases. Though the two parts require a separate tuning, both share the same principles: first some low-level checks are performed on the raw signals, in order to discriminate malfunctions like missing data, spikes, anomalous constant measurements and irregular duration of the phases. More sophisticated methods have then been developed to investigate the presence of anomalies that were not detected by the previous screening.

The fault detection problem has thus been treated either in terms of classification problem, testing different algorithms based on classification techniques such as binary trees, support vector machines (SVM) and principal component analysis (PCA), or as a forecasting problem, using an approach based on the Bayesian theory to predict the faulty or normal state of the process based on the previous records.

In the PCA-based algorithm, the classification method was improved with respect to the normal statistical indicators, like the Hotelling's T^2, by introducing variable fault thresholds and a moving window to keep track of the normal variability of WWTP operation due to seasonality, varying loading conditions, etc.

In the beginning, operational data obtained from a municipal plant were used to train the algorithm. However, the information which could be extracted from this dataset was fairly limited, so in order to set-up a more comprehensive data set an AC model was developed starting with the standard Benchmark simulation model with improved

nitrogen kinetics and seasonal temperature variations. Detailed sensor models were also included, so that the occurrence of faults could be totally controlled, both in kind and timing, reproducing the anomalies observed in the operational data.

The performance of the various methods was then assessed comparing the anomalies detected by the algorithm with those actually observed, either in case of actual operational data, or produced by the simulation model. While the great majority of the gross failures were successfully detected by the preliminary screening, differing performances of the subsequent algorithm were obtained, depending on both the data set and the detection method.

A poor detection performance was observed using the operational data, in part due to an insufficient characterization of the fault events and in part to the limited number of signals available. It was observed, instead, that the higher availability of measurements provided by the numerical model enhances the discrimination capabilities of the tested methods, especially for on the cubic kernel SVM, while the moving window PCA (MWPCA) approach and especially the Bayesian predictor, although less effective, results less affected by a change in the combination of diagnostic parameters used.

Apart from the numerical analysis on both the operational and simulation data, the MWPCA method, first developed in Matlab environment, was then ported in LabVIEW and tested on the alternate aerobic/anaerobic nitrogen removal tanks of the municipal wastewater treatment plant of Mantua, Italy.

Acknowledgments

This thesis certainly would not have been possible without the help and support of many people and I will try here to thank them.

First of all, I want to thank my family, Mohamed, Lia and Marco for their love and support and most of all for being there. These have been rough years and it is good to still be all together.

I also want to thank my friends, the old ones that accompany me since elementary school and those met along the way: Ste, Ila, Lollo, Giggi, Leo, Gaia, Beppe, Vero, Mac, Ire, Andre, Alessandra, Vale, Ale&Ale, Bea, Lore, Massi...

Thank you to the board of teachers that three years ago allowed me to join this doctorate and gave me the opportunity experience all of this and to meet so many new people. Thanks to my supervisors, professor Marsili-Libelli who first pushed me to apply for this adventure and following closer my work, professor Lubello and professor Dichtl for the suggestions and professor Matthies for the precious advises and for hosting me in his institute during the months spent in Braunschweig.

Thanks to all the people with whom I shared the desks and the laughs: first of all to Matteo and Giacomo, with whom I started this journey in the realm of the fault detection, my “neighbours” of the LabGOL in Florence, Alessandro, Federica, Francesco, Guido, Luca and the new arrived guys. Thanks to the colleagues of the WiRe, Cosima, the true boss of the institute (no offense prof. Matthies), Elmar, Bojana, Noemi, Oleksy, Jaroslav, Sadiq, Thilo, Ehsan and Joachim.

Thank you to my fellow colleagues of the XXIX doctorate cycle Pina, Dario, Federica, Vito, Andrea, Stefano, etc., who began with me this adventure, and thank to the sempai of the previous cycle, Chiara, Valentina, Ilaria, Giovanni, Irene, Alessandro, Tommaso, etc. that helped me in so many ways during these years.

Finally, I am grateful to all the other people that helped me in any way along the road, forgive me if I forgot to mention you here, it was not intentional.

Emanuele

Table of contents

Chapter 1 Introduction

1.1 Problem statement

Despite the promising theoretical and experimental results in the research field management and control of wastewater treatment (WWT), the methods proposed in literature are usually difficult to deploy on a real plant and therefore rarely employed in practice. Low reliability of sensors and actuators are often indicated as significant barriers between results in obtained in the research and development stages and the industrial practice. In addition, the biological processes such as the wastewater treatments present some characteristics which impedes a straightforward use of simple and non-adaptive control techniques. Some key aspects of WWTs must be considered: first of all, these biological processes are characterized by the simultaneous occurrence of several chemical, biochemical and physical processes characterized by complex causal relationships. On the other hand, however, it should be noted that the biological processes are typically slower than those in the chemical industry, with time scales ranging from the minute scale for the control of the dissolved oxygen (DO) concentration, to weeks or months for the growth of the biomass. This aspect of course lightens the computational burden demanded for the process supervision. The dependency of the characteristics of the biological process from the changing environmental conditions as well as the dependency from the past operations of the treatment processes should also be considered. One of the most influencing agent is given by the seasonal change in the water temperature but there are many others which are more difficult to quantify. In time, the measurement devices have become increasingly cheaper, resulting in an increased availability of data. However, the relationships among the measured variables, the process states and the process performance are not always understood well and many important process variables remain impossible to gauge. For example, the amount of some specific groups of bacteria (e.g. nitrifying bacteria, denitrifying bacteria) as well as the concentrations of some biochemical components concentrations, cannot be measured on-line. Exact knowledge of relationships between directly controlled variables and hidden process states is inherently absent hereby troubling the formulation and acceptance of automated control laws in practice.

Another important aspect to consider is that in contrast to other kind of industries, where it is possible for example to discard low-quality ingredients and it is possible to avoid the sale of the flowed products, all wastewaters have to be treated, irrespective of

their qualities. The contingency of highly polluted sewages must therefore be accounted for in the design phase since not being able to properly handle certain wastewaters can results in economic fines and restriction of licenses in most industrialized countries, and it is of the outmost importance to recognise the occurrence of such event and to discriminate them from the anomalies in the measurements due to a defect of the instrument. In this sense, the existing supervision and control techniques for wastewater treatment systems generally suffer from a lack of robustness that impairs their practical employment, often due to an inadequate or insufficient knowledge of the existing relationships between monitored and controlled variables or a lack of integration of several control loops and the inability to foresee disturbances to wastewater treatment plants.

Therefore, the main objectives of this dissertation are the evaluation, validation and improvement of techniques for process monitoring and diagnosis of wastewater treatment systems. More specifically this thesis will focus on the development and testing of fault detection methods for a particular class of treatment plants where the nitrogen removal efficiency is increased without building new tanks but exploiting the existing facility under an improved process configuration called intermittent cycles or alternated cycles (AC).

It is increasingly difficult to process municipal reject water by means of conventional wastewater treatment plants, given its increasing imbalance between the carbon and nitrogen fractions. There is now a consolidated trend of a decreasing chemical oxygen demand (COD) load and an increasing reduced nitrogen content, mostly in the ammonia form. For this reason, many municipal wastewater treatment plants are being retrofitted by changing their process flow-sheet in order to make nitrogen removal their primary goal. One efficient way of improving the efficiency of existing small-to-medium wastewater treatment plants (WWTPs) is to change their operation into a sequence of alternating aerobic/anoxic cycles by switching on and off the aeration in the oxidation tank (Santinelli et al., 2011; Gracia et al., 2013; Martín de la Vega et al., 2013). An extensive description of the AC process will be given in Chapter 2. The cheapest way of controlling the aeration switch in this kind of plant configuration consists in monitoring the oxidation reduction potential (ORP), whose value is related to the ammonium-N content. The principle is the same as in Sequencing Batch Reactors (SBR) (Kerrn-Jesperson & Henze, 1993; Wilderer et al., 2001), with the difference that in this case the switching is applied to a continuous-flow reactor. During the oxidation phase NH_4^+ is oxidized to nitrite (NO_2^-) and then to nitrate (NO_3^-), and the ORP signal rises to positive values until it stabilizes when all the NH_4^+ has been oxidized. At this point

the aeration is switched off and the denitrification phase begins, reducing, provided enough organic carbon is available, the oxidized forms of nitrogen (NO_2^- and NO_3^-) just produced. ORP signal has been largely used to control the process (Battistoni et al., 2007; Fatone et al., 2008; Guglielmi and Andreottola, 2011; Santinelli et al., 2011), however the limited reliability of the ORP probe has also been reported (de Gracia et al., 2013). With the growing reliability and availability of specific ion probes, the direct measurement of reduced (NH_4^+) and oxidized (NO_x) nitrogen is now a viable alternative to ORP, provided the sensor is kept in good order and that its malfunctioning is promptly detected.

Thus, sophisticated sensors such as Ion-Specific Electrodes (ISE) have become essential for a precise control policy but they require a careful maintenance. For this reason, a monitoring system including a fault detection (FD) method is instrumental for a successful management. This method should also be capable of discriminating a sensor failure to a process anomaly, such as overload or toxic spillage. Hence the need to design a smart FD algorithm with these capabilities.

1.2 State of the art

Generally, the fault detection and isolation (FDI) algorithms can be divided into three main categories (Venkatasubramanian et al., 2003): quantitative model-based methods, qualitative knowledge-based methods and process history-based methods.

Quantitative model-based methods require a deep knowledge of the process and proved to be extremely powerful tools, even though their application is limited to linear or very specific nonlinear problem at most. These kinds of methods can be, for example, represented by one or more observers, i.e. models used to provide an estimation of the relevant signals and allowing to detect abnormal process behaviours when the system-extracted signal is compared to the estimated one. Some successful recent applications can be found in (Nagy-Kiss and Schutz, 2013), where the diagnosis of a nonlinear system such as a WWTP is achieved with a multi-model approach using a set of nonlinear observers, and in (Xu et al., 2014), where the effectiveness of a new approach based on a bank of interval observers performing both fault detection and fault isolation is proven on a continuous-flow stirred-tank reactor (CSTR) case study. However, despite the potentials and the efforts made to make easier their use (Schraa et al., 2006), this kind of methods are still the least popular among the FDI techniques, due to the high system complexity and the lack of good data from which to develop an accurate model.

Qualitative model-based fault diagnostics is performed each time one possesses a fundamental understanding of the process behaviour but not so deep to allow the expression of the input-output relationships in terms of mathematical functions. Typical qualitative model-based methods, widely used in risk assessment studies, are the so-called fault trees i.e. logic trees that relate primary events with hazards. During the construction process the cause-effect relationships are derived by asking questions, to workers and experts in general, on what could produce the hazard and connecting the answers by logic nodes.

However real processes can rarely be represented as deterministic systems and most of the times the diagnostic task requires a quantitative assessment rather than a qualitative one. For this reason, instead of the model-based approach, that needs a complete a priori knowledge of the process, the vast majority of the FDI approaches rely on methods able to extract that knowledge from the history data, i.e. perform a feature extraction. The features can then be processed by non-statistical data mining methods, such as neural networks (NN), statistical methods, such as principal component analysis (PCA), widely used in the context of WWTP fault detection (Rosén and Olsson, 1998; Corominas et al., 2011; Garcia-Alvarez et al., 2011; Garcia-Alvarez et al., 2012; Villez et al., 2013), or partial least squares (PLS), classification methods such as classification trees and support vector machines (SVM), or by a combination of them (Fuente et al., 2012, 2011; Goode & Chow, 1994; Ruiz et al., 2011; Sainz, 2004).

Most of the methods mentioned are limited by the underlying assumption of a linear relationships between the quantities of interest which, however, especially in the environmental systems, this is not generally true. Thus, over the years many solutions have been proposed to adapt the linear algorithms to the nonlinear process conditions. As shall be seen more extensively in the fourth chapter, one popular strategy is to deal with the nonlinearities by using the so-called kernel functions to map the input space into a different space (called feature space) where the data vary linearly and therefore the linear methods can be successfully employed (Lee et al., 2004; Tong et al., 2013). This is computationally cheap and may help to find an appropriate way to transform the original, nonlinearly related, data, though the choice of the kernel parameters can be difficult and it has a limited applicability for diagnostic purposes, since it does not directly operate on the original measurements but rather on their transformations, the features, for which no explicit meaning is generally available (Jia et al., 2012; Liu et al., 2014).

It has been demonstrated that detection and isolation of anomalies in dynamic systems is possible, however this application conflicts with the assumption of time-

independence of most of the methods used, and therefore returning possibly misleading results, such as an excess of false alarms. Various methods were specifically introduced to account for the autocorrelations in the measurements of dynamic systems, such as the 'dynamic PCA' (DPCA) proposed by (Ku et al., 1995), which investigates the presence of dynamical relationships among the features at different time lags though using a fixed (time-invariant) linear model. However, most of the dynamic processes do not preserve the relationships among the variables but rather they describe gradual changes of these relationships, thus the covariance structure of data varies accordingly. In order to deal with such task some efficient strategies were proposed, considering shorter portions of the available datasets in which the time-invariance assumption still holds and investigating how the information extracted in this portions vary in time. This can be achieved using the so-called 'moving window' (MW) approach, meaning that the covariance matrix is computed at each time instant using the last N samples available as in (Rosén and Lennox, 2001), or similarly introducing a 'forgetting factor', a weighting coefficient decreasing the relative importance of the older samples, as in (Li et al., 2000).

These strategies are at the basis of the adaptive FD methods, where a reference model is computed using a portion of the data and this reference is updated in time (sliding the window or revising the weight of the available samples) according to some acceptance criteria. The underlying assumption is that the abnormal changes in the process occur faster than the updating speed of the reference model. Indeed, if the anomalous event is slow enough to satisfy the acceptance conditions, then its information will result in an undesired adaptation of the model (Lennox and Rosén, 2002; Baggiani and Marsili-Libelli, 2009).

In Chapter 4 an extensive review of the statistical and classification methods used in this work is provided, considering both their theoretical foundations and their application to the FD of the biological systems reported in literature.

1.3 Main results and contributions

The main purpose of this thesis lies is comparing the performances of different algorithms and strategies for the detection of the anomalies occurring in a WWTP that uses the intermittent aeration process configuration to improve the removal efficiency of the nitrogen.

A new detection framework is proposed, separating the detection of the gross and most evident malfunctions, such as sudden spikes and constant signals, from the investigation of the finer faults, i.e. those related to drifts and noise affecting the gauging instruments.

A set of simple checks on the signals was devised to address the first class of faults, and this screening shows very high detection performances, being able to identify the majority of the observed anomalies.

As for the detection of the finer faults, some adaptive methods have been employed, consisting of a statistical approach, based on the Bayes' theorem, and other techniques based on machine learning algorithms such as binary trees, principal component analysis (PCA) and support vector machines (SVM). The theoretical background of the methods and their application to the FD problems shall be found in Chapter 4, and an adaptive version of them, specifically tailored for the characteristics of the specific plant configuration investigated is described in Chapter 5. In the fifth chapter, can be also found the results of the tests performed under different conditions of data availability, both in terms of monitored variables and in terms of variability of the data used for the training. The results show that when trained with a relatively short set of data (less than a year, therefore without considering the whole seasonal variability of the processes) and moderately diversified set of parameters, the PCA-based algorithm displays a good compromise between detection performance and reliability of the fault identification and this is also true in case a small number of observed variable is available, However, if a relatively higher number of signals is available for the training or a larger set of instances covering the whole seasonal variability of the process is provided, the use of a nonlinear SVM method appears as the most sensible choice. On the other hand, the fault detection provided by the Bayesian predictor, although generally identifying a lower number of faulty events, is the method that provides the highest accuracy under all the different combinations of diagnostic parameters chosen, regardless the amount of historic data avaliable.

Chapter 2 The biological wastewater treatment

Wastewater treatment is defined as the ensemble of processes used to remove contaminants from household and industrial sewage, that is to say a water which is no longer suitable for use due to the presence of organic and inorganic pollutants. It consists in a succession of chemical, physical and biological processes in which the undesired compounds are removed from the water by concentrating them in the form of sludge and creating an effluent that can be either returned to the water cycle with minimal environmental issues or reused in the industrial process that generated it. The sludges, by-product of the treatment, are often contaminated with toxic compounds and therefore must be properly treated in order to be safely disposed in special landfills or even reused in agriculture, according to the origin of the sludge and the quality of the treatment.

Wastewater is a term that covers a broad spectrum of contaminated waters coming from different origins but that can mainly be summarized in two categories: municipal wastewater (or sewage) and industrial wastewater. Municipal wastewater is typically generated by the households but it may include liquid waste from small industries and storm water runoff or urban runoff. Households wastewater are rich in fat, proteins, urea and cellulose while runoff waters usually have a different concentration of the same pollutants which add up to some micro-pollutants such as hydrocarbons, pesticides, detergents, etc. It is becoming common, in the developed countries, to separate the household wastewater into blackwater and greywater and treat them separately, so to make appealing the reuse of treated greywaters for public green watering of flushing toilets. Industrial wastewater has a composition that varies according to its origin. Traditional wastewater treatment plants can only treat industrial wastewater that can be qualitatively considered tantamount to domestic sewage. This type of wastewater can be subject to a series of preliminary treatments before discharge to the sewer, in order to remove the pollutants incompatible to the further treatments. In fact, some compounds can result toxic or dramatically interfere with the microbial growth which is the base of the traditional biological treatments. For all wastewater that does not fulfil the requirements to be discharged in the sewer or by their nature would be insensitive to a biological treatment, special treatment to be performed *in situ* in the production place shall be provided.

2.1 Wastewater characterization

The elements composing the wastewater can be divided in a few main categories according to their composition and the effect on the environment if released without a proper treatment.

Table 2.1 Main constituents of a domestic wastewater (Henze, 2008)

Constituent	Source	Effect
Microorganisms	Pathogenic bacteria, virus, antibiotics, worm eggs	Risk when bathing and eating shellfish
Biodegradable organic materials	Oxygen depletion in rivers, and lakes	Fish death, odours
Other organic materials	Detergents, pesticides, fat, oil and grease, colouring, solvents, phenols, cyanide	Toxic effect, aesthetic inconvenience, bio-accumulation in the food chain
Nutrients	Nitrogen, phosphorous, ammonium	Eutrophication, oxygen depletion, toxic effect
Metals	Hg, Pb, Cd, Cr, Cu, Ni	Toxic effect, bio-accumulation
Other inorganic materials	Acids (typically hydrogen sulphide) bases	Corrosion, toxic effect
Thermal effect	Hot water	Changing living conditions for flora and fauna
Odour (and taste)	Hydrogen sulphide	Aesthetic inconvenience, toxic effect
Radioactivity		Toxic effect, accumulation

Generally organic materials are the major pollutants in sewage. Their content is traditionally measured in terms of BOD and COD, respectively acronyms for biological oxygen demand and chemical oxygen demand. The first usually requires a slow procedure while the latter can be quicker but dirty (if mercury is used). With COD, expressed in [mg O_2/L], one indirectly measures the organic content in wastewater by evaluating the amount of oxygen required to chemically oxidise the organic matter using dichromate in an acid solution. It measures the majority of the organic matter and it can be divided in fractions, for example suspended and soluble, useful for further considerations according to the treatment process. It is possible to determine the theoretical COD of a known substance from is oxidation equation, for example for ethanol:

$$C_2H_6O+3O_2 \rightarrow 2CO_2+3H_2O. \qquad (2.1)$$

from which the theoretical COD can be computed as the ratio between the mass of the three molecules of oxygen required and the mass of one molecule of ethanol.

Another indirect measurement of the organic pollution load in the wastewater is the BOD, representing the oxygen consumption per litre needed by the aerobic organisms

to oxidise the organic materials at 20°C in a given amount of time. It is commonly used in its BOD_5 form, where the subscript indicates that the computation of the oxygen consumption is referred to a 5 days (120 hours) period. According to specific constraints of time, accuracy or economical convenience other evaluation periods can be used, computing for example BOD_1, BOD_{25} or BOD_7, however BOD_5 is the standard and when a BOD measurement is reported without subscript the '5' is understood. This research makes no exception and all BOD measurement reported are to be intended BOD_5, unless otherwise stated.

When referring to the treatment capacity of a WWTP, a measure often used is person equivalent (PE). This unit characterize the pollutant load delivered to the treatment facility in a 24 hours period and it is defined both in terms of volume or organic load:

- 1 PE = 0.2 m^3/d (200 L/d)
- 1 PE = 60 g BOD_5/d

These two values are fixed and internationally accepted and should not be confused with the amount of organic pollution delivered by an average person the sewers, the so-called person load (PL). The latter, in fact, represents the average contribution in terms of pollutants and depend on the lifestyle, the households and industrial activities and more generally the socio-economic factors of the catchment area considered and therefore varies from country to country.

Wastewater variations in terms of flowrates and pollutant load are expected on variegated range of time bases and this is an important factor to account for in the design of the treatment processes and their control. A diurnal pattern of the water consumption related to the human activities can be observed (Figure 2.1), typically showing minimum flows in the early hours of the morning, a first peak around lunchtime and a second in correspondence of the dinner time. The amplitude of these variations is mainly related with the size of the community served, with smaller communities having the higher difference between the lower and the upper discharge peak.

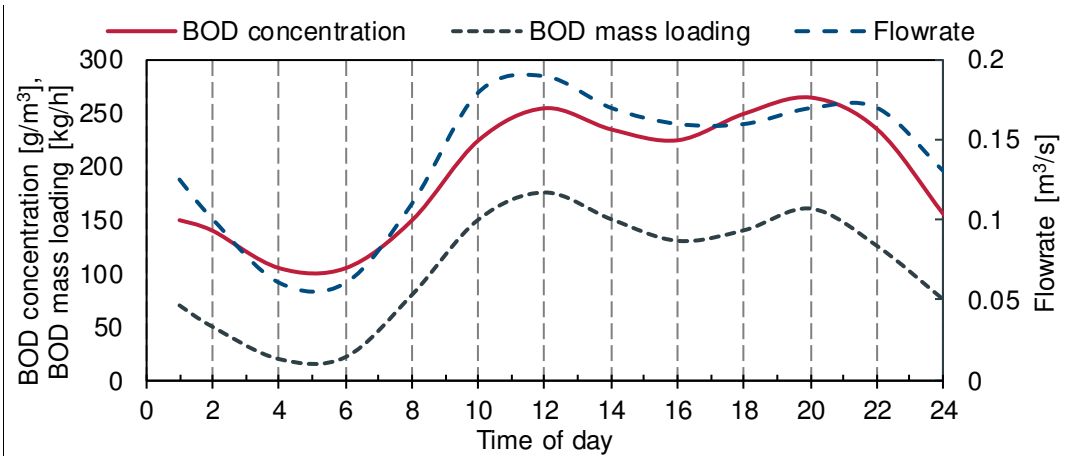


Figure 2.1 Typical hourly variations in domestic wastewater flowrates and concentrations (Metcalf & Eddy Inc. et al., 2004)

Similarly, a weekly pattern can be generally identified in function of the industrial activities and the working days. Seasonal variations of the organic load should also be taken into account when designing a treatment plant, especially in the regions interested by touristic activities focused in specific periods of the year.

A significantly influencing factor for the design of the sewage treatments to implement is represented by the ratio among the various components of the wastewater, for example high nitrogen to carbon ratio may suggest the need of an external carbon supply to favour the denitrification process. The ratios are supposed to be almost constant for a given plant, with values falling in a limited range that should not change at least on a seasonal basis. Industrial discharges, however, are likely to cause discrepancies in these ratios when they are collected together with the domestic sewage, due to the possible releasing of high quantities of pollutants in a short time frame that can cause "shock loadings" to the treatment plant. Moreover, the pollutant content of a sewage and the relationship among its constituents is only partially related to the contribution of the industries and households served, a conspicuous part may also come from rain-washing water and ground infiltration, which are typically events limited in time but may cause significant alterations in the constituents' ratios. The analysis of the deviances of these ratios from their usual or expected values can be used to investigate their source, distinguishing whether they are caused by an actual change in the influent composition or they are just apparent and the anomaly is due to unreliable measurements.

2.2 Operational configuration of a wastewater treatment plant

Wastewater treatment plants consists in a sequence of tanks, generally made of concrete, each with a specific function. One can however distinguish two separate

"production lines" of the treatment: the water line and the sludge line. For what concerns the water line the typical sequence of treatment in an activate sludge plant is composed by some preliminary treatments, deputed to the removal of the heaviest solid materials and oil separation, followed by the secondary treatments, which consist in the actual biological treatment where the suspended organic and pollutant compounds are removed and finally the tertiary treatments which allow to polish the biological treatment before delivering the water to the river body. Another line, devoted to the treatment of the sludge produced during the process, develops parallel to the aforementioned treatments and its main goal is to reduce the water content, the volume and the microbial content in this kind of waste. In the following paragraphs will be presented the typical treatments of a traditional biological treatment plant.

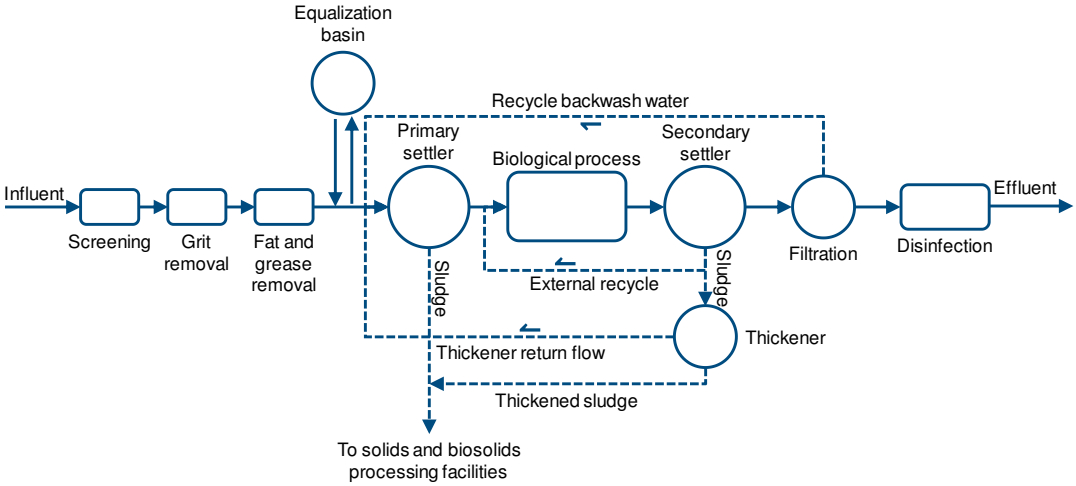


Figure 2.2 Location of physical unit operations in a wastewater treatment plant flow diagram (Metcalf & Eddy Inc. et al., 2004)

2.2.1 Pre-treatments

A series of preliminary treatments, or pre-treatments, is required on the top of all the other treatments in order to remove the gross materials and all the other substances that could damage the efficiency and operational of the following treatment stages.

2.2.1.1 Screening

The objective of the screening is to hold the largest settling and non-settling solids such as branches, dead leaves, rocks, trash or rugs. The screening grid is always installed in the entering channel with a 1:3 slope. This channel widens in correspondence with the grid so that, accounting for the envelope of the grid bars, the water speed in the section remains the same as the entrance speed. Such speed shall be high enough to avoid settling before the greed but at the same time not too low. Typical speed values are between 0.6 m/s and 0.9 m/s. The screening process can be classified either according to the space between the bars of the grid or to the cleaning system used. In

the first case we distinguish between coarse (space between bars 5÷10 cm), medium (2.5÷5 cm) and thin screening (1÷2.5 cm). Most of the plants employ a coarse screening followed by a thinner one. The cleaning system of the grids can be manual, for coarse screening or smaller plants where the amount of retained solid is low, or mechanical in all other cases.

2.2.1.2 Grit removal

The grit removal is mainly provided in case of unitary sewer (black and runoff water) in order to remove all the inorganic suspended materials with a diameter smaller than 0.2 mm, for example sand, grit, stones and broken glass. This treatment is performed in sand or grit chambers, tanks in which the velocity of the water is adjusted to allow the settlement in a reasonable time of all the solid particles characterised by a specific weight higher than water.

2.2.1.3 Fat and grease removal

Fat or grease removal is introduced in the treatment cycle after the screening and the grit removal when the treated fluid presents a quantity of suspended oil and grease such to affect the efficiency of the following treatment, especially the biological ones. In fact, oil compounds have tendency to shroud with a thin film the biological particles limiting their contact with oxygen and consequently their oxidation. In ordinary plant configuration fat and grease, floating on the water surface, are in most part retained by skimmers placed before the primary clarifier entrance while in larger plants a specific fat removal compartment can be required. Here a deeper fat removal efficiency can be achieved blowing air from the bottom of these tank which destabilises the suspended fat particles forming a foam on the surface that can be easily collected.

2.2.2 Primary treatments

The primary treatments are designed to reduce the variations in the organic load of the influent and at the same time possibly reduce it by gravity separation in specific tanks. These primary treatments alone are able to reduce the BOD_5 of the treated wastewater up to 30 %.

2.2.2.1 Flow equalization

The biological treatments are more efficient under uniform flow conditions therefore in case the influent has a highly variable flowrate or the pollutant load is expected to have strong fluctuations, typically because of overloads or in correspondence of storm events, the incoming sewage is routed to a large tank, the equalization basin, designed

to damp the hydraulic peaks and the organic load and make them sufficiently constant. Strong mixing and aeration are used to favour the dilution and homogenization of the water as well as to avoid septic conditions

2.2.2.2 Primary settling

During the primary sedimentation stage the wastewater is retained in open large tanks called primary settlers or clarifiers. In these quiescent settling basins are used to further separate the solids from the liquid suspension by gravity separation in an environment where the dispersion due to turbulence is reduced. For this reason, the tanks are designed to be quite shallow, but their height is not less than 1.80 m, this to avoid that the wind lift the already settled sludge, and with a length and width such that the influent is retained for at least 2-3 hours but without creating dead spots in the corners where putrescence phenomena can occur. The settled sludge is collected at the centre of the tank and sent to the "sludge line" for further treatment.

2.2.3 Secondary treatments

Secondary treatments represent the most strictly chemical-biological part of the treatment process. Here the biodegradable organic compounds are demolished and nutrient removal takes place with the help of specific bacterial populations. The sewage at this stage is still turbid due to the colloidal and putrescible suspensions, therefore the organic particles are first oxidised, i.e. made rot proof, and subsequently removed. Another important objective of this stage is the abatement of nitrogen, present either in the form of ammonia ions (NH_4^+) or in the form of nitrites (NO_2^-) and nitrates (NO_3^-) ions. Nitrogen, as well as phosphorous, is the main responsible for water eutrophication. If released in a natural water body the ammonium nitrogen can be toxic and can also generate anoxic phenomena due to the high oxygen consumption required for its oxidation (4.3 g O_2/g N).

2.2.3.1 Oxidation-Nitrification

In the activate sludge plants, given the high concentration of microorganisms and the density of the sewage that does not allow the growth of algae, not all the oxygen necessary to the bacteria development can derive from the interexchange with the atmosphere, and it is therefore necessary to provide the tank with an artificial aeration system for the air supply.

During oxidation of sewage the organic colloidal compounds are flocculated so that they can be easily removed by settling. With aerations (or biological oxidation) the suspended solids incapable of settling and the biodegradable dissolved solids are

converted into sludge that can be separated from the water in the settling stage that always follows this treatment.

The ammonium nitrogen (NH_3) is also oxidised in this stage. This compound usually represents the most relevant source of nitrogen in the influent (60 % of all nitrogen) and is originated from the urea and the biological degradation of the nitrogenous organic substances that form the cellular material of the organisms. This process can be described by the following simple relation:

$$C_5H_7O_2N+5O_2 \rightarrow 5CO_2+2H_2O+NH_3. \quad (2.2)$$

The ammonium is first oxidised into nitrites then the nitrites are oxidised into nitrates thanks to the work of two different aerobic autotrophic bacteria, respectively called *Nitrosomonas* and the *Nitrobacter* (Figure 2.3), using carbon dioxide as carbon source.

(a) (b)

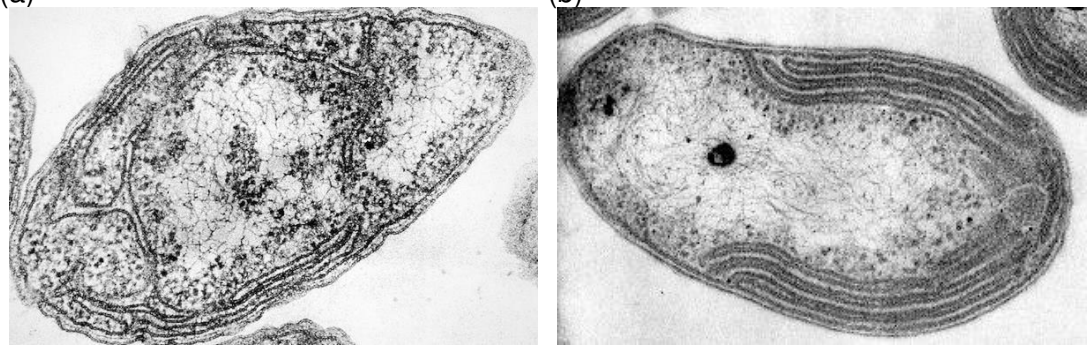


Figure 2.3 The *Nitrosonomas europaea* (a) and *Nitrobacter agilis* (b) bacteria

The oxidation reactions are:

$$NH_4^+ + \tfrac{3}{2}O_2 \rightarrow NO_2+H_2O+2H^+ \quad (2.3)$$

$$NO_2^- + \tfrac{1}{2}O_2 \rightarrow NO_3^- \quad (2.4)$$

The first reaction represents the control stage as the oxidation velocity of ammonium is much slower than the oxidation velocity of nitrites into nitrates, therefore from an operational point of view the objective should be to favour the growth of ammonium-oxidizing bacteria (AOB), the *Nitrosomonas*. The variables that mostly influence this process are the carbonaceous substrate concentration, the level of DO, the temperature and pH (Ruiz et al., 2003). Among the different operational condition that cause the increase of nitrite concentrations due to the inhibition of the nitrite-oxidizing bacteria (NOB) the following are the more relevant (Alleman, 1984; Turk and Mavinic, 1989; Abeling and Seyfried, 1992):

– the concentration of dissolved oxygen: it is usually not preferable to have DO concentration below 1-2 mg/L. AOB have more affinity to the oxygen and for concentrations 0.3 mg/L < DO < 1 mg/L the growth of NOB is difficult and nitrite become the principal by-product of the process;

- the pH level: for alkalinity values of pH > 7 with the increase of pH the AOB activity increases as well, while the growth of NOB decreases. Optimal pH values are between 8.5 and 9;
- the BOD_5/N ratio: this value identifies the fraction of nitrifying bacteria with respect to the heterotrophic bacteria employed for the organic load removal. The optimal value is around 5;
- the hydraulic retention time: for equal temperature AOB can operate with lower HRT than NOB (see Figure 2.4).

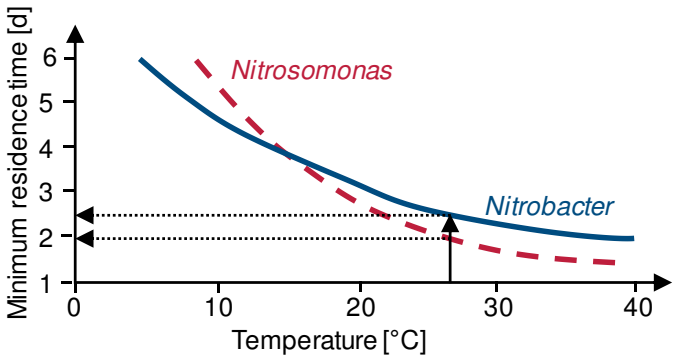


Figure 2.4 Autotrophic growth according to the temperature of the sewage

The nitrification process, however, can grant a substantial abatement of the nitrogen content only if coupled with a denitrification system: once the ammonia is transformed in nitrates it is necessary to complete its transformation into the nitrogen gas form by creating anoxic environmental conditions.

2.2.3.2 Denitrification

The last step of the nitrogen removal process is called denitrification and usually takes place in specific tanks placed between the primary treatments and the aerobic compartment. Here nitrites and nitrates are transformed by the activity of a special class of heterotrophic facultative anaerobes bacterial populations called *Pseudomonas*. In anoxic conditions these bacteria use the carbonaceous substrate of the organic particles present in the sewage for their respiration and perform the cellular synthesis of the oxygen contained in the nitrates according to the reaction:

$$2NO_3^- \rightarrow 3O_2 + 2N_2 \qquad (2.5)$$

The denitrification speed is influenced by the alkalinity of the solution (the optimal range is between pH = 6 and pH = 8), by the temperature and by the characteristics of the sewage: the higher is the amount of rapidly biodegradable organic materials, the quicker is the denitrification process. This implies that the denitrification treatment also contributes to a reduction of the organic load. Theoretically *Pseudomonas* consume 4.5 kg BOD_5/kg N_{den} but only 2.5 kg BOD_5/kg N_{den} are actually used for the respiration

process, while the rest is stocked inside the bacteria and is therefore available for the subsequent oxidation.

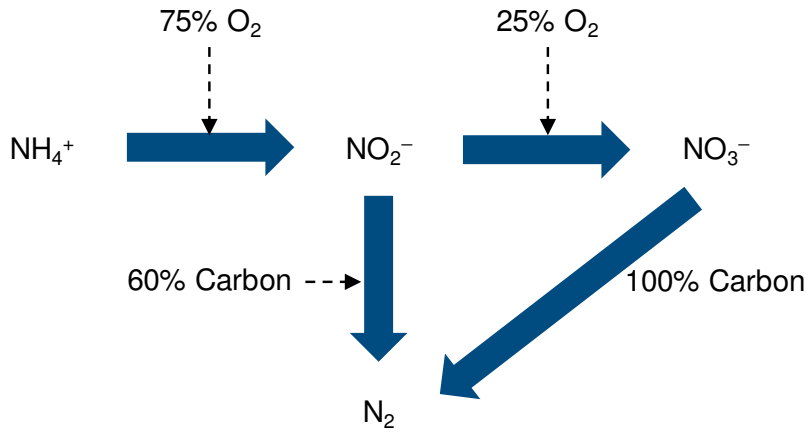


Figure 2.5 Nitrification and denitrification scheme with oxygen and organic carbon consumption percentages (Pambrun et al., 2008)

Usually the nitrification and denitrification processes take place in separate reactors each dedicate to one process. In the most common configuration the denitrification compartment precedes the aerobic one, and this is because this way it is possible to exploit the a so-called 'internal recycle' and 'external recycle' to maximise the denitrification efficiency. The first recycle provides the nitrates obtained from the oxidation of the ammonia in the aerobic tank and the second supplies the anoxic process with the necessary carbonaceous substrate from the secondary settler, without the need of an external carbon supply. Other nitrification/denitrification strategies and plant configuration have been designed in time, in the attempt of maximising the efficiency of the nutrient removal process. Among these a popular one, especially employed in the retrofitting of existing facilities, is the alternated cycles configuration where, as shall be seen in the next chapter, the aerobic and anoxic conditions for the nitrification and denitrification reactions are achieved, by means of an accurate control of the aeration, in the same reactor.

2.2.3.3 Secondary settling

After the biological treatment, the effluent is subject to a settling stage, necessary to separate by gravity the water from the suspended activate sludge of the previous treatment. It is important to reach a neat separation of the liquid and solid phases, so not to compromise the work done so far: the so called active sludge, in fact, contains all the pollutants initially present in the swage, not to mention that the presence of suspended solids in the effluent would impair the clarity of the receiving water body. The water inside the settler must have a slow speed so that the sludge can deposit on the bottom where, as in the primary settler, they are collected. Part of them is sent to for the final

treatment of the “sludge line” and another percentage is pumped, or recycled, to the beginning of the secondary treatments in order to supply the denitrifying bacteria with a sufficient carbonaceous substrate. The clarified water is collected in a circular channel on the borders of the tank and sent to the subsequent compartments for the final treatments.

2.2.4 Tertiary treatments

These final treatments have the objective of completing and perfecting the water treatment process removing possibly present suspended solid or pathogens.

2.2.4.1 Filtration

A filtration session is performed if there is the possibility of solids still suspended in the clarified water: in order to remove these compounds and reduce the turbidity of the effluent, the water can be filtered using textiles that would retain the small particles.

2.2.4.2 Disinfection

This aim of this treatment is to reduce the number of microorganisms, especially pathogens, in the water discharged, in particular when its later use is for bathing or irrigating. The most common way to disinfect the water is using chlorine in the liquid or gas form or as sodium hypochlorite, while the use of chloramine for non-drinkable waters is not advised due to its persistence. Other common sanitation methods involve the use of ozone or ultraviolet (UV) light. In general, the effectiveness of the sanitation procedure is directly proportional to the dosage of disinfectant and the contact time but also the turbidity of the treated water plays a key role especially with the UV treatment.

2.2.5 Sludge treatment

The main purpose of this compartment of a WWTP is to stabilize the organic material by-product of the water treatment process, which is typically characterised by a high concentration of pollutants and organic material, and at the same time minimize the costs associated with its disposal. The water extracted in then generally recycled at the top of the water line for further treatment. The treatment processes can be chemical, biological or physical/thermal and are divided into two main groups:

- separation processes, which are used to separate the solid phase from the liquid;
- conversion processes, which aim to modify the characteristics of the sludge in order to facilitate further treatments.

The treatments can belong either to one or both the above categories.

2.2.5.1 Thickening

The sludge produced by the secondary treatment is rich in water and, since the price of its disposal is computed on the basis of its weight, it is convenient to increase the amount of mass of sludge per volume unit, or in other words, to reduce its water content. In the thickener, a vertical tank, the phase separation is achieved by gravity and the inter-particle water, the most relevant amount of water retained by the sludge flakes, is removed. Here it is possible to obtain a 70-75% sludge humidity reduction. The thickened sludge deposited on the bottom is then collected and sent to another compartment for the drying.

2.2.5.2 Drying

The thickened sludge still contains a non-negligible amount of water that can be naturally or mechanically removed by appropriate treatments. The mechanical drying of the sludge is achieved, for instance, with the use of belt filter presses or by centrifugation.

2.2.5.3 Disposal

The dewatered sludge can be either disposed or reused, according to its characteristics and the normative requirements. When disposed, it is transported to a landfill, sometimes after being incinerated, but sludge with compatible characteristics can find a new use as amendment for agricultural or non-agricultural soils or being employed in composting plants for energy production.

Chapter 3 Characterization of the wastewater treatment structure and operation

In this chapter will be presented the materials used for the research. First the particular plant configuration will be illustrated together with some of the main control strategies. Since the data used are drawn both from operational data gathered at the municipal treatment plant of Mantua (Italy) and from a modified Benchmark Simulation Model (BSM), their respective operational characterization and the main features of their dataset will be described in two separate paragraphs. Finally, the parameterization used for the data will be presented.

3.1 The alternate cycle processes

In this paragraph is described the plant configuration of the so-called “alternate cycles” (AC) nitrogen removal process. This is an operational solution which can be employed in existing plants increasing the efficiency of the removal process without raising the costs.

3.1.1 The process

The need for ever better results in terms of efficiency together with the rising concerns over energy saving brought many small-to-medium plants to invest in the optimization of the nitrification-denitrification process. As already mentioned the biological nitrogen removal process traditionally achieved with a succession nitrification and denitrification steps performed in separate reactors. It is however an ever more frequently adopted solution to increase the nitrogen removal efficiency by changing the current plant configuration into an AC processes configuration, where the nitrification and denitrification steps develop according to time instead of space. This retrofitting is encouraged by the possibility to entirely reuse the existing structures, moreover without the need of building new ones, and by the versatility of this technology towards different plant configuration and incoming influent load conditions.

Alternate cycle is a continuous process that can be realized in a single tank where the aeration and anoxic phases, respectively responsible for the nitrification (ammonium oxidation) and denitrification of nitrogen, are automatically alternated. In other words, it is no longer necessary to continuously supply air to an aerobic compartment. Moreover, it is also possible to completely exploit the oxygen released by the reduction

of the nitrates, as in eq. (2.5) for the oxidation of the organic compounds, reducing the amount of water to supply during the aerobic phase. Since by using this configuration one can avoid to have different sections performing the anoxic denitrification and the aerobic oxidation separately, the recycle of the mixed liquor from the oxidation can also be avoided, resulting in further energetic savings and simplifying the management.

A simplified mathematical model of the process can be found in (Battistoni et al., 2003) under the following assumptions:

- a zero-order kinetic for nitrification and denitrification;
- negligible nitrate concentration in the influent, whose consequence is that all the incoming nitrogen is assumed to be in the ammonia form;
- the incoming ammonia volumetric loading is negligible (three times lower) when compared to the oxidation capability of the autotrophic organisms.

Under these assumptions a simplified mass balance is found for both ammonia and nitrates either for the aerobic (eqs. (3.1)-(3.2)) and the anoxic phase (eqs. (3.3)-(3.4)).

$$\mathrm{NO}_x^-\text{-N} = K_n \cdot X \cdot t \tag{3.1}$$

$$\mathrm{NH}_4^+\text{-N} = \mathrm{NH}_4^+\text{-N}_{t_1} + \frac{\mathrm{NH}_4^+\text{-N}_{in} \cdot \dot{V}}{V_{ox}} \cdot t - \left(\frac{\mathrm{NH}_4^+\text{-N}_{t_1} \cdot \dot{V}}{V_{ox}} + K_n \cdot X \right) \cdot t \tag{3.2}$$

$$\mathrm{NO}_x^-\text{-N} = \mathrm{NO}_x^-\text{-N}_{t_0} - \left(\frac{\mathrm{NO}_x^-\text{-N}_{t_0} \cdot \dot{V}}{V_{ox}} + K_d \cdot X \right) \cdot t \tag{3.3}$$

$$\mathrm{NH}_4^+\text{-N} = \frac{\dot{V}}{V_{ox}} \cdot \mathrm{NH}_4^+\text{-N}_{in} \cdot t \tag{3.4}$$

$$tc = \left(\frac{K_n \cdot X}{K_n \cdot X - \frac{\mathrm{NH}_4^+\text{-N}_{in} \cdot \dot{V}}{V_{ox}}} \right) \cdot \frac{\mathrm{NO}_x^-\text{-N}_{\max}}{K_d \cdot X} = \frac{\mathrm{NO}_x^-\text{-N}_{\max}}{\frac{\mathrm{NH}_4^+\text{-N}_{in} \cdot \dot{V}}{V_{ox}}} \tag{3.5}$$

where

NO_x^--N = nitrogen concentration in form of nitrites and nitrates ions [mg/L]

NO_4^+-N = nitrogen concentration in form of ammonia ion [mg/L]

K_n = maximum nitrification constant [d^{-1}]

K_d = maximum denitrification constant [d^{-1}]

$\dot{V}$ = influent flow-rate [$m^3 \cdot d^{-1}$]

V = volume of the reactor [m^3]

X = suspended solids in the mixed liquor [mg/L]

t = time [d]

tc = length of the cycle (sum of the aerobic and anoxic phases) [d]

and the meaning of the subscripts used is:

in = influent

t_0 = beginning of the anoxic phase

t_1 = beginning of the aerobic phase

These equations are used by (Battistoni et al., 2003) to predict the variations in the ammonia and nitrate concentrations in the reactor or in the influent. The total time of the cycle *tc* is derived in the ideal situation in which it can be only related to the kinetic constants and the influent characteristics. It is clear that according to this model the duration of the cycle depends on the volume of the reactor and the characteristics of the influent while the ammonia and nitrate concentrations are related only to the kinetic constants K_d and K_n.

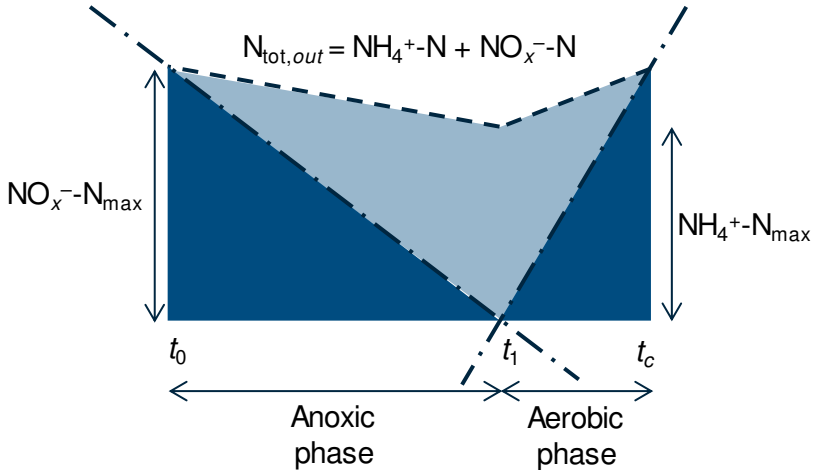


Figure 3.1 Nitrogen forms during an alternate cycle (Battistoni et al., 2003)

3.1.2 The performances

The intermittent aerobic-anoxic process was experimentally employed first in pilot-scale plants and subsequently in real plants, where the performances were constantly monitored in order to compare them with the traditional configuration (Battistoni et al., 2006; Dairi et al., 2010).

3.1.2.1 Nitrogen removal

AC process showed higher nitrogen removal performances with respect to the traditional processes. The total nitrogen removal efficiency achieved can be up to 90% (Nardelli et al., 2009), and this is mainly because all the nitrified ammonium to be denitrified is already inside the reactor, decreasing the nitrate output of the aerobic compartment of at least 30%.

The potentialities of the treatments in the tests of are evaluated by (Nardelli et al., 2009) according the balance:

$$\mathrm{LTN}_{den} = \mathrm{LTN}_{in} - \mathrm{LTN}_{qw} - \mathrm{LTN}_{out} \tag{3.6}$$

where LTN is the total nitrogen mass loading denitrified (subscript *den*), in the influent (subscript *in*), in the waste activated sludge (subscript *qw*) and in the effluent (subscript *out*) all measured in [kg/d].

Using the notation LTN_{nit} = total nitrogen mass loading nitrified; $\mathrm{LNO}_x\text{-}\mathrm{N}_{in} = \mathrm{NO}_x\text{-N}$ mass loading in the plant influent; $\mathrm{LNO}_x\text{-}\mathrm{N}_{out} = \mathrm{NO}_x\text{-N}$ mass loading in the plant effluent; LTKN_{in} = total Kjendahl nitrogen mass loading in the influent; LTKN_{ras} = total Kjendahl nitrogen mass loading in the return activated sludge and $\mathrm{LN}_{nb,org,out}$ = non-biodegradable organic nitrogen mass loading in the influent, all measured in kg/d, the performances were assessed evaluating the indicators of:

the nitrification efficiency of all the incoming nitrogen E_n:

$$\mathrm{E}_n\,[\%] = \frac{\mathrm{LTN}_{nit}}{\mathrm{LTN}_{in}} \cdot 100 = \frac{\mathrm{LTN}_{den} + \mathrm{LNO}_x\text{-}\mathrm{N}_{out} - \mathrm{LNO}_x\text{-}\mathrm{N}_{in}}{\mathrm{LTN}_{in}} \cdot 100, \tag{3.7}$$

the removal efficiency of all the incoming nitrogen E_d:

$$\mathrm{E}_d\,[\%] = \frac{\mathrm{LTN}_{den}}{\mathrm{LTN}_{in}} \cdot 100, \tag{3.8}$$

the nitrification efficiency refed to the nitrogen that is actually nitrified E_{nn}:

$$\mathrm{E}_{nn}\,[\%] = \frac{\mathrm{LTN}_{nit}}{\mathrm{LTKN}_{in} + \mathrm{LTKN}_{ras} + \mathrm{LTN}_{qw} - \mathrm{LN}_{nb,org,out}} \cdot 100, \tag{3.9}$$

and the removal efficiency of the nitrogen that is actually denitrified E_{dd}:

$$\mathrm{E}_{dd}\,[\%] = \frac{\mathrm{LTN}_{den}}{\mathrm{LTN}_{den} + \mathrm{LNO}_x\text{-}\mathrm{N}_{out}} \cdot 100. \tag{3.10}$$

3.1.2.2 Costs savings

The energetic saving is probably the most interesting advantage of the AC configuration. It comes from the plant optimization achieved in three different aspects: first the possibility to reduce the amount of oxygen to supply exploiting the oxygen of the nitrates to perform the oxidation of the organic compounds; second the switch from a continuous aeration system into an intermittent one and finally the economic saving due to the fact that the internal recycle is no longer needed. The AC allows an energy saving in the range 15-30% with the respect to a traditional plant configuration.

Some tests were also performed to investigate the impact of the AC process in the sludge production. The assessment, conducted in plants with specific characteristics such as the presence of a filtration system at the end of the sludge line, the absence of

storm water inputs and sludge from third party (e.g. landfill leachate), revealed a reduction of both wet and dry sludge between 15% and 40%. This reduction can be explained considering the aerobic/anoxic stress of the biomasses, which produces the diminishing of their yield coefficients, i.e. the production of new biomass. It is also worth notice that these tests confirmed that the alternated processes configuration does not affect the settling capability of the sludge, which results equal or even higher with respect to a traditional activated sludge plant configuration.

Another advantage of the AC retrofitting of existing plants is represented by the biological removal of phosphorus, in particular for influents characterized by higher rapidly biodegradable COD values. In fact, having the nitrification and denitrification process happening in the same reactor favours the growth of phosphorus accumulating organisms (PAOs) to a point that sometimes it is possible to avoid the addition of chemicals for the phosphorus precipitation.

3.1.3 The control

The key factor in the feasibility and convenience of the AC retrofitting of a plant is represented by the automatic control system responsible of the switching between the two phases. Most of the controllers, especially in smaller plants, regulate the activation of the aerators (which create the aerobic conditions) or the mixers (to uniform the concentration during the anoxic phase) by elaborating the profiles of two commonly measured quantities, the DO and the oxidation-reduction potential (ORP), which are linked to the chemical-physical and biological phenomena. In particular, the flex of the dissolved oxygen profile indicates the end on the aerobic phase and the flex of the ORP profile during the anoxic phase is used as an indicator of the end of the denitrification process (see Figure 3.2).

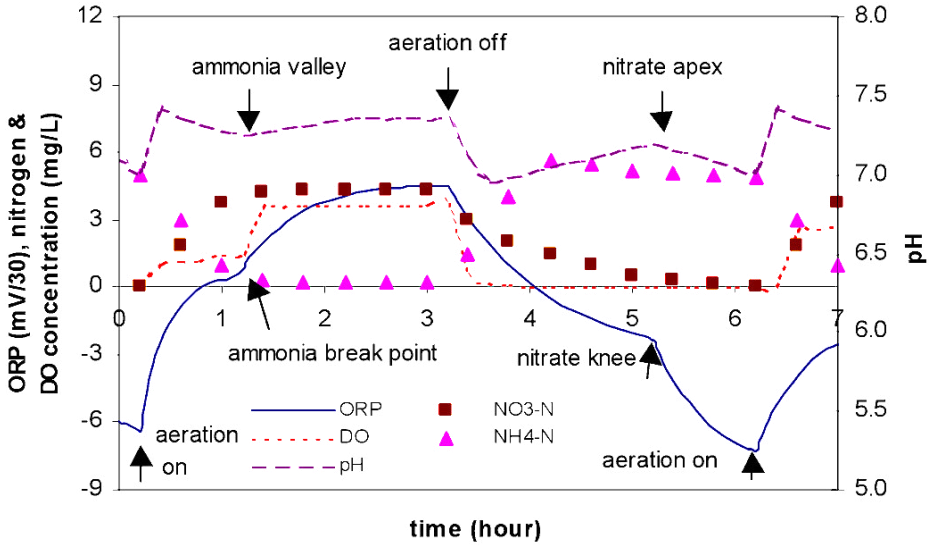


Figure 3.2 ORP, NH_3-N, NO_3^--N, DO and pH profrfiles in a nitrogen removal bioreactor (Olsson et al., 2005)

The first condition, the ammonia breakpoint, is a usually sharp rise of the DO due to the depletion of ammonia and is also observable in the ORP profile, normally in the range 60 - 150 mV, while the nitrates breakpoint, or 'knee', usually happens in the – 40- – 60 mV (Wareham et al., 1993; Héduit et al., 1996), a higher range than the one observed in the traditional plant configurations.

The pH in the reactor also varies periodically but these variations are more associated to the biochemical reactions that take place in the two phases than to the DO concentrations. Nitrification is generally characterised by higher values of pH, its profile first suddenly increases when the aeration is switched on, than very gently decrease until the oxygen reaches the ammonia breakpoint. This point in the pH signal is called the 'ammonia valley'. After this a rather slow but constant increase of the alkalinity is observed until the aeration is turned off again. When the aeration phase ends the pH at first rapidly drops, rising again until the nitrate breakpoint is hit: this is the so called 'nitrate apex', which marks the end of the denitrification process. In the time between the breakpoint and the activation of the blowers a slight decrease of the pH can be observed. These considerations over the possibility to infer the nitrification and denitrification processes using the measurements of instruments usually already present in the reactors, led to several, often patented, control systems.

As for what concerns the control strategies based on the alkalinity and the oxidation-reduction potentials, they can be split into two categories according to whether they rely on the absolute value of the measurements or on the detection of the aforementioned bending points.

In the first case the idea is to set a minimum and a maximum value for either the pH or ORP and switch the phase whenever the threshold is exceeded. This kind of control strategy however, can result difficult to implement especially with respect to the ORP, since its measurement can be subject to significant drifts and it strongly depends on the initial treatment of the platinum probes used, making the threshold found quite plant-specific. The alkalinity measurement instead is prone to less instrument-related uncertainties, nevertheless the pH mostly depends on the influent concentration and can therefore be subject to major changes in time.

Since implementing control strategies that are based on the absolute values of the measurements results difficult dealing with a fluid which is expected to have strong variations in its constituents, it makes sense to use the previously illustrated relations among pH, ORP and the biochemical processes of interest. Therefore, pattern recognition and feature extraction techniques have been used to identify the bending points of ORP ('ammonia breakpoint' and 'nitrate knee') and those od pH ('ammonia

valley' and 'nitrate apex') and set the activation and deactivation of the aerators accordingly. It is however a difficult task to determine these points, especially on the ORP curve, during overload episodes or in presence of other strong variations affecting the influent, so control strategies that compensate for this have been developed.

One particularly popular in the north of Italy is outlined in Figure 3.3 and described in (Battistoni et al., 2003; Nardelli et al., 2009). It essentially integrates a control based on the previously described relations between the nitrification/denitrification process and the pH/ORP profiles in time, which are no longer reliable in case of over-aerated or overload plant, with two additional controllers that, based on the simplified model of equations (3.1)-(3.5), compute the maximum and minimum duration of the phases, and ORP and DO maximum and minimum set-points.

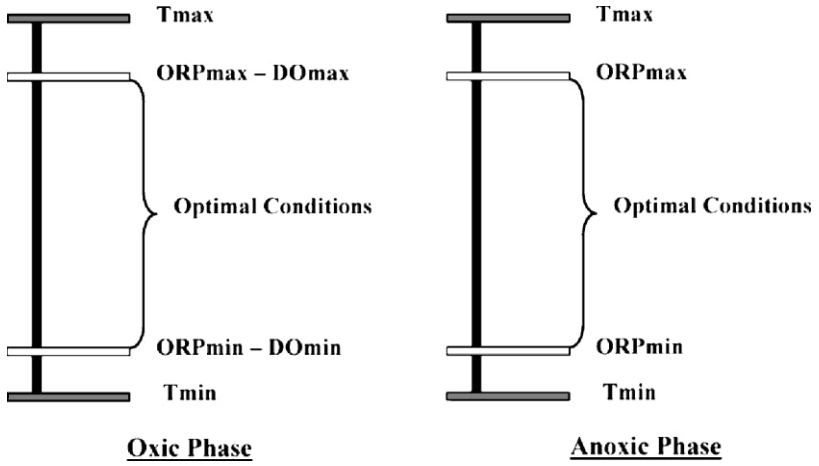


Figure 3.3 Example of AC control logic scheme (Nardelli et al., 2009)

Despite their popularity, the control strategies presented remain based on indirect measurements of the nitrification/denitrification processes and the by using them the actual state of the process can only be inferred. This is the reason that brought the larger plants to install ammonia and nitrate sensors to better monitor and control the removal process, but nowadays this tendency is followed also by smaller ones since the probes became cheaper. A well-studied control strategy is the one employed in the *BioDenipho* plant configuration, which in is former version (known as *BioDenitro*) consisted in an activated sludge process where the nitrification/denitrification was achieved by intermittently changing the flow path through two parallel reactors also intermittently aerated. This strategy allowed for two main phases and a series of other combinations designed to tackle specific operational conditions.

In Figure 3.4 it is shown a four-phases strategy for the *BioDenipho* configuration, where the addition of an anaerobic reactor before the nitrification/denitrification tanks allow the phosphorous removal and marks the difference with the *BioDenitro* process.

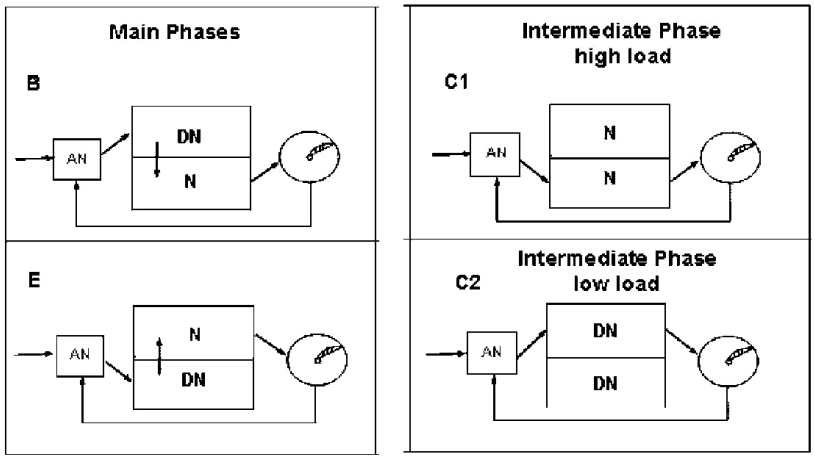


Figure 3.4 A simple four-phase *BioDenipho* process (DN: denitrification; N: nitrification). B and E are the two main phases and C1 and C2 are the two intermediate phases (Olsson et al., 2005)

A more sophisticated control strategy is shown in Figure 3.5 and is composed of six phases. The transition from one phase to the other obeys the following rules:

Rule 1: Transition from phase A to phase B takes place when the ammonia concentration in reactor R1 reaches NH_{max};

Rule 2: Transition from phase B to phase C takes place when the nitrate concentration in reactor R1 reaches NO_{min};

Rule 3: Transition from phase C to phase D takes place when the ammonia concentration in reactor R2 reaches NH_{min};

Rule 4 to 6 regulating the transition from D to E, E to F and F to A mirror the ones above.

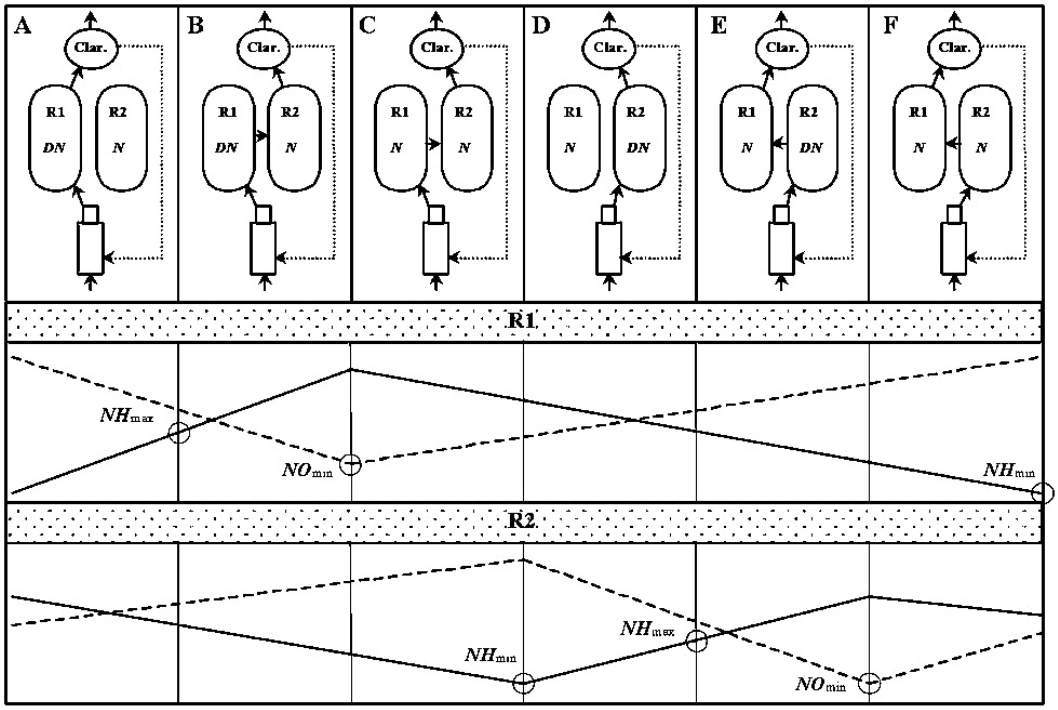


Figure 3.5 Flow path, aeration status and typical variations of ammonia and nitrate in a six-phase *BioDenipho* process cycle. N: nitrification; DN: denitrification; solid and dashed curves are respectively the ammonia and nitrate nitrogen concentrations (Olsson et al., 2005)

The same concerns expressed before on the use of absolute values of pH and ORP to set the thresholds that regulate the activation and deactivation of the blowers apply also when using direct measurements of the nutrient concentrations. For this reason, the idea of 'criteria function' was developed, expressing the lower thresholds not as constant values but dynamically determining them based on the reactor status as for example in (3.11):

$$\begin{aligned} \mathrm{NO}_{min} &= \alpha \cdot \mathrm{NH}_4^+\text{-N} + \beta, \\ \mathrm{NH}_{min} &= \gamma \cdot \mathrm{NO}_3^-\text{-N} + \delta \end{aligned} \tag{3.11}$$

where α, β, γ and δ are fixed parameters and NH_4^+-N and NO_3^--N are the measured concentrations. These 'criteria functions' can as well have a nonlinear form, and for example, with reference to Figure 3.5, they can make such that if the first reactor receives a nitrogen overload in phase B, then the NO_{min} is adjusted and raised so that the denitrification phase can end earlier and leaving more time for to oxidise the ammonia.

3.2 Analysis of the operational data

The data used in this work were drawn from the operational records of a WWTP treating a mixed domestic/industrial effluent from the city of Mantua, in north-east Italy. The plant with primary sedimentation, has a hydraulic capacity of 100,000 PE and an organic loading capacity of 40,000 PE and presents four biological reactors working in parallel but with common sludge recycle from the secondary settler. It is managed by TEA s.r.l. and in all four tanks the nitrification and denitrification processes are achieved by adopting the intermittent aeration process configuration. In the period interested by the study a timed logic controlled the activation/deactivation of the blowers for three of the tanks, while in the other the switch between the aerobic and anoxic phase was decided by the commercial process controller named OSCAR ('optimal solutions for cost abatement in nutrients removal', ETC-ENG s.r.l., Trento, Italy).

This study focuses on the data collected in the reactor managed via the OSCAR process controller. These data consist in the measurement produced by an ANISE (Hach-Lange, GmbH, Dusseldorf, Germany) probe composed by an 'ammonium-N ion-specific electrode' (AISE) and a 'nitrate ion-specific electrode' (NISE) and a S::can (Scan Messtechnik GmbH, Vienna, Austria) probe measuring ammonia, potassium, pH and temperature.

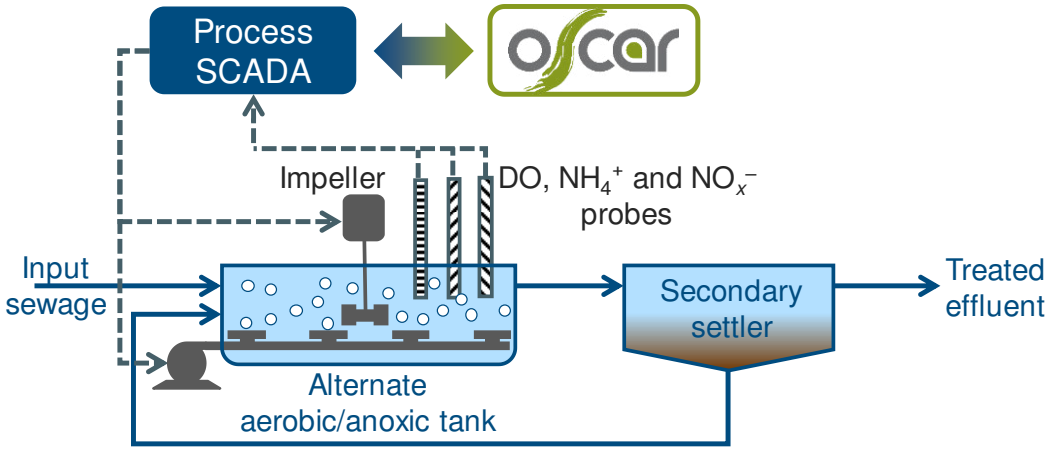


Figure 3.6 Process control scheme: the process SCADA manages the field measurements and actuators, serving as an interface between the process and the OSCAR controller

The measurements provided by the S::can probe spanned a relatively short period compared to the others, from 19th July to 27th November 2012 (Figure 3.7), and presented a low quality.

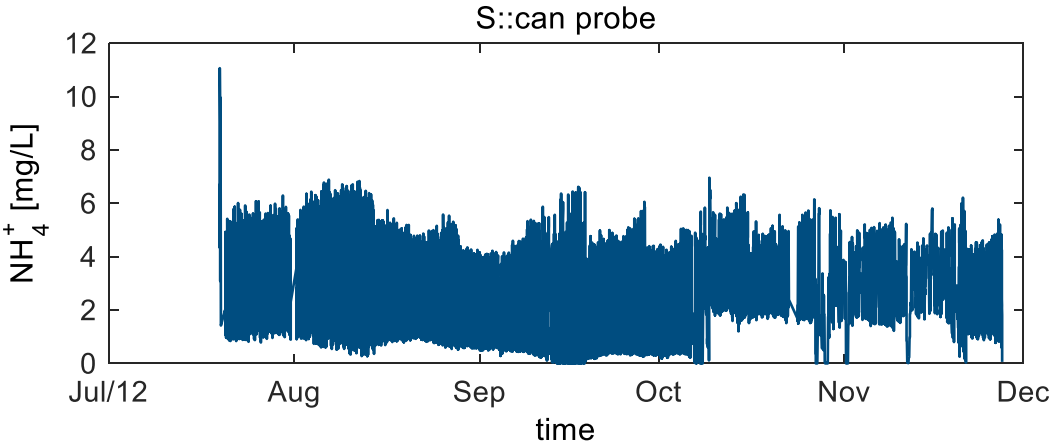


Figure 3.7 The concentrations measured by the NH_4^+ S::can probe

The measurements of the ANISE probe, instead, although consisting in the sole concentration values measured by the ammonia and nitrate probes, were available from the 1st July 2012 to the 22nd October 2013 (Figure 3.8). However, the data actually used for this work span from the 18th Sept. 2012 to the 7th May 2013. The starting time has been chosen so to begin the test right after the cleaning of the probes, due to the fact that the signal has a different behaviour accordingly to the state of the probes as we shall see next. The data after 7th May are instead in most part incomplete and show an erratic behaviour, therefore they were not used.

Some further consideration shall be made concerning the differences between the measurements provided by the two probes installed. It can be observed, for instance, that the sensors although installed in the same reactor, present significantly different concentration profiles. The ANISE probe measures concentration values that varies between 2 and 6 mg/L in the summer period and maximum values that raise to 8-9 mg/L

in wintertime. The S::can probe, instead, shows more irregular concentration behaviour that varies between 1 and 6 mg/L at the end of July 2012 but the range widens afterwards.

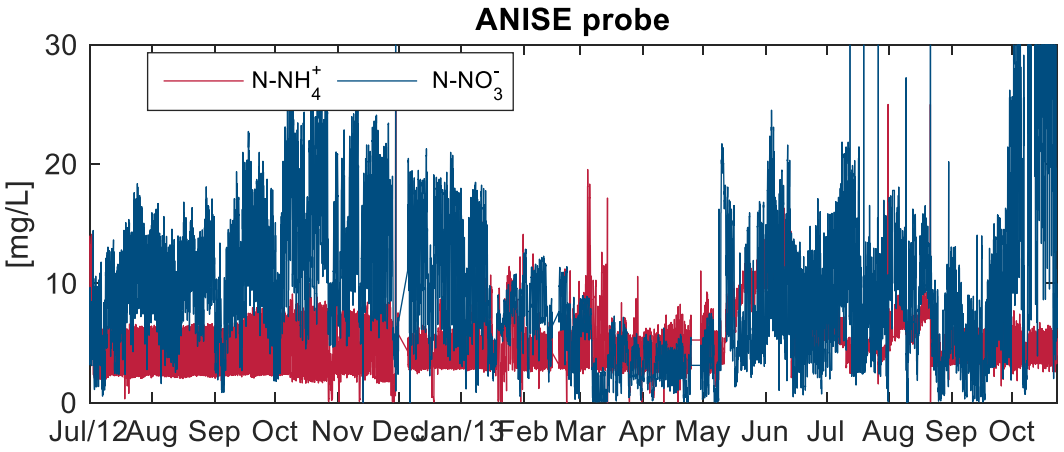


Figure 3.8 The concentrations measured by the NH_4^+/NO_3^- ANISE probe

A closer comparison of the two signals also reveals other discrepancies. As mentioned the range of the measurements provided is different, with those provided by the S::can probe usually 2-3 mg/L lower than those provided by ANISE. This is probably due to a suboptimal calibration and maintenance of the two instruments. The time labels of the measurements in the offline dataset used are also different. While the data stamp of the ANISE measurements resulted aligned with the one of other measurements of the plant a problem in the data acquisition software anticipated the time label the S::can probe measurements by almost 10 minutes (Figure 3.9).

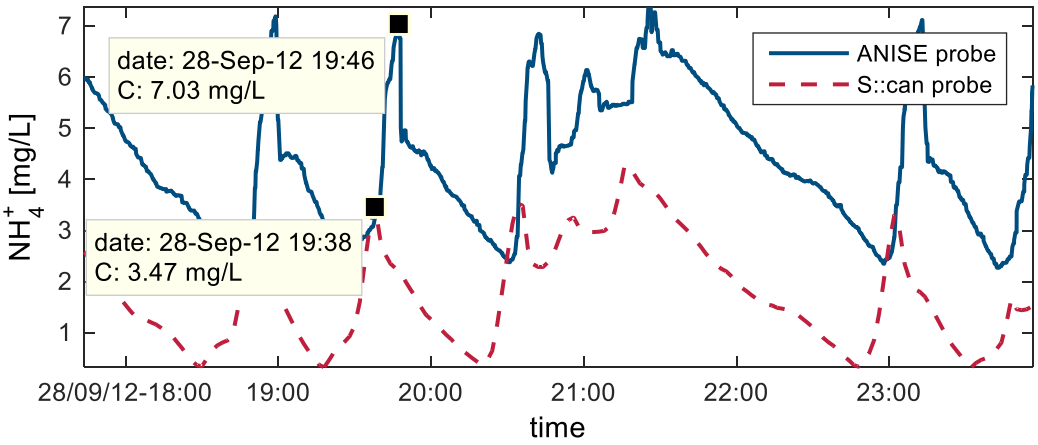


Figure 3.9 Differences in the measured signals of the two probes

These considerations brought to the decision of discarding the S::can probe measurements, therefore the following considerations only apply to the measurements provided by the ANISE probe.

The dataset available presents an extremely variable sampling interval (Figure 3.10a) which is between 10 and 70 seconds, with mean value at 30 s in the first part (Figure 3.10b) and between 60 and 70 s in from February 2013 to the end of the dataset.

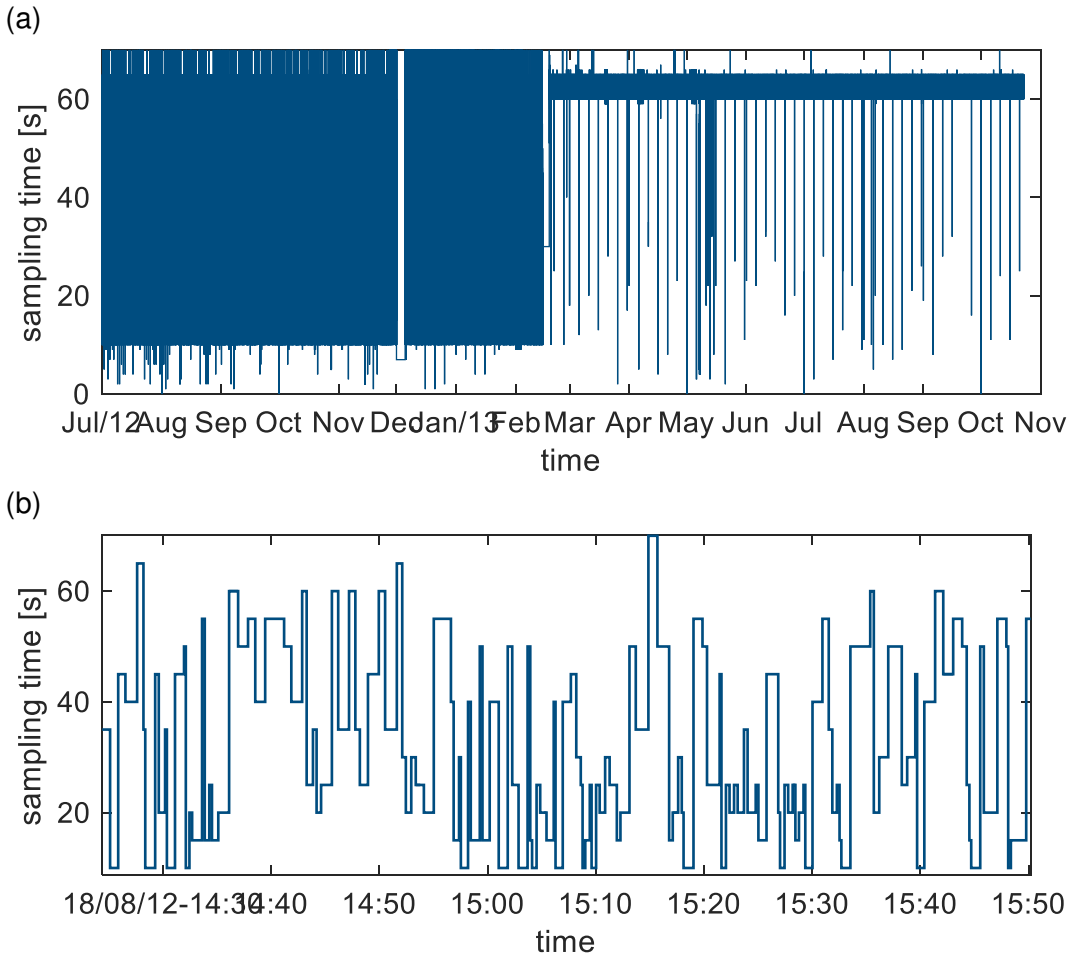


Figure 3.10 Example of irregular sampling of the ammonia concentrations (ANISE probe)

It should be also noticed that the data-logger shows an error, revealed by a negative sampling time, whenever the time switches between daylight saving time (DST) and summer time (Figure 3.11).

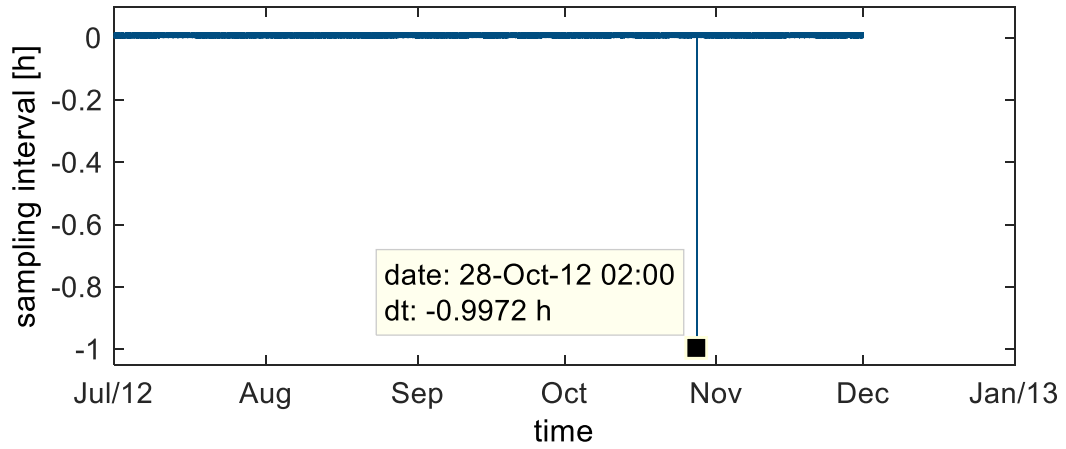


Figure 3.11 Example of negative sampling time revealing the data logging error when switching from summer time to DST

A particular diagnostic feature observed in both the ammonia and the nitrate signals is the ripple caused by water movements produced by the mixer. When the probe is clean the wavelets caused by the impellers hit the sensor and cause a considerable ripple,

whereas a fouled probe is shielded from the eddies and produces much less ripple. Figure 3.12 compares the signal of the NH_4^+ probe in the clean and foul cases.

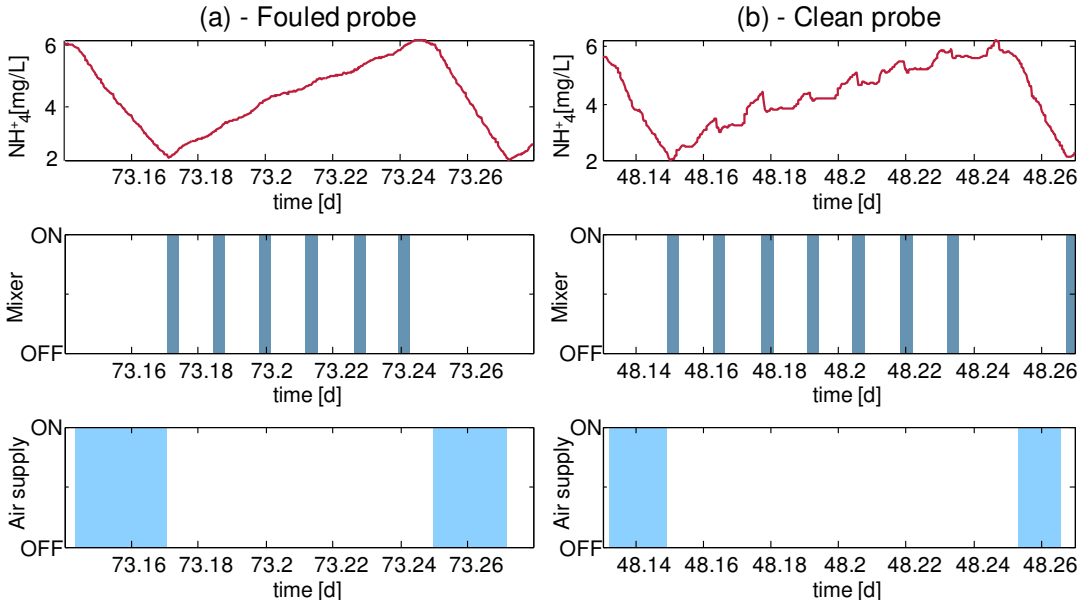


Figure 3.12 The ripple during the anoxic phase is an indicator of the progressive fouling of the probe. The left plot (a) is from a foul probe (almost no ripple), whereas the right plot (b) shows a considerable ripple

The analysis on the operational data of the Mantua plant revealed that a calibration and cleaning phase of the instrumentation are scheduled almost every 14 days and required up to 2 hours. It can be noticed in Figure 3.13 that the shielding effect of the fouling on the nitrate and ammonia signal is generally different, however it was not possible to establish which signal is more affected by the fouling.

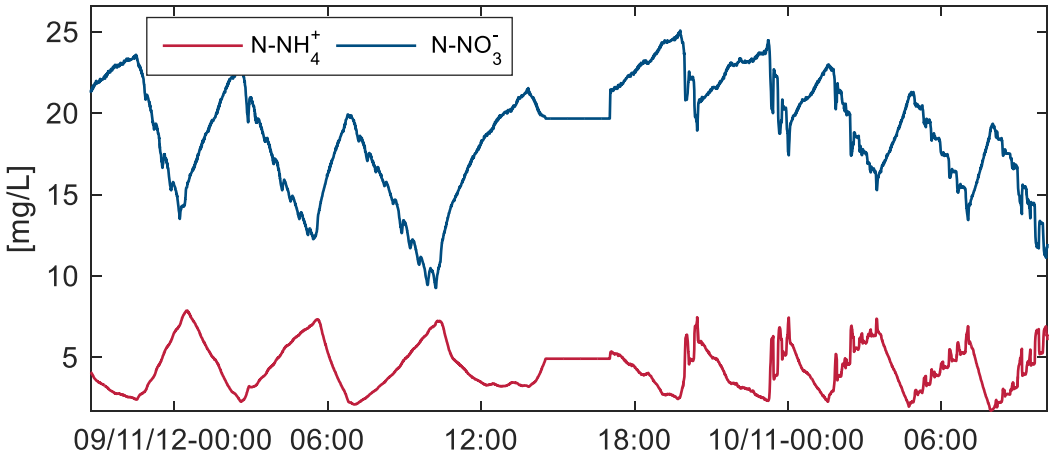


Figure 3.13 A particular of the measurements signals before and after the clean and calibration phase, a more disturbed signal after the procedure denotes a clean probe which is more sensitive to water buffeting caused by the mixer

Other anomalies of the nitrate and ammonia signals emerged from the analysis of the operational data. Their characterization and exhaustive description is demanded to a following paragraph for a more direct comparison with the anomalies introduced in the numerical model used to generate synthetic data and a straightforward connection to the diagnostic tools introduced.

3.2.1.1 Blower switching logic

Another signal retrieved from the instrumental data is the one governing the activation/deactivation of the blowers. This will be of extreme importance because it will be used to identify the aerobic and anoxic state of the process. The Mantua plant manager provided two aeration control signals, respectively 411i and 411, where the number identifies the reactor: the first governs the main aeration system while the other seems to manage a supplementary aeration system intervening in specific conditions, as shown in Figure 3.14. For the purpose of the determining whether the current state is in the aerobic or anoxic phase the signal of blower 411i was considered sufficient.

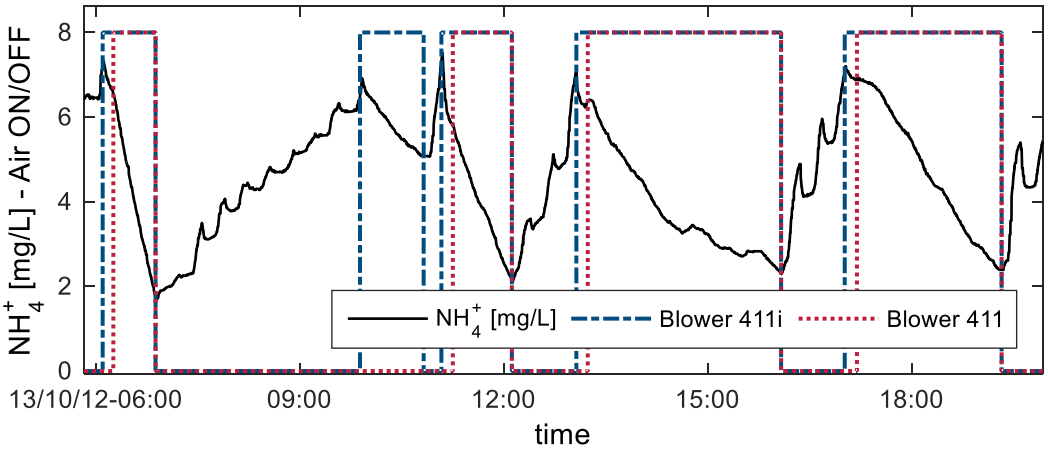


Figure 3.14 Comparison between the blower activation signals 411 and 411i in the municipal WWTP of Mantua

From the I/O signal of the blower it is possible to draw some considerations about the state of the system in the aerobic and anoxic phases. Examining for instance the length of both the nitrification and denitrification phase one can identify a threshold that separates the possibly suspect phases, those that lasts the most, from the reasonably normal ones. In general, the aerobic phases are characterised by longer durations, between 1 and 3 hours, but as can be seen in Figure 3.15 many extensive aeration periods were registered, with aerobic phases that sometimes outlasted the day.

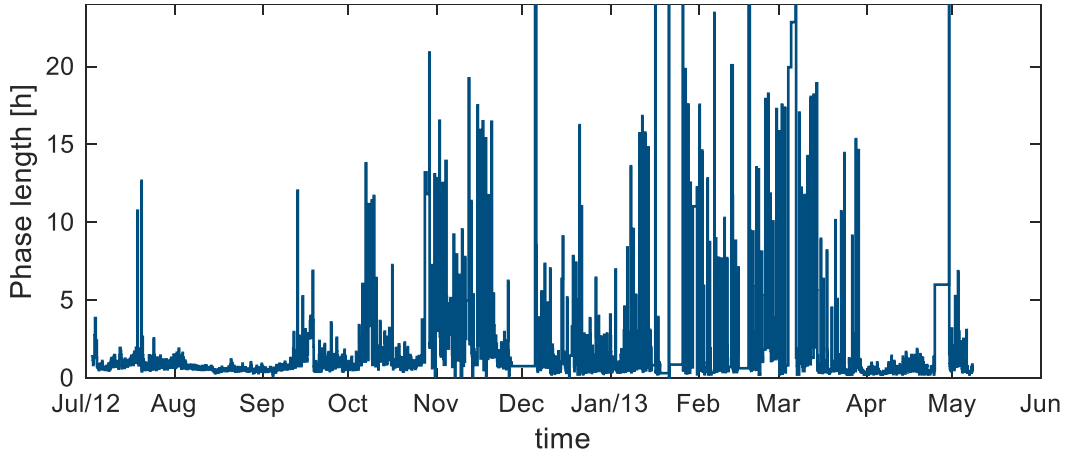


Figure 3.15 Length of the aerobic phases in the studied period

As for what concerns the anoxic phases it can be noticed that their durations are usually constrained in the 1-3 h interval and the phases that exceed this limit are a few. The difference in behaviour with respect to the aerobic phases can be explained assuming a different control strategy of the OSCAR controller which seems to favour the complete oxidation of the ammonia.

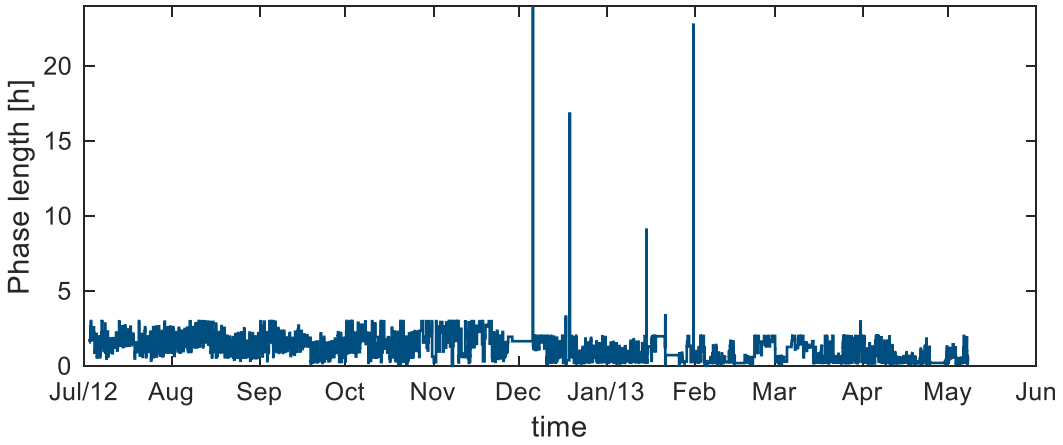


Figure 3.16 Length of the anoxic phases in the studied period

The limited amount of data available and the general poor quality of them made the objective of defining a FDI algorithm extremely difficult to reach. As shall be seen in the chapter of the results, the lack of information regarding the source of the fault/anomaly and its scarce characterization, both in terms of timely extent and quantification, prevented the possibility to create a reliable and robust detection method and most of all did not provided a sufficient dataset to perform the necessary validation of the methods implemented. In order to overcome these shortcomings, it was decided to synthetically produce the data needed using a mathematical model structure of a WWTP. The choice naturally fell on the benchmark framework, a valuable and reliable protocol used by more than two decades to implement new plant layouts or test the response of existing ones to different control strategies.

3.3 A modified Benchmark simulation model for fault detection

The simulation of the internal processes of a WWTP can be effectively obtained using the Benchmark Simulation Model 1 (BSM1) (Copp, 2002; Alex et al., 2008) that resulted from the Activated Sludge Models (ASM) (Henze et al., 2000) developed within the framework of the European Co-operation in the field of Scientific and Technical research (COST) action 682 and 624. Initially sharing the ASM limitation to a 14 days simulation interval the BSM1 an extended version was later proposed (the BSM1 long term, or BSM_LT (Rosén et al., 2004)) in the attempt to allow the

assessment of the long term effect of the implemented control strategies. An adapted BSM1 using a modified ASM3_2N (Iacopozzi et al., 2007) modelling of the biological treatment and the 'dry weather' input of the BSM1_LT was used in this work to produce, provided the necessary modifications, a series of synthetic data used for the training and assessment of the fault detection procedures objective of this research. In the following paragraphs the benchmark model will be described in it essential parts and the modification introduced to adapt it to the intermitted aeration plant configuration will be presented.

3.3.1 Definition and scope of the Benchmark Simulation Models

The BSM simulation protocol is an accurate model of the biochemical processes of a WWTP that comes with a comprehensive description of a typical plant layout and configuration, a number of variables describing the time inputs and a series of performance indexes. All these elements concur in making it the most adequate tool for studying the methodological and operative aspects of the WWT systems.

The Activated Sludge Model n. 1, or ASM1 (Henze et al., 1987), model represented the first tool used for the research, development and optimization of the biological nitrogen and organic carbon removal. From its experience the ASM2 was subsequently developed (Henze et al., 1995), including the bacterial kinetics of the phosphorus accumulating organisms (PAOs), useful to model the treatment of sewages rich in phosphorous. They both represented the first step in the direction of a reference system for the design and the assessment of advanced control strategies which later evolved in the ASM3.

The three models composing the ASM family only differ on the modelling approach towards the simulation of the biological processes but they share the same standardized plant layout. The main features of the plant configuration implemented are the following:

- the plant configuration is composed of 5 tanks in series followed by a secondary settler (Figure 3.17);
- tank 1 and 2 are unaerated and fully mixed, while tanks 3 to 5 are aerated;
- the first two reactors have volume 1'000 m^3 each and the other three 1'333 m^3 each, for a total volume of 6'000 m^3;
- the aerated tanks are supplied with a fixed air flow-rate U_a = 18'000 m^3/d;
- the secondary settler has a volume of 6'000 m^3 (area of 1'500 m^2 and a depth of 4 m) subdivided in 10 layers, each 0.4 m high, and the feed point is set at

2.2 m from the bottom (6th layer). The settler is modelled according to (Takács et al., 1991);

- the plant configuration has two internal recycles, a recycle from the 5th to the 1st reactor with a default flowrate of 55'338 m^3/d flowing through a pipe with internal volume 100 m^3 and a recycle from the settler to the 1st tank of fixed flow-rate 18'446 m^3/d flowing through a pipe with internal volume 300 m^3;
- the sludge waste is continuously pumped from the settler with a constant discharge of 385 m^3.

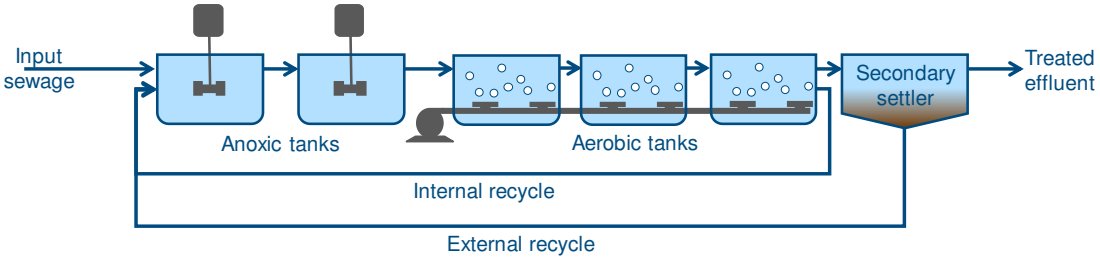


Figure 3.17 Schematic representation of the 'simulation benchmark' configuration

For the modelling of the settling process the choice has fallen upon the one proposed by (Takács et al., 1991) and it is based on the concept of double-exponential settling velocity function, which has the advantage of being applicable to both the flocculants and the hindered settling conditions. The basis of the biological model of the BSM1 is represented by the ASM models family, whose main differences will be highlighted in a qualitative way. The details of the governing equations are left to their relative publications, however in the next paragraphs an extensive description of the biological model used in this work will be provided. Let's just observe here that the biological reactors are here modelled in the ASM framework under the assumption of continuous-flow stirred-tank reactor (CSTR), which means the concentrations are assumed to be timely constant in all the tank, and that the volume V is constant for all the time of the simulation, so that the flowrate Q_{in} that enters the tank is the same as the output, Q_{out} i.e. $Q_{in} = Q_{out}$. A direct consequence is also that the concentration of the compounds exiting the reactor is the same as the concentration inside the tank.

The BSM protocol also comes with a set of three data files each representing a different kind of influent disturbances, namely the 'dry weather', the 'storm weather' and the 'rain weather' events, and are publicly available for download. These files, intended to allow the testing of the process configurations and control strategies under different influent conditions, contain the equivalent of 14 days of influent data sampled every 15 minutes and the input variables provided are time, S_S, $X_{B,H}$, X_S, X_I, S_{NH}, S_I, S_{ND}, X_{ND}, Q, S_O, $X_{B,A}$, X_P, S_{NO} and S_{ALK}. The meaning of most of these abbreviations will be explained in the following paragraph, except for those that have been modified such as

X_P, the representing the particulate products arising from the biomass decay, $X_{B,H}$ and $X_{B,A}$, respectively representing the heterotrophic and autotrophic active biomass and S_{ND} and X_{ND}, respectively the soluble and particulate fraction of the biodegradable organic nitrogen. Notice that the input values of S_O, $X_{B,A}$, X_P and S_{NO} are assumed to be zero while the alkalinity input S_{ALK} is set constantly at 7 mol/m^3 for the entire period. The 14 days period considered incorporates the diurnal variations of the organic load as well as the weekly trends, however such a limited time span makes negligible to account for the temperature-dependency of the kinetic parameters and once experimentally determined they can be considered constants. The 'simulation benchmark' protocol recommends to first run the model in a steady state and then use that as a starting point for the dynamical simulations, in order to limit the influence of the starting condition on the dynamical process.

An important aspect of the benchmark protocol consists in the implementation of sensors and actuators that allow the testing of the custom control strategies designed. The formalization of their realistic modelling within the benchmark framework is subsequent to its release BSM1 and is based on two papers (Alex et al., 2003; Rieger et al., 2003). The latter in particular provides a classification (Table 3.1) sufficiently comprehensive to encompass all the different sensor types and aimed to ease the comparison of the simulated results by discouraging the users to implement custom sensors unless necessary. The classification contains both continuous (A, B_0, C_0) and time-discrete (B_1, C_1, D) sensor models. The main parameter describing is the response time, a parameter that describes the dynamic response of a sensor to a step input Figure 3.18.

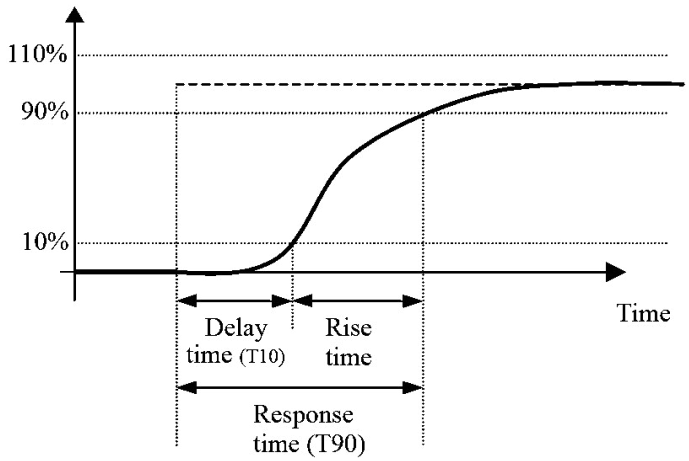


Figure 3.18 Definition of the response time

The response time is defined as the sum of the delay time T_{10}, defined as the time to reach 10% of the final value of a step response, and the rise (or fall) time. Together they represent the time to reach and not leave a band between 90% and 110% of the final step response.

Table 3.1 Sensor classes according to (Rieger et al., 2003)

Sensor class	T_{90}	Examples
A	1 min	Ion-sensitive, optical without filtration
B_0	10 min	Gas-sensitive + fast filtration
B_1	10 min	Photometric + fast filtration
C_0	20 min	Gas-sensitive + slow filtration
C_1	20 min	Photometric+ slow filtration, sedimentation
D	30 min	Photometric od titrimetric for total components

Under the assumption that the real-time behaviour of a sensor is a combination of a first, or higher, order dynamics and a transport delay time, the sensors are modelled as a combination of first-order delay transfer functions G_{sensor} in the form:

$$G_{sensor}(s) = \frac{K}{1+Ts}, \tag{3.12}$$

where K is the gain and s the Laplace operator. Table 3.2 shows for each class of continuous sensors the number of transfer function suggested for the modelling and the value of T.

Table 3.2 Parameters of the response time modelling (Rieger et al., 2003)

Sensor class	T_{90}	Number of transfer fcns	T	$R_{T10/T90}$
A	1 min	2	0.257	0.133
B_0	10 min	8	0.849	0.392
C_0	20 min	8	1.699	0.392

The example in eq. (3.13) shows in form of differential equation the sensor model in case of class A sensor, where two first-order delay transfer functions are used:

$$\begin{aligned} \dot{x}_1(t) &= \frac{K}{T}u(t) - \frac{1}{T}x_1(t) \\ \dot{u}_2(t) &= \frac{K}{T}x_1(t) - \frac{1}{T}u(t), \end{aligned} \tag{3.13}$$

where $u(t)$ is the ideal signal of the coming from the simulation, $x_1(t)$ is the internal state of the sensor and $u_2(t)$ is the delayed measurement signal.

The sensor model proposed for the BSM1 does not consider any attenuation on the signal, nor the presence of systematic errors, nevertheless it suggests to add a simplified model of the noise affecting the measurements. The noise is introduced as a normally distributed random signal, the white noise $\eta(t)$, multiplied by the product of the measurement range ($z_{max} - z_{min}$) and the noise level (nl = 2.5 %):

$$n(t) = \eta(t) \cdot (z_{max} - z_{min}) \cdot nl, \tag{3.14}$$

with $\eta(t) \sim \mathcal{N}(0,1)$. Both the noise level and the measurement range are information typically provided by the instrument producers, however in the BSM1 framework, it is recommended to use the values in Table 3.3 to facilitate the comparison of results.

Table 3.3 Recommended BSM1 sensor parameters range (Alex et al., 2008)

Measured variable	Measurement range
Flow-rate (high range) [m^3/d]	0-100'000
Water level [m]	0-5
Temperature [°C]	5-25
pH	5-9
S_O [g COD/m^3]	0-10
Sludge blanket level [m]	0-5
S_{NO} [g N/m^3]	0-20
S_{NH} (low range) [g N/m^3]	0-20
S_{NH} (high range) [g N/m^3]	0-50
S_{ALK} [mol HCO_3/m^3]	0-20
Mixed-liquor suspended solids [g/m^3]	0-10'000
Effluent total suspended solids [g/m^3]	0-200
COD [g COD/m^3]	0-1'000
OUR [g COD/m^3·d]	0-2'000

The noise signal described above is to be considered an additive component to the delayed measurement signal, so following the example of eq.(3.13) the intermediate disturbed measurement $z_1(t)$ can be computed as:

$$z_1(t) = u_2(t) + n(t)\,. \tag{3.15}$$

Finally, a further consideration on the sensor behaviour should consider that the instrument can only return measurements that fall inside its $[z_{max}\text{-}z_{min}]$ range, which can be obtained by applying a simple saturation filter to the signal z_1:

$$z(t) = \begin{cases} z_{max}, & \text{if } z_1(t) > z_{max} \\ z_1(t), & \text{if } z_{min} \le z_1(t) \le z_{max} \\ z_{min}, & \text{if } z_1(t) < z_{min} \end{cases} \tag{3.16}$$

3.3.1.1 Properties and limits of the earlier activated sludge models

The evolution of the ASM family was justified by the need to create a more advanced and efficient model support and by the limitations of the earlier ASM versions Here below a list of the main characteristics and shortcoming of the ASM1 and ASM2 model is presented.

In both models the kinetics of the heterotrophic organisms were not depending on the variation of alkalinity in the mixed liquor nor on the nitrogen concentration.

In the ASM1 model the soluble biodegradable organic nitrogen (S_{ND}) and the particulate one (S_{NI}) were distinct, where in real case their different contribution results extremely difficult to measure separately. For this reason, the later ASM2 model only defined a soluble substrate composed by nitrites and nitrates in the state variable S_{NO3}. Also, the ammonification kinetic defined in ASM1, resulted difficult to quantify, and

was substituted in ASM2 by a more correct expression, but the quickness of the process made it difficult to predict its effects anyway.

The inert particulate substrate X_I was differentiated in the ASM1 model according to its origin, the type of influent or the biomass, while the ASM2 model cancelled these distinctions because difficult to apply in real case scenarios;

The lysis, combined with hydrolysis, is used in both models to describe the coupled effects of the endogenous respiration of the stored substrate, the death, the predation and the lysis of the biomass. However, the hydrolysis process has a dominant effect on the quantification of the oxygen consumption and the denitrification of the heterotrophic organisms and the coupling of it with the lysis process brought some difficulties in the choice of the kinetic parameters. Moreover, the description of the fermentation in the ASM2 model resulted complex and without good results.

The ASM1 model does not includes the possibility to differentiate the decay rates of the nitrifying organisms in aerobic and anoxic conditions. An elevated solid retention time and the formation of anoxic zones in the aerobic reactor made it difficult to identify the maximum nitrification rate.

Finally, in the assumption of both the ASM1 and the ASM2 models the biomass only grows according to the amount of external substrate and the consumption of oxygen after the end of the substrate causes the decay of the biomass. In conventional activated sludge processes, instead, the feeding regime of the biomass is variable (a ‘feast phase’ when the substrate is present and a ‘famine phase’ when is absent) and in this case the storage inside the cells plays an important role.

3.3.1.2 Properties of the activated sludge model n.3

Those considerations and the earlier applications of the models brought the need of a model that could overcome their limitations, such as the model of the endogenous respirations and the storage, eliminating the properties that did not produce relevant results, such as the lysis and the fermentation process. The result is the ASM3 model (Gujer et al., 2000), which includes all the corrections above and eliminates the variables related to the phosphorus, resulting in a more slender model yet not less complete. The ASM3 results actually simpler to calibrate with respect to the ASM1 model since the growth-decay metabolic processes are replaced by the concept of birth-death-regeneration. This way the state variables are no longer directly influenced by the process parameters as in ASM1. In ASM1 model the respiration processes are associated either to the growth phase and to the cellular decay and it is therefore not able to explain the rapid variation in oxygen consumption experimentally observed. In ASM3, instead,

the respiration processes are only related to the aerobic growth and the decay process is replaced by endogenous processes: the cellular material is released by the cellular lysis, without oxygen consumption.

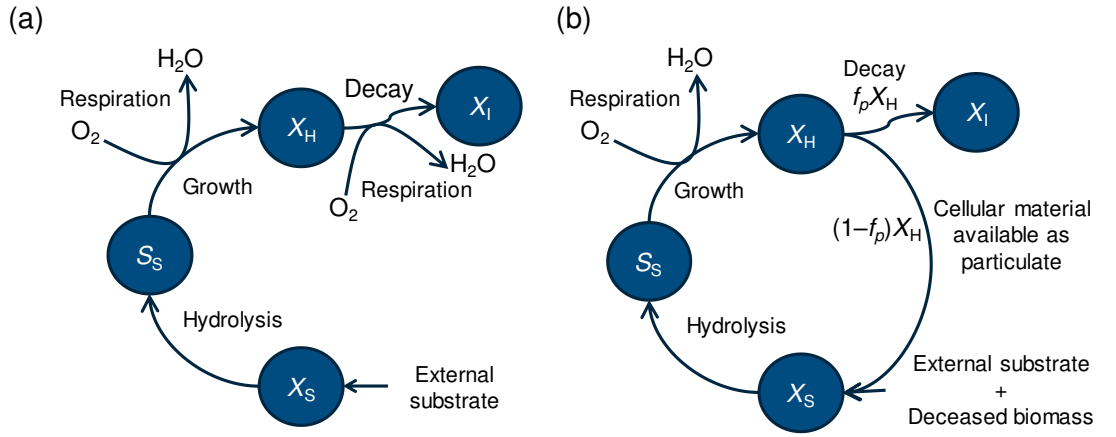


Figure 3.19 Schematic representation of the growth-decay model in ASM1 (a) and the death-regeneration model in ASM3 (b)

The non-biodegradable fraction ($X_I = f_p \cdot X_H$) remains as inert material, while the remaining part ($(1 - f_p) \cdot X_H$) is considered available as slowly biodegradable substrate (Figure 3.19).

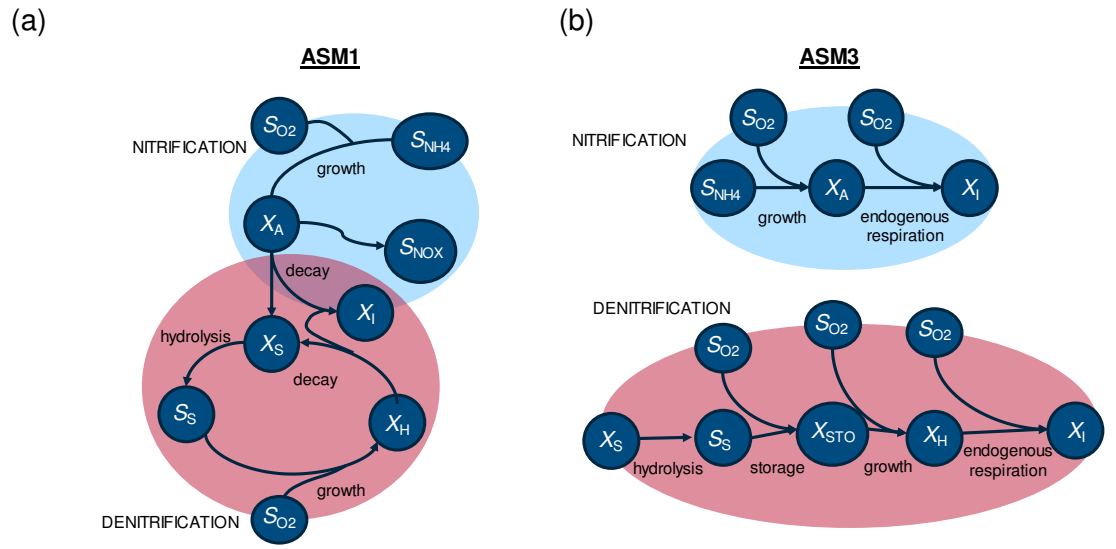


Figure 3.20 Differences between the ASM1 (a) and the ASM3 models (b)

Another relevant difference between the ASM1 and the ASM3 model is that the latter introduces the concept of the storage. The assumption is that all the rapidly biodegradable substrate (S_S) in absence of external organic load is directly converted in storage material (X_{STO}). These compounds represent the carbonaceous source for the growth of the heterotrophic biomass for all the famine phase. This change is reflected by the fact that while in the ASM1 model all the state variables are influenced by the change of the in the microbial parameters, in the ASM3 model this relationship is weaker, increasing the identifiability of the parameters. Finally, it can be observed (Figure 3.20) that the 'carbon path' and the 'nitrogen path' in the ASM3 models are de-coupled, and a separation in introduced between the endogenous metabolism of the

heterotrophic and autotrophic organisms without the recycling of the substrate, defined in the ASM1 as the decay of the autotrophic biomass.

3.3.2 A modified long- term benchmark simulation model based on the activated sludge model n.2 with two-steps nitrification

The present study uses as basis for the modified BSM1_LT an slightly updated version of the ASM3_2N model (Iacopozzi et al., 2007), presented in (Giusti et al., 2011) plus some adaptation due to the AC configuration examined The changes that are going to be presented only concern the biological processes, while for the settling process the model proposed by (Takács et al., 1991) and the specification of the BSM1 still apply.

While the BSM1 and the later BSM1_LT base the modelling of the biological processes mainly on the ASM1 model, due to its widespread diffusion at the time of the release rather than to its performances, the present work used as theoretical base of the biological process a revised version of the ASM3_2N model. The decision is motivated by the need to have a coherent and detailed description of the process of interest, which in this case is the nitrification/denitrification process, and it has been demonstrated in the previous paragraph that the ASM3 provides a better description for that than the ASM1. The ASM3_2N model is a modified version of the ASM3 characterised by the addition of the so called '2 steps nitrification process'. The modifications introduced further improve the description of the complete kinetic of the nitrifying bacteria, distinguishing the two stages of the process and accounting for the intermediate by-products (the nitrites). The state variables of the nitrogen substrate are therefore separated into ammonium (S_{NH4}), nitrites (S_{NO2}) and nitrates (S_{NO3}), while those of the autotrophic biomass are split into the one of ammonium-oxidating bacteria (X_{ns}), where the subscript denotes the *Nitrosomonas* bacteria), responsible for the oxidation of ammonia into nitrites, and that of nitrite-oxidating bacteria (X_{nb}, from the *Nitrobacter* bacteria), which are responsible for the transformation of the nitrites in nitrates. The complete list of the state variables is reported in Table 3.4.

Table 3.4 State variable of the ASM3_2N model (Iacopozzi et al., 2007)

Type	Variable number	Symbol	Units	Description
Soluble components	x_1	S_{O_2}	[g $O_2 \cdot m^{-3}$]	Dissolved oxygen
	x_2	S_S	[g COD$\cdot m^{-3}$]	Readily biodegradable substrates
	x_3	S_{N_2g}	[g N$\cdot m^{-3}$]	Dinitrogen released by denitrification
	x_4	S_{NH_4}	[g N$\cdot m^{-3}$]	Ammonium concentration
	x_5	S_{NO_2}	[g N$\cdot m^{-3}$]	Nitrite nitrogen concentration
	x_6	S_{NO_3}	[g N$\cdot m^{-3}$]	Nitrate nitrogen concentration
	x_7	S_I	[g COD$\cdot m^{-3}$]	Soluble inert organic material
	x_8	S_{ALK}	[mol $HCO_3^- \cdot m^{-3}$]	Alkalinity
Particulate components	x_9	X_I	[g COD$\cdot m^{-3}$]	Particulate inert organic material
	x_{10}	X_S	[g COD$\cdot m^{-3}$]	Slowly biodegradable substrates
	x_{11}	X_H	[g COD$\cdot m^{-3}$]	Heterotrophic biomass
	x_{12}	X_{STO}	[gCOD$\cdot m^{-3}$]	Organic material stored by the heterotrophs
	x_{13}	X_{ns}	[g COD$\cdot m^{-3}$]	Ammonia-oxidizing autotrophic biomass (nitrosomonas)
	x_{14}	X_{nb}	[g COD$\cdot m^{-3}$]	Nitrite-oxidizing autotrophic biomass (nitrobacter)

The assumption of a single nitrification step is acceptable when modelling a generic sewage treatment process, since the nitrite is considered as an intermediate by-product only present in the plant for a limited time. However, there are situations in which is important to quantify the amount of nitrites in the system, for example when there are issues with the nitrification capability of the plant, when there is a limit on the nitrite concentration of the effluent, when the characteristics of the influent facilitate the inhibition of the nitrite-oxidising bacteria or when the operative conditions limit the oxidation to the first step. In case the nitrification/denitrification processes have a strong importance, as in AC plant configurations, it is therefore preferable to have a more detailed model of them, so that the control strategies can be more effectively tested and that the process description results more accurate.

Table 3.5 Difference of the description of the nitrification process between the ASM3 model and the ASM3_2N

ASM3	ASM3_2N
$NH_4^+ + 2O_2 \xrightarrow{\text{autotrophic}} NO_3^- + 2H^+ + H_2O$	$NH_4^+ + \frac{3}{2}O_2 \xrightarrow{\text{ammonia-oxidising}} NO_2^- + 2H^+ + H_2O$ $NO_2^- + \frac{1}{2}O_2 \xrightarrow{\text{nitrite-oxidising}} NO_3^-$ $NH_4^+ + 2O_2 \xrightarrow{\text{autotrophic}} NO_3^- + 2H^+ + H_2O$

The introduction of the two-step nitrification process brings the need to revise also the denitrification model of the ASM3 since the growth of the facultative heterotrophic organisms in anoxic conditions uses either nitrite or nitrate as final electron acceptor. Therefore, also the denitrification process in ASM3_2N is split in two steps. The denitrification step can be analysed in two different ways: one considers the denitrification process according to a scheme where the nitrate is first reduced to nitrite and later the nitrite is reduced to dinitrogen; the other considers a parallel reaction where either the nitrates and the nitrites are reduced to gas dinitrogen. For the present work the latter assumption is adopted, despite the debate over which of the two is the best modelling approach is still open (Abdul-Talib et al., 2002; Sin et al., 2008).

In the ASM3_2N model presented in (Iacopozzi et al., 2007) when the nitrate concentration is null the equation related to the anoxic storage is not, and so are those of the endogenous respiration processes, which means that there is storage of organic substrate (and consequently a reduction of the concentration in the mixed liqueur) even when the nitrate is not present. Also, according to other kinetic rates the stored biomass auto-consumes by respiration reducing the nitrate even in absence of this compound. For this reason the kinetic ratios have been modified in (Giusti et al., 2011) splitting the storage to distinguish if it develops on the nitrites (ρ_2) or on the nitrates (ρ_3), and therefore it tends to decrease as the concentrations of nitrates decrease. In this updated version of the ASM3_2N model also the equation governing anoxic endogenous respiration have been modified, splitting them in one equation for each nitrogen oxide. Moreover, with respect to the work of (Iacopozzi et al., 2007), two partial ratios have been introduced, $S_{NO2}/(S_{NO2} + S_{NO3})$ and $S_{NO3}/(S_{NO2} + S_{NO3})$, accounting for the nitrites and nitrates concentration fraction with respect to the cumulative concentration of the NO*x*. This addition provides a better description of the evolution of the denitrification process, allowing to explicitly identify which process (ρ_5 or ρ_6) is favoured based on the nitrites or nitrates concentration and consequently which is the most influential.

The following table (Table 3.6) lists the transformation processes simulated by the numerical model together with their kinetic rates ρ_i, where it can be observed that with respect to the biological transformations induced by the endogenous respiration both in the aerobic (ρ_7, ρ_{15} and ρ_{16}) and anoxic phase (ρ_8, ρ_9, ρ_{17} and ρ_{18}) already present in the ASM3 model it is added the separation in two steps of the denitrification (ρ_5-ρ_6) and denitrification process (ρ_{13}-ρ_{14}).

Table 3.6 Kinetic rates of the processes involved in the aerobic tank

Proc. Num.	Process rate equation	Description
ρ_1	$k_{\mathrm{STO}}(temp(t)) \cdot \frac{x_1}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_2}{K_{\mathrm{S}} + x_2} \cdot x_{11}$	Aerobic storage of S_S
ρ_2	$k_{\mathrm{STO}}(temp(t)) \cdot \eta_{\mathrm{NO_x}} \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_6}{K_{\mathrm{NO_x}} + x_6} \cdot \frac{x_6}{x_5 + x_6} \cdot \frac{x_2}{K_{\mathrm{S}} + x_2} \cdot x_{11}$	Anoxic storage of S_S on NO_3
ρ_3	$k_{\mathrm{STO}}(temp(t)) \cdot \eta_{\mathrm{NO_x}} \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_5}{K_{\mathrm{NO_x}} + x_5} \cdot \frac{x_5}{x_5 + x_6} \cdot \frac{x_2}{K_{\mathrm{S}} + x_2} \cdot x_{11}$	Anoxic storage of S_S on NO_2
ρ_4	$\mu_{\mathrm{H}}(temp(t)) \cdot \frac{x_1}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_4}{K_{\mathrm{NH_4}} + x_4} \cdot \frac{x_8}{K_{\mathrm{ALK}} + x_8} \cdot \frac{x_{12} / x_{11}}{K_{\mathrm{STO}} + (x_{12} / x_{11})} \cdot x_{11}$	Aerobic growth of heterotrophic organisms
ρ_5	$\mu_{\mathrm{H}}(temp(t)) \cdot \eta_{\mathrm{NO_x}} \cdot \frac{x_4}{K_{\mathrm{NH_4}} + x_4} \cdot \frac{x_5}{K_{\mathrm{NO_x}} + x_5} \cdot \frac{x_5}{x_5 + x_6} \cdot \frac{x_8}{K_{\mathrm{ALK}} + x_8} \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_{12} / x_{11}}{K_{\mathrm{STO}} + (x_{12} / x_{11})} \cdot x_{11}$	Nitrite denitrification (aerobic growth on S_{NO2})
ρ_6	$\mu_{\mathrm{H}}(temp(t)) \cdot \eta_{\mathrm{NO_x}} \cdot \frac{x_4}{K_{\mathrm{NH_4}} + x_4} \cdot \frac{x_6}{K_{\mathrm{NO_x}} + x_6} \cdot \frac{x_6}{x_5 + x_6} \cdot \frac{x_8}{K_{\mathrm{ALK}} + x_8} \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_{12} / x_{11}}{K_{\mathrm{STO}} + (x_{12} / x_{11})} \cdot x_{11}$	Nitrate denitrification (aerobic growth on S_{NO3})
ρ_7	$b_{\mathrm{H,O_2}}(temp(t)) \cdot \frac{x_1}{K_{\mathrm{O_2}} + x_1} \cdot x_{11}$	Aerobic endogenous respiration
ρ_8	$b_{\mathrm{H,NO_x}} \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_6}{K_{\mathrm{NO_x}} + x_6} \cdot \frac{x_6}{x_5 + x_6} \cdot x_{11}$	Anoxic NO_3 endogenous respiration
ρ_9	$b_{\mathrm{H,NO_x}} \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_5}{K_{\mathrm{NO_x}} + x_5} \cdot \frac{x_5}{x_5 + x_6} \cdot x_{11}$	Anoxic NO_2 endogenous respiration
ρ_{10}	$b_{\mathrm{STO,O_2}}(temp(t)) \cdot \frac{x_1}{K_{\mathrm{O_2}} + x_1} \cdot x_{12}$	X_{STO} aerobic respiration
ρ_{11}	$b_{\mathrm{STO,NO_x}}(temp(t)) \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_6}{K_{\mathrm{NO_x}} + x_6} \cdot \frac{x_6}{x_5 + x_6} \cdot x_{12}$	X_{STO} NO_3 anoxic respiration
ρ_{12}	$b_{\mathrm{STO,NO_x}}(temp(t)) \cdot \frac{K_{\mathrm{O_2}}}{K_{\mathrm{O_2}} + x_1} \cdot \frac{x_5}{K_{\mathrm{NO_x}} + x_5} \cdot \frac{x_5}{x_5 + x_6} \cdot x_{12}$	X_{STO} NO_2 anoxic respiration
ρ_{13}	$\mu_{\mathrm{ns}}(temp(t)) \cdot \frac{x_1}{K_{\mathrm{A,O_2,ns}} + x_1} \cdot \frac{x_4}{K_{\mathrm{A,NH_4}} + x_4} \cdot \frac{x_8}{K_{\mathrm{A,ALK}} + x_8} \cdot x_{13}$	Nitrification. I step (aerobic growth of X_{ns})

Proc. Num.	Process rate equation	Description
ρ_{14}	$\mu_{nb}(temp(t)) \cdot \frac{x_1}{K_{A,O_2,nb} + x_1} \cdot \frac{K_{NH_4,I}}{K_{NH_4,I} + x_4} \cdot \frac{x_8}{K_{A,ALK} + x_8} \cdot \frac{x_5}{K_{A,NO_x} + x_5} \cdot x_{14}$	Nitrification. II step (aerobic growth of X_{nb})
ρ_{15}	$b_{ns,O_2}(temp(t)) \cdot \frac{x_1}{K_{A,O_2,ns} + x_1} \cdot x_{13}$	X_{ns} aerobic endogenous respiration
ρ_{16}	$b_{nb,O_2}(temp(t)) \cdot \frac{x_1}{K_{A,O_2,nb} + x_1} \cdot x_{14}$	X_{nb} aerobic endogenous respiration
ρ_{17}	$b_{A,NO_2}(temp(t)) \cdot \frac{K_{A,O_2,ns}}{K_{A,O_2,ns} + x_1} \cdot \frac{x_5}{K_{A,NO_x} + x_5} \cdot x_{13}$	X_{ns} anoxic endogenous respiration
ρ_{18}	$b_{A,NO_3}(temp(t)) \cdot \frac{K_{A,O_2,nb}}{K_{A,O_2,nb} + x_1} \cdot \frac{x_6}{K_{A,NO_x} + x_6} \cdot x_{14}$	X_{nb} anoxic endogenous respiration
ρ_{19}	$k_h(temp(t)) \cdot \frac{x_{10} / x_{11}}{K_X + (x_{10} / x_{11})} \cdot x_{11}$	Hydrolysis

The meaning of the model parameters used in Table 3.4 and that of the parameters used further in the text is clarified in Table 3.7, distinguishing them in three classes depending on whether they belong to the autotrophic biomass, the heterotrophic biomass or the hydrolysis process. In the context of the ASM models and of the standard BSM the parameters are usually considered constants, since the it is assumed that they do not change in the short period analysed (14 days). However, in this case, having chosen a long term approach brings the need to account for the temperature-dependency of the model parameters and that is achieved by computing their value at each time-step according to the Arrhenius law (3.17), where the current value of the parameter π_T depends on the nominal value π_{20} relative to a temperature of 20°C (except when is specified otherwise) and on its specific coefficient θ:

$$\pi_T = \pi_{20} e^{\theta \cdot (T-20)} \tag{3.17}$$

Notice that some of the parameters used were experimentally determined in a previous work since they were not included in the standard ASM3 list of parameters.

Table 3.7 Kinetic parameters and their relative description

Kinetic parameter	Description	Value (20°C)	θ	Units	Reference
		Hydrolysis			
k_h	Hydrolysis rate constant	3	0.041	g COD_{XS}· (g COD_{XH})$^{-1}$ ·d^{-1}	(Henze et al., 2000)
K_X	Hydrolysis saturation constant	1	-	g COD_{XS}· (g COD_{XH})$^{-1}$	(Henze et al., 2000)
Ka	Aeration system efficiency	0.15	-	m^{-3}	(Iacopozzi et al., 2007)
		Heterotrophic organisms, X_H			
k_{STO}	Storage rate constant	5	0.069	g COD_{SS}· (g COD_{XH})$^{-1}$ ·d^{-1}	(Henze et al., 2000)
η_{NOx}	Reduction factor for denitrif.	0.6	-	-	(Henze et al., 2000)
K_{O2}	Sat./inhibition constant for S_{O2}	0.2	-	g O_2·m^{-3}	(Henze et al., 2000)
K_{NOx}	Saturation constant for S_{NOx}	0.5	-	g N-NO_3^-·m^{-3}	(Henze et al., 2000)
K_S	Sat./inhibition constant for S_S	3	-	g COD_{SS}·m^{-3}	(Henze et al., 2000)
K_{STO}	Sat./inhibition constant for X_{STO}	1	-	g COD_{XSTO}· (g COD_{XH})$^{-1}$	(Henze et al., 2000)
μ_H	Maximum growth rare for X_H	2	0.069	d^{-1}	(Henze et al., 2000)
K_{NH4}	Saturation constant for S_{NH4}	0.05	-	g N-NH_4^+·m^{-3}	(Henze et al., 2000)
K_{ALK}	Saturation constant for X_H	0.1	-	mole HCO_3^-·m^{-3}	(Henze et al., 2000)
$b_{H,O2}$	Aerobic endogenous respiration rate for X_H	0.2	0.069	d^{-1}	(Henze et al., 2000)
$b_{H,NOx}$	Anoxic endogenous respiration rate for X_H	0.1	0.069	d^{-1}	(Henze et al., 2000)
$b_{STO,O2}$	Aerobic endogenous respiration rate for X_{STO}	0.2	0.069	d^{-1}	(Henze et al., 2000)
$b_{STO,NOx}$	Anoxic endogenous respiration rate for X_{STO}	0.1	0.069	d^{-1}	(Henze et al., 2000)
		Autotrophic organisms, X_A			
μ_{ns}	Maximum growth rare for X_{ns}	0.37*	0.105	d^{-1}	(Giusti et al., 2011)
μ_{nb}	Maximum growth rare for X_{nb}	0.87*	0.105	d^{-1}	(Giusti et al., 2011)
$K_{A,NH4}$	Saturation constant for S_{NH4}	1	-	g N-NH_4^+·m^{-3}	(Henze et al., 2000)
$K_{NH4,I}$	Ammonia inhibition of nitrate oxidation	5	-	g N-NH_4^+·m^{-3}	(Henze et al., 2000)
$K_{A,O2}$	Sat./inhibition constant for S_{O2}	0.5	-	g O_2·m^{-3}	(Henze et al., 2000)
$K_{A,NO2}$	Saturation constant for S_{NO2}	0.5	-	g N-NO_2^-·m^{-3}	(Henze et al., 2000)
$K_{A,NO3}$	Saturation constant for S_{NO3}	0.5	-	g N-NO_3^-·m^{-3}	(Henze et al., 2000)
$K_{A,ALK}$	Saturation constant for X_H	0.5	-	mol HCO_3^-·m^{-3}	(Henze et al., 2000)
$b_{ns,O2}$	Aerobic endogenous respiration rate for X_{ns}	0.15	0.11	d^{-1}	(Henze et al., 2000)

Kinetic parameter	Description	Value (20°C)	θ	Units	Reference
$b_{nb,O2}$	Aerobic endogenous respiration rate for X_{nb}	0.151	0.11	d^{-1}	(Henze et al., 2000)
$b_{A,ns}$	Anoxic endogenous respiration rate for X_{ns}	0.05	0.092	d^{-1}	(Henze et al., 2000)
$b_{A,nb}$	Anoxic endogenous respiration rate for X_{nb}	0.05	0.092	d^{-1}	(Henze et al., 2000)

* parameter at 15°C

A detail of the conversion rates r_i relating the state variables with the process can be found in Table 3.8 while the description of the stoichiometric parameters involved can be found in Table 3.9.

Table 3.8 Detail of the kinetic conversion rates for the state variables

State variable	Symbol	Kinetic conversion rate r_i
x_1	S_{O_2}	$r_1 = (Y_{STO,O2} - 1)\cdot\rho_1 + (1 - (1/Y_{H,O2}))\cdot\rho_4 + (f_{XI} - 1)\cdot\rho_7 - \rho_{10} + (1 - (3.43/Y_{A,ns}))\cdot\rho_{13} + (1 - (1.14/Y_{A,nb}))\cdot\rho_{14} + (f_{XI} - 1)\cdot\rho_{15} + (f_{XI} - 1)\cdot\rho_{16}$
x_2	S_S	$r_2 = -\rho_1 - \rho_2 - \rho_3 + (1 - f_{SI})\cdot\rho_{19}$
x_3	$S_{N_2 g}$	$r_3 = [(1 - Y_{STO,NOx})/2.86]\cdot\rho_2 + [(1 - Y_{STO,NOx})/1.72]\cdot\rho_3 + [((1/Y_{H,NOx}) - 1)/1.72]\cdot\rho_5 + [((1/Y_{H,NOx}) - 1)/2.86]\cdot\rho_6 + [(1 - f_{XI})/2.86]\cdot\rho_8 + [(1 - f_{XI})/1.72]\cdot\rho_9 + (1/2.86)\cdot\rho_{11} + (1/1.72)\cdot\rho_{12} + [(1 - f_{XI})/1.72]\cdot\rho_{17} + [(1 - f_{XI})/2.86]\cdot\rho_{18}$
x_4	S_{NH_4}	$r_4 = i_{N,SS}\cdot\rho_1 + i_{N,SS}\cdot\rho_2 + i_{N,SS}\cdot\rho_3 - i_{N,BM}\cdot\rho_4 - i_{N,BM}\cdot\rho_5 - i_{N,BM}\cdot\rho_6 + [i_{N,BM} - (f_{XI}\cdot i_{N,XI})]\cdot\rho_7 + [i_{N,BM} - (f_{XI}\cdot i_{N,XI})]\cdot\rho_8 + [i_{N,BM} - (f_{XI}\cdot i_{N,XI})]\cdot\rho_9 + [- i_{N,BM} - (1/Y_{A,ns})]\cdot\rho_{13} - i_{N,BM}\cdot\rho_{14} + [i_{N,BM} - (f_{XI}\cdot i_{N,XI})]\cdot\rho_{15} + [i_{N,BM} - (f_{XI}\cdot i_{N,XI})]\cdot\rho_{16} + (i_{N,BM} - f_{XI}\cdot i_{N,XI})\cdot\rho_{17} + (i_{N,BM} - f_{XI}\cdot i_{N,XI})\cdot\rho_{18} + [i_{N,XS} - i_{N,SS}\cdot(1 - f_{SI}) - (f_{SI}\cdot i_{N,SI})]\cdot\rho_{19}$
x_5	S_{NO_2}	$r_5 = [(Y_{STO,NOx} - 1)/1.72]\cdot\rho_3 + [(1 - (1/Y_{H,NOx}))/1.72]\cdot\rho_5 + [(f_{XI} - 1)/1.72]\cdot\rho_9 - (1/1.72)\cdot\rho_{12} + (1/Y_{A,ns})\cdot\rho_{13} - (1/Y_{A,nb})\cdot\rho_{14} + [(f_{XI} - 1)/1.72]\cdot\rho_{17}$
x_6	S_{NO_3}	$r_6 = [(Y_{STO,NOx} - 1)/2.86]\cdot\rho_2 + [(1 - (1/Y_{H,NOx}))/2.86]\cdot\rho_6 + [(f_{XI} - 1)/2.86]\cdot\rho_8 - (1/2.86)\cdot\rho_{11} + (1/Y_{A,nb})\cdot\rho_{14} + [(f_{XI} - 1)/2.86]\cdot\rho_{18}$
x_7	S_I	$r_7 = f_{SI}\cdot\rho_{19}$
x_8	S_{ALK}	$r_8 = (i_{N,SS}/14)\cdot\rho_1 + [(i_{N,SS} + (1 - Y_{STO,NOx} - 1)/2.86)/14]\cdot\rho_2 + [(i_{N,SS} + (1 - Y_{STO,NOx} - 1)/1.72)/14]\cdot\rho_3 - (i_{N,BM}/14)\cdot\rho_4 + [(- i_{N,BM} + (1 - (1/Y_{H,NOx}))/1.72)/14]\cdot\rho_5 + [(- i_{N,BM} + (1 - (1/Y_{H,NOx}))/2.86)/14]\cdot\rho_6 + [(i_{N,BM} - (f_{XI}\cdot i_{N,XI}))/14]\cdot\rho_7 + [((i_{N,BM} - (f_{XI}\cdot i_{N,XI})) + (1 - f_{XI})/2.86)/14]\cdot\rho_8 + [((i_{N,BM} - (f_{XI}\cdot i_{N,XI})) + (1 - f_{XI})/1.72)/14]\cdot\rho_9 + [1/(2.86\cdot 14)]\cdot\rho_{11} + [1/(1.72\cdot 14)]\cdot\rho_{12} + [- (1/7)\cdot(1/Y_{A,ns}) - (i_{N,BM}/14)]\cdot\rho_{13} - (i_{N,BM}/14)\cdot\rho_{14} + [(i_{N,BM} - (f_{XI}\cdot i_{N,XI}))/14]\cdot\rho_{15} + [(i_{N,BM} - (f_{XI}\cdot i_{N,XI}))/14]\cdot\rho_{16} + [(i_{N,BM} - f_{XI}\cdot i_{N,XI}) - (f_{XI} - 1)/1.72)/14]\cdot\rho_{17} + [(i_{N,BM} - f_{XI}\cdot i_{N,XI}) - (f_{XI} - 1)/2.86)/14]\cdot\rho_{18} + [(i_{N,XS} - i_{N,SS}\cdot(1 - f_{SI}) - (f_{SI}\cdot i_{N,SI}))/14]\cdot\rho_{19}$
x_9	X_I	$r_9 = f_{XI}\cdot\rho_7 + f_{XI}\cdot\rho_8 + f_{XI}\cdot\rho_9 + f_{XI}\cdot\rho_{15} + f_{XI}\cdot\rho_{16} + f_{XI}\cdot\rho_{17} + f_{XI}\cdot\rho_{18}$
x_{10}	X_S	$r_{10} = -\rho_{19}$
x_{11}	X_H	$r_{11} = \rho_4 + \rho_5 + \rho_6 - \rho_7 - \rho_8 - \rho_9$
x_{12}	X_{STO}	$r_{12} = Y_{STO,O2}\cdot\rho_1 + Y_{STO,NOx}\cdot\rho_2 + Y_{STO,NOx}\cdot\rho_3 - (1/Y_{H,O2})\cdot\rho_4 - (1/Y_{H,NOx})\cdot\rho_5 - (1/Y_{H,NOx})\cdot\rho_6 - \rho_{10} - \rho_{11} - \rho_{12}$
x_{13}	X_{ns}	$r_{13} = \rho_{13} - \rho_{15} - \rho_{17}$
x_{14}	X_{nb}	$r_{14} = \rho_{14} - \rho_{16} - \rho_{18}$

Table 3.9 Values of the stoichiometric constants

Stoichiometric constants	Description	Value	Units	Reference
f_{XI}	Production of X_I in endogenous respiration	0.20	g COD_{XI}·(g COD_{XBM})$^{-1}$	(Henze et al., 2000)
f_{SI}	Production of S_I in endogenous respiration	0.00	g COD_{SI}·(g COD_{XS})$^{-1}$	(Henze et al., 2000)
$Y_{STO,O2}$	Aerobic yield of stored product for S_S	0.85	g COD_{XSTO}·(g COD_{SS})$^{-1}$	(Henze et al., 2000)
$Y_{STO,NOx}$	Anoxic yield of stored product for S_S	0.80	g COD_{XSTO}·(g COD_{SS})$^{-1}$	(Henze et al., 2000)
$Y_{H,O2}$	Aerobic yield of X_H	0.63	g COD_{XH}·(g COD_{XSTO})$^{-1}$	(Henze et al., 2000)
$Y_{H,NOx}$	Anoxic yield of X_H	0.54	g COD_{XH}·(g COD_{XSTO})$^{-1}$	(Henze et al., 2000)
$Y_{A,ns}$	Aerobic yield of X_A	0.10	g COD_{Xns}·(g N-NO_2^-)$^{-1}$	(Marsili-Libelli et al., 2001)
$Y_{A,nb}$	Aerobic yield of X_A	0.14	g COD_{Xnb}·(g N-NO_3^-)$^{-1}$	(Marsili-Libelli et al., 2001)
$i_{N,SI}$	Nitrogen content of S_I	0.01	g N·(g COD_{SI})$^{-1}$	(Henze et al., 2000)
$i_{N,SS}$	Nitrogen content of S_S	0.03	g N·(g COD_{SS})$^{-1}$	(Henze et al., 2000)
$i_{N,XI}$	Nitrogen content of X_I	0.02	g N·(g COD_{XI})$^{-1}$	(Henze et al., 2000)
$i_{N,XS}$	Nitrogen content of X_S	0.04	g N·(g COD_{XS})$^{-1}$	(Henze et al., 2000)
$i_{N,BM}$	Nitrogen content of biomass X_H, X_A	0.07	g N·(g COD_{BM})$^{-1}$	(Henze et al., 2000)

It is however generally difficult to grasp the mutual relationships among all the processes and the variables involved by reading them in the form of Table 3.8 and it is instead preferable to express them using the so called 'stoichiometric matrix' (Table 3.10). Moreover, using this matrix it is easy to retrieve the conversion rates r_i since:

$$r_i = \sum_j w_{i,j}\rho_j \,, \tag{3.18}$$

where $w_{i,j}$ represents the stoichiometric relationship between the i-th state variable and the j-th process ρ_j.

Table 3.10 Stoichiometric matrix for the aerobic processes

States → / (x_i) → / ↓ Processes (ρ_j) ↓		S_{O2} (x_1)	S_S (x_2)	S_{N2g} (x_3)	S_{NH4} (x_4)	S_{NO2} (x_5)	S_{NO3} (x_6)	S_I (x_7)	S_{ALK} (x_8)	X_I (x_9)	X_S (x_{10})	X_H (x_{11})	X_{STO} (x_{12})	X_{ns} (x_{13})	X_{nb} (x_{14})
Aerobic storage of S_S	(ρ_1)	$Y_{STO,O2} - 1$	−1	0	$i_{N,SS}$	0	0	0	$i_{N,SS}/14$	0	0	0	$Y_{STO,O2}$	0	0
Anoxic storage of $S_{S,NO3}$	(ρ_2)	0	−1	$(1 - Y_{STO,NOx})/2.86$	$i_{N,SS}$	0	$(Y_{STO,NOx} - 1)/2.86$	0	$(i_{N,SS} + (1 - Y_{STO,NOx} - 1)/2.86)/14$	0	0	0	$Y_{STO,NOx}$	0	0
Anoxic storage of $S_{S,NO2}$	(ρ_3)	0	−1	$(1 - Y_{STO,NOx})/1.72$	$i_{N,SS}$	$(Y_{STO,NOx} - 1)/1.72$	0	0	$(i_{N,SS} + (1 - Y_{STO,NOx} - 1)/1.72)/14$	0	0	0	$Y_{STO,NOx}$	0	0
Aerobic growth of heterotrophic org.	(ρ_4)	$1 - (1/Y_{H,O2})$	0	0	$-i_{N,BM}$	0	0	0	$-(i_{N,BM}/14)$	0	0	1	$-(1/Y_{H,O2})$	0	0
Nitrite denitrification on S_{NO2}	(ρ_5)	0	0	$((1/Y_{H,NOx}) - 1)/1.72$	$-i_{N,BM}$	$(1 - (1/Y_{H,NOx}))/1.72$	0	0	$(-i_{N,BM} + (1 - (1/Y_{H,NOx}))/1.72)/14$	0	0	1	$-(1/Y_{H,NOx})$	0	0
Nitrate denitrification on S_{NO3}	(ρ_6)	0	0	$((1/Y_{H,NOx}) - 1)/2.86$	$-i_{N,BM}$	0	$(1 - (1/Y_{H,NOx}))/2.86$	0	$(-i_{N,BM} + (1 - (1/Y_{H,NOx})/2.86)/14$	0	0	1	$-(1/Y_{H,NOx})$	0	0
Aerobic endogenous respiration	(ρ_7)	$f_{XI} - 1$	0	0	$i_{N,BM} - (f_{XI} \cdot i_{N,XI})$	0	0	0	$(i_{N,BM} - (f_{XI} \cdot i_{N,XI}))/14$	f_{XI}	0	−1	0	0	0
Anoxic N_{O3} endogenous respiration	(ρ_8)	0	0	$(1 - f_{XI})/2.86$	$i_{N,BM} - (f_{XI} \cdot i_{N,XI})$	0	$(f_{XI} - 1)/2.86$	0	$((i_{N,BM} - (f_{XI} \cdot i_{N,XI})) + (1 - f_{XI})/2.86)/14$	f_{XI}	0	−1	0	0	0
Anoxic N_{O2} endogenous respiration	(ρ_9)	0	0	$(1 - f_{XI})/1.72$	$i_{N,BM} - (f_{XI} \cdot i_{N,XI})$	$(f_{XI} - 1)/1.72$	0	0	$((i_{N,BM} - (f_{XI} \cdot i_{N,XI})) + (1 - f_{XI})/1.72)/14$	f_{XI}	0	−1	0	0	0
X_{STO} aerobic respiration	(ρ_{10})	−1	0	0	0	0	0	0	0	0	0	0	−1	0	0
X_{STO} N_{O3} anoxic respiration	(ρ_{11})	0	0	1/2.86	0	0	−(1/2.86)	0	$1/(2.86 \cdot 14)$	0	0	0	−1	0	0
X_{STO} N_{O2} anoxic respiration	(ρ_{12})	0	0	1/1.72	0	−(1/1.72)	0	0	$1/(1.72 \cdot 14)$	0	0	0	−1	0	0
Nitrification, I step/aerobic growth of X_{ns}	(ρ_{13})	$1 - (3.43/Y_{A,ns})$	0	0	$-i_{N,BM} - (1/Y_{A,ns})$	$1/Y_{A,ns}$	0	0	$-(1/7) \cdot (1/Y_{A,ns}) - (i_{N,BM}/14)$	0	0	0	0	1	0
Nitrification, II step/aerobic growth of X_{nb}	(ρ_{14})	$1 - (1.14/Y_{A,nb})$	0	0	$-i_{N,BM}$	$-(1/Y_{A,nb})$	$1/Y_{A,nb}$	0	$-(i_{N,BM}/14)$	0	0	0	0	0	1
X_{ns} aerobic endogenous respiration	(ρ_{15})	$f_{XI} - 1$	0	0	$i_{N,BM} - (f_{XI} \cdot i_{N,XI})$	0	0	0	$(i_{N,BM} - (f_{XI} \cdot i_{N,XI}))/14$	f_{XI}	0	0	0	−1	0
X_{nb} aerobic endogenous respiration	(ρ_{16})	$f_{XI} - 1$	0	0	$i_{N,BM} - (f_{XI} \cdot i_{N,XI})$	0	0	0	$(i_{N,BM} - (f_{XI} \cdot i_{N,XI}))/14$	f_{XI}	0	0	0	0	−1
X_{ns} anoxic endogenous respiration	(ρ_{17})	0	0	$(1 - f_{XI})/1.72$	$i_{N,BM} - (f_{XI} \cdot i_{N,XI})$	$(f_{XI} - 1)/1.72$	0	0	$(i_{N,BM} - f_{XI} \cdot i_{N,XI}) - (f_{XI} - 1)/1.72)/14$	f_{XI}	0	0	0	−1	0
X_{nb} anoxic endogenous respiration	(ρ_{18})	0	0	$(1 - f_{XI})/2.86$	$i_{N,BM} - (f_{XI} \cdot i_{N,XI})$	0	$(f_{XI} - 1)/2.86$	0	$(i_{N,BM} - f_{XI} \cdot i_{N,XI}) - (f_{XI} - 1)/2.86)/14$	f_{XI}	0	0	0	0	−1
Hydrolysis	(ρ_{19})	0	$1 - f_{SI}$	0	$i_{N,XS} - i_{N,SS} \cdot (1 - f_{SI}) - (f_{SI} \cdot i_{N,SI})$	0	0	f_{SI}	$(i_{N,XS} - i_{N,SS} \cdot (1 - f_{SI}) - (f_{SI} \cdot i_{N,SI}))/14$	0	−1	0	0	0	0

Finally, the conversion rates so determined can be used to write the general mass balance equations of the biological reactor:

$$\dot{x}_i = \frac{1}{V}\left[Q_{in}(t)\cdot x_i^{in}(t) - Q_{out}(t)\cdot x_i(t) + r_i(x(t))\right] \text{ for } i = 2,\ldots,14. \tag{3.19}$$

Remembering that the reactors are modelled under the assumptions of being CSTRs with constant volume, and therefore it can be assumed that

$$Q_{in} = Q_{out}, \tag{3.20}$$

it is possible to define a quantity q with the dimension of the inverse of time:

$$q = \frac{Q_{in}}{V} = \frac{Q_{out}}{V}. \tag{3.21}$$

which allow to rewrite equation (3.19) in a simpler form:

$$\dot{x}_i = q\left[x_i^{in}(t) - x_i(t) + r_i(x(t))\right] \text{ for } i = 2,\ldots,14 \tag{3.22}$$

This is valid for all the state variables of the biological model except for x_1, i.e. S_{O2} (dissolved oxygen), where the contribution of the exchange with the atmosphere must be accounted for:

$$\dot{x}_1 = q\left[x_1^{in}(t) - x_1(t) + r_1(x(t))\right] + k_L a(T(t))\cdot\left[C_{sat}(T(t)) - x_1\right] \tag{3.23}$$

with C_{sat} given by the law:

$$C_{sat} = \exp\left[7.7117 - 1.31403\log(T(t) + 45.93)\right] \tag{3.24}$$

and

$$k_L a(t) = \left[K_a \cdot Air_{in}(t)\right]\cdot 1.024^{\,T(t)-20}. \tag{3.25}$$

The BSM1_LT model used as basis for this work is also provided, like the BSM1, with a dataset of synthetic data to be used as input, designed to account for different scenarios. Unlike its predecessor, though, only one dataset is provided for the long-term benchmark model, because in this case the simulation time spans over a year and there is room to include rain events and storm events alongside dry weather periods. The dataset is composed by the same variables of the files provided for the BSM1 protocol: time, S_S, $X_{B,H}$, X_S, X_I, S_{NH}, S_I, S_{ND}, X_{ND}, Q, S_O, $X_{B,A}$, X_P, S_{NO} and S_{ALK}, plus the addition of the temperature. The total period covered by the measurements is 609 days and, as in the BSM1/ASM files, the measurements are assumed to be sampled every 15 minutes. The temperature profile over the year is modelled as a sinusoidal trajectory with its peak value at the start of August and the minimum at the beginning of February (based on the hypothesis of a plant located in the northern hemisphere):

$$T[^{\circ}\mathrm{C}] = 15 + 5\cdot\cos\left[2\pi/365(\mathrm{time[d]} - 28)\right]. \tag{3.26}$$

In order to increase the realism of the temperature model another sinusoidal function with period of 1 day and amplitude of 1°C is summed to eq. (3.26), to account to the circadian variation of it.

The plant layout previously described for the ASM family and the BSM1 model has also been modified to account for the different configuration induced by the AC control strategy. In particular, with the intermittent aeration configuration the reactors are no more used exclusively for the aerobic or the anoxic processes but both can take place in the same reactor provided the aeration is active or not. This distinction of use has been therefore removed and instead of five separate tanks the biological reactor is assumed to consist in a larger one with a volume double than the total volume of the biological reactors of BSM1 (i.e. 12'000 m^3). The modelling of the settling process, instead, remains unchanged. Here below the main features of this plant configuration are listed:

- the plant configuration is composed of 1 biological reactor followed by a secondary settler (Figure 3.21);
- the biological reactor is fully mixed, both in the aerated and in the anoxic phase;
- the total volume of the biological reactor is 12'000 m^3 (but the recirculation flow rates and the excess sludge are kept equal to the original BSM);
- the biological reactor is intermittently supplied, according to the concentrations of ammonia and nitrates measured, with a fixed air flow-rate U_a = 18'000 m^3/d;
- the secondary settler has a volume of 6'000 m^3 (area of 1'500 m^2 and a depth of 4 m) subdivided in 10 layers, each 0.4 m high, and the feed point is set at 2.2 m from the bottom (6^{th} layer). The settler is modelled according to (Takács et al., 1991);
- the plant configuration has 1 recycle from the settler to the biological reactor with a fixed flow-rate 18'446 m^3/d flowing through a pipe with internal volume 300 m^3;
- the sludge waste is continuously pumped from the settler with a constant discharge of 385 m^3.

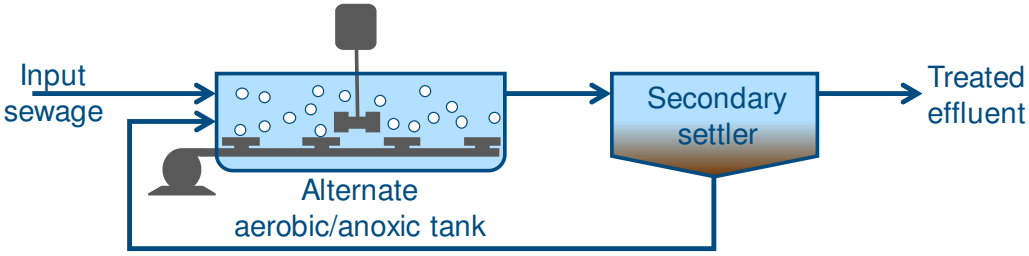


Figure 3.21 Schematic plant layout of the BSM model adapted for the intermittent aeration coniguration

In the real plant of Mantua, whose characteristics were presented in a previous paragraph, the intermittent aeration was governed by the OSCAR control supervisor based on the concentration measured in the aerobic tank. Unfortunately, the underlying control logics of OSCAR are unavailable because patented, therefore in the attempt of achieving similar results a simpler switching logic is implemented in the simulation model. The activation/deactivation of the blowers is regulated by a simple double relay controller depending on the concentrations of nitrogen and ammonia measured in the tank. Given the concentration thresholds for the NO_x^- and NH_4^+ $Th_{NOx,max}$, $Th_{NOx,min}$, $Th_{NH4,max}$, $Th_{NH4,min}$, and defining z_4 as the measurement of the ammonia (state x_4) and z_x as the sum of the NO_2^- and NO_3^- contribution,

$$z_x = f(x_x),\ x_x = x_5 + x_6 \tag{3.27}$$

the air is supplied according to the following rule:

$$Air_{in}(t) = \begin{cases} U_a, & \text{if } R_{NH_4}(t) \vee R_{NO_x}(t) \neq 0 \\ 0, & \text{if } R_{NH_4}(t) \vee R_{NO_x}(t) = 0 \end{cases} \tag{3.28}$$

where U_a is the fixed flow-rate used in the aerated reactors of the regular ASM models and $R_{NH4}(t)$ and $R_{NOx}(t)$ are the responses of the passive relays respectively:

$$R_{NH_4}(t) = \begin{cases} 0, & \text{if } z_4(t) < Th_{NH_4,min} \\ R_{NH_4}(t-1), & \text{if } Th_{NH_4,min} \leq z_4(t) \leq Th_{NH_4,max} \\ 1, & \text{if } z_4(t) > Th_{NH_4,max} \end{cases} \tag{3.29}$$

and

$$R_{NO_x}(t) = \begin{cases} 1, & \text{if } z_x(t) < Th_{NO_x,min} \\ R_{NO_x}(t-1), & \text{if } Th_{NO_x,min} \leq z_x(t) \leq Th_{NO_x,max} \\ 0, & \text{if } z_x(t) > Th_{NO_x,max} \end{cases} \tag{3.30}$$

In Table 3.11 are reported the threshold values that regulate the switching of the blowers.

Table 3.11 Threshold values regulating the activation of the blowers

Threshold	Value [mg/L]
$Th_{NH4,max}$	5
$Th_{NH4,min}$	2
$Th_{NOx,max}$	0.5
$Th_{NOx,min}$	10

As the performance of the AC process implemented depends on the switching logic of (3.28), it is crucial to recognise whether the concentration measured are reliable or not, which is exactly the purpose of this study. Therefore, when building a model with the scope of producing data to be used in this kind of analysis, one must not disregard a careful and exhaustive modelling of the sensors, especially those directly influencing

the control system, and of the disturbances affecting them. The implementation of the sensors in this work started from the one recommended for the simulation benchmark protocol and described in the previous paragraph. Some adjustments were required to adapt such a model to the examined case, mainly concerning the addition of a model for the drift and of the fouling phenomena which are considered important source of unreliability. For the sake of simplicity, the more refined sensor modelling just focused on the measurements directly affecting the control system: the concentration of NH_4^+ and NO_x^- in the biological reactor. The model wants to mimic the behaviour of a plant implementing an AC control configuration similarly to the one implemented in the Mantua plant, therefore it is assumed that the same category of sensors is used to feed the control logic, that is to say the ion-specific sensors. According to the classification of (Rieger et al., 2003) reported in Table 3.1 the ion-specific probes fall in the class A category and their transport function is modelled as a series of two first-order linear transport functions (Table 3.2) as in eq.(3.13). A noise component $n(t)$ is added to the sensor transformed signal following the relation (3.14) but with respect to the suggested values of noise level and measurement range of Table 3.3 some changes are introduced. It was observed that a noise level of 2.5 % of the measurement range resulted in a more disturbed signal compared to the one of the concentrations measured in the Mantua plant, therefore it was opted for a smaller coefficient, $nl = 0.5$ %. On the contrary the measurement range of NO_x^-, the measured variable accounting for both the nitrites and nitrates contributions (3.27), is extended from the 0-20 interval to 0-50 (see Table 3.12), consistently with the higher nitrate concentration observed in the operational data of the Mantua plant.

Table 3.12 Measurement range of the nitrate and ammonia sensors

Measured variable	Measurement range
S_{NOx} [g N/m^3]	0-50
S_{NH4} (low range) [g N/m^3]	0-20

As anticipated before the intent of the model is to produce a dataset of measurements as realistic as possible so that they can be used to train and assess the fault detection algorithms that will be discussed later, and to this purpose is instrumental a realistic description of the disturbances that affect the sensors. Therefore the sensor model proposed in (Rieger et al., 2003) is here improved incorporating a drift component $d(t)$ and a fouling component $fl(t)$, so that the disturbed measurement signal $z(t)$ returned by the modelled sensor be the result of the following combination:

$$z(t) = u_2(t) + d(t) + n(t) \cdot fl(t), \tag{3.31}$$

where $u_2(t)$ and $n(t)$ respectively are the output of the transfer function and the noise component previously described.

Moreover, it was also implemented a 'calibration phase' of the sensors. Every 14 days, the average calibration frequency observed in the Mantua plant, a step signal resets the coefficients of the simulated disturbances, i.e. resetting the random values used to modulate the effect of the perturbations induced by the drift and the fouling in a realistic way. This intended to mimic the effect of extracting and recalibrating the probe, cleaning it, repairing or substituting parts of it.

As well as the noise, the drift effect in the measurements is modelled as an additive component to the transformed signal. To incorporate the randomness of real measurements the drift has been designed (eq.(3.32)) as composed by a constant value $k_d = 0.08$, representing the maximum deviation that can affect the instrument, modulated by a coefficient $\delta(t)$ drawn from a uniform distribution $\mathcal{U}(-1,1)$ after every calibration:

$$\dot{d}(t) = k_d \cdot \delta(t) . \tag{3.32}$$

The fact that the random coefficient $\delta(t)$ can assume both positive and negative values accounts for the drift effect on the measurement to either increase or decrease the actual value of the measured variable. Furthermore, by randomly picking a value that can be either close or far from the previous one can account for the different kind of maintenance that can take place during the calibration procedure (cleaning, calibration, substitution of one or more components, etc.)

The effect of the fouling on the measurements, that is to say the disturbance on the measurement due to a reduced sensitivity and accuracy due to layers of dirt depositing on the sensor surface, is here modelled only in its macroscopic effect observed in the operational data of Mantua, i.e. the attenuation of the noise of the measurements. The effect of fouling on the sensor is therefore simulated as an exponential decay function $fl(t)$ that multiplies the random noise component of the concentration measured:

$$fl(t) = e^{-\alpha(t)\cdot t} , \tag{3.33}$$

where the exponent α(t) is the product of a constant and a stochastic value:

$$\alpha(t) = k_{fl} \cdot \phi(t) . \tag{3.34}$$

The constant coefficient k_{fl} is obtained assuming a maximum of 95% reduction of the measurement noise after 14 days since the last calibration of the sensor:

$$k_{fl} = -\ln(1-0.95)/\Delta t_{calib} \tag{3.35}$$

while the random ϕ is drawn after every calibration phase from a uniform distribution $\mathcal{U}(0,1)$. Also in this case, as for the drift, the idea is use to the constant coefficient to

impose a maximum attenuation level but, lacking a more detailed model of the fouling process that link the effect to the state of the biological reactor, account for the different possible fouling realizations by means of a stochastic variable (Figure 3.22).

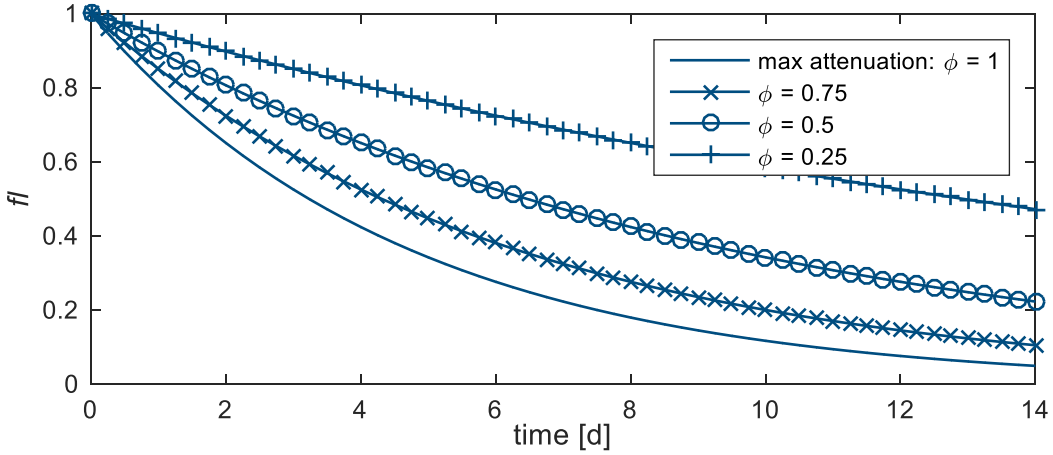


Figure 3.22 Example of attenuation levels during the 14 days period

3.4 Definition and detectability of faults and anomalies

The aim of this research is to design a FD method for the WWTPs implementing the intermittent aeration configuration for the optimal removal of the nitrogen concentrations. In this paragraph, will be first described the main characteristics of the faults and anomalies encountered in the analysis of the operational data and those introduced in the simulation model. Subsequently the diagnostic parameters extracted from the signals will be presented.

3.4.1 Fault definitions

The analysis of the historical operational data provided by the municipal plant of Mantua allowed to determine a set of recurring anomalies in the signals. The origin of some of them can be quite easily related to a malfunction of one or multiple measuring instruments, for some others, instead, is not straightforward to determine whether the disturbance is caused by an anomalous organic load concentration in the influent or by a deviance of the measurement device. Further on in the paragraph the terms ‘gross faults’ and ‘finer faults’ will be used to refer respectively to the first category of faults and to the latter. Notice that some of the anomalies observed in the operational data have been also introduced and simulated in the numerical model.

3.4.1.1 Gross faults

In the category of gross faults are collected all those anomalies in the measurements caused by a fault of the instrument and that mostly manifest themselves with a distinctive signature alteration of the signal. Here below are reported the most relevant.

The signal breakdown (Figure 3.23) is a discontinuity in the data flow that can be related to a power failure or to a malfunction of the communication channel. In a real-time scenario a warning should be issued if the acquisition system does not receive a data from the signals within a previously determined time frame threshold. In the off-line data analysis this malfunction can be easily revealed with a similar approach by comparing the timestamp label of two consecutive measurements.

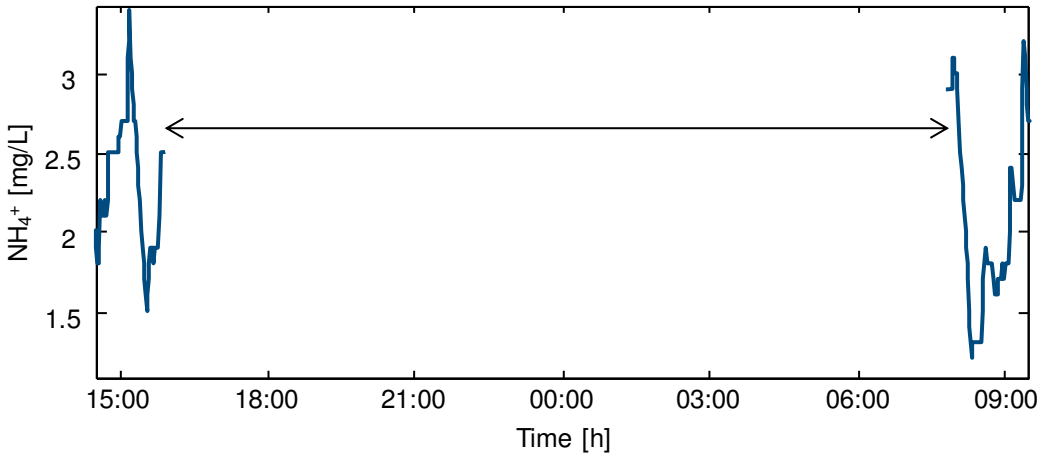


Figure 3.23 Example of signal breakdown fault

The presence of a fault is reported if the samples have a constant value over the same time-horizon as in (Figure 3.24). In this category, we include either the malfunctions related to ranging and out-of-scale problems of the instrument but also the communication anomalies that produce measurements ‘stuck’ on the same value for a long time.

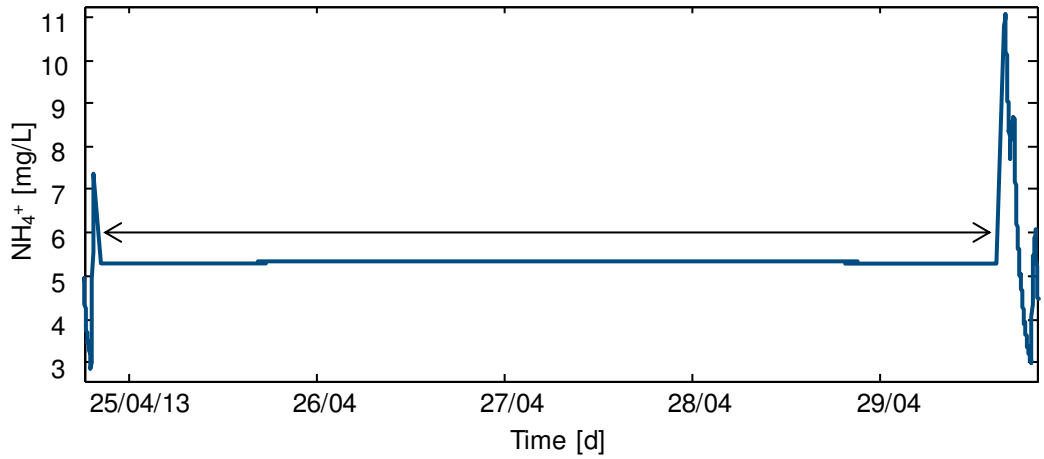


Figure 3.24 Example of constant signal (zero-scale) fault

A frequently encountered class of fault is represented by the spikes on the signal (Figure 3.25), short-lived artefacts possibly caused by electric disturbance. The

observed amplitude of the spikes is typically greater than 3 or 4 mg/L in the ammonia and nitrate measurements of the operational data. In the sensor model developed for the numerical model this kind of disturbances was not introduced.

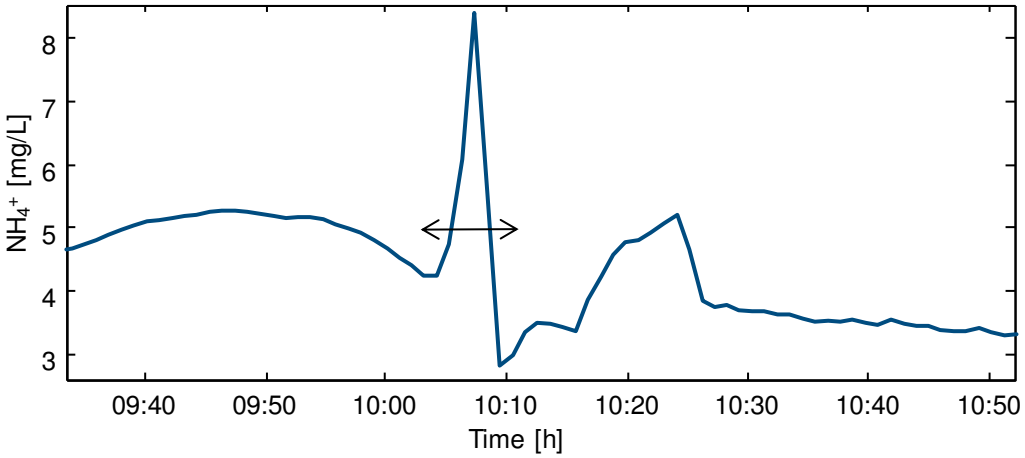


Figure 3.25 Example of spike on the ammonia concentrations measured

3.4.1.2 Finer faults

The anomalies that go under this category are characterized by a more complex affection on the signal, such that it is a hard task to recognise 'at glance' whether they are caused by a proper malfunction of the measurement instrument, and thus manifestation of an actual fault, or they are related to an alteration of the incoming pollutant concentrations. Generally, the only way to recognise a departure of the measurements from the actual concentration level of a compound would be to use multiple sensors or compare the measurements returned by the instrument with laboratory analysis, but unfortunately such comparison was not available for the data used in the present study. Being able to identify these faults and distinguish them from the process anomalies on a statistical basis suitable to provide a real-time estimation of the state of the instrument and its reliability was the driving motive of this research, therefore a proper identification of the fault episodes revealed to be crucial in the development of the FD procedure. Despite that, the off-line operational data from the Mantua plant. on which the detection algorithm has been first tested presented the inconvenience of poor description of the fault events, which prevented the possibility of a thorough training of the method as shall be seen in a few lines. The need of an accurate list of the instrument malfunction has been one of the main reasons that brought to the use of a numerical model for the data generation, where it has been possible to keep track of the anomalies introduced in the modelled sensors and easily distinguish their effect from those induced by the influent. Here below are presented some considerations over the finer faults observed.

One recurring example of finer anomaly is represented by a prolonged aeration phase (Figure 3.26), characterised, especially in the operational data, by high nitrate concentrations while the ammonia levels stay in their normal range; in fact, the proprietary logic of the control supervisor OSCAR prevent the ammonium overloads automatically switching the aeration on whenever the NH_4^+ concentration exceeds a given threshold. These anomalies can be either related to the influent pollutant content or an effect of a malfunction of the sensors. In the first case the prolonged aeration period would be a natural consequence of a high ammonium content in the influent, such to require an unusual amount of time to decrease its concentration below the set-point threshold for the deactivation of the blowers. In the latter case, however, a malfunction of the instrument resulting in an overestimation of the actual concentration in the tank would lead the control system to continue the air supply beyond the necessary with the consequence, among other factors, of an increase of the operational costs.

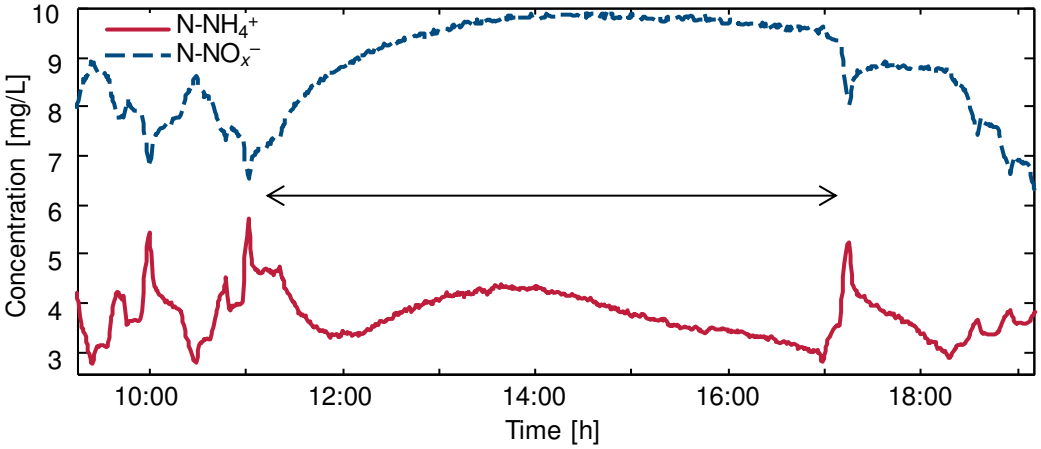


Figure 3.26 Example of prolonged aeration phase leading to high nitrate concentrations

As anticipated, distinguish between the operational and the instrumental origin of the anomaly observed is not trivial and requires the availability of a set of independent evaluations of the quantities of interest. Unfortunately, this kind of information was not provided for the data retrieved from the municipal plant of Mantua and this distinction could not be made. Therefore, as shall be seen later in section 5.1, the occurrence of this kind of anomaly in the signal could not be properly investigated with the more refined methods set up for the detection of the complex faults, and, similarly to the gross faults a simple method, based on a quite easy to determine feature, the duration, was implemented to filter it out from the finer detection. This inconvenience of course does not affect the data generated by the numerical model, since the disturbances perturbing the measurements are exactly known and separable from those affecting the input concentrations.

Another very common anomaly, observed in particular in the experimental plant data, is represented by periods of extremely low nitrogen concentrations in the tank (nitrate below 2 mg/L). In its typical pattern (Figure 3.27) the anoxic phases result longer than in the normal operative conditions and the nitrate concentrations slowly decrease towards almost zero and then increase again until the normal behaviour is restored, while the aerobic phases result generally extremely short. Again, this can be either related to a lower nitrogen content of the incoming influent or a drift of the ion-specific probe underestimating the concentrations measured but despite the apparent economic advantage of having minimum aeration periods the prolonged anoxic phase could introduce in the system an undesirably high amount of nitrogen and result in the violation of the normative constraints for the effluent quality.

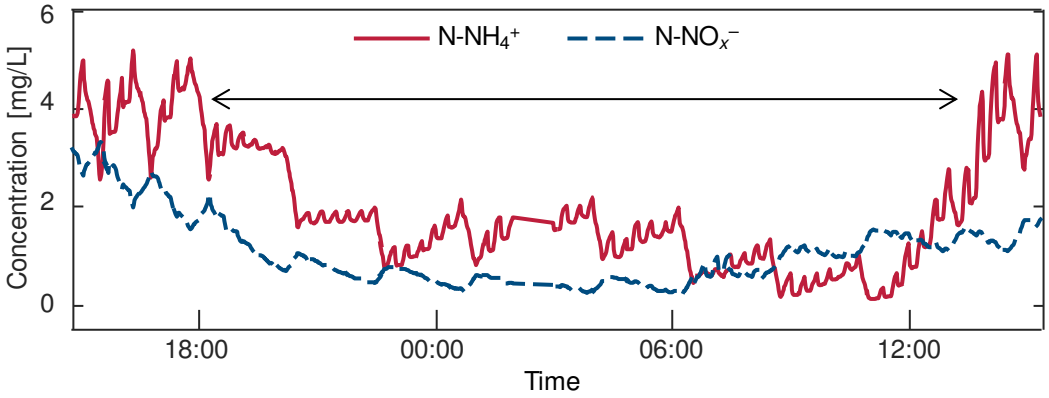


Figure 3.27 Example of low nitrate concentrations.

Other alterations affecting the instrument result in less evident effect on the measured signal, even more in a process operating with intermittently rising and decreasing concentrations. This made particularly hard to list the malfunctions occurring in the experimental data.

3.4.2 Preliminary screening of the gross malfunctions

Starting from the classification introduced in paragraph 3.4.1 and the considerations about the detectability of the different anomalies observed, a differentiated detection strategy for the ‘gross faults’ and the ‘fine faults’ is proposed. Gross faults are easier to identify and their detection is for the most part unrelated to the aerobic-anoxic phase discretization approach needed for the detection of the finer class of anomalies. This is particularly advantageous since it allows to issue a fault warning without waiting for the end of the operational phase. For these reasons the screening procedure to detect the gross anomalies is inserted in the FD framework proposed before the one devoted to the assessment of the fine faults. Here below is presented the detail of the preliminary check on the signal implemented for the different class of gross anomalies.

– Signal breakdown: given an average sampling interval of 30 s for the field measurements and 1 minute for those generated by the numerical model, a signal breakdown fault is diagnosed if no samples are received for ten minutes (approximately twenty samples).
– Constant signal: if the samples have a constant value over the same time-horizon as in the previous case (~10 minutes) a fault is reported. Both proper constant signal faults and zero-scale faults are detected using this approach.
– Spikes: from the analysis of the field data of the nitrate and the ammonia concentrations it was observed an average amplitude of the spikes between 3 and 4 mg/L while the maximum variation between consecutive measurements is less than 1 mg/L, therefore a detection threshold is set at a conservative value of 2 mg/L. So, if two subsequent samples differ by more than 2 mg/L, this is attributed to a spike disturbance, because it is very unlikely that such a sudden variation corresponds to an actual concentration change. This detection tool also allowed to discriminate the anomaly observed in the measurements of last hours of 7th of May 2013 (Figure 3.28), where it is present a strong measurement anomaly which has the characteristics of a signal overlapping, probably caused by a malfunction of the data logging system. In the data generated by the numerical model a more general criterion is used to set the detection threshold. For each type of sensor, the limit difference to detect the presence of a spike in the signals is set at 10 % of the respective recommended range of the measurements defined in Table 3.3.

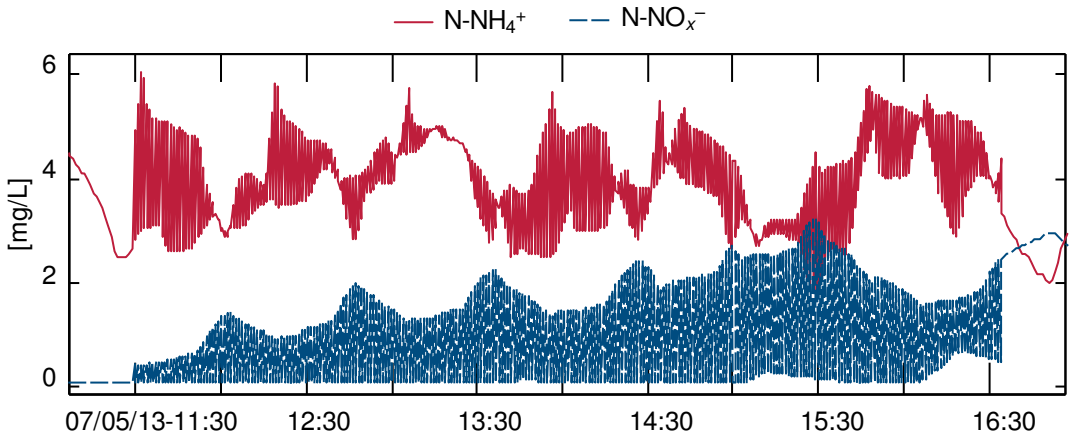


Figure 3.28 Signal overlapping on 7th May 2013.

– Phase duration anomaly: from the analysis of the field measurements emerged that the average duration of the aerobic and anoxic phases is between 1 and 2 hours but some episodes of longer phase duration were observed. In the operational data available, however, no clear indication was provided on

whether the anomalous duration of the phases was originated by an anomalous organic load of the influent or by an overestimation or underestimation of the actual concentration in the reactor inducing the process controller to respond with longer aerobic or anoxic durations. In order to filter out these cases and leave the refined fault detection methods to deal only with the correctly labelled anomalies, a preliminary screening is implemented to isolate the phases lasting more than 3 hours.

3.4.3 Trend analysis and parameter extraction

Differently from a traditional plant configuration, where the concentrations measured in the reactors are expected to change only according to changes in the influent and in the environmental conditions, in alternated cycles plant configurations the same tank periodically acts as aerobic and anoxic reactor resulting in ammonia and nitrate concentrations periodically increase and decrease according to the activation/deactivation of the blowers imposed by the process controller. The trends of the main process variables during the basic alternate cycle are shown in Figure 3.29. During the oxygenation phase ammonium nitrogen is oxidized to nitrate, whereas during the anoxic phase nitrate is reduced to molecular nitrogen and new ammonium enters the reactor. The state of the aerator is signalled by an On/Off Boolean signal.

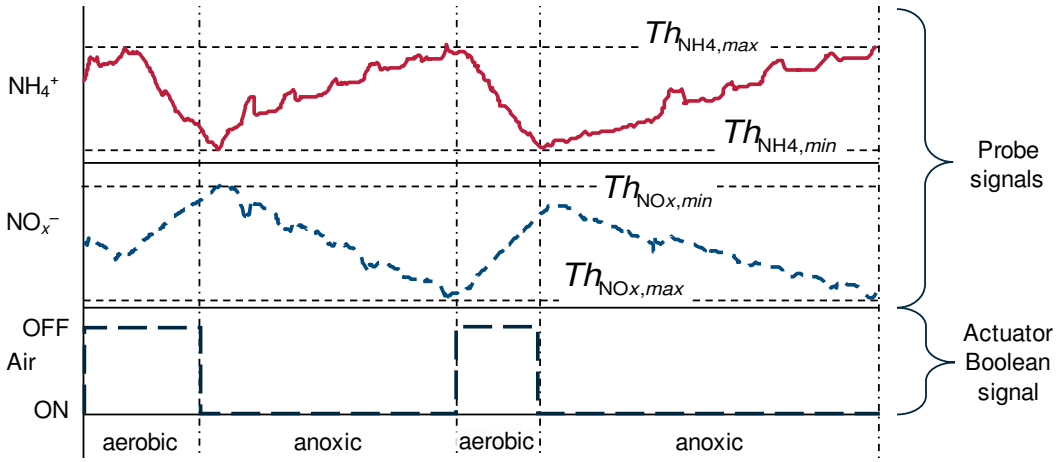


Figure 3.29 Basic alternate cycle and related signals. Two sensor outputs (NH_4^+, NO_x^-) come from the sensors and an actuator signals the state with a Boolean output.

3.4.3.1 Differing parametrization depending on the available data

This intermittent behaviour of the concentrations in an AC plant configuration naturally makes more difficult the task to infer the state of the sensor and the reliability of the measurements from the relation among the punctual value of a set of parameters. The mutually-exclusive switching of the aeration or of the mixer in the Mantua plant is

managed by a complex control system developed by E.T.C. engineering s.r.l. whose control logics are unavailable due to patent restriction and therefore are not taken into account, however from a general point of view the cause-effect relationship is such that the switching is related to the exceeding of the predefined thresholds for the NO_x^- and NH_4^+ concentrations. Based on this same principle a simplified switching rule (eq.(3.28)) is implemented in the numerical model to simulate the AC configuration. Starting from these considerations, a set of representative parameters from the trends of Figure 3.29, the probe signals are used as observed variables and the Boolean signals as phase delimiters.

As shall be seen further on in Chapter 5, given the intermittent nature of the process it has been decided to split the detection procedure of the more subtle anomalies in two parts, one for the aerobic phases and one for the anoxic ones, to be also performed intermittently. For this reason, the parameterization of the signals too was chosen in a way suitable to reflect this alternation, which means that the parameters computed are intended to be representative of the characteristics of the signal in the current phase.

The four parameters, shown in Figure 3.30, extracted in each phase, are reported in Table 3.13.

Table 3.13 Parameters extracted from the operational plant data of Mantua in each process phase

Description	Symbol
Average ammonia concentration in the biological reactor	$m_{\cdot,\mathrm{NH4+}}$
Growth or decay rate of the ammonia concentration in the biological reactor (slope $NH_4^+{}_{out}$)	$s_{\cdot,\mathrm{NH4+}}$
Average nitrates concentration in the biological reactor	$m_{\cdot,\mathrm{NO}x-}$
Growth or decay rate of the nitrate concentration in the biological reactor (slope $NO_x^-{}_{in}$)	$s_{\cdot,\mathrm{NO}x-}$

The rationale behind this choice lies in the physical meaning of the parameters computed: the slopes of the ammonia and nitrate signals in the aerobic phase represent the two aspects of the nitrification process, the decrease of the ammonia concentrations due to the oxidation process and the consequent increase in the nitrates concentrates; the meaning of the two average concentration values is instead straightforward. The same rationale just explained for the aerobic phase also applies for the anoxic phase, with the only difference of course that the two slopes computed in his case represents the two sides of the denitrification process. Despite the perturbations given by the incoming ammonia from the influent and nitrogen from the recycle, the idea is that there is a close relationship between the two increase/decrease rates and the average concentrations of the quantities of interest in the reactor, and this relationship is expected not to change, at least not quickly and not in the middle term time range. Here below the detail of the

computational procedure for the parameter extraction is presented, focusing in particular on the computation of the concentration rations.

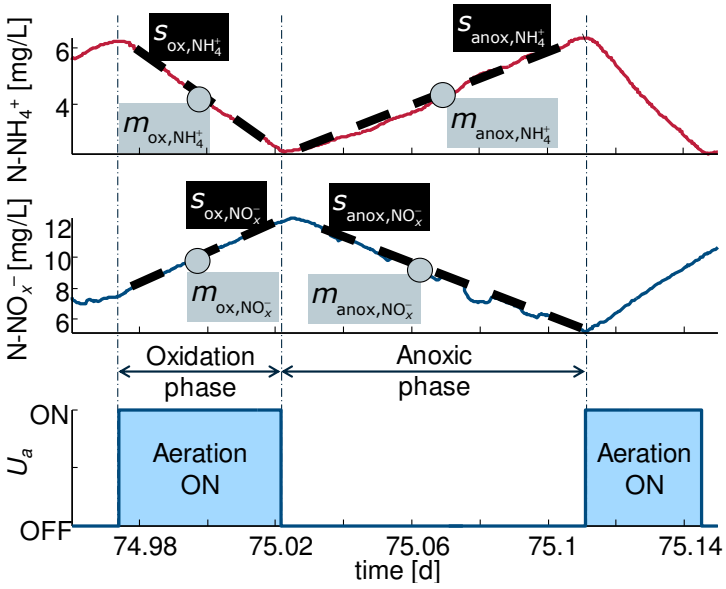


Figure 3.30 Definition of the fault detection parameters and related computing scheme. The parameters indicated have the following meaning: $m_{ox,NH4+}$= NH_4^+ nitrification average concentration; $m_{ox,NOx-}$ = NO_x^-, nitrification average concentration; $s_{ox,NH4+}$ = NH_4^+ nitrification slope; $s_{ox,NOx-}$ = NO_x^- nitrification slope; $m_{anox,NH4+}$ = NH_4^+ denitrification average concentration; $m_{anox,NOx-}$ = NO_x^- denitrification average concentration; $s_{anox,NH4+}$ = NH_4^+ denitrification slope; $s_{anox,NOx-}$ = NO_x^- denitrification slope

A pre-sampling delay Δt at the beginning of each phase is necessary to avoid computing the slope with uncertain data, before the true slope can be reliably computed. The initial part of each phase has a variable latency depending on the conditions of the cycle. To discard the initial data a moving window of M samples is considered and the slopes are computed as

$$s_{w,N} = \frac{1}{M_{\cdot,N}} \sum_{i=1}^{M_{\cdot,N}-1} \frac{N(i+1)-N(i)}{t(i+1)-t(i)}, \qquad N = \mathrm{NH_4^+},\ \mathrm{NO_x^-}. \tag{3.36}$$

The moving window is advanced by one sample and the slopes are compared with pre-set threshold values $\bar{S}$. When the conditions of eq. (3.37) are satisfied,

$$\begin{aligned} &\text{Aerobic phase} && \mathrm{NH_4^+}: && \bar{S}_{\mathrm{ox,NH_4^+}} > s_{w,\mathrm{NH_4^+}} \\ &&& \mathrm{NO_x^-}: && \bar{S}_{\mathrm{ox,NO_x^-}} < s_{w,\mathrm{NO_x^-}} \\ &\text{Anoxic phase} && \mathrm{NH_4^+}: && \bar{S}_{\mathrm{anox,NH_4^+}} < s_{w,\mathrm{NH_4^+}} \\ &&& \mathrm{NO_x^-}: && \bar{S}_{\mathrm{anox,NO_x^-}} > s_{w,\mathrm{NO_x^-}} \end{aligned} \tag{3.37}$$

the slope is computed as the regression line based on all the remaining phase samples, as shown in Figure 3.31. The blower on/off signal was used to detect the phase edges.

Table 3.14 Parameters of the moving windows used to compute the signal slopes

Parameter	Aerobic phase	Anoxic phase
Ammonia window length ($M_{\cdot,NH4+}$)	10	5
Ammonia slope threshold level ($\bar{S}_{\cdot,NH4+}$)	–20	40
Nitrates window length ($M_{\cdot,NOx-}$)	10	5
Nitrates slope threshold level ($\bar{S}_{\cdot,NOx-}$)	20	–40

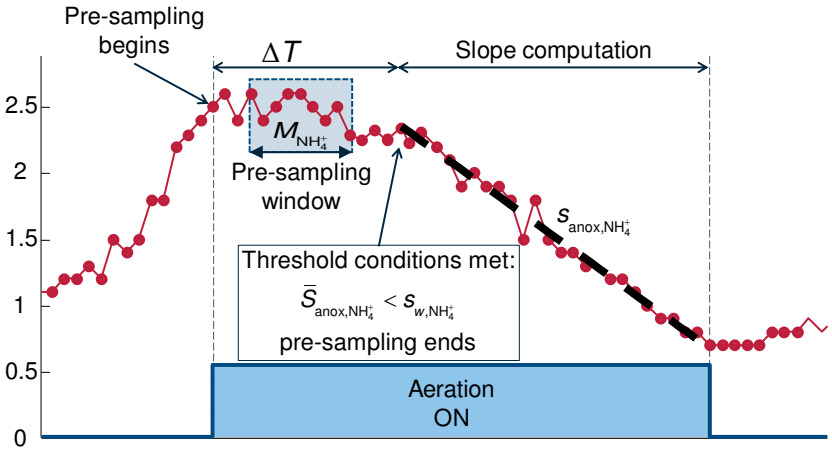


Figure 3.31 An example of slope computation in the anoxic NH_4^+ case. The moving window is M_{NH4+} samples long and scans the data until the conditions of Eq. (3.37) are satisfied. The slope computation begins from that point using all the remaining data until the end of the phase.

As already observed, layers of dirt tend to deposit on the surface of the measurement instruments causing their fouling. The most evident effect of this phenomenon is that the signal of a fouled probes appears smoother, less sensitive to the small concentration variations, when compared with the more coarse shape of a clean one. Moreover, although a thorough quantitative investigation of the cause-effect relationship was not the objective of this research, it is safe to assume that a smoother, and therefore fouled, signal is likely to return biased measurements, thus the evaluation of the smoothness of the concentration profiles can be used as a qualitative assessment of the state of the instrument. The particular parameterization chosen returns an interesting by-product that can be used as well as a diagnostic feature for the mentioned purpose: the correlation coefficient r^2, associated to the computation of the regression lines. Later on, in the chapter describing the details of the FD method implemented, it shall be presented how this parameter is linked to the state of the sensor.

One of the major limitations in the field measurements used consisted in the availability of just two signals: the nitrate and ammonia concentrations in the biological reactor, and in the consequently limited amount of information that could be extracted from. Using the numerical model, instead, it has been possible to simulate the sensors at will and thus to draw on a higher number of measurements, consequently increasing the discriminatory power of the methods devised. Keeping the same parameterization framework used for the plant data, i.e. the division in aerobic and anoxic phases, a

number of sensors have been introduced, being careful that they gauge physical quantities related to the nitrification/denitrification process and are actually measurable in a real plant. As for what concerns the ammonia and nitrogen concentrations in the aerobic tank, the same parameterization presented for the operational data (average values and slopes) is maintained. It is worth mentioning, also, that the NO_x^- measurements are simulated summing the contributions of the state variables describing the nitrites and the nitrates, respectively S_{NO2} (x_5) and S_{NO3} (x_6), as in eq. (3.27). The ammonia measurements considered instead are relative to its specific state variable: S_{NH4} (x_4).

All the other sensors used have been modelled using the simpler approach proposed by the benchmark protocol, which means without other disturbances than the noise, as in eq.(3.15). The additional monitored variables considered are the temperature, the concentration of dissolved oxygen (state variable x_1, S_{O2}) and the concentration of suspended solids (state variable x_2, S_S). These variables have been parameterized using just the average value and the difference between the maximum and minimum measured value in the time spanned by the respective process phase, since no other meaningful relation was found. Table 3.15 reports the list of the parameters retrieved for each phase from the numerical model.

Table 3.15 Parameters extracted for each process phase from the data generated by the numerical model

Description	Symbol
Average ammonia concentration entering the biological reactor	$m_{\cdot,NH4+,in}$
Difference between the maximum and minimum ammonia concentration entering the biological reactor (range $NH_4^+{}_{in}$)	$rg_{\cdot,NH4+,in}$
Average ammonia concentration in the biological reactor	$m_{\cdot,NH4+,out}$
Difference between the maximum and minimum ammonia concentration in the biological reactor (range $NH_4^+{}_{out}$)	$rg_{\cdot,NH4+,out}$
Growth or decay rate of the ammonia concentration in the biological reactor (slope $NH_4^+{}_{out}$)	$s_{\cdot,NH4+,out}$
Average nitrates concentration entering the biological reactor	$m_{\cdot,NOx-,in}$
Difference between the maximum and minimum nitrates concentration entering the biological reactor (range $NO_x^-{}_{in}$)	$rg_{\cdot,NOx-,in}$
Average nitrates concentration in the biological reactor	$m_{\cdot,NOx-,out}$
Difference between the maximum and minimum nitrates concentration in the biological reactor (range $NO_x^-{}_{,out}$)	$rg_{\cdot,NOx-,out}$
Growth or decay rate of the nitrate concentration in the biological reactor (slope $NO_x^-{}_{out}$)	$s_{\cdot,NOx-,out}$
Average temperature	$m_{\cdot,Temp}$
Average dissolved oxygen concentration entering the biological reactor	$m_{\cdot,DO,in}$
Difference between the maximum and minimum dissolved oxygen concentration entering the biological reactor (range DO_{in})	$rg_{\cdot,DO,in}$
Average dissolved oxygen concentration in the biological reactor	$m_{\cdot,DO,out}$
Difference between the maximum and minimum dissolved oxygen concentration in the biological reactor (range DO_{out})	$rg_{\cdot,DO,out}$

Description	Symbol
Average suspended solids concentration entering the biological reactor	$m_{\cdot,\mathrm{TSS,in}}$
Difference between the maximum and minimum suspended solids concentration entering the biological reactor (range TSS_{in})	$rg_{\cdot,\mathrm{TSS,in}}$
Average suspended solids concentration in the biological reactor	$m_{\cdot,\mathrm{TSS,out}}$
Difference between the maximum and minimum suspended solids concentration in the biological reactor (range TSS_{out})	$rg_{\cdot,\mathrm{TSS,out}}$

The subscript ‘in’ refers to the variables measured before entering the aerobic reactor while the subscript ‘out’ referring to the concentrations measured inside the tank is justified by the assumption that the reactor behaves like a CSTR with constant volume (eq.(3.20)) therefore the concentrations measured inside the tank are the same that exit.

The fault detection algorithms proposed have been tested using different combinations of the parameters in Table 3.15, as shall be seen in the following chapters.

Chapter 4 Review of methods for the fault detection

Sophisticated sensors such as Ion-Specific Electrodes (ISE) are essential for a precise control policy but they require a careful maintenance. For this reason, a monitoring system including a fault detection device is instrumental for a successful management. This device should also be capable of discriminating a sensor failure to a process anomaly, such as overload or toxic spillage. Hence the need to design a smart fault detection algorithm with these capabilities.

Generally, the Fault Detection and Isolation (FDI) algorithms can be divided into three main categories (Venkatasubramanian et al., 2003): quantitative model-based methods, qualitative knowledge-based methods and process history-based methods.

Quantitative model-based methods require a deep knowledge of the process and proved to be extremely powerful tools, even though their application is limited to linear or very specific nonlinear problem at most. These kinds of methods can be, for example, represented by one or more observers, i.e. models used to provide an estimation of the relevant signals and allowing to detect abnormal process behaviours when the system-extracted signal is compared to the estimated one. Some successful recent applications can be found in (Nagy-Kiss and Schutz, 2013), where the diagnosis of a nonlinear system such as a WWTP is achieved with a multi-model approach using a set of nonlinear observers, and in (Xu et al., 2014), where the effectiveness of a new approach based on a bank of interval observers performing both fault detection and fault isolation is proven on a CSTR case study. However, despite the potentials and the efforts made to make easier their use (Schraa et al., 2006), this kind of methods are still the least popular among the FDI techniques, due to the high system complexity and the lack of good data from which develop an accurate model.

Qualitative model-based fault diagnostics is performed each time one possesses a fundamental understanding of the process behaviour but not so deep to allow the expression of the input-output relationships in terms of mathematical functions. Typical qualitative model-based methods, widely used in risk assessment studies, are the so-called fault trees i.e. logic trees that relate primary events with hazards. During the construction process the cause-effect relationships are derived by asking questions, to workers and experts in general, on what could produce the hazard and connecting the answers by logic nodes.

However real processes can rarely be represented as deterministic systems and most of the times the diagnostic task requires a quantitative assessment rather than a qualitative one. For this reason, instead of the model-based approach, that needs a complete a priori knowledge of the process, the vast majority of the FDI approaches rely on methods able to extract that knowledge from the history data, i.e. perform a feature extraction. The features can then be processed by non-statistical data mining methods, such as neural networks (NN), statistical methods, such as principal component analysis (PCA) or partial least squares (PLS), classification methods such as classification trees and support vector machines (SVM), or by a combination of them (Fuente et al., 2012, 2011; Goode & Chow, 1994; Ruiz et al., 2011; Sainz, 2004).

As anticipated in the introduction it was decided to treat the detection problem as a classification problem, therefore most of the methods tested are based on well-known classifiers. In this chapter is presented the theoretical basis of the methods employed in the next chapter to investigate the presence of faults in the measurements. After some general considerations on the data structure and the notation used the PCA method is introduced either in its classical formulation and in the subsequent extensions to the nonlinear, dynamical or batch processes. The second paragraph will be focused on the methods based on Bayes' Theorem and their application to the FD in terms of classification. The third class of methods treated will be the based on the classification trees, especially concentrating on the binary classification trees. Finally, the methods based on the SVM will be examined.

In order to facilitate the comparison among the different approaches the methods will be applied to the same set of data. Let $\boldsymbol{V} \in \Re^{N \times L}$ be a matrix containing the noise-free values $v_{i,j}$ of a process described by L variables (or features) in N samples. Typically, the number of monitored variables are usually much less than the number of samples, therefore hereafter it will be assumed that $L \leq N$. It is then possible to define a data matrix $\boldsymbol{X} \in \Re^{N \times L}$ containing the measurements of $\boldsymbol{V}$. The relationship between the values and their respective measurements is given by:

$$x_{i,j} = v_{i,j} + \varepsilon_{i,j}, \tag{4.1}$$

where $\varepsilon_{i,j}$ are the deviations between the true value and its corresponding measurement, i.e. the measurement errors.

Generally, only the measurements $x_{i,j}$ are known and, in absence of further information, it is assumed that the mean error is zero:

$$\mathbb{E}[\varepsilon_{i,j}] = 0. \tag{4.2}$$

In order to perform statistical inference, it is necessary to satisfy two additional requirements: the process variables must follow a multivariate (MV) normal distribution:

$$\boldsymbol{v}_{i,\cdot} \sim \mathrm{MV}\,\mathcal{N}(\mu,\sigma) \tag{4.3}$$

and the measurement errors must be independent and drawn from the same normal distribution:

$$\varepsilon_{i,j} \sim \mathcal{N}(0,\sigma_\varepsilon)\,. \tag{4.4}$$

The measurements of $\boldsymbol{X}$ often belong to different physical variables having different units of measurements or orders of magnitude. For this reason, a new matrix of standardized data $\tilde{\boldsymbol{X}}$ can be created, centring and scaling the values of $\boldsymbol{X}$ as follows:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - u_j}{s_j}, \quad i = 1,\ldots,N;\ j = 1,\ldots,L, \tag{4.5}$$

Usually u_j is chosen as the mean of the measurements of the variable j and s_j as its variance or its standard deviation. According to this choice, the notation of u_j and s_j will be henceforth changed respectively in μ_j and σ_j and they will be defined as:

$$\begin{gathered} u_j = \mu_j, \quad \mu_j = \frac{1}{N}\sum_{i=1}^{N} x_{i,j}, \quad j = 1,\ldots,L; \\ s_j = \sigma_j, \quad \sigma_j = \sqrt{\frac{\sum_{i=1}^{N}(x_{i,j} - \mu_j)^2}{N}}, \quad j = 1,\ldots,L\,. \end{gathered} \tag{4.6}$$

Therefore, the centring and scaling of $\boldsymbol{X}$ will become:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}, \quad i = 1,\ldots,N;\ j = 1,\ldots,L\,. \tag{4.7}$$

A very common practice is the so called "standardization" of $\boldsymbol{X}$, where μ_j and σ_j, the mean and the variance over the variable j, are set respectively to 0 and 1.

Once the measurement matrix has been centred and scaled it is possible to compute the covariance matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \mathrm{cov}\ \tilde{\boldsymbol{X}} = \frac{1}{L-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}} \in \Re^{L\times L}. \tag{4.8}$$

Note that if $\tilde{\boldsymbol{X}}$ is standardized than the covariance matrix is the same as the correlation matrix, i.e. $c_{i,i} = 1, \forall\ i = 1,\ldots,L$. Coherently with this notation, the following paragraphs, the matrix $\boldsymbol{X}$ will indicate the matrix of the measurements without scaling and $\tilde{\boldsymbol{X}}$ the standardized version of $\boldsymbol{X}$.

4.1 Fault detection based on the principal component analysis

First introduced by (Pearson, 1901) and subsequently developed by (Hotelling, 1947), PCA (also known as Karhunen-Loève transform) is a non-parametric method (Shlens, 2014) widely used to emphasize the informative content of a large number of variables, where the information may be masked by noise and data cross-correlation. It provides a linear transformation of the original variables into a new set of optimally uncorrelated orthonormal features, called principal components (PCs), converting a data-rich and information-poor data set into new data for which the information to data ratio is higher. FDI methods based on PCA have been widely used in the context of WWTP fault detection (Rosén and Olsson, 1998; Corominas et al., 2011; Garcia-Alvarez et al., 2011; Garcia-Alvarez et al., 2012; Villez et al., 2013). Here below the definitions of the classical PCA method and some extensions to the nonlinear problems are presented. Unless otherwise specified the PCA methods will be applied to the standardized version of the measurements matrix, so that all the features have the same order of magnitude.

4.1.1 Identification of the principal components

Once $\boldsymbol{X}$ has been standardized the aim of the PCA is to find a transform of the scaled measurement set $\tilde{\boldsymbol{X}}$ that maximize the variance, represented by the diagonal elements of the covariance matrix $\boldsymbol{C}$, and at the same time reduce de redundancy, represented by the elements outside the diagonal.

Considering a set of K new variables (with $K \leq L$), linear combination of the variables of $\tilde{\boldsymbol{X}}$:

$$\begin{aligned} z_{i,1} &= \tilde{\boldsymbol{x}}_{i,\cdot} \cdot \boldsymbol{w}_{\cdot,1} = \tilde{x}_{i,1} \cdot w_{1,1} + \tilde{x}_{i,2} \cdot w_{2,1} + \cdots + \tilde{x}_{i,L} \cdot w_{L,1} \\ z_{i,2} &= \tilde{\boldsymbol{x}}_{i,\cdot} \cdot \boldsymbol{w}_{\cdot,2} = \tilde{x}_{i,2} \cdot w_{1,2} + \tilde{x}_{i,2} \cdot w_{2,2} + \cdots + \tilde{x}_{i,L} w_{L,2} \\ &\vdots \\ z_{i,K} &= \tilde{\boldsymbol{x}}_{i,\cdot} \cdot \boldsymbol{w}_{\cdot,K} = \tilde{x}_{i,1} \cdot w_{1,K} + \tilde{x}_{i,2} \cdot w_{2,K} + \cdots + \tilde{x}_{i,L} \cdot w_{L,K} \end{aligned} \tag{4.9}$$

the new variance and covariance matrices can be obtained from the covariance of the scaled measurements:

$$\begin{aligned} \operatorname{var}(\boldsymbol{z}_{\cdot,k}) &= \boldsymbol{w}_{\cdot,k}^T \cdot \boldsymbol{C} \cdot \boldsymbol{w}_{\cdot,k} \quad \text{with } k = 1,\ldots,K\,, \\ \operatorname{cov}(\boldsymbol{z}_{\cdot,h}, \boldsymbol{z}_{\cdot,k}) &= \boldsymbol{w}_{\cdot,k}^T \cdot \boldsymbol{C} \cdot \boldsymbol{w}_{\cdot,k} \quad \text{with } h,k = 1,\ldots,K\,. \end{aligned} \tag{4.10}$$

The PCA transform is thus obtained as the linear combination that maximizes the variance while at the same time removing their correlation, i.e. satisfying the relation:

$$\text{cov}(\boldsymbol{z}_{\cdot,h}, \boldsymbol{z}_{\cdot,k}) = 0 \quad \text{with } h,k = 1,\ldots,K \,|\, h \neq k \,. \tag{4.11}$$

The first principal component (PC_1) is computed as the vector of the new base that maximizes the variance of $\boldsymbol{z}_{\cdot,1}$ and has unit norm:

$$\max_{\boldsymbol{w}_{\cdot,1}^T \cdot \boldsymbol{w}_{\cdot,k}} \text{var}(\boldsymbol{z}_{\cdot,1}) = \max_{\boldsymbol{w}_{\cdot,k}^T \neq 0} \text{var}(\boldsymbol{w}_{\cdot,1}^T \cdot \tilde{\boldsymbol{X}}) \,. \tag{4.12}$$

The second PC (PC_2) is then obtained as the non-null vector $\boldsymbol{w}_{\cdot,2}$ that maximizes $\text{var}(\boldsymbol{z}_{\cdot,2})$ while being orthogonal to the first principal component, $\boldsymbol{w}_{\cdot,1}$, and constrained to have unit norm:

$$\max_{\boldsymbol{w}_{\cdot,2}^T \cdot \boldsymbol{w}_{\cdot,2}=1,\, \text{cov}(\boldsymbol{z}_{\cdot,1}, \boldsymbol{z}_{\cdot,2})=0} \text{var}(\boldsymbol{z}_{\cdot,2}) = \max_{\boldsymbol{w}_{\cdot,2}^T \cdot \boldsymbol{w}_{\cdot,2}=1,\, \text{cov}(\tilde{\boldsymbol{X}} \cdot \boldsymbol{w}_{\cdot,1}, \tilde{\boldsymbol{X}} \cdot \boldsymbol{w}_{\cdot,2})=0} \text{var}(\tilde{\boldsymbol{X}} \cdot \boldsymbol{w}_{\cdot,2}) \,. \tag{4.13}$$

All other components can be determined following this procedure, maximizing the variance of additional linear combination and with the constraint of being orthogonal to the previous PCs. In general, the k-th component can be determined as:

$$\max_{\substack{\boldsymbol{w}_{\cdot,k}^T \cdot \boldsymbol{w}_{\cdot,k}=1, \\ \text{cov}(\boldsymbol{z}_{\cdot,h}, \boldsymbol{z}_{\cdot,k})=0}} \text{var}(\boldsymbol{z}_{\cdot,k}) = \max_{\substack{\boldsymbol{w}_{\cdot,k}^T \cdot \boldsymbol{w}_{\cdot,k}=1, \\ \text{cov}(\tilde{\boldsymbol{X}} \cdot \boldsymbol{w}_{\cdot,h}, \tilde{\boldsymbol{X}} \cdot \boldsymbol{w}_{\cdot,k})=0}} \text{var}(\tilde{\boldsymbol{X}} \cdot \boldsymbol{w}_{\cdot,k}) \,, \quad \forall h < k \,. \tag{4.14}$$

Note that one can at most determine a number principal components equal to the number of variables composing the measurements matrix, i.e. $K \leq L$

The so obtained PCs $\boldsymbol{w}_{\cdot,k}$ represent the eigenvectors of the covariance matrix $\boldsymbol{C}$ while the corresponding eigenvalues λ_k are the variance of the linear combinations:

$$\lambda_k = \text{var}(\boldsymbol{z}_{\cdot,k}) = \text{var}(\boldsymbol{w}_{\cdot,k} \cdot \tilde{\boldsymbol{X}}) \,. \tag{4.15}$$

Notice that since the covariance $\boldsymbol{C}$ is assumed to be positive definite, the eigenvalues are positive. Sorting the eigenvalues by descending order and sorting accordingly the corresponding eigenvectors the result is the same as following the procedure above described. Moreover, the PCs can be more easily obtained by diagonalizing the covariance matrix for the maximum possible number of components, i.e. for $K = L$:

$$\text{cov}(\tilde{\boldsymbol{X}}) = \boldsymbol{W} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{W}^T \,, \tag{4.16}$$

where $\boldsymbol{\Lambda} \in \Re^{L \times L}$ is the diagonal matrix of the L eigenvalues and $\boldsymbol{W} \in \Re^{L \times L}$ is the matrix of corresponding eigenvectors.

In PCA it is common to refer to the principal components, the eigenvectors $\boldsymbol{w}_{\cdot,k}$, as 'loadings'. Their linear combinations $\boldsymbol{z}_{\cdot,k}$, instead, are usually called 'principal scores', or simply 'scores', and represent the coordinates of the measurement $\tilde{\boldsymbol{X}}$ in the new system:

$$\boldsymbol{Z} = \tilde{\boldsymbol{X}} \cdot \boldsymbol{W} \in \Re^{N \times L} \,. \tag{4.17}$$

A direct consequence of the transform applied is that the covariance matrix compute for $\boldsymbol{Z}$,

$$\mathrm{cov}(\boldsymbol{Z}) = \frac{1}{L-1}\boldsymbol{Z}^T \cdot \boldsymbol{Z} = \frac{1}{L-1}\boldsymbol{\Lambda}, \tag{4.18}$$

is diagonal, thus any information redundancy in the original data have been eliminated, although they can be retrieved by the inverse transform $\tilde{\boldsymbol{X}} = \boldsymbol{Z} \cdot \boldsymbol{W}^T$.

4.1.2 Principal component analysis for dimensionality reduction

From the descending order chosen to sort the eigenvalues, and their respective eigenvectors, comes the possibility to reduce the dimensionality of $\boldsymbol{W}$ to an acceptable degree of information loss. Choosing the first K largest eigenvalues ($K < L$), and therefore the corresponding K eigenvectors, it is possible to build the matrix $\boldsymbol{W}_K \in \Re^{L \times K}$, where the subscript indicates that is a reduced version of $\boldsymbol{W}$ containing only the K retained components. It is now possible to project the measurements onto the lower-dimensional space, obtaining the scores $\boldsymbol{Z}_K$:

$$\boldsymbol{Z}_K = \tilde{\boldsymbol{X}} \cdot \boldsymbol{W}_K, \text{ with } \boldsymbol{Z}_K \in \Re^{N \times K}. \tag{4.19}$$

Similarly as before, one can obtain the (approximated) measurements matrix by the inverse transform:

$$\hat{\tilde{\boldsymbol{X}}} = \boldsymbol{Z}_K \cdot \boldsymbol{W}_K^T. \tag{4.20}$$

However, the measurement matrix retrieved is not the same as $\tilde{\boldsymbol{X}}$ because of the dimensional reduction. In order to get the exact same measurement matrix, one must add the other components not retained:

$$\tilde{\boldsymbol{X}} = \sum_{k=1}^{K} \boldsymbol{z}_{:,k} \cdot \boldsymbol{w}_{:,k}^T + \sum_{h=K}^{L} \boldsymbol{z}_{:,h} \cdot \boldsymbol{w}_{:,h}^T. \tag{4.21}$$

The vector of residuals between the original and the reconstructed is defined as follows:

$$\boldsymbol{r}_{i,:} = \hat{\tilde{\boldsymbol{x}}}_{i,:} - \tilde{\boldsymbol{x}}_{i,:}. \tag{4.22}$$

4.1.2.1 Choice of the number of principal components to retain

From eq.(4.15) follows an important property concerning the relative variance captured by each component, which is to say that the variance captured by the k-th component is equal to ratio of the k-th eigenvalue over the sum of all the eigenvalues:

$$\mathrm{RV}(k) = \frac{\mathrm{var}(\boldsymbol{z}_{\cdot,k})}{\mathrm{tr}(\boldsymbol{\Lambda})} = \frac{\lambda_k}{\sum_{h=1}^{L} \lambda_h}. \tag{4.23}$$

This property can be used to select the appropriate number of principal components to retain by calculating the cumulated relative variance of the first K components:

$$\mathrm{CRV}(K) = \frac{\sum_{k=1}^{K} \mathrm{var}(\boldsymbol{z}_{\cdot,k})}{\mathrm{tr}(\boldsymbol{\Lambda})} = \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{h=1}^{L} \lambda_h}, \text{ with } K < L. \tag{4.24}$$

From eq.(4.24) it is clear that once the PCA transform is applied, the more components are retained the less information is lost. On the other hand, the informative contribution of the latter variables is increasingly lower and comes at the price of an increased computational burden. For this reason, one shall find a compromise between the dimensional reduction of the problem and the informative content retained.

It is no surprise that a number of method of selecting a convenient number of PCs have been developed. The most common methods, being the already mentioned PC selection by means of captured variance, the eigenvalue scree plot, data scrambling and data reconstruction method, are reviewed in the following sub-paragraphs.

4.1.2.1.1 Explained variance

Based on the CRV computed using eq.(4.24) this simple method is particularly suitable for problems in which the main goal is the dimensional reduction. It consists in retaining a number of components that explains a cumulative relative variance greater than a chosen threshold (usually 80% or 90%), i.e. above the minimal proportion of variance that needs to be retained by the model.

4.1.2.1.2 Eigenvalue scree plot

A more refined method to choose the most suitable number of components to retain is the so-called eigenvalue scree plot. The scree plot itself is simply the plot of the eigenvalues corresponding to the principal components, ordered, as aforementioned, from the largest to smallest. Given that the eigenvalues are proportional to their respective captured variance, one could equally choose to plot the RV values. In this selection method one focuses more on the marginal informative contribution gained by including an additional component: if the increase of CRV due to the addition of the component is meaningful than that component is retained, otherwise it is concluded that the additional component can be neglected and a slender model would be almost

equivalent. In other words, for each k-th component that one considers to retain the following percentage is computed:

$$\frac{\mathrm{RV}(k)}{\mathrm{RV}(k-1)} \cdot 100, \tag{4.25}$$

which gives an idea of the increase in CRV given by the inclusion of the k-th component.

A faster and more intuitive way of performing this method is to identify the large drops on the eigenvalues scree plot. The eigenvalues reflect the captured variance of each component, thus the gaps between them reflect the marginal gain given by retaining an additional component to the model.

Another common practice in using the eigenvalues scree plot to determine the entity of the model reduction consists in setting a threshold, usually at 0.7 (Dunteman, 1989; Jolliffe, 2002), and retaining all the components whose eigenvalue lie above this limit.

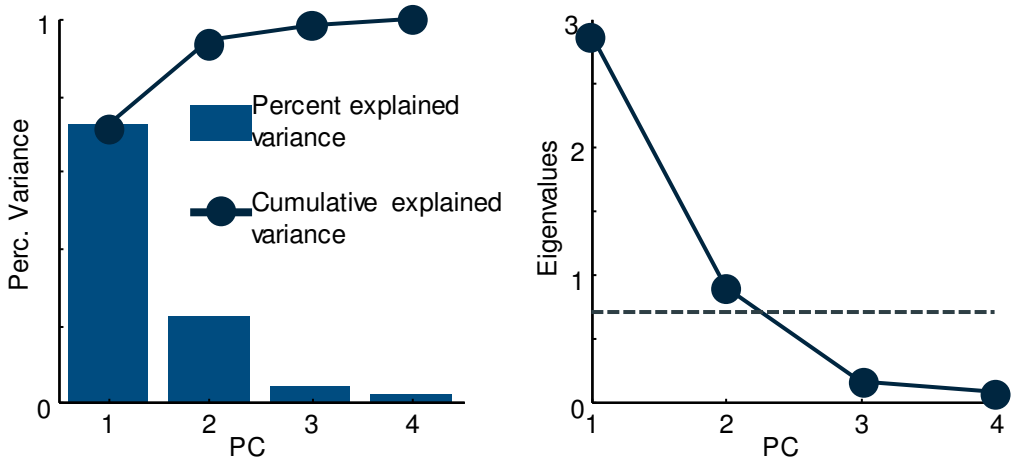


Figure 4.1 Example of eigenvalue scree plot (left) used to determine the model reduction

4.1.2.1.3 Scrambling

Another way to determine the number of PCs that should be retained is based on the so-called data scrambling. It consists of generating a secondary dataset with the same elements of the original one and permuting the values for each column, i.e. for each variable. This permutation, or scrambling, is supposed to significantly remove or modify the relationships among the measurements. A PCA is then performed on this secondary dataset. The method compares each eigenvalue of the original dataset with the correspondent eigenvalue of the new one. If the eigenvalue of the original non-scrambled dataset is greater than its scrambled counterpart, then the related PC of the original dataset is assumed to capture more variance than a PC computed for a non-informative dataset. Conversely, if the eigenvalue of a PC computed for the scrambled dataset is greater than the one of a non-scrambled dataset than it is possible to conclude that the captured RV of the PC in the original dataset is not meaningful and it can be neglected in building the reduced model.

4.1.3 Fault detection by principal component analysis using statistical control charts

In this paragraph are presented some common statistical control charts used to perform a fault detection analysis on the data processed by PCA. It is important to remark that these statistical charts require the variables to follow a multivariate normal distribution and that the measurement errors must be independent and normally distributed.

4.1.3.1 Univariate control charts

It is common in practice for industries to perform process monitoring by constructing univariate charts for the measurements of some diagnostic variables. Typical examples are the Shewhart control charts, consisting in a point representation of a measurement (or one of their statistic) at different times. The chart can include an indication of the mean value calculated on all the previous samples or on a part of them, a lower and an upper limit which indicate the thresholds within the process is statistically considered unlikely to be abnormal. However, if there is any correlation between the monitored variables, then both the rate of false alarms (henceforth called type I errors) and the rate of false acceptance (type II errors) for the given process are expected to increase, being not possible to identify some anomalies focusing on just one variable.

4.1.3.2 Multivariate control charts

In case of correlation among the variables, a more suitable way of assessing the normal or abnormal behaviour of the measurements is represented by the multivariate charts.

Considering a process whose data samples follow an H-variate distribution where $\boldsymbol{u}$ is the vector of the means and $\boldsymbol{S}$ the covariance matrix, it is possible to construct a X^2 statistic as follows:

$$X^2 = (\boldsymbol{y}_{i,\cdot} - \boldsymbol{u})\boldsymbol{S}^{-1}(\boldsymbol{y}_{i,\cdot} - \boldsymbol{u})^T . \tag{4.26}$$

The X^2 statistic represents the measure of the Mahalanobis distance of a sample from a χ^2-distribution with H degrees of freedom i.e. a special case of the gamma distribution which is a sum of the squares of H independent normal random variables. It is preferred to use this kind of statistic in a multivariate dataset instead of the classical evaluation of the standard deviation of the distance from the centre of mass because the latter assumes that the samples are spherically distributed around the centre of mass, so the belonging of the tested sample to the set is only dependent on its distance. On the other hand, in a multivariate distribution the fact that a tested sample statistically belong to the set or not

is expected to depend not only on the distance but also on the direction, since the position of the samples around the centre of mass typically assume a non-spherical shape. The Mahalanobis distance assumes that the set's probability distribution of an H-variate distribution is best represented by an H-dimensional hyper-ellipsoid estimated by building the covariance matrix of the set; any point lying outside the ellipsoid will therefore indicate an anomaly.

4.1.3.2.1 Hotelling's T^2 statistic

However, one can compute the X^2 statistic only when the true mean and covariance matrix are known. If this is not the case, one can estimate the means and covariance matrix of a new sample computing the Hotelling's T^2 statistic (Hotelling, 1947; Mason et al., 2001), which is a generalization of the Student's t statistic widely used in fault detection based on PCA decomposition (Garcia-Alvarez et al., 2009; Fuente et al., 2012; Garcia-Alvarez et al., 2012; Alferes et al., 2013).

In the simple case the t-test evaluates if the mean is equal to a given value μ:

$$t = \frac{\bar{y} - \mu}{\sigma / \sqrt{N}}, \tag{4.27}$$

which is a χ^2 distribution where $\bar{y}$ is the sample mean, σ is the standard deviation and N the number of samples. This expression is equivalent to:

$$\begin{aligned} t^2 &= \frac{(\bar{y} - \mu)^2}{\sigma^2 / N} \\ &= N(\bar{y} - \mu)(\sigma^2)^{-1}(\bar{y} - \mu) \sim \chi^2 \end{aligned} \tag{4.28}$$

Given a process $\boldsymbol{X} \sim \mathcal{N}_H(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e. following a H-variate normal distribution with $\boldsymbol{\mu}$ the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix, and defined the sample mean $\bar{\boldsymbol{y}} = (y_1, \ldots, y_N)/N$, the t^2 metric can be generalized to yeald:

$$T^2 = N \left(\bar{\boldsymbol{y}} - \boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1} \left(\bar{\boldsymbol{y}} - \boldsymbol{\mu}\right) \sim \chi^2_H. \tag{4.29}$$

To avoid misinterpretations, the transpose has been here indicated by the prime ($'$). The Hotelling's T^2 statistic is then a Mahalanobis measure of a χ^2 distribution with H degrees of freedom and is related to the F-distribution (also known as Fisher-Snedecor distribution) by:

$$T^2 \sim \frac{H \left(N-1\right)}{N-H} \cdot F^{\alpha}_{N-H,H}, \tag{4.30}$$

with $F^{\alpha}_{N-H,H}$ representing the Fisher statistic with $N - H$ and H degrees of freedom and for a suitably chosen significance level α (corresponding to a $100(1 - \alpha)\%$ percentile).

It is important to remark that this is only valid for new observations, i.e. for samples that were not used for the estimation of μ, σ or $\boldsymbol{\Sigma}$. The distribution of the Hotelling's T^2 statistic for samples included in the calibration set, instead, follows a β-distribution.

It is now possible to build a multivariate control chart in the reduced scores space context, defining the T^2 statistics of an observation vector $\tilde{\boldsymbol{x}}_{i,\cdot}$ as

$$\begin{aligned} T^2(i) &= \sum_{k=1}^{K} \frac{z_{i,k}^2}{\lambda_k} \\ &= \boldsymbol{z}_{i,\cdot} \boldsymbol{\Lambda}_K^{-1} \boldsymbol{z}'_{i,\cdot} \qquad \text{with } K < L, \\ &= \tilde{\boldsymbol{x}}_{i,\cdot} \boldsymbol{W}_K \boldsymbol{\Lambda}_K^{-1} \boldsymbol{W}'_K \tilde{\boldsymbol{x}}'_{i,\cdot} \end{aligned} \tag{4.31}$$

where $\boldsymbol{\Lambda}_K = \text{diag}(\lambda_1,\ldots,\lambda_K)$ is the diagonal matrix of the first K eigenvalues. The monitored variables are considered "normal" for a significance level α if the statistics defined by eq.(4.31) is lower than the upper control limit (UCL) defined as

$$T_{\text{lim}}^2 = \frac{K\ N-1}{N-K} \cdot F_{N-K,K}^{\alpha}, \tag{4.32}$$

where $F_{N-K,K}^{\alpha}$ is the Fisher statistics with $N - K$ and K degrees of freedom.

4.1.3.2.2 *Q* statistic

In case $K < L$, i.e. the PCA model retains less components than the maximum number, it is also possible to build a joint control chart statistic for the retained PCs, which are supposed to contain the largest part of the information of the current and future well-behaved samples. The statistics that evaluates to what extent the dimension reduction of the PCA model influences the new sample goes by the name of Q statistic, also referred to as 'squared prediction error' (SPE), and is defined as the sum of the squared residuals of the PCA transform, i.e. the components that are not retained. Its goal is to measure the variation share that the PCA model does not account for.

The Q statistics score of an observation vector $\tilde{\boldsymbol{x}}_{i,\cdot}$ is defined as:

$$\begin{aligned} Q(i) &= \boldsymbol{r}_{i,\cdot} \boldsymbol{r}'_{i,\cdot} \\ &= \tilde{\boldsymbol{x}}_{i,\cdot} \boldsymbol{W}_{L-K} \boldsymbol{W}'_{L-K} \tilde{\boldsymbol{x}}'_{i,\cdot} \quad \text{with } K < L, \\ &= \tilde{\boldsymbol{x}}_{i,\cdot} (\boldsymbol{I} - \boldsymbol{W}_K \boldsymbol{W}'_K) \tilde{\boldsymbol{x}}'_{i,\cdot} \end{aligned} \tag{4.33}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{W}_{L-K}$ is the eigenvector matrix of the residuals. Note that also Q is a χ^2 distribution with $L - K$ degrees of freedom, since $\boldsymbol{W}'_{L-K}\tilde{\boldsymbol{x}}_{\cdot,k} \sim \chi^2(0,\boldsymbol{\Lambda}_{L-K})$. The UCL of the Q control chart, which is called Q_{lim}, is defined as:

$$Q_{\lim} = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\left(1/h_0\right)}, \tag{4.34}$$

where c_α is the $(1 - \alpha)$ quantile of the normal distribution and the other quantities are defined as follows:

$$h_0 = 1 - 2\frac{\theta_1 \theta_2}{3\theta_3^2}; \ \theta_q = \sum_{i=K+1}^{L} \lambda_i^q, \text{ for } q = 1, 2, 3. \tag{4.35}$$

The geometrical interpretation of the two statistics follows from the projection of the considered samples onto the hyperplane identified by the retained PCs (Figure 4.2). Since the scores computed are mutually uncorrelated and their square values are divided by the respective eigenvalues (i.e. the contained variance), the Hotelling's T^2 is by definition the Mahalonobis distance between the estimated mean and the projected sample within the subspace defined by hyperplane of the PCs. The violation of the Hotelling's T^2 (in absence of violation of the Q statistic) indicates the occurrence of an abnormal event, though not violating the identified correlation structure. Conversely, a violation of the Q statistic indicates the estimated relationships, represented by the hyperplane, are not valid for the tested sample.

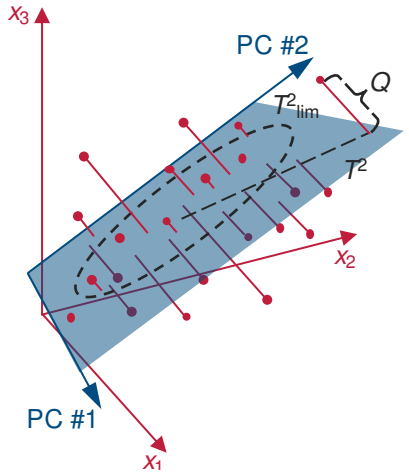


Figure 4.2 Geometrical interpretation of the T^2 and Q statistics in case of a PCA transform where a three-dimensional dataset is projected on a plane

These considerations are the basis of a possible use of the PCA method for FD or classification purposes. A hard classification of the tested sample in the 'normal'/'abnormal' categories can be achieved linking the assignment to one class or to the other based on the violation of the control limits. Considering the combination of the violation of the two UCT and their different meanings, on the other hand, one could expand the number of possible classes or design a more refined FD method for the classification of the anomalies in different severity classes. However, T^2 and the Q statistics remain essentially two global indicators of the state of a process but do not directly provide any information on the source of the anomaly. A valuable tool in this

sense is provided by the computation of the so-called 'contribution variables' (CV). Typically, whenever one of the statistical thresholds is violated by an observation $\tilde{\boldsymbol{x}}_{i,\cdot}$ its projection $\boldsymbol{z}_{i,\cdot}$ onto the PCs plane is processed as:

$$\mathrm{CV}(i) = \boldsymbol{z}_{i,\cdot} \cdot \boldsymbol{\Lambda}^{-1/2} \cdot \boldsymbol{W}^T, \tag{4.36}$$

where i is the index of the currently tested observation. Regardless of the diagonal matrix $\boldsymbol{\Lambda}$ and the loadings matrix $\boldsymbol{W}$ being full ranked or belonging to the reduced PCA model, the transformation (4.36) yelds a vector CV with as many elements as the original variables of $\tilde{\boldsymbol{x}}_{i,\cdot}$. Its largest value will identify the feature that most influenced the projection and therefore the one most probable responsible for the violation.

4.1.4 Extensions to the Principal Component Analysis

The classical PCA method described in the previous paragraphs is limited by the underlying assumption of a linear relationships between the measured and the transformed variables, whatever the dimensions of the latter. However, especially in the environmental systems, this is not generally true, thus applying PCA to nonlinear or dynamic processes would not help to put the underlying relations between the variables in a better perspective but would rather provide a linear static approximation. For this reason, over the years many solutions have been proposed to adapt the PCA-based algorithms to nonlinear process conditions.

4.1.4.1 Methods for nonlinear processes based on the principal component analysis

The simpler way to handle the nonlinearities is to perform a transformation on the data that turns the nonlinear relationships into linear. Operations like log-transform, roots or powers of the original dataset are typical choices. However, these nonlinear relationships are typically not known *a priori*, so that the optimal choice of the transform operation is not possible. Moreover, the data are expected to contain error and, even when the nonlinearities are known, the transformation is likely to introduce a bias in the model.

A nonlinear extension of PCA developed by (Schölkopf et al., 1998) and called 'kernel PCA' (KPCA) is often applied in the context of process monitoring (Lee et al., 2004; Tong et al., 2013). The basic idea is to deal with the nonlinearities by using kernel functions to map the input space into a different space (called feature space) where the data vary linearly and therefore can be linearly separated. In other words, the original data are first transformed into the so-called features, then a standard PCA is performed

on these features. A major benefit of KPCA is that, thanks to the so-called 'kernel trick', the models are computed very efficiently. The trick can be applied for all the transformations (i.e. the kernel functions) satisfying the Mercer's theorem and allows one not to explicitly compute all the coordinates in the new space, but rather to simply compute the inner products between the images of all pairs of data in the feature space. This is computationally cheap and may help to find an appropriate way to transform the original, nonlinearly related, data, though it does not provide a solution for the nonlinear transformation of the measurements errors and the choice of the kernel parameters can be difficult (Jia et al., 2012; Liu et al., 2014). Moreover, KPCA has a limited applicability for diagnostic purposes, since it does not directly operate on the original measurements but rather on their transformations, the features, for which no explicit meaning is generally available. A more exhaustive discussion on the kernel trick and its application to a different linear classifier in a nonlinear setting will be discussed in paragraph 4.4.2.

4.1.4.2 Methods for batch processes based on the principal component analysis

Some PCA methods specifically developed for batch processes are now presented. These techniques were initially developed for the monitoring of chemical processes but proved to be effective also in the wastewater treatment processes due to the similarities between the two fields. The classical PCA method is designed to handle datasets where the information is organized in form of second-order arrays, while batch processes are instead usually described by third-grade tensors, hence the motivation for the development of dedicated methods. The three coordinates which describe the batch array are typically the measured variable ($j = 1,\ldots,L$), the time index at which the measurement was taken during the batch ($i = 1,\ldots,N$) and the index of the measured batch ($k = 1,\ldots,B$), thus producing a tensor $\boldsymbol{X}$, with dimensions $N \times L \times B$, whose elements are $x_{i,j,k} = \boldsymbol{X}(i,j,k)$.

Before computing the PCA model, MPCA converts the third-order tensor into a matrix though a transformation, called unfolding, of the original tensor, basically slicing the third-order tensor along a selected coordinate and arranging the slices side by side. Theoretically there are six ways of unfolding a tensor of rank three, but only two of them are of practical use for the development of a PCA-based monitoring system: the variable-wise unfolding and the batch-wise unfolding. The first approach consists in slicing the tensor so that each slice contains all the measurements for all the variables of the k-th batch and then sorting them as shown in Figure 4.3 producing a matrix of

dimensions $B \cdot N \times L$. The underlying assumption is that the measurements taken at the same time for the same batch are considered a single multivariate sample.

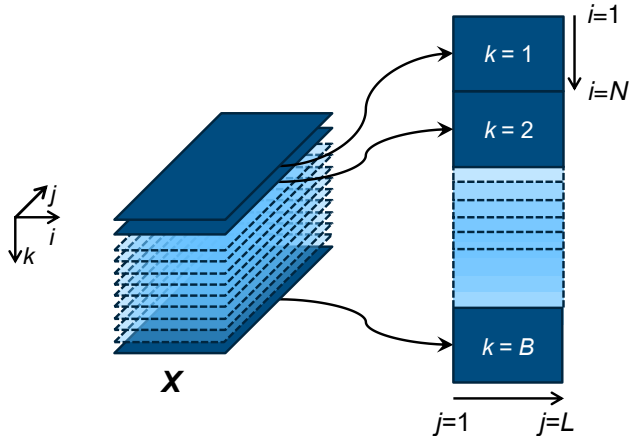


Figure 4.3 Example of batch-wise unfolding of a third-order tensor. The coordinate corresponding to the batches is *k* (Villez, 2007)

In the variable-wise unfolding a matrix is built by slicing the original tensor along the coordinate L and placing side by side the slices so to create a $B \times N \cdot L$ matrix as depicted in Figure 4.4.

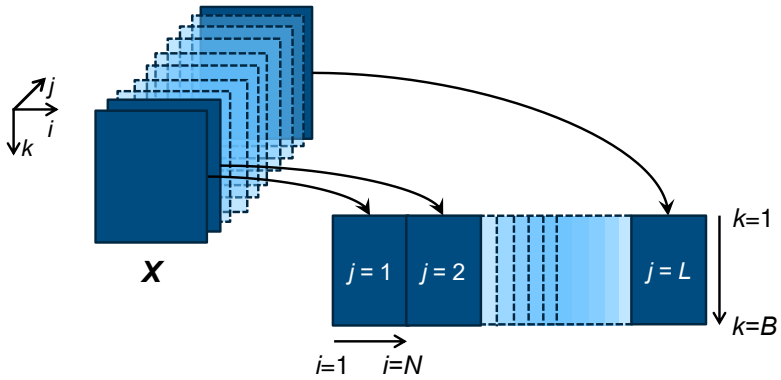


Figure 4.4 Example of variable-wise unfolding of a third-order tensor. The coordinate corresponding to the variable processes is *j* (Villez, 2007)(Villez, 2007)(Villez, 2007)(Villez, 2007)

When using the batch-wise unfolding a problem may arise in the monitoring task since one must wait for the batch process to be completed before building the PCA model, delaying the detection of the anomalies. Typically, three approaches are used to deal with this problem. One is called ‘zero deviation’. Since the data are usually centred and scaled, one can use the mean values as an expectation of the measurement yet to come, so that the PCA model can be built without delay. Another approach used in case of centred and scaled data is to assume that the deviation of the future data from the mean will be the same of the current deviation, which means that for each variable the empty places of the vector are filled with the last scaled measurements. A more refined technique is based on the data estimation. This is an iterative process that first approximates the missing values with the mean values (as in the first approach) then corrects the estimation using the PCA model built upon the historical measurements to

reconstruct the data until convergence is achieved. The condition to reach a unique solution is to have a number of measurements greater than the number of retained PCs.

The unfolding method of batch processes is a popular technique in PCA-based monitoring methods but is also employed in many non-PCA-based methods (Lee and Vanrolleghem, 2003; Villez et al., 2008; Aguado and Rosén, 2008; Garcia-Alvarez et al., 2012; Tong et al., 2013; Yao et al., 2014). Other methods involving the unfolding (hence the name U-PCA) of the dataset, typically a three-way matrix, show good results in continuous (Garcia-Alvarez et al., 2011) and transitional processes (Garcia-Alvarez et al., 2012).

4.1.4.3 Methods for dynamic processes based on the principal component analysis

Another variant of the principal component analysis, specifically introduced to account for the autocorrelations in the measurements of dynamic systems, was proposed by (Ku et al., 1995) and is called 'dynamic PCA' (DPCA). It has been demonstrated that detection and isolation of anomalies in dynamic systems is possible, however this application conflicts with the statistical basis of PCA by violating the assumption of time-independence, and therefore returning possibly misleading results, such as an excess of false alarms. Since for dynamic systems the current values depend on the past, in the DPCA method the first step is to build a matrix $\tilde{\boldsymbol{X}}_g$ which accounts for the dynamic relationships between the last g samples, with g representing the time lag:

$$\tilde{\boldsymbol{X}}_g = \left[\tilde{\boldsymbol{x}}_{i,\cdot} \quad \tilde{\boldsymbol{x}}_{i-1,\cdot} \quad \cdots \quad \tilde{\boldsymbol{x}}_{i-g,\cdot}\right] \in \Re^{G\times(L\cdot g+1)}, \text{ with } i = 1,\ldots,G;\ G < N. \tag{4.37}$$

For each time lag g a PCA is performed and the auto- and cross-correlation are evaluated. While the classical PCA method focuses on the largest eigenvalues, considering the smaller ones as a product of noise, DPCA regards the smaller eigenvalues as an indication of the existence of nonlinearities and dynamic relationships among the samples. For example, if from a first PCA on $\tilde{\boldsymbol{x}}_{i,\cdot}$ ($g = 0$) it results that two eigenvalues are smaller than a given threshold, that indicates the presence of two linear static relationships. Than one performs a PCA with lag $g = 1$ (so on $[\tilde{\boldsymbol{x}}_{i,\cdot}\ \tilde{\boldsymbol{x}}_{i-1,\cdot}]$) and if the number of small eigenvalues found is five that would reveal the presence of a dynamic relation. In fact, four of them represent the two static relationships found in the previous step (at each step they will be repeated $\Delta g+1$ times), and the fifth is related to the dynamic relation. A further step, performing PCA for a lag $g = 2$ $[\tilde{\boldsymbol{x}}_{i,\cdot}\ \tilde{\boldsymbol{x}}_{i-1,\cdot}\ \tilde{\boldsymbol{x}}_{i-2,\cdot}]$, will reveal the presence of other dynamic relations. If now the small eigenvalues are eight than the method can stop since six of them represents the linear relations (the two small eigenvalues found at the beginning, $g = 0$, multiplied $\Delta g+1 = 3$ times) and the

remaining two are related to the dynamic relation identified at the previous step ($\Delta g+1 = 2$ in this case since the dynamic relation was found at $g = 1$). No additional dynamic relationships are found so the algorithm can stop reporting that the processed data contain two linear relations and one first order (since it was identified at the first step) dynamic relation. It is to be noted that in the described method the PCA model is fixed beforehand, i.e. this method defines a time-invariant model.

However, most of the dynamic processes do not preserve the relationships among the variables but rather they describe gradual changes of these relationships, thus the covariance structure of data varies accordingly. Many authors proposed methods to take into account the changes in the covariance matrix. An efficient one, can be found in (Rosén and Lennox, 2001) and is called Adaptive PCA. It is based on the so called 'moving window' (MW), meaning that the covariance matrix is computed at each time instant t using the last Q samples, where Q is the fixed length of the moving window:

$$\boldsymbol{C}_t = \sum_{i=t-Q}^{t} \tilde{\boldsymbol{x}}_{i,\cdot} \tilde{\boldsymbol{x}}_{i,\cdot}^T . \tag{4.38}$$

Another method, developed by (Li et al., 2000) and called 'recursive PCA' (RPCA), proposes to update the covariance matrix each time new data are available: if the covariance matrix at instant t–1 is known, $\boldsymbol{C}_{t-1}$, once the new sample $\tilde{\boldsymbol{x}}_{t,\cdot}$ is available the covariance matrix is updated as follows:

$$\boldsymbol{C}_t = \varphi \cdot \boldsymbol{C}_{t-1} + (1-\varphi) \cdot \tilde{\boldsymbol{x}}_{t,\cdot} \tilde{\boldsymbol{x}}_{t,\cdot}^T , \tag{4.39}$$

where φ (> 0) is the forgetting factor, a coefficient that assigns an incremental relative weight on the older samples. A high value of φ will lead to a longer memory of past behaviours, while a low value leads to shorter memory of past behaviour.

Thus, for both methods the updating consists of building a new PCA model at each step, although using different criteria to achieve it. For example, with 'adaptive PCA' (APCA) the number of PCs to retain is fixed, it is determined at the first step and does not change as the window moves. Also, all the new samples are allowed to update the model, regardless they are regular or anomalous. In RPCA, instead, the number of past samples to consider for the computation of the PCA model is governed by the parameter φ but the method prevents the samples that violate the statistical limits to update the reference space, since generally one does not want to adapt to abnormal changes in the process behaviour. Also at each iteration, if a model reduction is pursued, the number or relevant PCs is updated and the statistical limits will change accordingly.

A method conceived to combine the positive aspects of both the adaptive PCA and the Recursive PCA is the one proposed by (Baggiani and Marsili-Libelli, 2009) in view of a real-time application. Here the reference PCA model is computed on the data of a

fixed-length window which slides according to the outcome of the statistical test of the new samples available. Whenever a new sample satisfies the statistical threshold it is allowed to join the reference dataset while at the same time the oldest sample is discarded. Conversely the samples whose statistical score is above the thresholds are labelled as anomalies and not allowed to update the reference space.

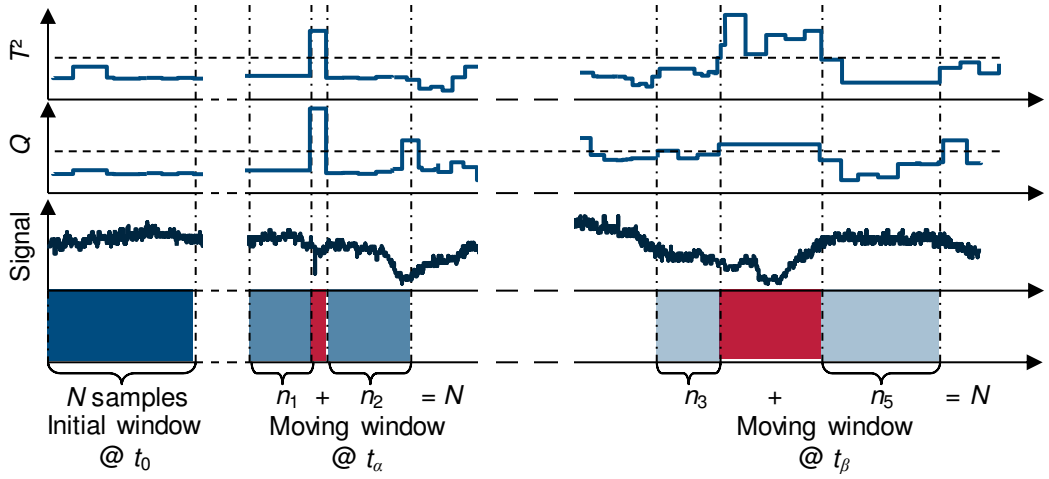


Figure 4.5 Initial and real-time window for mowing window PCA method proposed by (Baggiani and Marsili-Libelli, 2009). The red-shaded areas represent the anomalous samples excluded from reference updating

It should be noted that the underlying assumption of these updating methods is that the abnormal changes in the process are faster than the updating speed of the model. Indeed, if the change occurring is abnormal and it is slow enough to satisfy the acceptance conditions, then its information will result in an undesired adaptation of the model. The diagnosis task, already difficult in adaptive PCA models where the properties frequently change, will be even more challenging.

In their consecutive work (Lennox and Rosén, 2002) further developed their method adding to the adaptive part, which accounts for changes in the structure of the covariance, a multi-scale part which is devoted to account for the presence of significant dynamics with different time scales. They called the new method ‘adaptive multi-scale PCA’ (AdMSPCA). Using wavelet decomposition AdMSPCA produces for each variable band-limited signal (called details) and an approximation which are then projected onto scale-specific PCA models. A violation in the statistical limits at a given time scale identifies the presence of a significant dynamics and thus labels the data as significant. Only the significant data are thus used to reconstruct the signals and the reconstructed data becomes the basis on which an overall PCA model is built. The latter is the model eventually used for the monitoring task.

4.2 Classifiers based on the Bayes decision theory

From a probabilistic point of view, a classification problem can be seen as a forecasting problem, where the class of an observation is predicted, or inferred, based on the consideration that similar observations fell in that specific class. In other words, the observation is assigned to a specific class conditionally on the probability that the observation has a similar behaviour of those belonging to the same class. More formally, given a classification problem where an observation vector, or features vector, $\boldsymbol{x}$, belonging to a process $\boldsymbol{X}$, has to be classified using M classes ($\omega_1,\omega_2,...,\omega_M$), the objective is to determine the conditional probability $P(\omega_m|\boldsymbol{x})$ that satisfies the classification requirements. This is the so-called *a posteriori* probability, or posterior, of the well-known Bayes' theorem. Given a set of N events $\mathcal{A}_i$, with $i = 1,\ldots,N$ and such that

$$\sum_{i=1}^{N} P(\mathcal{A}_i) = 1, \tag{4.40}$$

the probability of occurrence of an arbitrary event $\mathcal{B}$ is given by:

$$P(\mathcal{B}) = \sum_{i=1}^{N} P(\mathcal{B} \mid \mathcal{A}_i) \cdot P(\mathcal{A}_i), \tag{4.41}$$

where $P(\mathcal{B}|\mathcal{A})$ is the conditional probability of $\mathcal{B}$ given the occurrence of $\mathcal{A}$ defined as:

$$P(\mathcal{B} \mid \mathcal{A}) = \frac{P(\mathcal{B}, \mathcal{A})}{P(\mathcal{A})}, \tag{4.42}$$

with $P(\mathcal{B},\mathcal{A})$ the joint probability of the two events. From the latter equation it is possible to derive the so-called Bayes' rule:

$$P(\mathcal{B} \mid \mathcal{A}) \cdot P(\mathcal{A}) = P(\mathcal{A} \mid \mathcal{B}) \cdot P(\mathcal{B}). \tag{4.43}$$

Re-arranging the previous relation and translating it to the previous example, it is now possible to formulate the classification task in Bayesian terms:

$$P(\omega_m \mid \boldsymbol{x}) = \frac{P(\boldsymbol{x} \mid \omega_m) \cdot P(\omega_m)}{P(\boldsymbol{x})}. \tag{4.44}$$

From (4.41) the probability $P(\boldsymbol{x})$ is equal to:

$$P(\boldsymbol{x}) = \sum_{m=1}^{M} P(\boldsymbol{x} \mid \omega_m) \cdot P(\omega_m). \tag{4.45}$$

Let us consider the simple case of a binary classification problem, i.e. $M = 2$ and ω_1 and ω_2 the only two possible classes. One can assume to know the *a priori* probabilities

$P(\omega_1)$ and $P(\omega_2)$, and even if it is not the case they can be estimated from the available data as the frequency of occurrence in each class. Given have N observations, they can be split in N_1 and N_2 observation focusing respectively on those belonging to the first and to the second class, so that $N = N_1 + N_2$. The priors are then computed by the frequentist approach as:

$$P(\omega_m) \approx \frac{N_m}{N}, \tag{4.46}$$

for $m = 1, 2$. It is also reasonable to assume as known the conditional probabilities $P(\boldsymbol{x}|\omega_m)$, which represent the probability density functions (pdf) describing the distribution of the feature vector in each class. It is common to refer to $P(\boldsymbol{x}|\omega_m)$ as the 'likelihood' of the m-th event, or class, ω_m, with respect to the observations $\boldsymbol{x}$. Also in this case, if the pdfs are not known they can be estimated from the known available dataset.

The classification task is then achieved by assigning each sample to the most probable class, i.e. computing the maximum of the M conditional probabilities $P(\omega_m|\boldsymbol{x})$ or the maximum of an appropriate function of them. In the simplest scenario:

$$\begin{aligned} &\text{if } P(\omega_1 \mid \boldsymbol{x}) > P(\omega_2 \mid \boldsymbol{x}), \boldsymbol{x} \text{ is classified to } \omega_1, \\ &\text{if } P(\omega_1 \mid \boldsymbol{x}) < P(\omega_2 \mid \boldsymbol{x}), \boldsymbol{x} \text{ is classified to } \omega_2. \end{aligned} \tag{4.47}$$

In case of equality the sample can be assigned to either class.

An important property of this classification method can be easily derived assuming the case of two equiprobable priors, i.e. $P(\omega_1) = P(\omega_2) = ½$. The rule of (4.47) corresponds to comparing the numerators of (4.44), i.e.:

$$P(\boldsymbol{x} \mid \omega_1)P(\omega_1) \gtrless P(\boldsymbol{x} \mid \omega_2)P(\omega_2), \tag{4.48}$$

where the $P(\boldsymbol{x})$ is not considered being the same for both members, but in case of equiprobable priors also the terms $P(\omega_m)$ can be neglected, becoming:

$$P(\boldsymbol{x} \mid \omega_1) \gtrless P(\boldsymbol{x} \mid \omega_2). \tag{4.49}$$

The further assumption that the sample $\boldsymbol{x}$ consists of a single feature, and thus can be written as x, leads to representing the classification problem as in Figure 4.6.

The vertical line separates the domain of x in two regions R_1 and R_2, respectively denoting the intervals in which an observation of x is assigned to ω_1 or ω_2 depending on whether it is smaller or larger than the threshold x_0. However, a classification error is unavoidable for the region where the two probabilities overlap. The mathematical expression of this probability error P_e is given by:

$$P_e = \frac{1}{2}\int_{-\infty}^{x_0} P(x \mid \omega_1)\, dx + \frac{1}{2}\int_{x_0}^{+\infty} P(x \mid \omega_2)\, dx. \tag{4.50}$$

It can be easily proved that if the threshold x_0 is placed in correspondence of the intersection of the two probabilities than the probability error P_e, graphically represented by the shaded area under the curves, is minimum. In fact, shifting left or right the threshold the error probability (shaded area) increases.

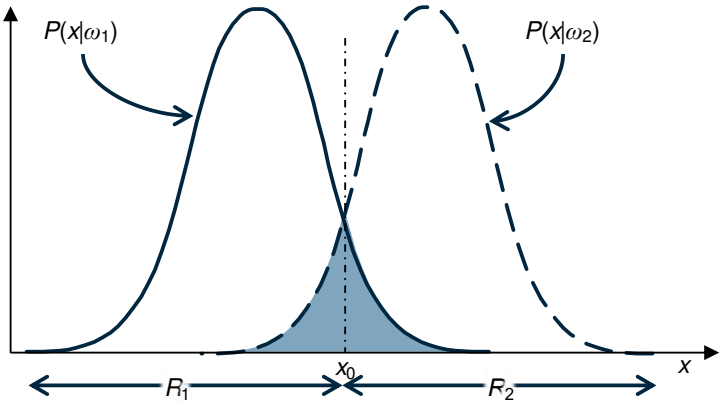


Figure 4.6 Example of Bayesian classifier in case of a single feature attribute *x* to be assigned to two equiprobable classes ω_1 (solid line) and ω_2 (dashed line)

Minimizing the classification error probability, however, is not always the best option, because it means that all the errors are given the same importance, whereas there are cases in which the consequences of a misclassification can be more serious than others. Let us consider a classification problem involving M classes, thus M regions R_m for each m-th class ω_m can be identified. The sample vector $\boldsymbol{x}$, actually belonging to a class ω_α will be assigned to the class ω_β if it lies in the R_β interval. It is then possible to define a weight, a penalty term $\lambda_{\alpha,\beta}$ called 'loss' associated to the wrong decision. The losses can be collected in a loss matrix and a risk associated to the misclassification of the α-th class can be computed:

$$r_\alpha = \sum_{\beta=1}^{M} \lambda_{\alpha,\beta} \int_{R_\beta} P(\boldsymbol{x} \mid \omega_\alpha)\, d\boldsymbol{x} \,. \tag{4.51}$$

In this framework, the objective is to minimize the average risk r defined as:

$$\begin{aligned} r &= \sum_{\alpha=1}^{M} r_\alpha P(\omega_\alpha) \\ &= \sum_{\beta=1}^{M} \int_{R_\beta} \left(\sum_{\alpha=1}^{M} \lambda_{\alpha,\beta} P(\boldsymbol{x} \mid \omega_\alpha) P(\omega_\alpha) \right) d\boldsymbol{x} \end{aligned} \tag{4.52}$$

which is equivalent to assign $\boldsymbol{x}$ to the partitioning region R_β if:

$$l_\beta \equiv \sum_{\alpha=1}^{M} \lambda_{\alpha,\beta} P(\boldsymbol{x} \mid \omega_\alpha) P(\omega_\alpha) < l_m \equiv \sum_{\alpha=1}^{M} \lambda_{\alpha,m} P(\boldsymbol{x} \mid \omega_\alpha) P(\omega_\alpha) \,. \tag{4.53}$$

Of course, this is the same as minimizing the classification error probability when the loss $\lambda_{\alpha,\beta}$ is equal to $1 - \delta_{\alpha,\beta}$, with the Kronecker's delta $\delta_{\alpha,\beta} = 0$ if $\alpha \neq \beta$ and1 if $\alpha = \beta$.

In the previous example of the two-class case one would have:

$$l_1 = \lambda_{1,1} P(\boldsymbol{x} \mid \omega_1) P(\omega_1) + \lambda_{2,1} P(\boldsymbol{x} \mid \omega_2) P(\omega_2) \text{ and}$$
$$l_2 = \lambda_{1,2} P(\boldsymbol{x} \mid \omega_1) P(\omega_1) + \lambda_{2,2} P(\boldsymbol{x} \mid \omega_2) P(\omega_2), \tag{4.54}$$

assigning the sample to the first class if the risk associated is lower than the alternative, i.e. $l_1 < l_2$,

$$(\lambda_{2,1} - \lambda_{2,2}) P(\boldsymbol{x} \mid \omega_2) P(\omega_2) < (\lambda_{1,2} - \lambda_{1,1}) P(\boldsymbol{x} \mid \omega_1) P(\omega_1). \tag{4.55}$$

It is then natural to assume that the loss of a correct assignment ($\lambda_{i,i}$) is weighted less than a wrong assignment ($\lambda_{i,j}$), and the rule (4.53) becomes

$$\boldsymbol{x} \in \omega_1 \text{ if } l_{1,2} \equiv \frac{P(\boldsymbol{x} \mid \omega_1)}{P(\boldsymbol{x} \mid \omega_2)} > \frac{P(\omega_2) \cdot (\lambda_{2,1} - \lambda_{2,2})}{P(\omega_1) \cdot (\lambda_{1,2} - \lambda_{1,1})} \text{ or}$$
$$\boldsymbol{x} \in \omega_2 \text{ if } l_{1,2} \equiv \frac{P(\boldsymbol{x} \mid \omega_1)}{P(\boldsymbol{x} \mid \omega_2)} < \frac{P(\omega_2) \cdot (\lambda_{2,1} - \lambda_{2,2})}{P(\omega_1) \cdot (\lambda_{1,2} - \lambda_{1,1})}. \tag{4.56}$$

This rule is known as likelihood ratio test, from the likelihood ratio $l_{1,2}$.

To better clarify the concept of associated to the evaluation of the misclassification risk it is possible to assume that the risk of wrong assignment for a sample belonging to the second class is higher, in other words $\lambda_{2,1} < \lambda_{1,2}$. In the hypothesis of equiprobable priors as before, this different weighting of the misclassification error has the effect of moving to the left the threshold x_0 of Figure 4.6, increasing the R_2 region and decreasing R_1.

Thus, either minimizing the classification error by minimizing the probability error P_e or the classification risk r, in the general case of M classes, can be seen as a partitioning problem where in the feature space M regions have to be determined, separated by a threshold, or 'decision surface'. It can be preferable in some cases to compute the decision surfaces using functions of the probabilities or of the risk functions instead of working directly with them. For example, using a monotonically increasing function $f(\cdot)$ of the probability it is possible to compute a function $g_\alpha(\boldsymbol{x}) = f(P(\omega_\alpha|\boldsymbol{x}))$ called 'discriminant function' where:

$$\boldsymbol{x} \in \omega_\alpha \text{ if } g_\alpha(\boldsymbol{x}) > g_\beta(\boldsymbol{x}), \forall\, \alpha \neq \beta. \tag{4.57}$$

4.2.1 Estimation of probability density functions

In the previous examples the prior pdfs were assumed to be known, however it seldom the case and in many problems they have to be estimated from the available data. Moreover, the distribution shape may be known (Gumbel, Gaussian, etc.) but its parameters are unknown or, more frequently, the variance and the mean are available but the specific distribution is unknown.

4.2.1.1 Parametric estimation of the probability density functions

In the following examples an M-class problem is considered, where the likelihoods $P(\boldsymbol{x}|\omega_m)$ are expressed in a parametric space Q and the vectors $\boldsymbol{q}_m$ of the defining, and unknown, parameters are to be determined based on the available data. In the hypothesis that the data belonging to one class do not affect the estimation of the parameters of the other classes, i.e. they are independent and identically distributed (iid), one can draw N samples $\boldsymbol{x}_{i,\cdot}$ from a pdf $P(\boldsymbol{x};\boldsymbol{q})$ and build the joint pdf $P(\boldsymbol{X};\boldsymbol{q})$, where $\boldsymbol{X} = \{\boldsymbol{x}_{1,\cdot},\boldsymbol{x}_{2,\cdot},\ldots,\boldsymbol{x}_{N,\cdot}\}$ as:

$$P(\boldsymbol{X};\boldsymbol{q}) \equiv P(\boldsymbol{x}_{1,\cdot},\boldsymbol{x}_{2,\cdot},\ldots,\boldsymbol{x}_{N,\cdot};\boldsymbol{q}) = \prod_{i=1}^{N} P(\boldsymbol{x}_{i,\cdot};\boldsymbol{q}), \tag{4.58}$$

which is also known as the likelihood function of $\boldsymbol{q}$ with respect to $\boldsymbol{X}$.

It is now possible to use (4.58) to compute the optimal parameter vector $\hat{\boldsymbol{q}}_{MLE}$ as the one maximizing:

$$\hat{\boldsymbol{q}}_{MLE} = \arg\max_{\boldsymbol{q}\in Q} \prod_{i=1}^{N} P(\boldsymbol{x}_{i,\cdot};\boldsymbol{q}). \tag{4.59}$$

This method is called the maximum likelihood estimate (MLE) of $\boldsymbol{q}$.

The maximum $\hat{\boldsymbol{q}}_{MLE}$, must satisfy the necessary condition

$$\frac{\partial \prod_{i=1}^{N} P(\boldsymbol{x}_{i,\cdot};\boldsymbol{q})}{\partial \boldsymbol{q}} = 0. \tag{4.60}$$

However, it can be preferable to consider its logarithm and define

$$L(\boldsymbol{q}) = \ln \prod_{i=1}^{N} P(\boldsymbol{x}_{i,\cdot};\boldsymbol{q}), \tag{4.61}$$

and then compute the value $\boldsymbol{q}$ that vanishes its derivative:

$$\frac{\partial L(\boldsymbol{q})}{\partial \boldsymbol{q}} = \sum_{i=1}^{N} \frac{\partial \ln P(\boldsymbol{x}_{i,\cdot};\boldsymbol{q})}{\partial \boldsymbol{q}} = \sum_{i=1}^{N} \frac{1}{P(\boldsymbol{x}_{i,\cdot};\boldsymbol{q})} \frac{\partial P(\boldsymbol{x}_{i,\cdot};\boldsymbol{q})}{\partial \boldsymbol{q}} = 0. \tag{4.62}$$

In the definition of the MLE method no prior assumption was made as to the shape of the pdf $\boldsymbol{q}$, which makes the estimate reasonably biased, i.e. its mean does not necessarily coincide with the true value. As a consequence of the central limit theorem, however, for sufficiently large N all distributions converge towards the Gaussian distribution and thus

$$\lim_{N\to\infty} \mathbb{E}[\hat{\boldsymbol{q}}_{MLE}] = \boldsymbol{q}_0, \tag{4.63}$$

with $\boldsymbol{q}_0$ the true value, which makes the MLE an asymptotically unbiased method.

MLE is also asymptotically consistent, in the sense that

$$\lim_{N\to\infty} \mathbb{P}\ \left\|\hat{\boldsymbol{q}}_{MLE} - \boldsymbol{q}_0\right\| < \epsilon\ = 1, \tag{4.64}$$

with ϵ arbitrarily small, but also

$$\lim_{N\to\infty} \mathbb{E}\left[\left\|\hat{\boldsymbol{q}}_{MLE} - \boldsymbol{q}_0\right\|^2\right] = 0\,. \tag{4.65}$$

The first relation means that for $N \to \infty$ the estimate converges to the true value $\boldsymbol{q}_0$ in probability, the second relation says that it also converges, according to the square of the mean value, which is the equivalent of saying that the estimate achieves the Cramer-Rao lower bound. The MLE is therefore, at least for sufficiently large values of N, unbiased, consistent and efficient normally distributed estimator.

A different approach for estimating the distribution consists in considering $\boldsymbol{q}$ not the vector of the unknown parameters of the distribution but a random variable that has to be estimated. From the Bayes' theorem (4.44) is:

$$P(\boldsymbol{q}\,|\,\boldsymbol{X}) = \frac{P(\boldsymbol{q})P(\boldsymbol{X}\,|\,\boldsymbol{q})}{P(\boldsymbol{X})}, \tag{4.66}$$

and the goal is now to estimate the optimal value of the random variable that maximise the probability P($\boldsymbol{q}$|X):

$$\hat{\boldsymbol{q}}_{MAP} = \boldsymbol{q}\left|\frac{\partial P(\boldsymbol{q}\,|\,\boldsymbol{X})}{\partial \boldsymbol{q}} = \frac{\partial\ P(\boldsymbol{q})P(\boldsymbol{X}\,|\,\boldsymbol{q})}{\partial \boldsymbol{q}} = 0\,.\right. \tag{4.67}$$

This is called 'maximum *a posteriori* probability' (MAP) method and its main difference with MLE is in the involvement of the prior $P(\boldsymbol{q})$ for the estimation. Both methods however, show similar results if the assumed distribution for $P(\boldsymbol{q})$ is uniform or has small variations, i.e. if a constant or nearly constant prior is assumed, otherwise in the estimations of the two methods tend to differ.

Another different method, called 'Bayesian inference' proposes to focus on the conditional probability $P(\boldsymbol{x}|\boldsymbol{X})$ instead of focusing on the conditional probability of $\boldsymbol{q}$.

From the extension of (4.45) to the continuous case and (4.66):

$$P(\boldsymbol{q}\,|\,\boldsymbol{X}) = \frac{P(\boldsymbol{X}\,|\,\boldsymbol{q})P(\boldsymbol{q})}{P(\boldsymbol{X})} = \frac{P(\boldsymbol{X}\,|\,\boldsymbol{q})P(\boldsymbol{q})}{\int P(\boldsymbol{X}\,|\,\boldsymbol{q})P(\boldsymbol{q})\ d\boldsymbol{q}}, \tag{4.68}$$

and still under the assumption of iid samples:

$$P(\boldsymbol{X}\,|\,\boldsymbol{q}) = \prod_{i=1}^{N} P(\boldsymbol{x}_{i,}\,|\,\boldsymbol{q})\,, \tag{4.69}$$

from which it is possible to estimate

$$P(\boldsymbol{x}\,|\,\boldsymbol{X}) = \int P(\boldsymbol{x}\,|\,\boldsymbol{q})P(\boldsymbol{q}\,|\,\boldsymbol{X})\ d\boldsymbol{q}\,. \tag{4.70}$$

For large numbers of N, as the likelihood product $\prod_i P(\boldsymbol{x}_{i,}|\boldsymbol{q})$ converges to a delta function around the mean $\hat{\boldsymbol{q}}$, the Bayesian inference estimator approximates the MLE

one. In this case the Bayesian inference also approximates the MAP estimator, since $P(\boldsymbol{x}|\boldsymbol{X}) \approx P(\boldsymbol{x}|\hat{\boldsymbol{q}})$, therefore one can claim that for $N \to \infty$ the three methods asymptotically converge to the same estimate.

From Shannon's information theory and its definition of information entropy it is possible to estimate the unknown density function $P(\theta)$ of a random variable θ from a set of given constraints such as mean, variance, etc. by maximizing its associated entropy. The information entropy H associated to $P(\theta)$ is:

$$H = -\int P(\theta) \ln P(\theta)\, d\theta \,. \tag{4.71}$$

When no specific constraints are known for the pdf one can simply assume that it is nonzero between two generic points θ_α and θ_β:

$$\int_{\theta_\alpha}^{\theta_\beta} P(\theta)\, d\theta = 1\,, \tag{4.72}$$

and minimize (4.71) with respect to $P(\theta)$. With such a method, one will obtain that $P(\theta)$ is estimated as a uniform distribution, i.e. all values of θ are equally possible, however adding mean and variance as constraints the approximating distribution that maximizes the randomness will result the Gaussian.

4.2.1.2 Non-parametric estimation of the probability density functions

A different approach to the estimation of the pdf comes from variations of the histogram approximation of the distribution of an unknown pdf when there is a lack of prior knowledge on the distribution form. In the one-dimensional case one will build the approximating histogram of a random variable $\boldsymbol{x}$ by dividing the abscissa into B successive bins of fixed length h and approximating the probability of the b-th bin as the frequency

$$P_b(\boldsymbol{x}) \approx \tilde{P}_b(\boldsymbol{x}) = \frac{1}{h}\frac{k_N}{N} = \frac{1}{h}\frac{\sum_i \boldsymbol{x}_{i,}}{N},\ i \in [(b\text{–}1)\cdot h,\, b\cdot h];\ b \in [1,\ldots,B]. \tag{4.73}$$

To guarantee the convergence towards the true pdf, the number of bins M must increase and at the same time their width h must reduce. However, this has some major shortcomings: the number and the width of the bins are deeply influenced by the number of samples falling in each bin, which means that one cannot infinitely reduce h because the number of sample N available will always be finite, this implies a finite number of bins too and consequently the fact that the estimated distribution will be a discrete one. Thus, in practice one will proceed with a non-parametric estimation of the pdf using a number "large enough" of bins with a width "small enough" depending on the degree of approximation judged satisfying (see Figure 4.7).

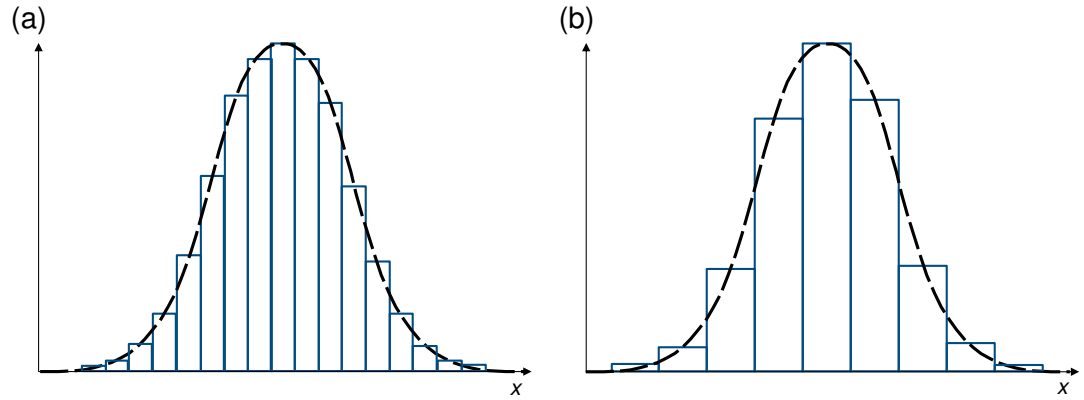


Figure 4.7 Probability density function approximation by histogram varying the bin width, large (a) and small (b) (Theodoridis and Koutroumbas, 2009)

In case of a multi-dimensional set of data, say L-dimensional, the distribution is estimated by dividing the space in the multi-dimensional equivalent of the bins, i.e. hypercubes of side h and volume h^L. If $x_{i,j}$, with $i = 1,\ldots,N$ and $j = 1,\ldots,L$ are the components of the available sample vector $\boldsymbol{x}_{i,\cdot}$ in the L-dimensional space, one can define the function $\phi(\boldsymbol{x}_{i,\cdot})$:

$$\phi(\boldsymbol{x}_{i,\cdot}) = \begin{cases} 1 & \text{for } |x_{i,j}| \leq 1/2 \\ 0 & \text{otherwise} \end{cases} \tag{4.74}$$

which is 1 for all the points in the hypercube of length 1 and 0 outside and can be used to rewrite (4.73) for the multi-dimensional case:

$$\tilde{P}(\boldsymbol{x}) = \frac{1}{h^L}\left(\frac{1}{N}\sum_{i=1}^{N}\phi\left(\frac{\boldsymbol{x}-\boldsymbol{x}_{i,\cdot}}{h}\right)\right) \tag{4.75}$$

Still the $P(\boldsymbol{x})$ is approximated as a sum of discontinuous step functions, which may be not desirable in some applications. For this reason, it was proposed (Parzen, 1962) to use a set of smooth functions $\varphi(\cdot)$, with the characteristic of being

$$\begin{aligned} &\varphi(\boldsymbol{x}) \geq 0 \text{ and} \\ &\int_x \varphi(\boldsymbol{x})\ d\boldsymbol{x} = 1 \end{aligned} \tag{4.76}$$

to approximate de pdfs, instead of using $\phi(\cdot)$. This class of functions are usually referred to as 'Parzen windows' or 'kernel functions' hence the name of this method 'kernel density estimation' (KDE). Thus, similarly to (4.75) the approximated pdf is

$$\tilde{P}(\boldsymbol{x}) = \frac{1}{h^L}\left(\frac{1}{N}\sum_{i=1}^{N}\varphi\left(\frac{\boldsymbol{x}-\boldsymbol{x}_{i}}{h}\right)\right) \tag{4.77}$$

for the general multidimensional case. Notice that using smooth function the parameter h assumes a new meaning. What was before the width of the bin or the side of the hypercube used to compute the discrete estimate of $P(\boldsymbol{x})$ becomes here a smoothing parameter called 'bandwidth'. Some remarks on its influence on the final approximation should be made: for a given set of N training samples a small h increases the variance. The estimated pdf will result from the superposition of extremely sharp functions

centred at the training points so that the final shape will be noisy, with high spikes around the samples and lower values in the rest of the domain. As the number of training points increases, however, the noise reduces because the spiky functions will be more closely located. A broader h, on the contrary, will provide a less detailed approximation of the pdf, probably underestimating some peaks in multimodal distributions, also it will be less affected by an increase in the number of available training samples (Figure 4.8).

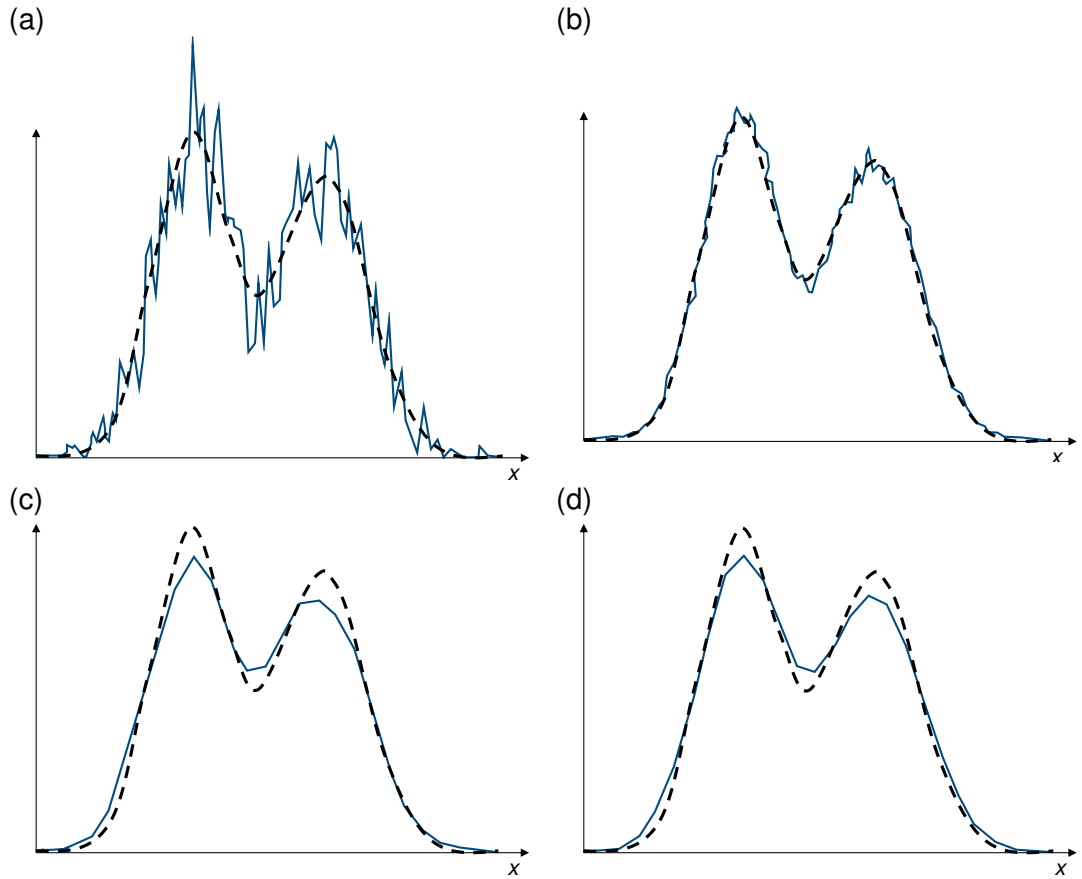


Figure 4.8 KDE approximation of a pdf (dotted line) using Gaussian kernels with (a) h = 0.1 and 1000 training samples, (b) h = 0.1 and 20'000 samples, (c) h = 0.8 and 1'000 samples, (d) h = 0.8 and 20'000 samples (Theodoridis and Koutroumbas, 2009)

In KDE the width/volume h^L around the points $\boldsymbol{x}$ is considered fixed, so the number of points k_N falling in each bin/hypercube can vary if their distribution on the domain is not uniform. The k nearest neighbour density estimation method instead, proposes to approximate the pdf using a fixed number k of points each time, and therefore a variable volume that from time to time can include them. Low-density areas will require a larger volume, that can be now written as a function of the training set $V(\boldsymbol{x})$ while in highly dense areas the width will be smaller. The general expression of the estimator is:

$$\tilde{P}(\boldsymbol{x}) = \frac{1}{V(\boldsymbol{x})}\frac{k}{N}. \tag{4.78}$$

This method can be also employed as a classifier. Assuming that a new sample $\boldsymbol{x}$ is available one can compute its distance, not necessarily a Euclidean one, from all the training vectors of the various classes ω_m. Calling r_1 and r_2 respectively the radius of the

hyper-sphere (or hyper-ellipsoid) centred in $\boldsymbol{x}$ and containing k points of ω_1 or ω_2, and denoting V_1 and V_2 their volume, the classification task is performed as:

$$\begin{aligned} \boldsymbol{x} \in \omega_1 \text{ if } l_{1,2} \approx \frac{k \cdot N_2 \cdot V_2}{k \cdot N_1 \cdot V_1} > \frac{P(\omega_2) \cdot (\lambda_{2,1} - \lambda_{2,2})}{P(\omega_1) \cdot (\lambda_{1,2} - \lambda_{1,1})}, \\ \text{or } \frac{V_2}{V_1} > \frac{N_1 \cdot P(\omega_2) \cdot (\lambda_{2,1} - \lambda_{2,2})}{N_2 \cdot P(\omega_1) \cdot (\lambda_{1,2} - \lambda_{1,1})}. \end{aligned} \tag{4.79}$$

If in the one-dimensional case N observations were considered enough to train estimate the M pdfs $P(\boldsymbol{x}|\omega_m)$ used by the Bayesian classifier to assign a new point to one of the M classes, in a L-dimensional space one should have an excessively high number of training observations N^L, which is often practically unfeasible. However, by assuming the L features of $\boldsymbol{x}$ to be iid the problem reduces to the estimation of L one-dimensional distributions for each class and a $L \cdot N$ points would be sufficient. The so-called 'naïve-Bayes' classifier works under these assumptions and assigns the unknown L-dimensional sample of $\boldsymbol{x}$ to the class:

$$\hat{\omega} = \arg\max_{\omega_m} \prod_{j=1}^{L} P(\boldsymbol{x}_{\cdot,j} \mid \omega_m). \tag{4.80}$$

This method proved to perform quite well even in case the statistical independence assumption is relaxed.

4.3 Binary Decision Trees

A popular class of nonlinear methods for the classification known as 'decision trees' is now considered. They are one of the most popular and intuitive methods for classification as they can treat mixed numeric and categorical datasets. Also, due to their structural simplicity, they are easily interpretable. In a classification tree the attributes, or features, characterising the observed phenomena are partitioned in different regions, or 'branches' by some points t_b of the attribute domain called 'nodes'. The nodes and branches of the different features are linked together in a sequential manner by a set of IF-THEN rules representing the decision process of the classification and the graphical portrait of the ensemble of these decisions paths displays the shape of the tree (Figure 4.9). Classifying an observation $\boldsymbol{x}$ with L features in M classes means then to climb a tree where at each node t_b the following branch is selected conditionally on whether the value of the attribute $\boldsymbol{x}_{\cdot,j}$ with $j = 1,\ldots,L$ is larger or smaller than some threshold value α_{tb}; the point where the climb stops, the leaf, represents the category to which the observation is assigned.

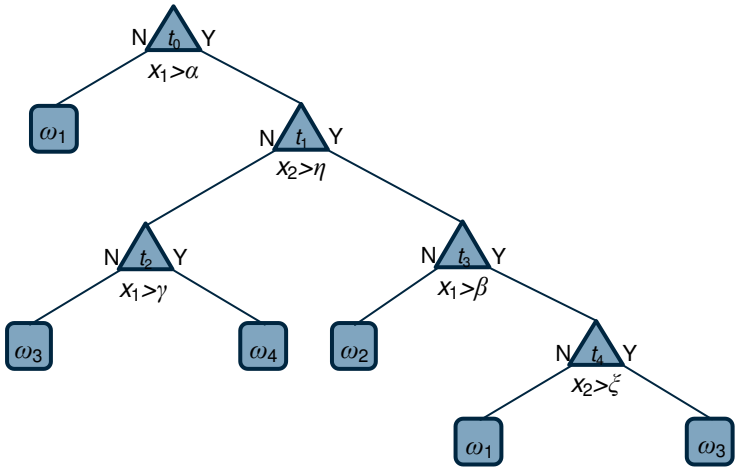


Figure 4.9 Example of binary classification tree for a dataset with two-dimensional features in four classes

This way of performing the classification, having only two possible outcomes for each node, is known as ordinary binary classification trees (OBCTs) and is also geometrically equivalent of splitting the space spanned by the L features into B hyperrectangles representing the different classes (Figure 4.10).

Other types of trees which split the space into convex polyhedral cells or into spheres are also possible.

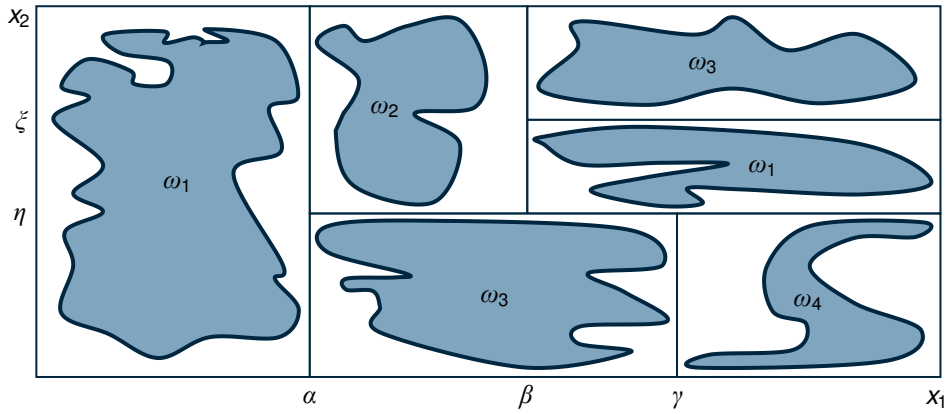


Figure 4.10 Spatial representation of the decision tree of Figure 4.9. The boxes represent the partitioning and the shaded areas are the regions occupied by the observations of the aggregated per classes

A key question is now which attribute to consider as the first feature in the training phase, which therefore represents the root, and how to build the tree from there. Since each node t_b is associated with a specific subset $\boldsymbol{X}_{tb}$ of the training set $\boldsymbol{X}$ then splitting of a node is equivalent to the split of the subset $\boldsymbol{X}_{tb}$ into two disjoint children subsets $\boldsymbol{X}_{tb}|_{\mathrm{Y}}$ and $\boldsymbol{X}_{tb}|_{\mathrm{N}}$, where the first subscript 'Y' corresponds to the answer 'YES' to the question posed by the inequality $\boldsymbol{x}_{i,\cdot} > \alpha_{tb}$ and the 'N' subscript stands for 'NO'. So, for every split b it is

$$\begin{aligned} \boldsymbol{X}_{t_b}\big|_{\mathrm{Y}} \cap \boldsymbol{X}_{t_b}\big|_{\mathrm{N}} &= \emptyset \text{ and} \\ \boldsymbol{X}_{t_b}\big|_{\mathrm{Y}} \cup \boldsymbol{X}_{t_b}\big|_{\mathrm{N}} &= \boldsymbol{X}_{t_b}. \end{aligned} \tag{4.81}$$

In the simple scenario where $\boldsymbol{X}$ has only scalar attributes there are infinite possible IF-THEN rules, i.e. the threshold α_{tb} of the b-th node can assume all the values in $\Re$ but only a finite number of splits is practically feasible. Indeed, since the training set $\boldsymbol{X}$ has a finite number of observations N, each j-th feature can take at most as many values as the cardinality of $\boldsymbol{x}_{\cdot,j}$ in the subset $\boldsymbol{X}_{tb}$. Calling $N_{tb,j}$ this cardinality the number of possible splits at the current node is

$$N_{t_b} = \sum_{j=1}^{L} N_{t_b,j}\,. \tag{4.82}$$

A splitting criterion is therefore required to select the optimal partition associated to the subset $\boldsymbol{X}_{tb}$. Common sense suggests that these children subsets $\boldsymbol{X}_{tb}|_{\mathrm{Y}}$ and $\boldsymbol{X}_{tb}|_{\mathrm{N}}$ originated by the split should present a higher preference for one or more specific classes with respect to the subset where the parent node was determined, for example one would expect that by choosing 'YES' at node t_b the subset $\boldsymbol{X}_{tb}|_{\mathrm{Y}}$ is composed by a group of observations belonging to a number $G < M$ of classes. From considering the probability $P(\omega_m|t_b)$ that a vector in the subset $\boldsymbol{X}_{tb}$, associated to the node t_b, belongs to class the m-th class, a commonly used measure of this preference is provided by the relation

$$H(t_b) = -\sum_{m=1}^{M} P(\omega_m \mid t_b) \log_2 P(\omega_m \mid t_b), \tag{4.83}$$

which is nothing but a different form of the Shannon's information entropy already seen in (4.71), with a base 2 logarithms instead of the natural logarithm. Thus, by computing this quantity one can measure the marginal decrease in 'information entropy', i.e. the increase in discriminatory power induced by the split t_b. It is possible to demonstrate that this measure is zero when all the remaining observations fall in a same class ω_α ($P(\omega_\alpha|t_b) = 1$) and is maximum when all the probabilities of assignment are the same ($P(\omega_m|t_b) = 1/M$, $\forall\ m = 1,\ldots,M$). To further clarify this concepts, let us call $N_{tb}|_Y$ the number of observations in the subset $\boldsymbol{X}_{tb}|_Y$ and $N_{tb}|_N$ those belonging to $\boldsymbol{X}_{tb}|_N$, so that $N_{tb} = N_{tb}|_Y + N_{tb}|_N$ will be the number of points affected by the split t_b. The increase in discriminatory power induced by the split is the given by:

$$\Delta H(t_b) = H(t_b) - \frac{N_{t_b}\big|_Y}{N_{t_b}} H\ t_{b+1}\big|_Y\ - \frac{N_{t_b}\big|_N}{N_{t_b}} H\ t_{b+1}\big|_N\ , \tag{4.84}$$

where $t_{b+1}|_Y$ and $t_{b+1}|_N$ are the children nodes generated by the parent node t_b respectively in the $\boldsymbol{X}_{tb}|_Y$ and $\boldsymbol{X}_{tb}|_N$ subsets. The best split will then be the one that leads to the minimum possible value of entropy, or in other terms the one that maximizes $\Delta H(t_b)$. As for the computation of the probabilities $P(\omega_m|t_b)$ they are generally computed as the ratio of all points $N_{tb}^{(m)}$ falling in the category ω_m over the total number N_{tb} of elements of the whole subset $\boldsymbol{X}_{tb}$. Other splitting criteria have been proposed, yielding better performances in some conditions, however their suitability essentially depends on the kind of problem.

Now that it is clear which path leads to the optimal splitting of the tree it is also important to provide a rule to indicate when to stop the branching. An overgrown tree can present the inconvenience of overfitting the training set, adhering too closely to some details, thus resulting in a lack of generalization capability. Of course, the most intuitive criterion is to stop when there is no more ambiguity in the remaining observations and they all belong to the same class ($P(\omega_\alpha|t_b) = 1$), but this may not always be feasible or may lead to unpractically large trees. One possibility is then to stop the split when a sufficient level of discrimination T is reached, i.e. if any possible split reduces the uncertainty of classification more than T, however this criterion is hardly ever a sensible choice. An alternative can be to impose a maximum number of splits or more commonly to perform a 'pruning' of the tree. In general terms, the latter consists in producing first a bigger classification tree and then, according to some criteria that are usually a compromise between complexity and classification error, prune some of its branches.

Regardless the stopping criteria chosen once a terminal node (a leaf) is reached it has to be assigned to a class. This is trivial if all the observations have the same class, in all other cases the simplest way is to assign it to the category ω_α which most the remaining observations belong to:

$$\boldsymbol{x} \in \omega_\alpha \text{ if } \alpha = \arg\max_m P(\omega_m \mid t_b). \tag{4.85}$$

Some similarities between the decision tree classifiers and the neural networks have been pointed out. Both methods aim at forming complex decision boundaries in the feature space, however a major difference lies in the way decisions are made. Decision trees employ a hierarchically structured decision function in a sequential manner while the NN work in a parallel fashion. From a performance point of view the NN offers generally better results with respect to the classification error, but the decision trees require much less training time. Despite the differences, however, it has been shown that linear tree classifiers can be adequately mapped to a multilayer perceptron structure. Construction of fuzzy decision trees have also been suggested, by allowing the possibility of partial membership of a feature vector in the nodes that make up the tree structure. The hierarchical nature of the decision trees is indeed important to rank the decision variables but often conceals the logical rules behind the decision and the resulting decision sequence often lacks generality and flexibility. Application of the fuzzy theory to the crisp decision trees dates back to the seventies and it was initially focused on the design of the optimal algorithm that given the fuzzy tree allow the extract the optimal class assignment for the inputs. Most of the current research on the coupling between decision trees and fuzzy inference, which regained popularity in the nineties, is mainly devoted to the problem of automatic induction of the 'soft' classification tree from a set of input data. Various approaches have been tested, some considering the decision trees as probabilistic classifiers (Quinlan, 1987; Quinlan, 1990), some others more than dealing with the statistical uncertainty used the fuzzy trees with the intent of representing and incorporating in the decision the cognitive uncertainty associated with the human thinking process (Yuan and Shaw, 1995). The idea is then to make a crisp decision tree more flexible by using its hierarchical structure to build a fuzzy inference engine (FIE), i.e. fuzzification is achieved by imposing a fuzzy structure over the basic skeleton of a standard decision tree. Many algorithms have been proposed in time for the construction of the translation of the classification decision trees in fuzzy terms all having in common some main characteristics. In (Suárez and Lutsko, 1999) and (Olaru and Wehenkel, 2003) each split is redefined in fuzzy terms leading to the so-called 'soft decision', which, opposed to the crisp splits provided by the classical decision trees, introduce the possibility of having overlapping regions where the input can be assigned

to both the successors. This ambiguity is modulated by optimizing the cut points and the width of the transition zones among the successors. In different terms this is what has been obtained in (Marsili-Libelli et al., 2012) where the crisp partitioning of the feature sets defined by a previously determined classification tree is replaced by a softer one. Here a parametrized membership function is assigned to each interval defined by two cut-points on the same variable, initially placing the maximum value in the middle point of the interval and the minimum are placed in the middle points of the adjacent intervals; this way the intersection point of the memberships, i.e. the highest degree of fuzziness results located exactly in correspondence of the breakpoints. A subsequent optimization procedure on the same training data used to define the crisp decision tree will then determine the best parameters of the membership functions rearranging their shape and thus defining the optimal width of the transition zones.

4.4 Support Vector Machine

In section 4.2 the goal was to design a classifier based on probability density or probability functions, while in section 4.3 decision trees were considered, whose resulting classifiers were equivalent to a set of linear discriminant functions, though hierarchically arranged. Here we will focus on the design of a particular category of linear classifiers, regardless of the underlying distributions describing the training data. The method, called support vector machine (SVM) will be first presented in the case of linearly separable classes. Then will be discussed the application of the linear method to tackle the nonlinearly separable classes without involving the manipulation of the spatial representation of the data. A method will be finally presented in case the nonlinearities are such that some minor adjustments are not sufficient to guarantee a good performance: an appropriate mapping of the data in some higher-dimensional space will ensure the possibility to linearly separate the classes and so to design the optimal linear classifier.

4.4.1 Support vector machine in the linear case

The major advantage of linear classifiers is their simplicity and computational attractiveness. The paragraph starts with the assumption that all the observations $\boldsymbol{x}$ can be effectively classified using a linear classifier, and will discuss the application of the SVM techniques for the computation of the corresponding linear discriminant functions. A more general problem where a linear classifier cannot correctly classify all vectors, is then discussed and some approach to design an optimal linear classifier to tackle the non-linarites are presented.

Let us once again focus on the two-class case and consider the previously introduced discriminant function to be linear. The decision hypersurface in the L-dimensional feature space of the generic observation $\boldsymbol{x} \in \Re^L$ is then a hyperplane

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0 = 0 \tag{4.86}$$

where $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_L]^T$ is known as the weight vector and w_0 is here used to indicate the intercept (in the special case that $w_0 = 0$, the hyperplane passes through the origin). Picking two points $\boldsymbol{x}_{1,\cdot}$, $\boldsymbol{x}_{2,\cdot}$ on the decision hyperplane, then

$$0 = \boldsymbol{w}^T \boldsymbol{x}_{1,\cdot} + w_0 = \boldsymbol{w}^T \boldsymbol{x}_{2,\cdot} + w_0 \tag{4.87}$$

thus,

$$\boldsymbol{w}^T (\boldsymbol{x}_{1,\cdot} - \boldsymbol{x}_{2,\cdot}) = 0 . \tag{4.88}$$

The difference vector $\boldsymbol{x}_{1,\cdot} - \boldsymbol{x}_{2,\cdot}$ lies too on the decision hyperplane, then the vector $\boldsymbol{w}$ is orthogonal to that hyperplane. Now it is easy to deduce that the norm $|g(\boldsymbol{x})|$ represents the Euclidean distance of the point $\boldsymbol{x}$ from the decision hyperplane and that it will assume positive values on one side of the plane and negative on the other.

4.4.1.1 Linearly separable classes

Assuming then that the two classes ω_1 and ω_2 are linearly separable the objective is then to determine the hyperplane $g(\boldsymbol{x})$ that classifies all the N training vectors $\boldsymbol{x}_{i,\cdot} \in \boldsymbol{X}$ ($\boldsymbol{X} \in \Re^{N \times L}$). It can be demonstrated, however, that this hyperplane is not unique, and the objective must then be adjusted in the research of the optimal separation hyperplane, one that not only is able to classify all the training data but also presents some characteristics that make it a potentially good classifier for the (unknown) samples yet to come. In other words, it should be not too much tailored on the training data and this property is called the 'generalization performance' of the classifier. From these considerations, it is now possible to affirm that a sensible choice for the optimal separation hyperplane is the one that maximizes the 'margin' from the classes. In more mathematical terms the goal can be stated as the research of the separation hyperplane identified by the direction $\boldsymbol{w}$ for which is maximum the distance d from the nearest points in class ω_1 and class ω_2. Being w_1 and w_2 the components of $\boldsymbol{w}$ on the spaces of the two classes, the distance of the generic point $\boldsymbol{x}_{i,\cdot}$ is computed as

$$d(\boldsymbol{x}_{i,\cdot}) = \frac{|\, g(\boldsymbol{x}_{i,\cdot})\,|}{\sqrt{w_1^2 + w_2^2}} = \frac{|\, g(\boldsymbol{x}_{i,\cdot})\,|}{\|\boldsymbol{w}\|}. \tag{4.89}$$

In order to separate the following considerations from the problem of scaling, it will be from now on assumed that the direction $\boldsymbol{w}$ and the threshold w_0 are such that $g(\boldsymbol{x}) = 1$ for the nearest points in ω_1 and -1 for those of ω_2 i.e.

$$\begin{aligned} \boldsymbol{w}^T \boldsymbol{x} + w_0 \geq 1, \; \forall\, \boldsymbol{x} \in \omega_1, \\ \boldsymbol{w}^T \boldsymbol{x} + w_0 \leq -1, \; \forall\, \boldsymbol{x} \in \omega_2. \end{aligned} \tag{4.90}$$

This way the margin will be

$$\frac{1}{\|\boldsymbol{w}\|} + \frac{1}{\|\boldsymbol{w}\|} = \frac{2}{\|\boldsymbol{w}\|}, \tag{4.91}$$

and calling y_i the class labels of each sample of the training set, equal to 1 for class ω_1 and -1 otherwise, the optimization problem can be formalized as the computation of the defining parameters $\boldsymbol{w}$ and w_0 of the hyperplane such that is minimized

$$J(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2, \tag{4.92}$$

subject to

$$y_i(\boldsymbol{w}^T\boldsymbol{x}+w_0)\geq 1 \text{ for } i = 1,\ldots,N, \tag{4.93}$$

which is a quadratic optimization problem with linear inequality constraints.

If the Karush–Kuhn–Tucker (KKT) necessary conditions:

$$\begin{gathered}\frac{\partial}{\partial \boldsymbol{w}}\mathcal{L}(\boldsymbol{w},w_0,\boldsymbol{\lambda})=0;\\ \frac{\partial}{\partial w_0}\mathcal{L}(\boldsymbol{w},w_0,\boldsymbol{\lambda})=0;\\ \boldsymbol{\lambda}\geq 0;\\ \lambda_i[y_i(\boldsymbol{w}^T\boldsymbol{x}_{i,\cdot}+w_0)-1]=0, \text{ for } i = 1,\ldots,N.\end{gathered} \tag{4.94}$$

are met, where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers and the Lagrangian function

$$\mathcal{L}(\boldsymbol{w},w_0,\boldsymbol{\lambda})=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}-\sum_{i=1}^{N}\lambda_i[y_i(\boldsymbol{w}^T\boldsymbol{x}_{i,\cdot}+w_0)-1], \tag{4.95}$$

the first two conditions of (4.94) can be respectively rewritten as

$$\begin{gathered}\boldsymbol{w}=\sum_{i=1}^{N}\lambda_i y_i\boldsymbol{x}_{i,\cdot}, \text{ and}\\ \sum_{i=1}^{N}\lambda_i y_i=0.\end{gathered} \tag{4.96}$$

Actually, since the Lagrangian multipliers λ_i can be either positive or zero, it is safe to assume that not all N of them give a contribution but in more general terms the direction is given by a combination of the $N_s \leq N$ feature vectors $\hat{\boldsymbol{x}}_{i,\cdot} \in \hat{\boldsymbol{X}}$ ($\hat{\boldsymbol{X}} \in \Re^{Ns\times L}$) corresponding to the non-zero multipliers:

$$\boldsymbol{w}=\sum_{i=1}^{N_s}\lambda_i y_i\hat{\boldsymbol{x}}_{i,\cdot}. \tag{4.97}$$

These are known as support vectors and the optimum hyperplane classifier as a support vector machine (SVM). They are the critical elements of the optimization task and in fact the optimization results insensitive to all other vectors for which $\lambda_i = 0$, which can lie in all the space associated to their class, while the support vectors instead can be only found in either of the hyperplanes

$$\boldsymbol{w}^T\hat{\boldsymbol{x}}+w_0=\pm 1. \tag{4.98}$$

In order to determine the parameters, the problem can be solved considering the equivalent dual form of (4.92) by maximizing

$$\mathcal{L}(\boldsymbol{w},w_0,\boldsymbol{\lambda}), \tag{4.99}$$

subject to the KKT conditions previously introduced:

$$\boldsymbol{w} = \sum_{i=1}^{N_s} \lambda_i y_i \hat{\boldsymbol{x}}_{i,\cdot}\,; \qquad \sum_{i=1}^{N_s} \lambda_i y_i = 0\,; \qquad \lambda_i \geq 0, \text{ for } i = 1,\ldots,N_s, \tag{4.100}$$

which have the advantage of presenting two equalities instead of inequalities and thus can be substituted in (4.99) and reformulate it as

$$\max_{\lambda} \left(\sum_{i=1}^{N_s} \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \left\langle \hat{\boldsymbol{x}}_{i,\cdot}, \hat{\boldsymbol{x}}_{j,\cdot} \right\rangle \right), \tag{4.101}$$

where the possibility to write the optimization problem in terms of the inner product $\langle \hat{\boldsymbol{x}}_{i,\cdot}, \hat{\boldsymbol{x}}_{j,\cdot} \rangle$ will come in handy later. Once the optimal λ_i are determined, the hyperplane can be computed from the first constraint and the classification is performed as

$$\boldsymbol{x} \in \omega_1\ (\omega_2) \text{ if } g(\boldsymbol{x}) = \sum_{i=1}^{N_S} \lambda_i y_i \left\langle \hat{\boldsymbol{x}}_{i,\cdot}, \boldsymbol{x} \right\rangle + w_0 > (<) 0\,. \tag{4.102}$$

Although the resulting optimal hyperplane is unique, no guarantee in this sense is given concerning the Lagrange multipliers, which means the expansion of $\boldsymbol{w}$ in terms of support vectors in may not be unique.

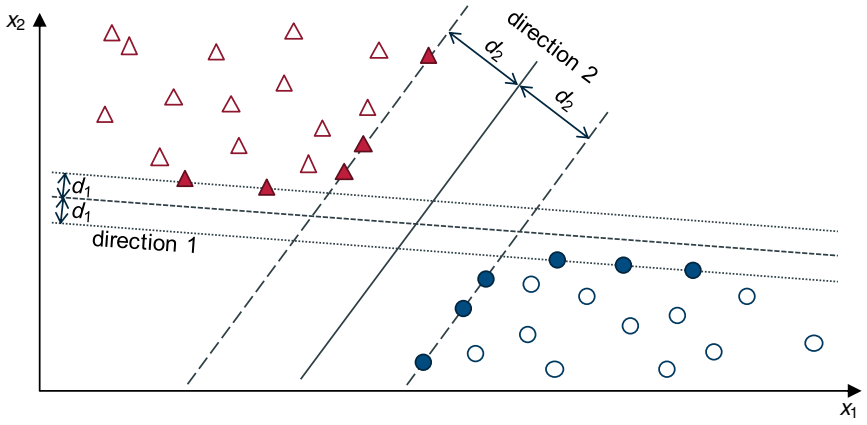


Figure 4.11 Different possibilities for the separation hyperplane of a linearly separable two-class dataset. The solid-coloured points are the support vectors. The direction 2 results the one that maximizes the separation margin

Some even more general considerations on the geometrical meaning of the SVM method can be made. Denoting as $\text{conv}\{\boldsymbol{X}\}$ the convex hull of the dataset $\hat{\boldsymbol{X}}$ containing the support vectors it can be demonstrated that it is formed by all the convex combinations of the N_s elements of the set $\boldsymbol{X}$, i.e.:

$$\text{conv}\{\hat{\boldsymbol{X}}\} = \left\{ \boldsymbol{y} : \boldsymbol{y} = \sum_{i=1}^{N_s} \lambda_i \hat{\boldsymbol{x}}_{i,\cdot} \right\}, \quad \hat{\boldsymbol{x}}_{i,\cdot} \in \hat{\boldsymbol{X}};\ \sum_{i=1}^{N_s} \lambda_i = 1;\ 0 \leq \lambda_i \leq 1\,. \tag{4.103}$$

Then for the linearly separable case solving the optimization problem in (4.101) is the equivalent of determining the hyperplane that bisects the linear segment joining two nearest points between the convex hulls of the data classes or, in terms of maximum. If

$y_i = 1$ and $y_i = -1$ are used to identify respectively the points classified in the first and in the second class, the problem becomes the minimization of the norm:

$$\min_{\lambda} \left\| \sum_{i:y_i=1} \lambda_i \hat{\boldsymbol{x}}_{i,\cdot} - \sum_{i:y_i=-1} \lambda_i \hat{\boldsymbol{x}}_{i,\cdot} \right\|^2 \tag{4.104}$$

under the constraints

$$\sum_{i:y_i=1} \lambda_i = 1, \quad \sum_{i:y_i=-1} \lambda_i = 1; \lambda_i \geq 0 \text{ and } i = 1,\ldots,N_s. \tag{4.105}$$

4.4.1.2 Non-separable classes

In case the two classes are not linearly separable it would not be possible to determine the optimal hyperplane as the one that maximizes the distance between the parallel hyperplanes of (4.98) because some of them will be falling inside the band or even cross the hyperplane and therefore are misclassified. Rewriting the condition (4.93) as

$$y_i(\boldsymbol{w}^T \hat{\boldsymbol{x}}_{i,\cdot} + w_0) \geq 1 - \xi_i, \text{ for } i = 1,\ldots,N_s, \tag{4.106}$$

where $N_s \leq N$ is again the number of non-zero Lagrangian multipliers and feature vectors $\hat{\boldsymbol{x}}_{i,\cdot}$ are their associated support vectors. The so-called 'slack variables' ξ_i are positive values used to discriminate the points that fall outside the band ($\xi_i = 0$) from those that fall inside the parallel hyperplanes ($0 < \xi_i \leq 1$) and those that across the border ($\xi_i > 1$). The optimization problem should then be reformulated as the research of the optimal separation hyperplane that maximizes the margin while at the same time keeping small the number of misclassified observations of the training set. In mathematics, the cost function (4.92) becomes

$$J(\boldsymbol{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N_s} \boldsymbol{I}(\xi_i), \tag{4.107}$$

where C is a positive coefficient that controls the relative influence of the two competing terms, $\boldsymbol{\xi}$ is the parameter vector and $\boldsymbol{I}(\xi_i)$ is such that

$$\boldsymbol{I}(\xi_i) = \begin{cases} 1, & \xi_i > 0 \\ 0, & \xi_i = 0 \end{cases} \tag{4.108}$$

$\boldsymbol{I}(\xi_i)$, however, is discontinuous, so it is preferable to express (4.107) as

$$J(\boldsymbol{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N_s} \xi_i, \tag{4.109}$$

and use (4.106) as constraint.

It is again possible to formulate the dual optimization problem in form of a maximization task of its Lagrangian expression, obtaining the formulation

$$\max_{\boldsymbol{\lambda}}\left(\sum_{i=1}^{N_s}\lambda_i-\frac{1}{2}\sum_{i,j}\lambda_i\lambda_j y_i y_j\left\langle\hat{\boldsymbol{x}}_{i,\cdot},\hat{\boldsymbol{x}}_{j,\cdot}\right\rangle\right), \tag{4.110}$$

subject to

$$\begin{gathered}0\le\lambda_i\le C\text{, for } i=1,\ldots,N_s;\\ \sum_{i=1}^{N_s}\lambda_i y_i=0.\end{gathered} \tag{4.111}$$

Notice that the multipliers λ_i

$$\lambda_i[y_i(\boldsymbol{w}^T\hat{\boldsymbol{x}}_{i,\cdot}+w_0)-1+\xi_i]=0\text{, for } i=1,\ldots,N_s, \tag{4.112}$$

associated to the points inside the band, even those that 'cross the border', i.e. all Lagrangian multipliers with $\xi_i<0$, are all equal to C, that is they all have the largest possible influence in determining the optimal hyperplane direction $\boldsymbol{w}$.

The main difference between the linearly separable and non-separable cases lies in the fact that for the latter one the Lagrange multipliers need to be bounded above by C, or that, from another point of view, the linearly separable case corresponds to the limit case when $C\to\infty$.

The extension of the just considered SVM methods to the multi-class case is straightforward. Given M classes the objective will be to determine the optimal separation hyperplane $g_\alpha(\boldsymbol{x})=0$, $\alpha\in[1,M]$, such that if the observation $\boldsymbol{x}$ belongs to the α-th class $g_\alpha(\boldsymbol{x})>g_\gamma(\boldsymbol{x})$ ($\alpha\neq\gamma$) and, as in (4.90), the linear function will return $g_\alpha(\boldsymbol{x})>0$ if $\boldsymbol{x}\in\omega_\alpha$ and $g_\alpha(\boldsymbol{x})<0$ otherwise, then

$$\boldsymbol{x}\in\omega_\alpha\text{ if }\alpha=\arg\max_m\{g_m(\boldsymbol{x})\}. \tag{4.113}$$

This kind of 'one-against-all' technique may lead to undetermined regions where more the one hyperplane is positive. An alternative is the 'one-against-one' approach where, $M\cdot(M-1)/2$ binary classifiers are trained on each pair of classes then perform the class assignment on the basis of a majority vote. A major drawback, however, is represented by the relatively large number of binary classifiers to be trained.

4.4.2 Support vector machine for the nonlinear case

In the previous paragraph the attention was focused on the optimal design of a linear SVM classifier both in case the classes are actually linearly separable and in the case when they are not. In the latter, a full separation of the observations belonging to different classes is not possible and the optimal partition is obtained by minimizing the misclassification error. This compromise, however, does not always lead to satisfactory results and a fully nonlinear classifier should be sought to tackle the problem.

One strategy could consist in performing a transformation of the observations from their L-dimensional space to a K-dimensional space where the classes can be efficiently separated by a hyperplane:

$$\boldsymbol{x} \in \Re^L \rightarrow \boldsymbol{z} \in \Re^K . \tag{4.114}$$

Since this new space $\Re^K$ usually has a higher dimensionality we might conclude that the mapping, although motivated by an increase of the separation efficiency, would also increase the computational complexity. However, this is not the case. Having described the optimization problem (4.101) in terms of inner products implies that in the linear case the expression of the optimal separation hyperplane was:

$$\begin{aligned} g(\boldsymbol{x}) &= \boldsymbol{w}^T \boldsymbol{x} + w_0 \\ &= \sum_{i=1}^{N_s} \lambda_i y_i \left\langle \hat{\boldsymbol{x}}_{i,\cdot}, \boldsymbol{x} \right\rangle + w_0 \end{aligned} \tag{4.115}$$

which means that ultimately the classification of the unknown observation $\boldsymbol{x}$ depends on the inner product with the support vectors. Supposing to map a two-dimensional observation $\boldsymbol{x}$ in a three-dimensional one, for example:

$$\boldsymbol{x} \in \Re^2 \rightarrow \boldsymbol{z} = \begin{bmatrix} x_1^2 \\ x_1 x_2 \sqrt{2} \\ x_2^2 \end{bmatrix}, \tag{4.116}$$

one would find that the correspondence of the inner products in the two spaces is

$$\left\langle \boldsymbol{z}_{i,\cdot}, \boldsymbol{z}_{j,\cdot} \right\rangle = \left\langle \boldsymbol{x}_{i,\cdot}, \boldsymbol{x}_{j,\cdot} \right\rangle^2 . \tag{4.117}$$

This is also known as ‘kernel trick’ and is consequence of the Mercer’s theorem, which states that it exists some feature mapping $\phi(\cdot)$ of a L-dimensional variable in a new Euclidean space Z:

$$\boldsymbol{x} \rightarrow \phi(\boldsymbol{x}) \in Z , \tag{4.118}$$

such that

$$\left\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{z}) \right\rangle = K(\boldsymbol{x}, \boldsymbol{z}) , \tag{4.119}$$

where $K(\boldsymbol{x},\boldsymbol{z})$ is a symmetric function such that $\forall\, \boldsymbol{x} \in \Re^L$

$$\begin{aligned} \int K(\boldsymbol{x}, \boldsymbol{z}) g(\boldsymbol{x}) g(\boldsymbol{z})\, d\boldsymbol{x}\, d\boldsymbol{z} \geq 0 , \\ \int g(\boldsymbol{x})^2\, d\boldsymbol{x} < +\infty . \end{aligned} \tag{4.120}$$

This class of functions $K(\boldsymbol{x},\boldsymbol{z})$are called ‘kernels’ and the space Z in which they define the inner product is known as ‘reproducing kernel Hilbert space’ (RKHS). Popular choices for the kernel functions are:

– polynomials

$$K(\boldsymbol{x},\boldsymbol{z}) = (\boldsymbol{x}^T\boldsymbol{z}+1)^q, \quad q>0; \tag{4.121}$$

– radial basis functions

$$K(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|^2}{\sigma^2}\right), \text{ with } \sigma \text{ known } a\ priori; \tag{4.122}$$

– hyperbolic tangent (the sigmoid transform of neural networks)

$$K(\boldsymbol{x},\boldsymbol{z}) = \tanh(\beta\boldsymbol{x}^T\boldsymbol{z}+\gamma), \text{ with appropriate } \beta \text{ and } \gamma. \tag{4.123}$$

Once the kernel function is selected the dual problem (4.110) subject to the constraints of (4.111) can be reformulated as

$$\max_{\lambda}\left(\sum_{i=1}^{N_s}\lambda_i - \frac{1}{2}\sum_{i,j}\lambda_i\lambda_j y_i y_j K(\hat{\boldsymbol{x}}_{i,\cdot},\hat{\boldsymbol{x}}_{j,\cdot})\right) \tag{4.124}$$

and the class assignment becomes

$$\boldsymbol{x}\in\omega_1\ (\omega_2) \text{ if } g(\boldsymbol{x}) = \sum_{i=1}^{N_s}\lambda_i y_i K(\hat{\boldsymbol{x}}_{i,\cdot},\boldsymbol{x}) + w_0 > (<) 0\,. \tag{4.125}$$

Due to the nonlinearity of the kernel function, the resulting classifier is a nonlinear one in the original $\Re^L$ space.

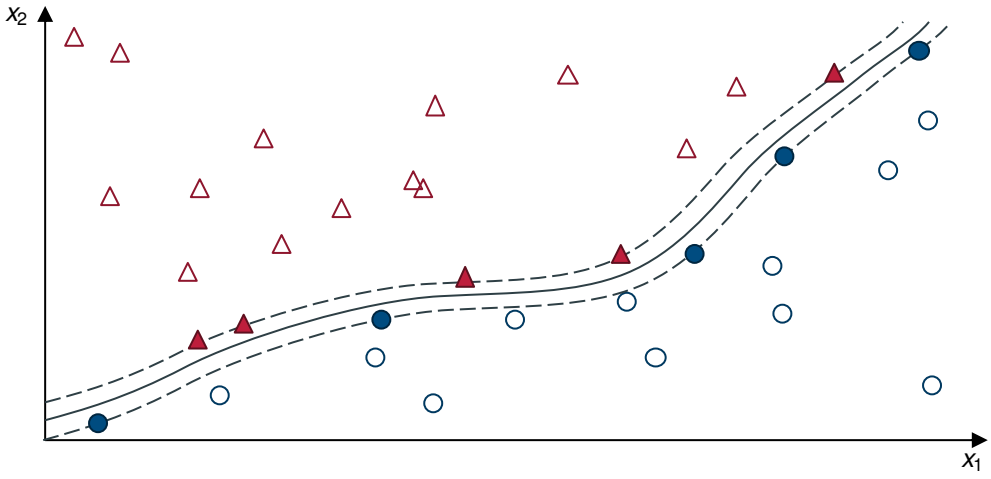


Figure 4.12 Example of nonlinear SVM classifier for the case of two non-linearly separable classes. A Gaussian radial basis function was used. the dotted lines represent the margins, the solid line is the separation hypersurface and the full marked points are the support vectors

Mercer's theorem guarantees the existence of such a space in case the conditions (4.120) are met but it does not provide any suggestion as to how to build this space or even about its dimension, so there is no hint on how to select the optimal kernel function. At most, once a kernel function is chosen, it is possible to set up yet another optimization problem in order to determine the best value of the parameter C of the cost function and the kernel parameters that minimize the classification error. The most common procedure solves the SVM task for different sets of hyperparameters, as the ensemble

of the parameters to be optimized is called, and finally select the classifier that best fit some requirements, for example maximizing the margin over the $\boldsymbol{w}$ and at the same time minimizing the bound over the hyperparameters.

A notable characteristic of the support vector machines is that the computational complexity is independent of the dimensionality of the chosen kernel space. The high-dimensional space can be designed without reference to explicit models, thus avoiding to use a large number of parameters, as would be required by the high dimensionality of the space. This property also influences the generalization performances (intended as the classification error of data outside the training set) of SVMs, which indeed are quite good. For this reason, SVM have been applied to a number of diverse applications, ranging from handwritten digit recognition, to medical imaging and person identification, usually ranking higher when compared with other classifiers (Meyer et al., 2003). Moreover, the success of the SVMs, in practice, spurred, whenever the computations could be expressed in terms of inner products, a research effort to extend a number of linear classifiers, including those based on PCA as we have seen, to nonlinear ones by embedding the kernel trick in their structure.

Chapter 5 Implementation of the fault detection methods and performance assessment

In the previous chapters an overview of the problem setting and the theoretical background of the methods chosen to tackle the fault detection problem for the present work were provided. In this chapter is discussed the implementation of these methods and their application to both the datasets introduced in Chapter 3: the operational plant data and the data generated by the numerical model.

The operational data used in the present work were provided by the plant managing society of the municipal treatment facility of Mantua, in Italy, and consisted of an almost one year long set of measurements of ammonia and nitrates concentrations measured by two ion-specific probes installed in one of the four reactors of the plant, one ANISE probe (Hach-Lange GmbH, Dusseldorf, Germany) and one S::can probe (Scan Messtechnik GmbH, Vienna, Austria). Due to the significantly low quality of the measurements provided by the latter, the data of the S::can probe were discarded and the design and testing of the detection algorithms was performed only on the data of the ANISE probe measured from the 18th September 2012 to the 7th May 2013 where can be counted 2102 aerobic phases and the same number of anoxic phases.

As already mentioned the proposed algorithm is designed to operate as a real-time fault detection (RTFD) tool supporting the supervising controller OSCAR (ETC eng, Trento, Italy) operating in the Mantua WWTP in conjunction with the plant SCADA (supervisory control and data acquisition) system, as shown in Figure 5.1. Based on the reliability assessment of the measurements produced by the RTFD, the plant manager or the automatic control supervisory system was able to adapt the control strategy to compensate for the possible loss in quality of the treatment or cost efficiency and direct the process SCADA to enact the necessary countermeasures. The choice of the optimal remedial action according to the different anomalies in the measurements quality is however beyond the scope of this research.

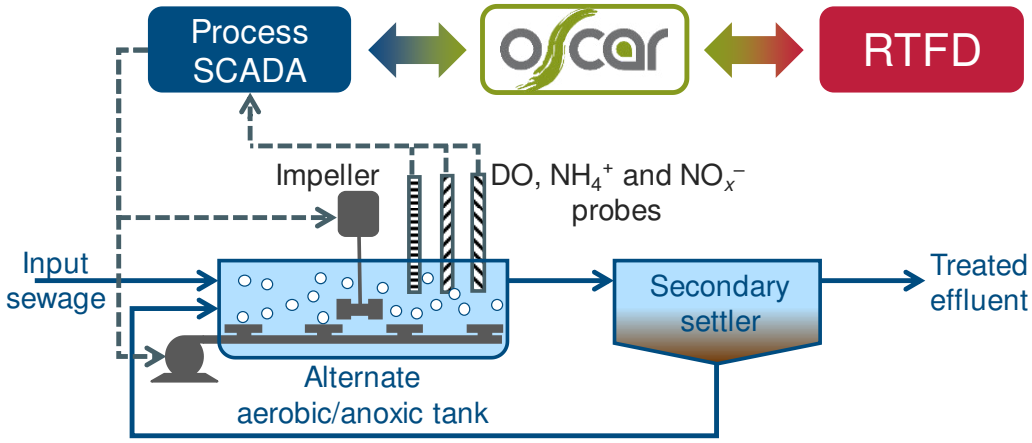


Figure 5.1 Schematic representation of the relationship among the different actors of the plant control in the configuration studied. The RTDF algorithm is developed as a support tool for the control supervisor providing it with reliability assessment of the measurements. The control supervisor will then enact the necessary control strategies through the process SCADA

In section 3.4 it was introduced and discussed the distinction between gross faults and finer anomalies affecting the measurements. The first category included all the most evident and easy-to-spot signal inconsistencies directly revealing the presence of malfunction in the instruments or in the acquisition system. With the term 'finer fault' we denote the anomalies for which a more refined investigation tool was needed in order to discriminate whether its origin lies in an instrument malfunction or in an anomalous organic loading of the influent. Thus, two different approaches to the detection were used, hierarchically developed to investigate the presence of faults with increasing levels of complexity: first the different signals are tested separately for the presence of the gross faults using the rationale discussed in section 3.4.2, then, in case nothing is found, different methods, whose implementation will be thoroughly discussed throughout this chapter, are employed for the detection of the finer faults. The prerequisite for the application of the more refined methods for the FD was the identification of a set of diagnostic parameters (see section 3.4.3) based upon the previously described data mining and classification techniques, with the aim to expose the hidden relationship characterizing the good (and by contrast the bad) behaviour of the sensors. A by-product of the parameterization of the signals, namely the correlation coefficient of the linear regression used to compute the concentration variation ratios of the ammonia, proved to have some diagnostic power by itself since it represents a measure of the roughness of a signal and thus an indirect indication of the fouling level of its measurement device. Finally, it is worth recalling that the considerations motivated by the trend analysis of the AC processes brought to the decision to split the finer FD check in two parts, one for the anoxic phase and one for the aerobic one. The two parts are based on the same methods and employ the same parametrization of the

signals but work in parallel each on its pertinent phase. A schematic representation of the FD procedure is shown in Figure 5.2.

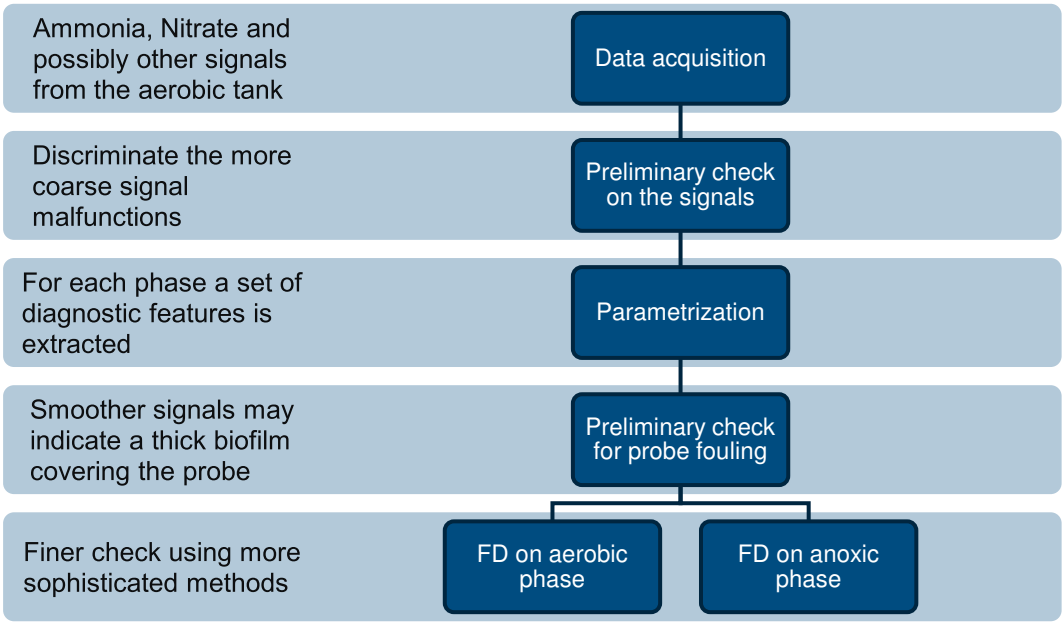


Figure 5.2 Outline of the main sequential steps RTFD algorithm

5.1 Implementation of the preliminary screenings

The main features of the treatment process and the characteristics of its quantities of interests have been already presented in Chapter 3, regarding both the operational measurements and the synthetic data. Chapter 3 also discussed the characterization and the detectability of the gross faults, using a slightly different screening approach depending on whether the data come from the operational plant or are numerically generated by the model. In the latter case, for example, there was no ambiguity as to the source of an inconsistent duration of the phases, therefore a preliminary screening to detect this kind of behaviour was not necessary. A summary of the type of faults investigated and their detection thresholds is reported in Table 5.1.

Table 5.1 Type of faults and their respective values for the plant operational data and the measurements generated by the numerical model

Type of fault	Operational data	Synthetic data
Signal breakdown	$\Delta t_{samp} \geq 10$ min (~20 samples)	$\Delta t_{samp} \geq 10$ min (~20 samples)
Constant signal / zero-scale fault	Constant values for $t \geq 10$ min (~20 samples)	Constant values for $t \geq 10$ min (~20 samples)
Spike	Δ consecutive meas. ≥ 2 mg/L (NH_4^+ and NO_x^-)	Δ consecutive meas. $\geq 10\%$ of meas. range (Table 3.3)
Anomalous phase duration	$\Delta t_{phase} \geq 3$ h	-

According to the methodology illustrated in section 3.4.3, at the end of each process phase a set of diagnostic parameters is extracted from the signals. Due to the limited

amount of data, mostly in terms of measured variables, a limited number of features from the operational data of the Mantua plant could be considered, namely the slopes and average concentration values of the nitrate and ammonia in the aerobic reactor (Table 3.13). Keeping in mind that the numerical model was developed as a surrogate of the real plant, we simulated a wider range of measurements with a precise physical meaning and which can be encountered in real-life treatment practice. The broader spectrum of possible measurements simulated with the model expanded therefore the set of features available for FD purposes, allowing to extend the group previously derived for the plant data, adding the average concentration values of the suspended solids and the dissolved oxygen, together with the temperature values (Table 3.15).

As previously mentioned, the water buffeting induced by the impellers on the probe during the anoxic phase produces ripples in the signals. These fluctuations are more pronounced when the sensor is clean, while as the level of fouling increases the roughness of the signal decreases (Figure 3.12) because the fouling layer acts like a damper to the incoming wavelets. The correlation coefficient r^2, associated to the regression line computing the slope of the nitrogen concentration, proved to be a good indicator of the probe fouling and by trial and error it was found that a value of $r^2 = 0.9$ provided a suitable alert level for probe cleaning. Figure 5.3 shows the changes in r^2 values before and after a probe cleaning. A clean probe is more sensitive to the small variations in concentrations induced by the turbulences and therefore the correlation coefficient of its linear regression coefficient r^2 is smaller. On the contrary, its increase may indicate a progressive fouling, as the signal becomes smoother.

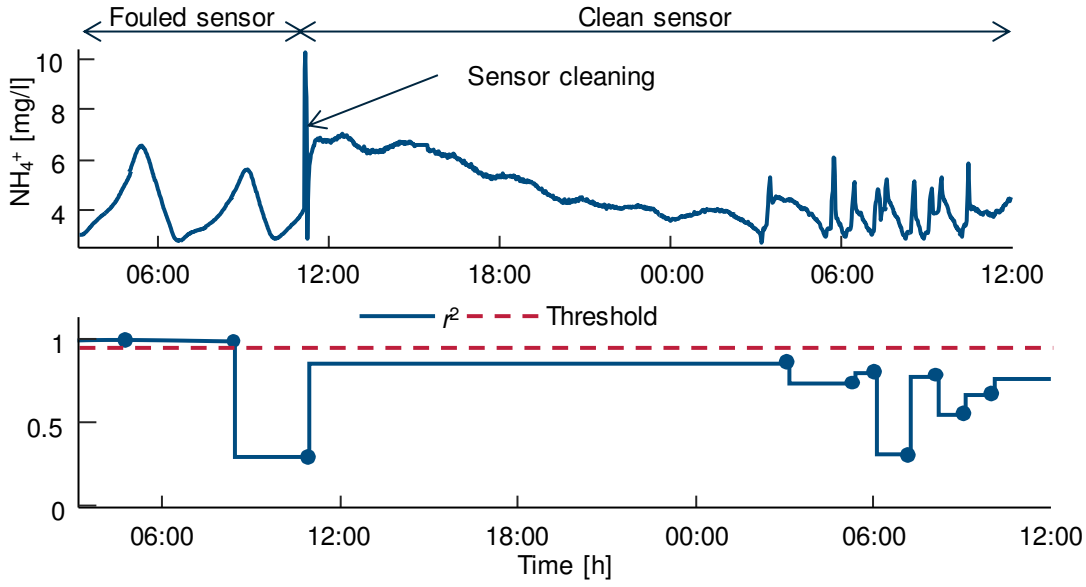


Figure 5.3 The fouling level of the ammonia probe is given by the square residuals of the regression line in the anoxic phase. The spike in the upper plot marks the probe cleaning

5.2 Implementation of the finer check methods

In this paragraph, we discuss the implementation of the FD methods used to identify the finer faults affecting the measurements. In particular, for the operational plant data two methods were employed, one is an adaptive version of the PCA algorithm discussed in section 4.1, while the other is based on the simple Bayes classifier presented in section 4.2. Some of the limitations encountered, due both to the limited set of measurements available and to the methods themselves, brought on one hand the need to test new detection methods on a wider dataset of numerically generated measurements and on the other promoted the implementation of two additional classification methods, one based on the classification trees (section 4.3) and the other on the support vector machines (section 4.4 The availability of more observed variables also allowed the use of different combinations of the diagnostic features for the training and testing of the different methods.

5.2.1 The real-time mowing window principal component analysis method

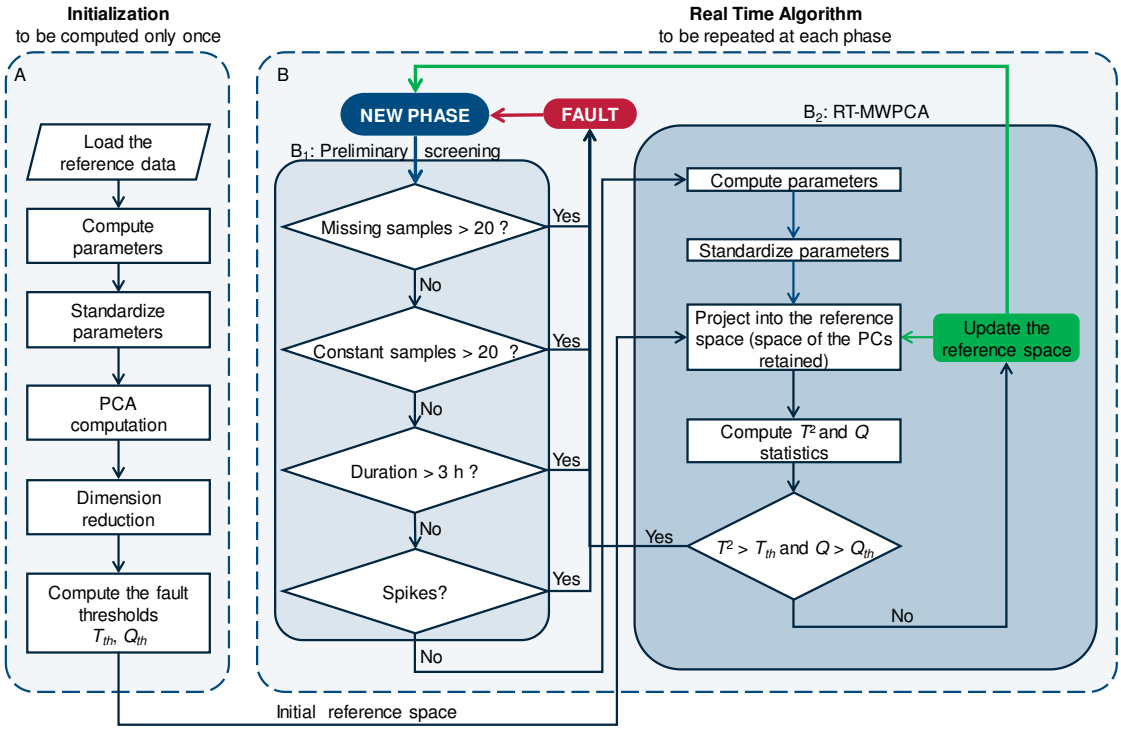


Figure 5.4 Workflow of the MWPCA algorithm

The fault detection method implemented in this work continues the idea introduced with the MWPCA by (Baggiani and Marsili-Libelli, 2009) of merging the main advantages of the recursive PCA method and the adaptive PCA method in a new methodology suitable for the application to a real-time (RT) detection process. In the RT-MWPCA method here implemented the use of a 'moving window' containing the

reference space of the PCs, sliding to adapt for the changes in time of the relationships among the parameters, is maintained and so is the idea to update this space conditionally on the respect of the statistical thresholds assessing 'how similar' is the new sample to those used to build that space. The novelty introduced here is that the two statistical thresholds are still based on those defined in the literature but multiplied by a coefficient determined during the training phase in order to maximise the detection performances. Figure 5.4 shows the complete workflow of the RT-MWPCA algorithm implemented.

5.2.1.1 Algorithm initialization

First, the fault detection algorithm is initialized (Box A in Figure 5.4) by selecting a set of reference data from which the parameters are extracted. The initial reference dataset should represent a fault-free operation period sufficiently extended to represent the weekly variability of the organic loading and is normally selected right after a probe cleaning and plant overhaul. The initial reference period chosen for each phase is therefore composed of the parameters computed for the first 100 fault-free contiguous phases for the Mantua plant data and 26 phases in the synthetic data case. A PCA is then performed on these parameters at first normalizing theme and subsequently computing the eigenvalues of the correlation matrix. The choice of the number of principal components to retain without losing relevant information is based on the eigenvalue scree plots by selecting the PCs whose associated eigenvalues are greater than 0.7 (Dunteman, 1989) and the space of the retained PCs is the initial reference space of the PCA model.

5.2.1.2 Threshold definition and optimization

Once this initial reference is computed, the FD method can be used for the real-time FD. As the measurements are collected, they are promptly tested by the preliminary screening section for the presence of the most evident sensor malfunctions (Box B_1 in Figure 5.4). After a phase is completed, without the preliminary screening detecting any fault, the phase is further analysed by the RT-MWPCA algorithm, as shown in Box B_2 of Figure 5.4.

The algorithm is conceived to follow the evolving process conditions through a real-time conditional update strategy starting with a reference set for fault detection and updating it to track the time-varying process characteristics. The fault thresholds T_{th} and Q_{th}, to be described later, are based on the Hotelling's T^2 and Q statistics. Given the reduced dimensions in the PC space, both thresholds are important, because while T^2 measures the similarity of the tested sample with respect to the subspace of the retained

components (PC_1, PC_2), Q monitors the importance of the information embedded in the excluded components.

A process anomaly is detected whenever both the current T^2 and Q scores exceed the thresholds T_{th} and Q_{th}, otherwise the phase is considered 'normal' and its parameters are allowed to update the reference space: the newest parameters are included in the dataset, while the oldest are discarded. In this way, the moving window has a constant size (the same as the initial reference set i.e. 100 samples for the real plant and 26 for the simulated measurements) and may be composed of not necessarily consecutive data.

The theoretical thresholds T^2_{lim} and Q_{lim} provided a too severe criterion for the identification of the faults in the current work and led to an undesirably high number of anomalies detected and therefore proving to be unsuitable for a reliable fault detection in our problem. More realistic discriminating values are thus determined by introducing the modified thresholds:

$$\begin{aligned} T_{th} &= k_T \cdot T^2_{lim}; \\ Q_{th} &= k_Q \cdot Q_{lim}, \end{aligned} \tag{5.1}$$

where the k_T and k_Q coefficients are estimated by minimizing the error function

$$E = |A - C| + |D - C|, \tag{5.2}$$

in which A is the number of actual faults, D is the number of faults detected by the algorithm and C is the number of matches between the detected and observed faults. Therefore, the first terms represent the number of false negatives, while the second counts the false positive detections so that the whole error function eq. (5.2) has the meaning of a classification error minimization, which was performed by the genetic algorithm `'ga'` in the Matlab® Statistics Toolbox. For practical purposes both k_T and k_Q were constrained in the $(0 \div 10)$ interval, allowing the threshold T_{th} (or Q_{th}) to be lower than the theoretical control limits or at most ten times higher.

5.2.2 The Bayesian method

Another method tested for the detection of the finer class of faults is based on the simple Bayesian classifier of paragraph 4.2. The algorithm proposed is also in this case divided in two parts, one devoted to the detection of irregular aerobic phases and one for the anoxic ones and both of course are employed only if no anomalies are detected by the preliminary check on the state of the measurements.

Each part of the algorithm is composed of an initialization step, which corresponds to the initial training of the Bayesian predictive model. Then the prediction procedure, to be repeated at each phase, can begin. This is composed of two sequential steps:

1) Preliminary screening, to detect the most evident malfunctions (same as in Box B_1 in Figure 5.4);
2) Subsequent finer fault detection, based on Bayesian updating.

At each new phase, if the preliminary screening reveals the presence of one or more gross anomalies the phase is considered faulty and the algorithm waits for the next phase. When no anomaly is detected by the preliminary check the current Bayesian predictive model, built using the parameters and the normal or abnormal state of the phases available so far, is used to estimate whether the parameters of the current phase most likely belong to a normal or an anomalous phase. At each next step the prediction is then compared to the actually observed state and the information is used to update the predictive model for the subsequent group of parameters.

5.2.2.1 Initialization step

In order to set up the initial Bayesian predictive model it is required to have a dataset composed by at least one regular phase and at least one phase in which is observed an anomaly not detected by the preliminary screening. This dataset is used to compute, for each parameter, two likelihood distributions: one is the distribution of the samples in the fault-free phases, the other in the complementary anomalous phases.

Given a set of data it is possible to use the Bayes' theorem to compute the conditional probability of detecting a fault given the current parameters as:

$$P(Fault \mid Data) = \frac{P(Data \mid Fault)P(Fault)}{P(Data)}, \tag{5.3}$$

where the first term in the numerator is the likelihood and the second is the prior.

The idea is to use the Bayes' theorem to predict the state of the process whenever a new observation is available, then at each step compare that prediction to the actual state and use this new information to update both the likelihoods and the priors in order to obtain a better prediction in the following steps. In other words, this means that given the new observation at step $t + 1$, represented by the set of parameters $\boldsymbol{x}(t+1)$, two conditional probabilities are computed: respectively eq.(5.4) and eq.(5.5). The first is the probability that the current observation represents a faulty state, i.e. belongs to a class ω_F, while the other is the probability that the process is normal, i.e. the observation $\boldsymbol{x}(t+1)$ belongs to the class ω_N:

$$P_{t+1}\ \omega_F \mid \boldsymbol{x}(t+1) = \frac{P_t(\omega_F)P_t\ \boldsymbol{x}(t+1) \mid \omega_F}{P_t(\omega_F)P_t\ \boldsymbol{x}(t+1) \mid \omega_F + P_t(\omega_N)P_t\ \boldsymbol{x}(t+1) \mid \omega_N}, \tag{5.4}$$

$$P_{t+1}\ \omega_N \mid \boldsymbol{x}(t+1) = \frac{P_t(\omega_N)P_t\ \boldsymbol{x}(t+1) \mid \omega_N}{P_t(\omega_F)P_t\ \boldsymbol{x}(t+1) \mid \omega_F + P_t(\omega_N)P_t\ \boldsymbol{x}(t+1) \mid \omega_N}. \tag{5.5}$$

Here $P_t(\boldsymbol{x}(t+1)|\omega_F)$ and $P_t(\boldsymbol{x}(t+1)|\omega_N)$ represent the likelihoods, provided the knowledge at time t, of obtaining the parameters of $\boldsymbol{x}(t+1)$ given the faulty (or normal) state of the process. So, at time t the densities of the parameters both in case of fault or normal process conditions are estimated, then they are used to evaluate the conditional probabilities of the parameters in the next sample at $t+1$.

5.2.2.2 Priors' estimation

Regarding the estimation of the priors $P_t(\omega_F)$ and $P_t(\omega_N)$ at each step t, they are computed as the frequencies of observing (or not observing) a fault, i.e. the number of anomalous or regular phases encountered divided by the total number of phases investigated so far N_t.

$$P_t(\omega_F) = \frac{1}{N_t}\sum_i \boldsymbol{x}(i), \quad i \in [1,t] \big| \boldsymbol{x}(i) \in \omega_F \tag{5.6}$$

$$P_t(\omega_N) = \frac{1}{N_t}\sum_i \boldsymbol{x}(i), \quad i \in [1,t] \big| \boldsymbol{x}(i) \in \omega_N \tag{5.7}$$

5.2.2.3 Likelihoods estimation

In principle, the probabilities density function can be either parametric distribution functions (such as normal, uniform, etc.), determined on the base of some *a priori* knowledge, or functions estimated without assuming a parametric form of the distribution. Due to the lack of prior knowledge on the distribution form of the parameters we opted out for the second approach.

As seen in paragraph 4.2.1.2 one way to approximate the densities is to use the histogram of the values. After dividing the sample space into N intervals (bins) the probability density of the sample x_i can be approximated by eqs (4.73) and (4.75).

However, estimating the densities of the likelihoods with the histogram has some major shortcomings: the number and the width of the bins deeply influences the shape of the distribution and most of all the distribution function is discrete. It was decided instead to use kernel density estimation (KDE) techniques to estimate continuous densities. The pdf is then built by creating an individual probability density curve for each parameter j of $\boldsymbol{x}$ in form of known kernel functions (which are symmetric distribution functions such as normal, etc.) for each data value, then the smooth curves are summed. From eq. (4.77) the approximated form of the likelihood of observing the j-th parameter of $\boldsymbol{x}(t+1)$ in the state $\omega_{\cdot}$ given the knowledge acquired until the time t is:

$$\tilde{P}_t\ x_j(t+1) \,|\, \omega_{\cdot} = \frac{1}{h^L}\left(\frac{1}{N_t}\sum_i K\left(\frac{\boldsymbol{x}_j(t+1) - \boldsymbol{x}_j(i)}{h}\right)\right), \quad i \in [1,t] \big| \boldsymbol{x}(i) \in \omega_{\cdot}, \tag{5.8}$$

where $K(\cdot)$ is the kernel function chosen and h is a smoothing parameter also called bandwidth.

In this study the KDE of the likelihoods is achieved using the `'fitdist'` function of the Matlab® Statistics and Machine Learning Toolbox™, choosing the normal distribution as kernel function. Using the first 150 samples of the datasets we compared the cumulative distribution functions (CDFs) obtained for different values of bandwidth in order to select the optimal smoothing parameters. The results found are approximately the same for all the parameters in both the aerobic and the anoxic phase, with a few exceptions represented by the slopes, therefore a smoothing factor $h = 0.5$ is used for all the average values and $h = 5$ for the slopes of nitrates and ammonia. Table 5.2 shows the bandwidth for the parameters extracted from the synthetic dataset but the same considerations apply also for the less populated set of parameters of the plant measurements.

Table 5.2 Parameters extracted for each process phase from the data generated by the numerical model

Parameter	KDE smoothing factor
$m_{\cdot,NH4+,in}$	0.5
$rg_{\cdot,NH4+,in}$	0.5
$m_{\cdot,NH4+,out}$	0.5
$rg_{\cdot,NH4+,out}$	0.5
$s_{\cdot,NH4+,out}$	5
$m_{\cdot,NOx-,in}$	0.5
$rg_{\cdot,NOx-,in}$	0.5
$m_{\cdot,NOx-,out}$	0.5
$rg_{\cdot,NOx-,out}$	0.5
$s_{\cdot,NOx-,out}$	5
$m_{\cdot,Temp}$	0.5
$m_{\cdot,DO,in}$	0.5
$rg_{\cdot,DO,in}$	0.5
$m_{\cdot,DO,out}$	0.5
$rg_{\cdot,DO,out}$	0.5
$m_{\cdot,TSS,in}$	0.5
$rg_{\cdot,TSS,in}$	0.5
$m_{\cdot,TSS,out}$	0.5
$rg_{\cdot,TSS,out}$	0.5

5.2.2.4 Prediction

Each time a new phase is completed and its parameters are available, the distributions, estimated as seen in the previous paragraph, are used to compute the likelihoods $P_t(\boldsymbol{x}(t+1)|\omega_F)$ (or $P_t(\boldsymbol{x}(t+1)|\omega_N)$) and then, together with the priors, are used to predict the state of the process. Assuming the likelihoods of the L parameters to be iid, thus their joint distribution can be computed as:

$$P_t\ \boldsymbol{x}(t)\,|\,\omega_{\cdot}\ = P_t\ x_1(t),\ldots,x_L(t)\,|\,\omega_{\cdot}\ = \prod_{j=1}^{L} P_t\ x_j(t)\ , \tag{5.9}$$

then eqs. (5.4) and (5.5) become:

$$P_{t+1}\ \omega_F \big| \boldsymbol{x}(t+1)\ = \frac{P_t(\omega_F)\cdot \prod_j P_t\ x_j(t+1)\big|\omega_F}{P_t(\omega_F)\cdot \prod_j P_t\ x_j(t+1)\big|\omega_F\ + P_t(\omega_N)\cdot \prod_j P_t\ x_j(t+1)\big|\omega_N} \tag{5.10}$$

and

$$P_{t+1}\ \omega_N \big| \boldsymbol{x}(t+1)\ = \frac{P_t(\omega_N)\cdot \prod_j P_t\ x_j(t+1)\big|\omega_N}{P_t(\omega_F)\cdot \prod_j P_t\ x_j(t+1)\big|\omega_F\ + P_t(\omega_N)\cdot \prod_j P_t\ x_j(t+1)\big|\omega_N}\cdot \tag{5.11}$$

The classification is then determined by the largest between the two probabilities:

$$State_{t+1} = \arg\max_{\omega_F,\omega_N}\ P_{t+1}\ \omega_N \big| \boldsymbol{x}(t+1)\ , P_{t+1}\ \omega_F \big| \boldsymbol{x}(t+1)\quad . \tag{5.12}$$

This Bayesian approach to the problem is intrinsically evolutionary. At each step, the features extracted from the measurements are compared to the previous ones, transformed into the probability distributions of the parameters and of the states, in order to predict the most probable condition of the sensors. In this method, it is not possible to isolate a training phase from a validation phase, since it recursively trains on the available data to forecast the state observation that will be part of the training set in the next iteration.

5.2.3 The binary classification trees

As introduced in paragraph 4.3 among the elements characterizing the binary classification trees the most important are the splitting criterion used and the stopping criterion determining the degree of complexity of the tree. In this study the implementation of the binary trees for classification are based on the `'fitctree'` function of the Matlab® Statistics and Machine Learning Toolbox™ using the ‘Gini diversity index’ (GDI) instead of the Shannon entropy of eq. (4.83) as a splitting criterion. This is a transformation of the ‘Simpson index’ λ representing the probability that two elements randomly drawn from a dataset of interest share the same class; its complement $1-\lambda$ therefore is the probability that the two entities belong to different classes, and is known in ecology as the ‘Gini diversity index’ defining the probability of encounter of different species. Keeping the same notation of the paragraph 4.3, the expression of the GDI is:

$$1-\sum_{m=1}^{M} P(\omega_m \mid t_b)^2 . \tag{5.13}$$

where the sum is over the classes m at the node t_b, and $P(\omega_m|t_b)$ is the probability that a vector in the subset $\boldsymbol{X}_{tb}$, associated to the node t_b, belongs to class the m-th class. Similar to the entropy measure, the GDI of a node leading to just one class is zero ($P(\omega_\alpha|t_b) = 1$ if ω_α is the one class), while is maximum (and positive) if all the classes are equally possible after the node. So, our decision trees are built choosing at each node the split that maximises the entropy (or impurity, as it is sometimes called) gain ΔGDI.

As for the complexity of the decision trees it was decided to use a stopping criterion based on the maximum number of decision splits (or branch nodes). In order to test the effect on the classification performance in trees of diverse complexity, it was chosen to use two classifiers (Table 5.3), one of low complexity, stopping the growth at 4 splits, and the other of medium complexity, with a maximum of 20 nodes.

Table 5.3 Characteristics of the classification trees implemented

Classifier	Splitting criterion	Stopping criterion
Simple binary classification tree	max ΔGDI	max 4 splits
Medium binary classification tree	max ΔGDI	max 20 splits

Following the same criteria of the previous methods, also in this case two separate classification trees were determined, one for the aerobic part and one for the anoxic one. The main differences of this approach, however, consists in a more explicit tackling of the fault detection in terms of classification problem and in the absence of updating of the classifiers. In fact, the decision trees are trained only once at the beginning using the historic data and are subsequently employed to detect the possible faulty state of the next phases as they are, without using the outcome of the past detections to update their structure.

5.2.4 The support vector machine classifiers

The potentials of the SVM method discussed in section 4.4 led us to test this approach in the fault detection task. As for the classification trees, the implementation of this method followed a detection approach oriented towards the classification task, aiming at finding the best classifier/predictor in a non-adaptive framework. Therefore, the SVM based classifiers for the anoxic and the aerobic phase (the same idea of separating the detection algorithms for the two phases is applied here) are first trained on a set of representative data and then used for the following phases without changes.

Further, it was decided to test the efficiency of the SVM as a detection tool either in its classical form of linear classifier, i.e. assuming the faulty and normal state of the

process to be linearly separable, either removing this assumption and assessing the discrimination performance of the method when a cubic polynomial transformation is applied to the data. In both cases, to decrease the possibility of bias in the detection using features of different scales, the training sets used are standardized, i.e. each variable is centred and scaled as in eq.(4.7) according to their mean and variance respectively. The training of the SVM classifiers is achieved in the Matlab® environment using the `'fitcsvm'` function of the Statistics and Machine Learning Toolbox™.

Table 5.4 Characteristics of the SVM classifiers implemented

Classifier	Kernel function	Kernel degree	Standardized data
Linear SVM	Linear	-	yes
Cubic ker. SVM	Polynomial	3	yes

5.2.5 The confusion matrix

In order to allow an exhaustive and yet synthetic assessment of the results obtained by the different method used for the finer fault detection and facilitate the comparison among them, their performance will be summarized in the form of a confusion matrix (Figure 5.5)

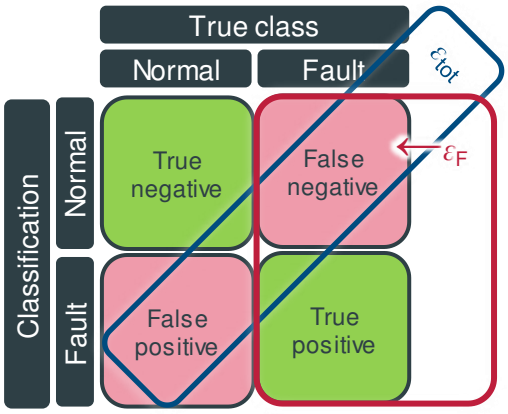


Figure 5.5 Example of confusion matrix. The elements on the secondary diagonal represent the misclassified instances and their percentage over the whole dataset indicates the global detection error ε_{tot}. The percentage of misclassified fault ε_F indicates instead the fault detection error

On the confusion matrix diagram, the rows correspond to the output of the classifier, therefore they represent the detected/predicted instances, while the columns show the true class of the observations. The principal diagonal counts how many (and in what percentage) of the tested observations are correctly classified, while the off-diagonal cells show the instances which were incorrectly classified. The cell in the extreme south west position will report both the classification success and classification fail percentages. Moreover, the south-east cell contains the accuracy for each classified category and for each observed class respectively The column on the far right of the

matrix shows the accuracy for each predicted class, while the row at the bottom of the plot shows the accuracy for each true class. The terms 'negative' and 'positive' are here used putting the stress on the detection purpose of this research, therefore a 'true positive' detection means that the algorithm succeeded in the identification of a faulty event while a 'false positive' indicates when a normal state is wrongly classified as an anomaly. For the same reason, in the confusion matrix representation of the results it will be used '0' to denote the normal class and '1' to denote the fault category.

5.3 Fault detection using the plant operational data

In this paragraph are presented the fault detection performances of the RT-FD algorithms on the operational data of the Mantua plant following the scheme of Figure 5.2. the detection methods used for this dataset are RT-MWPCA and the Bayesian classifier, implemented as described in the previous paragraph. For each phase the signals are parametrized as in the table below (same as Table 3.13).

Table 5.5 Parameters extracted from the operational plant data of Mantua in each process phase

Description	Symbol
Average ammonia concentration in the biological reactor	$m_{\cdot,\mathrm{NH4+}}$
Growth or decay rate of the ammonia concentration in the biological reactor (slope $NH_4^+{}_{out}$)	$s_{\cdot,\mathrm{NH4+}}$
Average nitrates concentration in the biological reactor	$m_{\cdot,\mathrm{NO}x-}$
Growth or decay rate of the nitrate concentration in the biological reactor (slope $NO_x^-{}_{in}$)	$s_{\cdot,\mathrm{NO}x-}$

The main characteristics of the dataset used are summarized in Table 5.6.

Table 5.6 Characteristics of the operational plant measurements dataset

Characteristics	Value
Length	~9 months
Number of phases	4404 (2202 aerobic + 2202 anoxic)
Number of phases with gross faults	391 (249 aerobic + 142 anoxic)
Number of phases with finer faults	70 (45 aerobic + 25 anoxic)

5.3.1 Preliminary screening

As already mentioned, regardless the method used to detect the finer faults, a preliminary screening to detect the presence of gross faults is performed as soon as the measurements are available. Table 5.7 shows the detail of the faults detected by the preliminary screening: in the whole tested period a total of 391 phases interested by gross faults were detected, of which 249 are aerobic and 142 anoxic (Table 5.6). In these instances, at least one fault was detected by the preliminary screening described in

paragraph 3.4.2, though some phases tested positive for more than one kind of fault. Most of them consisted of excessive length, typically associated with a nitrate overload in the aerobic phase. In addition, a smaller number of long phases may occur during the anoxic phases, caused by an ammonium build-up. The equal number of no data detected for both the ammonia and the nitrate signal are due to electrical failures involving both probes.

Table 5.7 Detailed partition of preliminary screening faults

Aerobic phase	No data	NH_4^+	6
		NO_3^-	6
	Constant signal	NH_4^+	21
		NO_3^-	39
	Anomalous duration		162
	Spikes	NH_4^+	28
		NO_3^-	18
	Inconsistent slope	NH_4^+	58
		NO_3^-	31
Anoxic phase	No data	NH_4^+	6
		NO_3^-	6
	Constant signal	NH_4^+	9
		NO_3^-	35
	Anomalous duration		32
	Spikes	NH_4^+	26
		NO_3^-	25
	Inconsistent slope	NH_4^+	8
		NO_3^-	53

From the plant management viewpoint, it is important that the fault detection is as timely as possible. During the preliminary screening the detection may occur earlier, depending on the nature of the fault, since is not necessary to wait for the end of the phase to begin the detection task. Figure 5.6 and Figure 5.7 show the detection delay produced by the preliminary screening for both phases, which often yield a response before the end of the phase. In the worst case the alarm is raised after 3 hours since the beginning of the phase, having set that as the maximum allowable phase length.

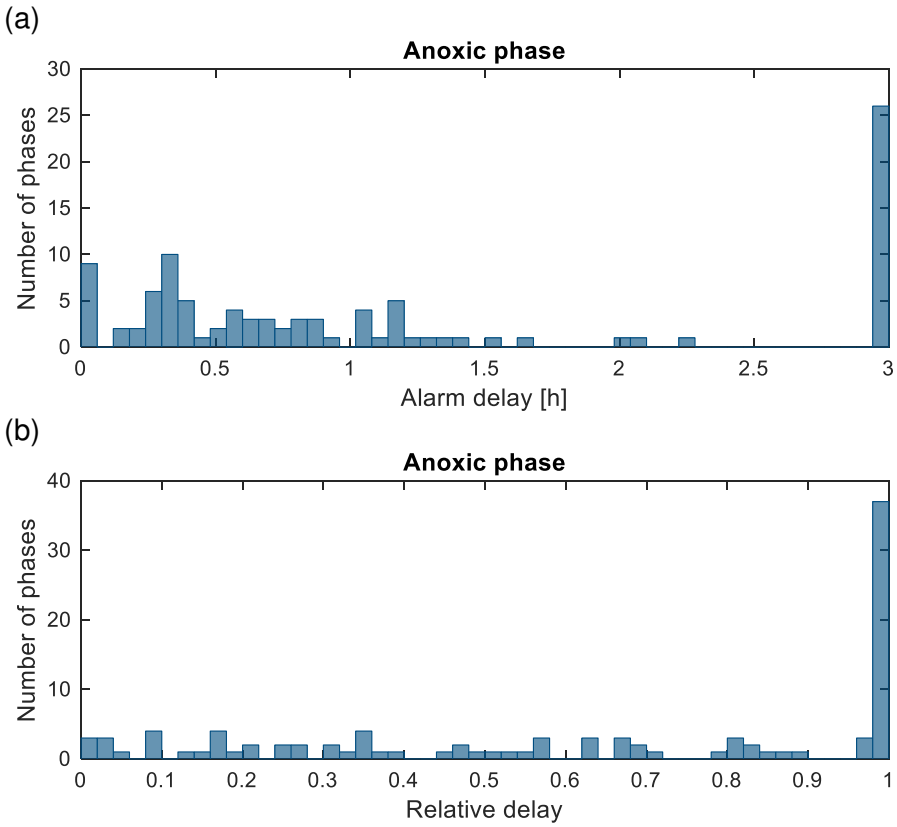


Figure 5.6 Percentage of alarm delay during the preliminary screening with respect to the phase duration

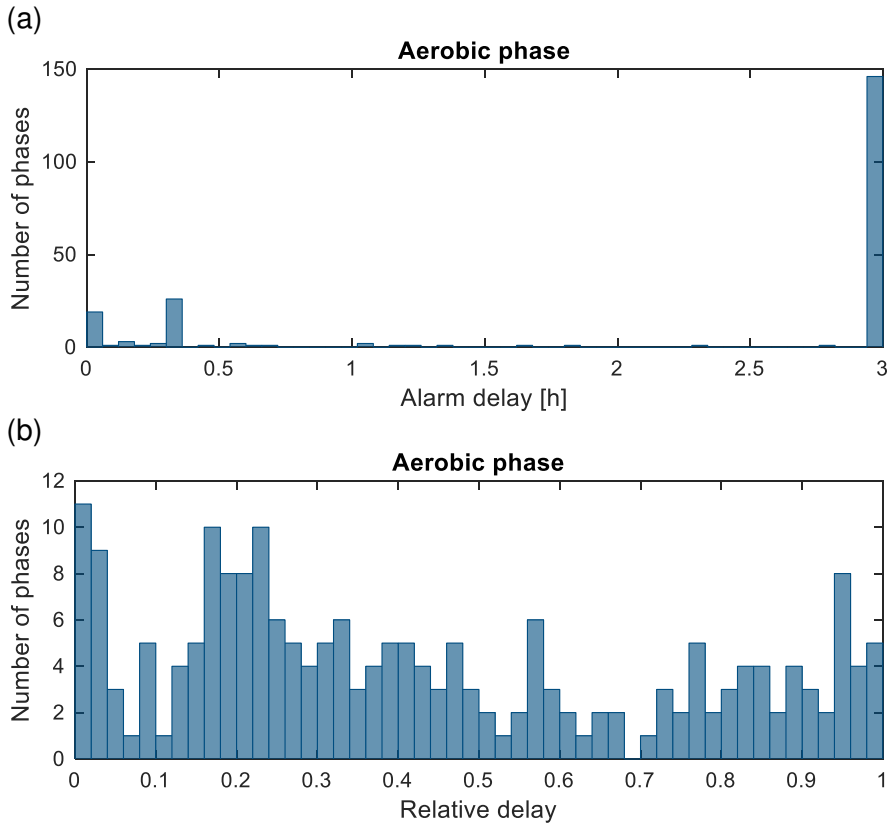


Figure 5.7 Percentage of alarm delay during the preliminary screening with respect to the phase duration

5.3.2 Test of the real-time mowing window principal component analysis method

The first step of the RT-MWPCA method as outlined in Figure 5.4 consists in the definition of the initial reference PCA models for the detection of the more subtle faults, which are here built using the first 100 samples for each phase starting from the 18th September 2012 and computing the reduced models retaining the most relevant components. For the plant data two components out of four were retained for each phase (Figure 5.8).

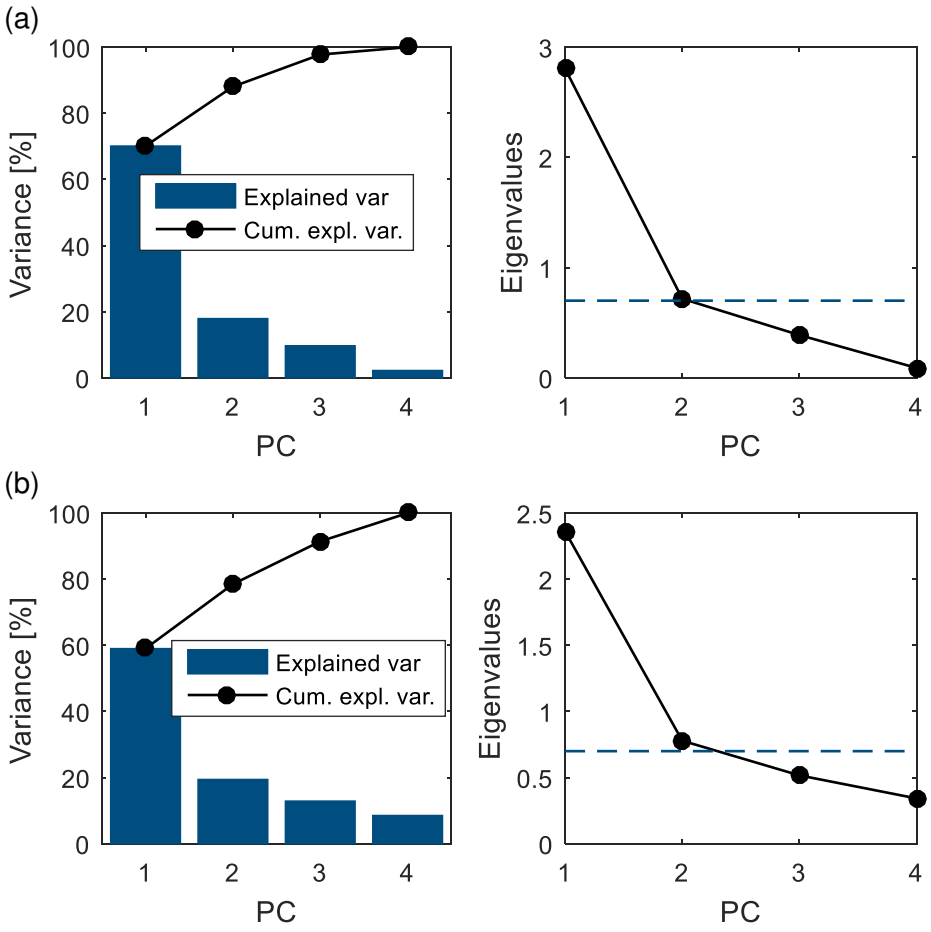


Figure 5.8 Scree plot and explained variance of the principal components of the initial reference dataset for the aerobic (a) and anoxic (b) phases. The reduced model retains in both cases the first two PCs, whose eigenvalues are greater than the given threshold of 0.7, as suggested by (Dunteman, 1989)

Once the reduced models are determined both the RT-MWPCA algorithm for the aerobic phase and for the anoxic phase are trained to maximize the detection performance over the remaining phases by computing the optimal threshold that minimize eq. (5.2). Of course, in this optimization only the phases that did not present any gross fault are considered.

Table 5.8 shows the optimal values of the threshold coefficients for the two phases.

Table 5.8 Optimized thresholds for the two phases.

		Aerobic phase	Anoxic phase
T^2	T^2_{lim}	6.241	6.241
	k_T	6.963	5.280
	T_{th}	43.456	32.952
Q	Q_{lim}	0.332	0.510
	k_Q	1.451	6.760
	Q_{th}	0.481	3.448

The calibrated thresholds are quite different: in general, it is possible to observe that the optimal threshold level falls between five and seven times the theoretical control limits, however the optimal level for the Q statistic in the aerobic phase is much lower, almost one and a half times the Q_{lim}.

5.3.2.1 Fault detection performances

In this section are presented the detection results of the RT-MWPCA algorithm at first separately for the aerobic and the anoxic phase and subsequently the combined detection performance for a more comprehensive assessment. Table 5.9 summarizes the characteristics of the dataset after the initialization of the method.

Table 5.9 Characteristics of the operational plant measurements dataset used to test the RT-MWPCA method

Characteristics	Value
Length	>9 months
Number of phases (after initialization)	4187 (2094 aerobic + 2093 anoxic)
Number of phases with gross faults (after initialization)	376 (243 aerobic + 133 anoxic)
Number of phases with finer faults	70 (45 aerobic + 25 anoxic)

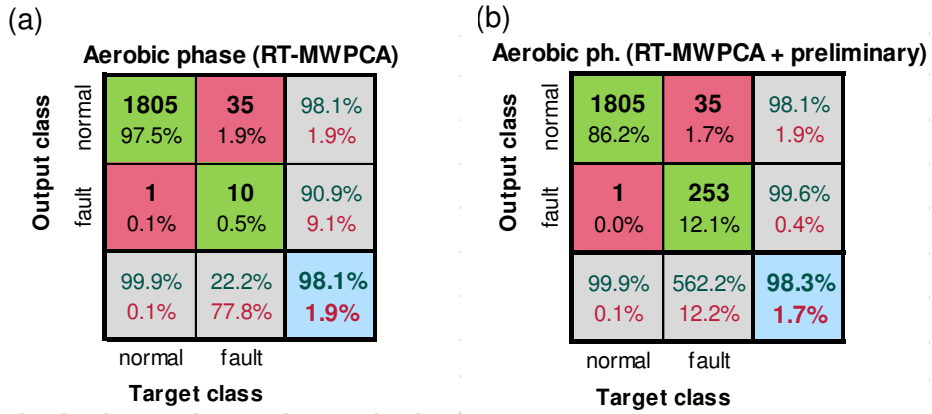


Figure 5.9 Confusion matrices describing the performances of the RT-MWPCA method in the aerobic phases either limited to the detection of the finer faults (a) and in its complete form including the preliminary screening (b)

The confusion matrices in Figure 5.9 show that globally the classification performances of the method are quite high, having a success above 98%. This result is however deeply influenced by the higher number of normal phases with respect to the anomalous phases. The algorithm is in fact tested on the 2094 aerobic phases that remain after the initialization phase of the method and comprise 1806 phases classified as

normal and 288 faults, of which 243 are 'gross faults' and 45 belong to the 'finer' category (Table 5.9). Considering only this last class of faults the algorithm shows poor detection performances, being able to identify only the 22% of them. However, counting also the contribution given by the preliminary screening to the detection task (Figure 5.9b) the implemented method proves to be a valuable tool for the FD.

A detailed assessment of the adaptive detection performances of RT-MWPCA method in time can be found in Figure 5.10 and. Here the solid lines represent the classification error over the number of phases investigated and the dotted line is the FD error, i.e. the classification error limited to the number of faults. We notice generally the algorithm has a low global classification error but as the time progressed it shows a positive trend. On the other hand, the identification error decreased in time, although it never steps below the 50%.

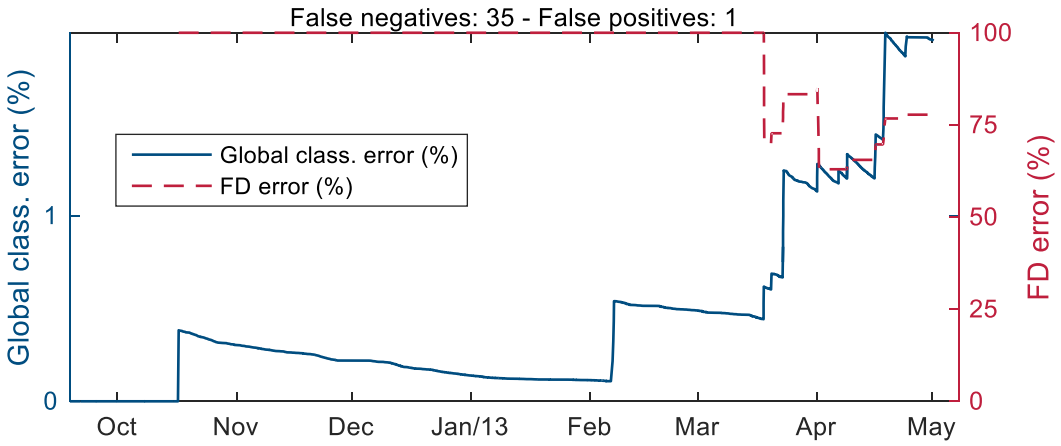


Figure 5.10 Detection performance of the RT-MWPCA algorithm, aerobic part. The solid line (left axis) represents the global classification error, while the dashed line (right axis) is the fault identification error

Similar considerations can be made for the RT-MWPCA algorithm designed to the anoxic phases. Out of the 2093 phases remaining after the initialization of the method and the early preliminary checks 1935 phases result 'normal', 133 present a gross fault and 25 finer faults (Table 5.9).

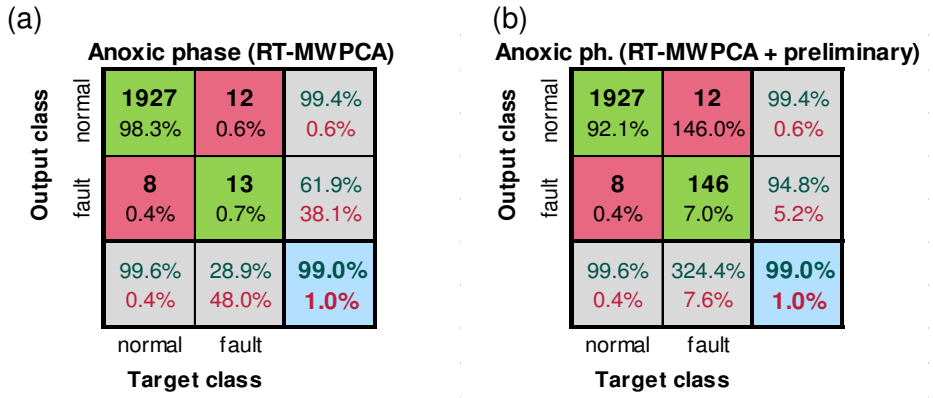


Figure 5.11 Confusion matrices describing the performances of the RT-MWPCA method in the anoxic phases either limited to the detection of the finer faults (a) and in its complete form including the preliminary screening (b)

In this case also, the global classification performance of the method shows good results (Figure 5.11), even if the excellent score is affected by the same disproportion between normal and faulty phases.

Despite that, even considering only the fault identification performance of the finer faults, the method has a better behaviour in the anoxic phase, decreasing the detection error to 48% by the end of the tested period (Figure 5.12).

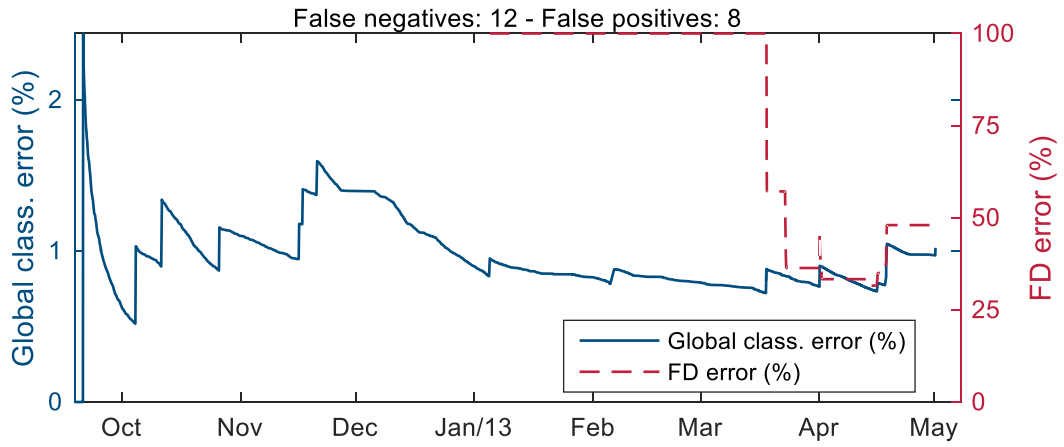


Figure 5.12 Detection performance of the RT-MWPCA algorithm, anoxic part. The solid line (left axis) represents the global classification error, while the dashed line (right axis) is the fault identification error

Though separate algorithms were developed for each phase, the combined results are summarized in Figure 5.13 in order to draw the necessary remarks on the whole algorithm. In terms of global performance, the algorithm shows quite satisfying efficiency levels however the results in terms of detection of the quantities of interest are more disappointing. The majority of the observed malfunctions are spotted by the preliminary screening on the signal, while the subsequent MWPCA could globally identify less than half of the 70 faults and process anomalies observed.

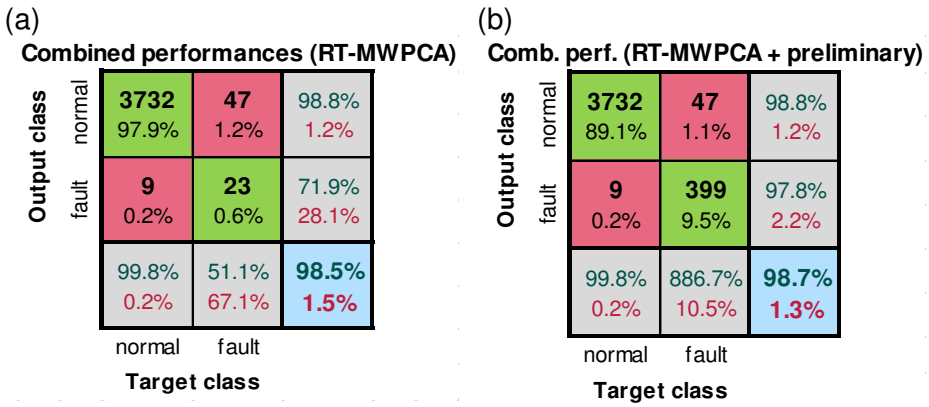


Figure 5.13 Confusion matrices describing the performances of the RT-MWPCA method in the anoxic phases either limited to the detection of the finer faults (a) and in its complete form including the preliminary screening (b)

The comparatively underachievement of the RT-MWPCA method can be explained by considering the calibration formula used for the training, which is meant to minimize the sum of classification error and wrong fault detection. This favours the creation of a

detection method with a conservative approach, especially in case the number of elements belonging to one class is one order bigger than the other. Thanks to the cautious behaviour of the algorithms, though, the number of false positive detections is extremely low and the great majority of the detected faults are actual process anomalies.

5.3.3 Testing of the Bayesian approach

Given the not so exciting performances of the RT-MWPCA algorithm from the point of view of the fault identification, to was decided to test the Bayesian classifier previously described on the same set of data of the Mantua plant. As anticipated the algorithm is characterized by an initialization phase in which at least one sample from each 'class' must be collected in order to compute the initial prior and likelihood distributions. This could reveal a major shortcoming in a dataset where the examples of one class are much less than the other and they are not uniformly distributed along the direction of the training. In fact, in the dataset of the operational plant data used the initialization phase of the Bayesian predictor for the aerobic phase can end at phase 392 (26th October 2012) when the first observed finer fault is encountered and can be used to compute the distributions of interests. For the anoxic phase, however, these conditions are met only at phase 1007 (21st January 2013). This is particularly evident from the plots in Figure 5.14 and Figure 5.15 where the prediction performances of the two algorithms are reported.

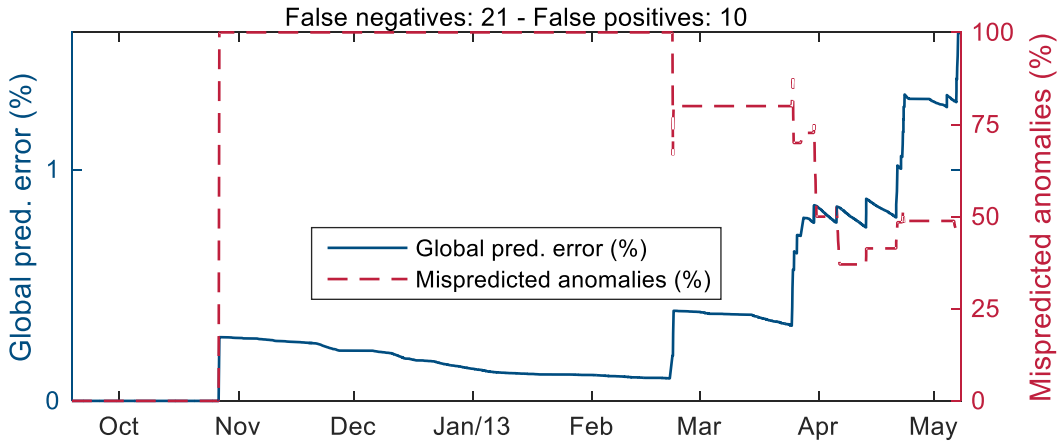


Figure 5.14 Prediction performance of the Bayesian algorithm, aerobic part. The solid line (left axis) represents the global prediction error, while the dashed line (right axis) is the anomaly prediction error

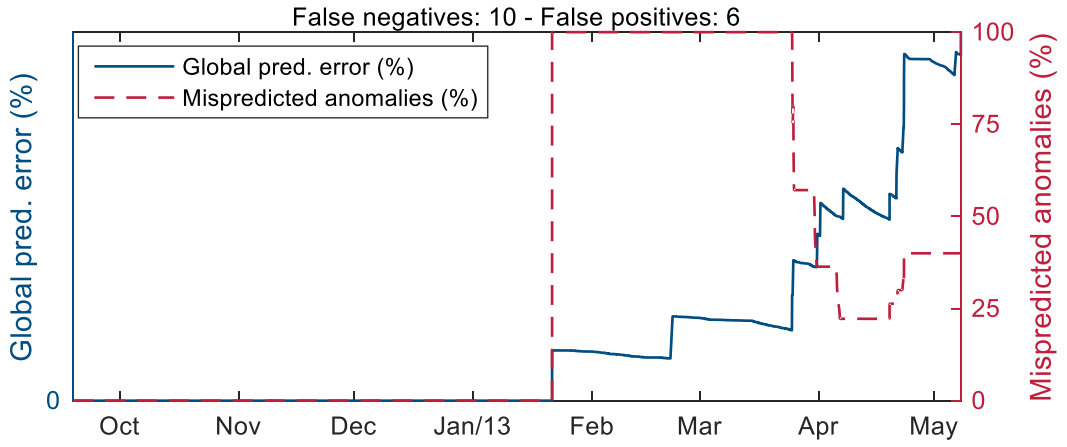


Figure 5.15 Prediction performance of the Bayesian algorithm, anoxic part. The solid line (left axis) represents the global prediction error, while the dashed line (right axis) is the anomaly prediction error

Despite the inconvenience of requiring a potentially long time for the initialization phase, this Bayesian approach proves to have a better detection performance than the RT-MWPCA method discussed in the previous paragraph.

The preliminary screening on the signals identifies 249 faults in the aerobic phases and 142 in the anoxic ones, therefore of the 2202 observations composing each part of the available dataset the Bayesian method is used to investigate 1953 aerobic and 2060 anoxic phases (Table 5.10).

Table 5.10 Characteristics of the operational plant measurements dataset used to test the Bayesian method

Characteristics	Value
Length	~9 months
Number of phases (after initialization)	4404 (2202 aerobic + 2202 anoxic)
Number of phases with gross faults (after initialization)	391 (249 aerobic + 142 anoxic)
Number of phases with finer faults	70 (45 aerobic + 25 anoxic)

However, since the detection of the gross faults is not the objective of this method the results reported in Figure 5.16 only concerns the identification of the finer faults.

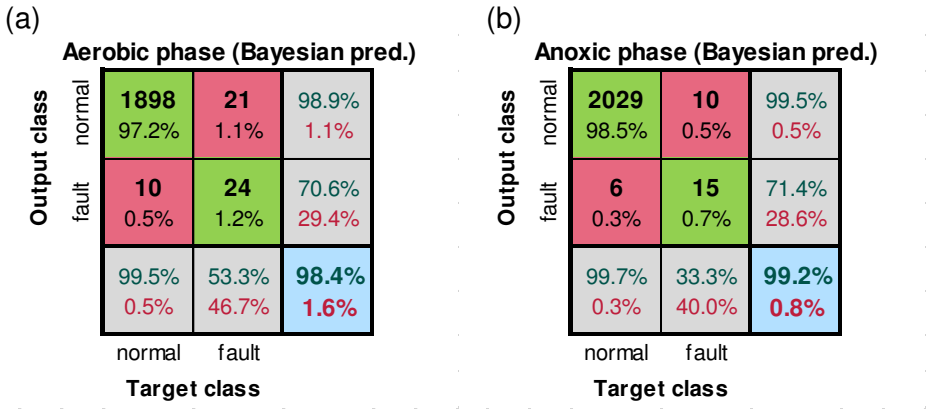


Figure 5.16 Confusion matrices describing the performances of the Bayesian method in both the aerobic (a) and the anoxic phases (b) limited to the detection of the finer faults

In both cases the algorithm shows a lower classification error than the RT-MWPCA, especially in terms of fault identification error. The global prediction error is quite similar to the other method, ~98% for the aerobic phase and ~99% for the anoxic, and this is certainly due to the higher number of 'normal' phases. In this case, however, the method is able, by the end of the tested period, to correctly identify more than 50% of the faults observed and at the same time the number of false alarms is kept very low, less than 30% of the total number of faults detected. This is more evident when the combined detection performance is considered, as in Figure 5.17.

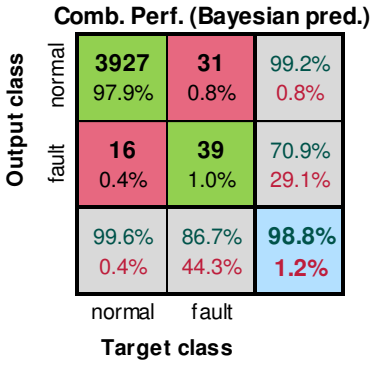


Figure 5.17 Confusion matrices describing the combined performances of the Bayesian methods

5.3.4 Comparison between the real-time mowing window principal component analysis and the Bayesian algorithm.

Considering the particular plant configuration, the proposed algorithms have been divided in two parts, one devoted to the detection of the aerobic phases and one for the anoxic ones. The two parts share the same principles: first some low-level controls are performed on the raw signals, in order to discriminate malfunctions like signals data interruptions, anomalous constant measurements and irregular duration of the phases, then for each investigated phase the portions of signal of both ammonia and nitrate are parametrized and four parameters are extracted: the two concentration growth or decay rates and the two averages of the measured concentrations. Other anomalies cannot be detected by simple controls on the data, therefore a more refined method is required. For this task two different approaches have been compared, one based on the PCA analysis and the other based on the Bayes' theory. In the PCA-based method developed the parameters are projected onto a reference space that maximize the variance among them and two thresholds based on the Hotelling's T^2 and Q statistics are used to discriminate the phases with an irregular behaviour: if the tested parameters produce scores that are higher than the respective thresholds in both statistics the presence of a process anomaly in the phase is reported, otherwise it is considered "normal" and its

parameters are used to update the reference space. In the Bayesian approach first two groups of parameters corresponding to the two possible states of the process are isolated and used to build the initial prior knowledge, then at each new step this prior knowledge is used to predict the most probable state of the process given the new parameters available. After comparing the prediction to the actual state of the process the results are used to update the prior knowledge.

The algorithm was tested with a nine-month time series from the municipal wastewater plant of Mantua (Italy) and from the comparison of the two methods emerged that the Bayesian algorithm generally performs better than the PCA-related one, showing a higher number of identifications of faults.

A direct comparison between the two approaches, simply putting together the previously shown results is not possible, mainly because their respective detection/prediction task starts at different times. Due to the differences in the initialization procedures the number of phases investigated by the two approaches is very different, nevertheless the objective of both methods is the same: to correctly identify the 70 observed fault events (45 among the aerobic phases and 25 in the anoxic part). Therefore Table 5.11 summarizes the performances of the two methods focusing on this one aspect of the detection task.

Table 5.11 Global algorithm performance results for both approaches. We choose to keep the column headers of the RT-MWPCA algorithm for a more direct comparison although the Bayesian algorithm actually provides predictions instead of detections

Observed faults	Detected faults	Identified faults	False negatives	False positives
		RT-MWPCA		
70	32	23	47	9
		Bayesian predictor		
70	55	39	76	11

The PCA-based algorithm globally performs worse than the Bayesian one, returning a smaller number of detected anomalies, though most of them are actual malfunctions and the number of false alarms is lower. The Bayesian method, on the contrary, spots a higher number of anomalies, and correctly predicts nearly the 60% of the faults, which however is still too low. It is clear then that none of the methods tested has good results but one of the main reasons lies on the dataset used. These methods are mainly employed in the detection of probe progressive failures and process anomalies, which typically involve multiple aerobic and anoxic phases. The algorithm, however, is designed to investigate the presence of finer faults one phase at the time and this leads in most of the cases to a fragmented (discontinuous) detection in which, for example, a fault event interesting two consecutive aerobic or anoxic phases, one is correctly identified and the

other result negative to the statistical test. In these cases, of course, all the phases composing the fault event shall be counted in the manual observations, but the algorithm investigates one phase at the time, and this can very likely produce that of two consecutive aerobic or anoxic phases, one can be correctly identified and the other result negative to the statistical test. For this reason, perhaps a "low definition" method to assess the results can appear preferable, although this would imply waiting several hours or even days to have a report on the state of the process and that conflicts with the primary aim of the FD technique, which requires a timely detection of an anomaly from its very onset. Another aspect to be taken into account regards the dataset used, which includes a small number of fault examples, too little to properly train the methods and leading the adaptation of the models towards an excessively small rate of detected malfunctions. The extremely high variability of the data compared to the few number of anomalies observed, makes the calibrated RT-MWPCA model exceedingly conservative. The same scarce availability of observed anomalies, especially in the first half of the dataset, prevents the Bayesian approach to fully identify all the anomalies by the end of the available dataset.

In the attempt to overcome these shortcomings and extend the assessment of the detection methods proposed, a dataset of synthetic measurements drawn from a ASM3-like model has also been used. The results are presented in the following paragraph.

5.4 Fault detection on synthetic measurements

In the previous section we discussed the application of two fault detection methods, one based on the PCA and the other on the Bayes' theorem, to a set of real-life data obtained from a small municipal WWTP. The data spanned approximately a none month-long period and were characterized by a limited number of observed variables (only the nitrate and ammonia concentrations in the aerobic tank) and a poorly commented set of reported anomalies. This produced unsatisfactory results, especially for the detection methods designed to identify the less evident faults, suggesting the use of a different and more comprehensive dataset.

Driven by these consideration, a numerical model has been then developed, as described in Chapter 3. The model is based on the BSM protocol but the typical plant scheme has been adapted to the alternated cycles process configuration and a revisited and more detailed characterization of the nitrogen kinetics is included. Moreover, particular attention has been paid in modelling the sensors involved in the aerobic compartment, especially designing their deviance from the normal operating conditions (see paragraph 3.3.2). The use of a numerical model to generate the synthetic data and

simulate the malfunctions brought the possibility to create datasets of measurements where the state of the sensor that produced them is exactly known and therefore the separation of the training set in 'normal' and 'faulty' class is more coherent. Also, the number of observed variables is higher, resulting in an increased number of diagnostic parameters available (Table 3.15).

In this new setting, it made sense to expand the range of methods used to identify the sensor failures by implementing four more classifiers, two decision trees of different complexity level and two SVM classifier (one linear, the other nonlinear). The characteristics of the new methods have been already presented in the previous paragraph. In the following part of the Chapter the main characteristics of the dataset used to train and validate the models will be discussed before presenting the results of the detection performances achieved by the different detection methods implemented.

5.4.1 Characteristics of the synthetic dataset

Before starting to analyse detection efficiency of the FD methods implemented in the synthetic dataset it is worth taking some time to discuss its structure. As mentioned in Chapter 3 the numerical model uses as input data the historical series of the BSM_LT, covering more than 1 year and a half. These results refer to a realization of the model where in the 608 simulated days the process controller alternates 1544 aerobic and anoxic phases. Despite they had not been explicitly modelled, a number of phases presented some artefacts on the signal compatible with the definition of gross faults given for the other dataset: typically, they are represented by excessively long phases (sometimes lasting several days) and steady signal (especially in correspondence with the simulated cleansing/recalibration of the sensors). The number of 'gross faults' reported was 45 for the aerobic phases and 32 for the anoxic, therefore the remaining 1499 aerobic phases and 1512 anoxic phases were used for the finer faults (Table 5.12). As for what concerns the number of finer faults they alter the measurements in 491 of the aerobic phases and 495 of the anoxic ones and are the effect of the disturbances affecting the nitrate and ammonia sensors, modelled as in Chapter 3.

Table 5.12 Characteristics of the complete synthetic measurements dataset

Characteristics	Value
Duration	~1.5 years (608 days)
Total number of phases	3088 (1544 aerobic + 1544 anoxic)
Number of phases with gross faults	77 (45 aerobic + 32 anoxic)
Number of phases with finer faults	986 (491 aerobic + 495 anoxic)

In Figure 5.18 and Figure 5.19 are reported the detail of the disturbances affecting the measurement instruments: the different components combine with the true value of the variable as in eq. (3.31).

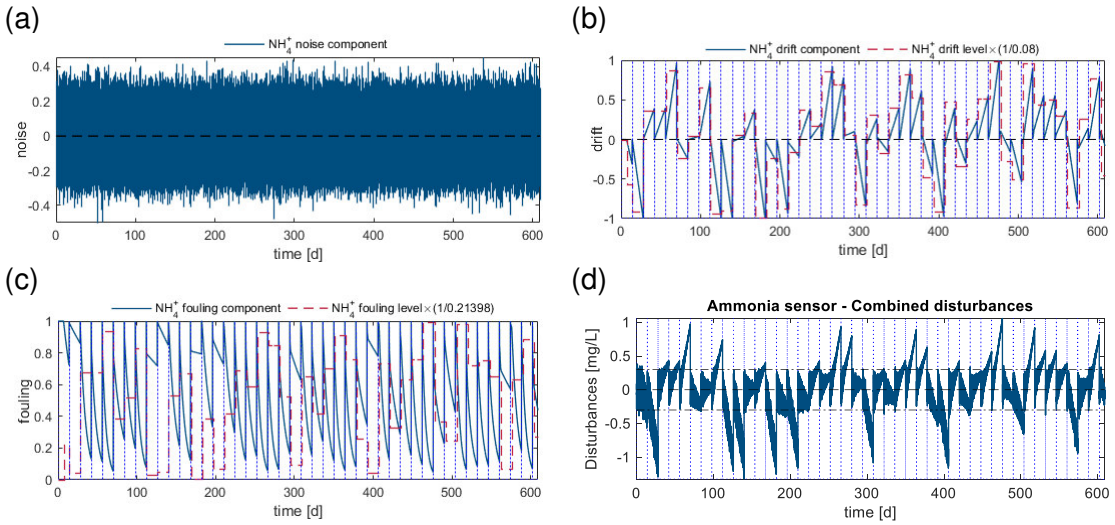


Figure 5.18 Detail of the modelled disturbances affecting the NH_4^+ sensor in the synthetic dataset. The measurement noise (a) together with the drift (b) and the fouling (c) combine in the global effect in (d) acting as an additive component on the signal. Beyond the band defined by the horizontal dashed lines in (d) the alteration of the measurement produced is considered relevant. The vertical dotted lines mark the periodical maintenance of the sensor

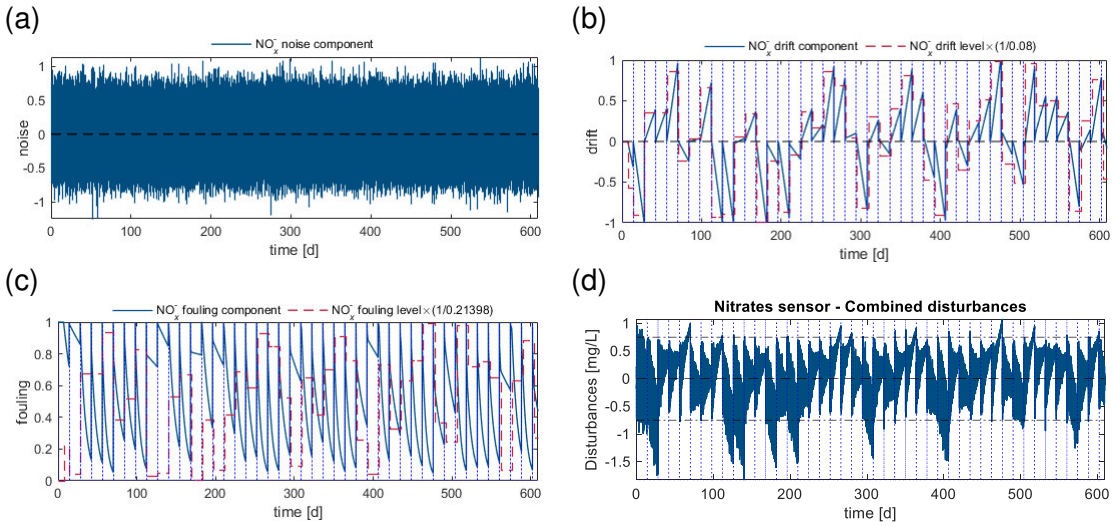


Figure 5.19 Detail of the modelled disturbances affecting the NO_x^- sensor in the synthetic dataset. The measurement noise (a) together with the drift (b) and the fouling (c) combine in the global effect in (d) acting as an additive component on the signal. Beyond the band defined by the horizontal dashed lines in (d) the alteration of the measurement produced is considered relevant. The vertical dotted lines mark the periodical maintenance of the sensor

The bottom-right element of both figures shows this combination of the three disturbances and represents the additive component that is going to be summed to the true concentration value and it is used to assess whether the distortion introduced in the measurements is acceptable or not. More precisely, it was chosen to consider as acceptable level of distortion only the white noise on the measurements, which is modelled as in eq.(3.14) using 0.5% as noise level and the values of Table 3.12 as measurement range for the sensors. Therefore, all the measurements violating the band

of ± 0.5% of the measurement range have been considered as affected by error. The faulty measurements affecting each sensor are reported in Figure 5.20.

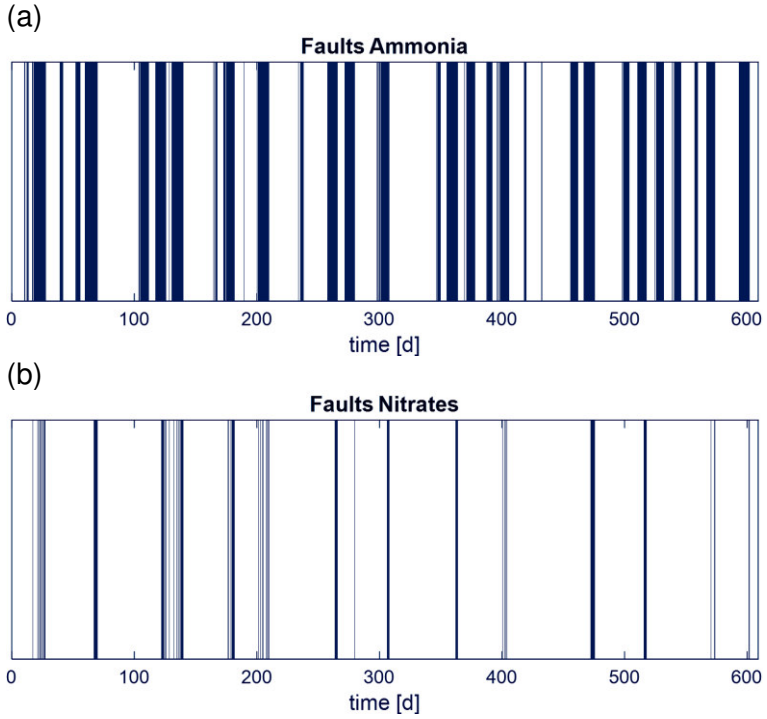


Figure 5.20 Time distribution over the whole simulated period of the faults affecting the ammonia (a) and the nitrates (b) sensors

However, investigating the reliability of each single measurements on a sample-to-sample basis is beyond the purpose of this research and, since a faulty instrument is likely to have a persistent effect on the measurements until it is fixed, we limit the reliability investigation to a single-phase level. For this reason, it was decided to consider all the phases where more than 50% of the nitrates and ammonia measurements as faulty, belonging to the 'finer fault' category.

Nevertheless, the FD algorithm designed are provided, as we mentioned above, with a preliminary check on the consistency of the measurements so that the most evident outliers and signal disturbances can be discriminated before applying the more refined methods. Both Figure 5.20 and Figure 5.21 show where the 'finer faults' are located along the simulation window. It can be noticed that they are more equally distributed with respect to the faults observed in the Mantua plant.

The availability of an extended set of measurements and the more even distribution of the anomalies in time allowed to set aside a portion of the dataset for the training, if required, and another portion for their validation. We recall that the RT-MWPCA method is composed of a first initialization phase, where a week-long set of fault-free measurements is used to build the starting PCA reference model, and a subsequent calibration phase, where a longer dataset, containing both normal and faulty phases, is used to determine the optimal value of the statistical thresholds. In order to compare the results of the different methods on datasets as similar as possible it was decided to train

and validate the binary classification trees and the SVM-based models respectively on the same calibration and validation sets used for the RT-MWPCA method, discarding the data used for the initialization of the algorithm, since those are the only ones containing the fault events. Thus, since in the synthetic dataset the initialization part of the RT-MWPCA is achieved using the first 26 observations of the dataset for each phase (~1 week), they will not be part of the set of data used to test the methods.

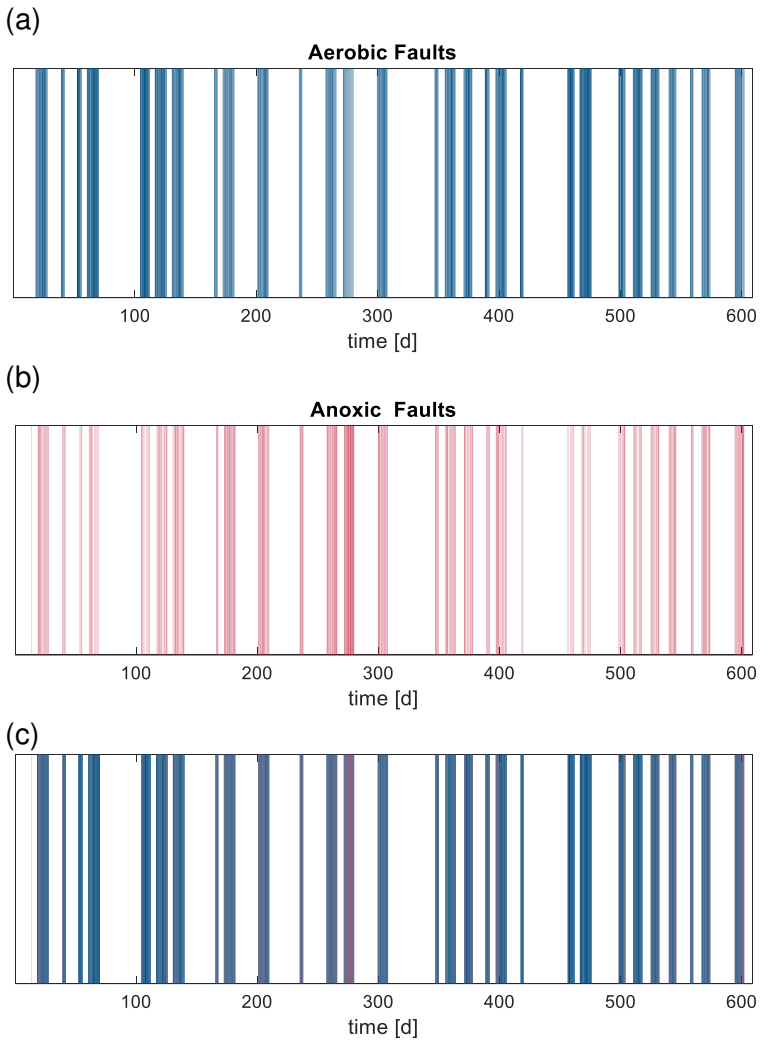


Figure 5.21 Time distribution over the whole simulated period of the faults affecting the ammonia and the nitrates sensors in the aerobic (a) and anoxic (b) phases and their combination (c). The blue shaded areas represent the aerobic phases and the red shaded areas the anoxic phases

The algorithms have been tested on two different partitioning of the dataset generated by the numerical model. The one whose characteristics are synthetized in Table 5.13 consists of a 6 months long data period reserved for the optimization of the statistical control limits (for the RT-MWPCA) or the training of the classification trees and SVM models; then another 6 months long data period (timely subsequent to the calibration one) has been used for the validation task.

Table 5.13 Characteristics of the dataset partitioning 1 for the calibration and validation of the RT-MWPCA, binary classification trees and SVMs methods

	Calibration set	Validation set
Duration	~6 months	6 months
Number of phases (after MWPCA init.)	682 (337 aerobic + 345 anoxic)	1178 (592 aerobic + 587 anoxic)
Finer faults	283 (140 aerobic + 143 anoxic)	375 (190 aerobic + 185 anoxic)

The main findings of this research will be reported in the subsequent sections with reference to the partitioning of Table 5.14. In this calibration/validation setting the algorithms are first trained using the data of a period approximately long 1 year and then validated using the remaining part of the dataset (~9 months). This is the most appropriate partitioning of the dataset since the methods can be trained on a period that encompasses all the seasonal variations of one full year, but some interesting characteristics can also emerge from the methods trained on a shorter data set.

Table 5.14 Characteristics of the dataset partitioning 2 for the calibration and validation of the RT-MWPCA, binary classification trees and SVMs methods

	Calibration set	Validation set
Duration	~1 year	~9 months
Number of phases (after MWPCA init.)	1864 (931 aerobic + 933 anoxic)	1092 (540 aerobic + 552 anoxic)
Finer faults	610 (304 aerobic + 306 anoxic)	364 (182 aerobic + 182 anoxic)

Regarding the Bayesian method, however, learning and prediction are intertwined at each iteration and cannot be separated using the same criterion. The Bayesian approach was therefore tested on a period which is the sum of the training and validation dataset of Table 4.13 used for the other methods, and whose characteristics are shown in Table 5.15.

Table 5.15 Characteristics of the synthetic measurements dataset used to test the Bayesian method

Characteristics	Value
Duration	~1 year
Number of phases (without gross faults)	1861 (929 aerobic + 932 anoxic)
Number of phases with finer faults	986 (330 aerobic + 328 anoxic)

Furthermore, as anticipated in the last paragraph of Chapter 3, the use of a synthetic dataset implied an extended number of observed variables, which reflected on a higher number of available diagnostic parameters. It was then decided to test different combinations of the parameters extracted from the observed variables related to the nitrification-denitrification process (ammonia, nitrates, DO, temperature, suspended solids, see Table 3.15) to test the different detection methods implemented. In other words, the diagnostic parameters have been grouped into five combinations, labelled from A to E and reported in Table 5.16 to Table 5.20; these combinations should be

intended as an example of the different descriptors that can be available to represent the samples tested with the different methods.

Table 5.16 Combination A of the parameters extracted for each process phase from the data generated by the numerical model

	Parameter	Name
Combination A	$m_{\cdot,NH4+,out}$	'NH4_out average'
	$s_{\cdot,NH4+,out}$	'NH4_out slope'
	$m_{\cdot,NOx-,out}$	'NOx_out average'
	$s_{\cdot,NOx-,out}$	'NOx_out slope'

Combination A represents the most basic configuration of the diagnostic parameters used and corresponds to the one used for the real data analysis. Despite it generally leads to poorer results in the training and testing of the FD methods, this combination allows the comparison of the results of the methods used for the synthetic dataset (the classification threes and the SVM models) with those obtained in the real plant dataset.

Table 5.17 Combination B of the parameters extracted for each process phase from the data generated by the numerical model

	Parameter	Name
Combination B	$m_{\cdot,NH4+,out}$	'NH4_out average'
	$s_{\cdot,NH4+,out}$	'NH4_out slope'
	$m_{\cdot,NOx-,out}$	'NOx_out average'
	$s_{\cdot,NOx-,out}$	'NOx_out slope'
	$m_{\cdot,Temp}$	'Temp average'

Combination B is similar to Combination A, save for the addition of the temperature. Using this combination on the same datasets tested with Combination A it is possible to assess how much this additional information influences the performances of the algorithms.

Table 5.18 Combination C of the parameters extracted for each process phase from the data generated by the numerical model

	Parameter	Name
Combination C	$m_{\cdot,NH4+,in}$	'NH4_in average'
	$m_{\cdot,NH4+,out}$	'NH4_out average'
	$s_{\cdot,NH4+,out}$	'NH4_out slope'
	$m_{\cdot,NOx-,in}$	'NOx_in average'
	$m_{\cdot,NOx-,out}$	'NOx_out average'
	$s_{\cdot,NOx-,out}$	'NOx_out slope'
	$m_{\cdot,Temp}$	'Temp average'

Based on the previous combination, in this group of diagnostic parameters are also included the average concentrations of the nitrates and ammonia in the influent. Combination C is intended to represent the most essential set of ingredients from which the classifiers, especially the nonlinear ones, can derive some implicit mass balance of the process that can help the FD task.

Table 5.19 Combination D of the parameters extracted for each process phase from the data generated by the numerical model

	Parameter	Name
Combination D	$m_{\cdot,NH4+,in}$	'NH4_in average'
	$rg_{\cdot,NH4+,in}$	'NH4_in range'
	$m_{\cdot,NH4+,out}$	'NH4_out average'
	$rg_{\cdot,NH4+,out}$	'NH4_out range'
	$s_{\cdot,NH4+,out}$	'NH4_out slope'
	$m_{\cdot,NOx-,in}$	'NOx_in average'
	$rg_{\cdot,NOx-,in}$	'NOx_in range'
	$m_{\cdot,NOx-,out}$	'NOx_out average'
	$rg_{\cdot,NOx-,out}$	'NOx_out range'
	$s_{\cdot,NOx-,out}$	'NOx_out slope'
	$m_{\cdot,Temp}$	'Temp average'
	$m_{\cdot,DO,in}$	'DO_in average'
	$rg_{\cdot,DO,in}$	'DO_in range'
	$m_{\cdot,DO,out}$	'DO_out average'
	$rg_{\cdot,DO,out}$	'DO_out range'
	$m_{\cdot,TSS,in}$	'TSS_in average'
	$rg_{\cdot,TSS,in}$	'TSS_in range'
	$m_{\cdot,TSS,out}$	'TSS_out average'
	$rg_{\cdot,TSS,out}$	'TSS_out range'

Combination D is the combination that includes all the parameters extracted from the synthetic dataset (Table 3.15). This is the combination with the highest informative content and it will be seen that using this set of descriptors produces the most performing classifiers. It includes all the parameters of the previous combinations together with the average concentration of the dissolved oxygen and of the suspended solids, each both in the influent and inside the reactor. Combination D is also characterized by the presence of the parameters '$rg_{\cdot,\cdot}$' representing the difference between the maximum and minimum value measured in the time interval of the phase for each variable excluded the temperature. This easy-to-compute parameter is meant to help the assessment on the consistency of the measurements summarized in the other combination by the average values.

Table 5.20 Combination E of the parameters extracted for each process phase from the data generated by the numerical model

	Parameter	Name
Combination E	$m_{\cdot,NH4+,in}$	'NH4_in average'
	$m_{\cdot,NH4+,out}$	'NH4_out average'
	$s_{\cdot,NH4+,out}$	'NH4_out slope'
	$m_{\cdot,NOx-,in}$	'NOx_in average'
	$m_{\cdot,NOx-,out}$	'NOx_out average'
	$s_{\cdot,NOx-,out}$	'NOx_out slope'
	$m_{\cdot,Temp}$	'Temp average'
	$m_{\cdot,DO,in}$	'DO_in average'
	$m_{\cdot,DO,out}$	'DO_out average'
	$m_{\cdot,TSS,in}$	'TSS_in average'
	$m_{\cdot,TSS,out}$	'TSS_out average'

Combination D can, however, be a very expensive way to describe the process since it dramatically increases the dimensionality of the datasets used. A lighter configuration has then been tested, Combination E, not including the range parameters. This reduction seems to only marginally affect the performances achieved with the more complex configuration and therefore can be considered a valuable compromise between the efficiency of the detection and the use of resources.

In the following sub-sections the results will be presented in detail for the Combination A which corresponds to the combination used in the operational plant data, and Combination E, which is the one that provides the best performance. The results achieved with the other combinations will be however presented in an aggregate form for comparison.

5.4.2 Results of the fault detection using the Bayesian approach

The results for the Bayesian method are discussed separately, since in this case the training and prediction tasks alternate at each iteration along the whole year-long dataset of Table 5.15.

The results in Figure 5.22 show that the performances of the Bayesian method as a general predictor using the diagnostic features of Combination A are quite satisfactory, predicting the correct class of the instances almost 60% of the times. Particularly disappointing are both the detection rate of faults and the false alarm rate, especially for the anoxic phase. With respect to the total number of faults occurred, in fact, the method is able to identify only a small portion of them, between 20% and 30%, while three on five of the anomalies highlighted correspond to false alarms. This behaviour contradicts the good results observed when the method is employed to dataset of real measurements

(Figure 5.16) and seem to express a limit of the method itself in treating this kind of problem.

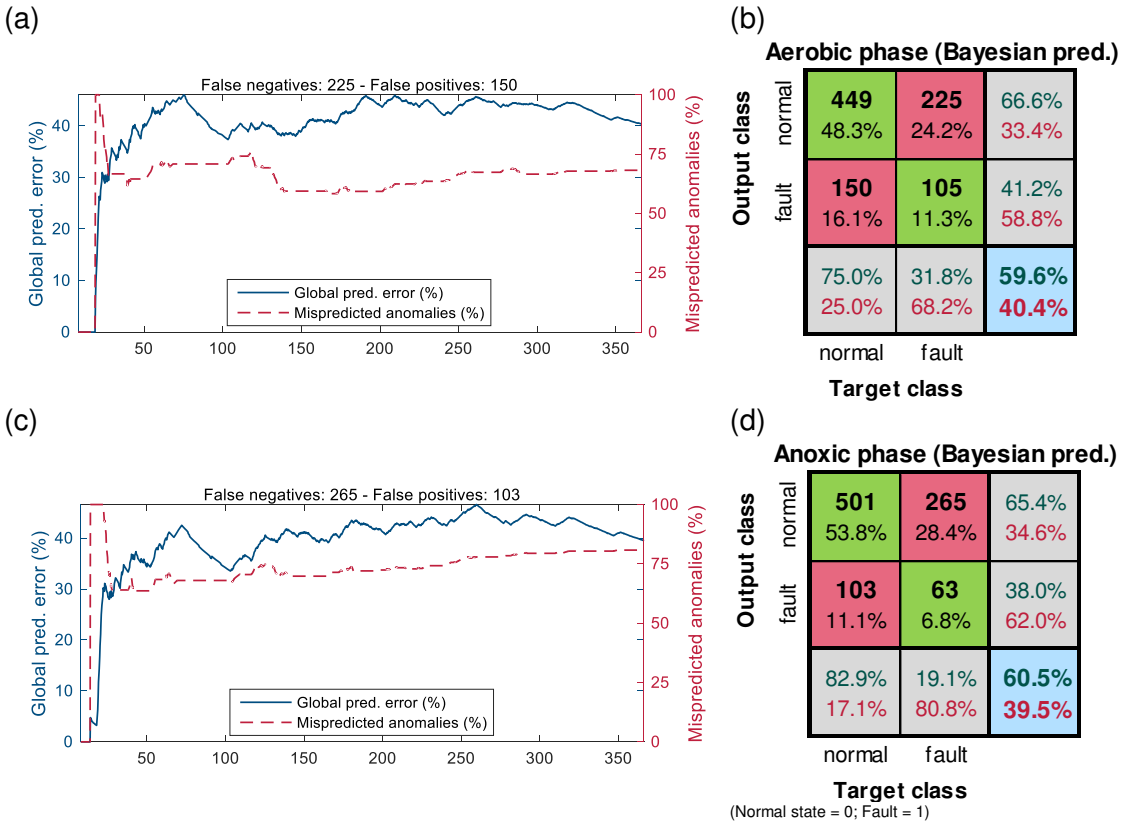


Figure 5.22 On the left column, the prediction performance of the Bayesian algorithm for the aerobic (a) and anoxic (c) part of the synthetic dataset using Combination A. The solid line (left axis) represents the percentage global prediction error, while the dashed line (right axis) is the percentage anomaly prediction error. On the right column, the confusion matrices describing the performances of the Bayesian method in both the aerobic (b) and the anoxic phases (d)

Quite similar, in fact, are the performances of the Bayesian method when Combination E is used (Figure 5.23). The global classification success is here a little lower while both the false alarm rate and the fault identification error have similar percentage (~60%).

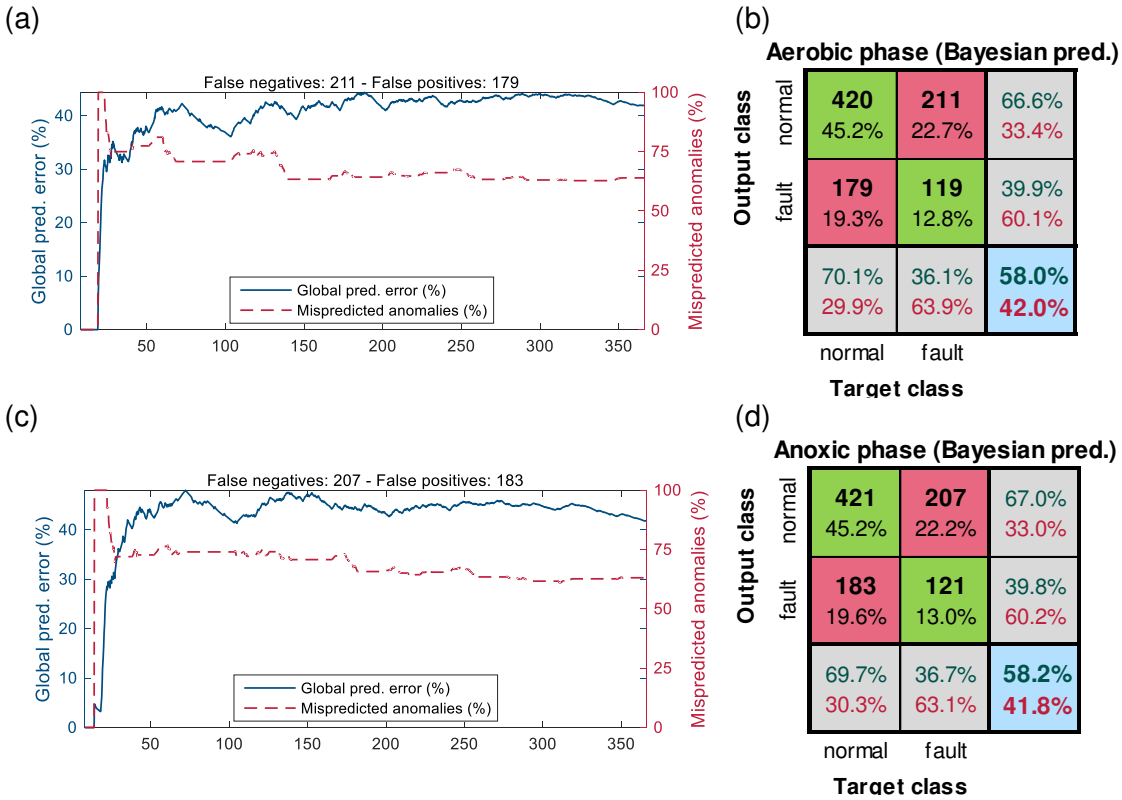
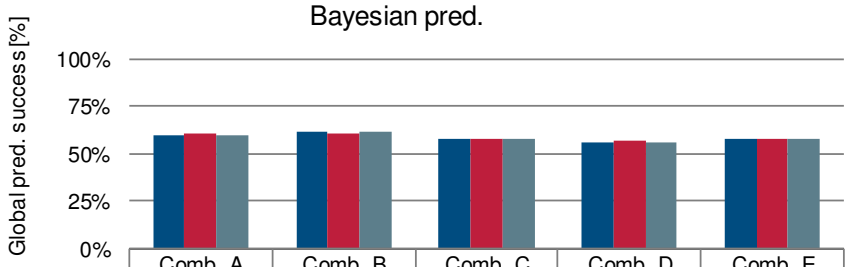


Figure 5.23 On the left column, the prediction performance of the Bayesian algorithm for the aerobic (a) and anoxic (c) part of the synthetic dataset using Combination E. The solid line (left axis) represents the percentage global prediction error, while the dashed line (right axis) is the percentage anomaly prediction error. On the right column, the confusion matrices describing the performances of the Bayesian method in both the aerobic (b) and the anoxic phases (d)

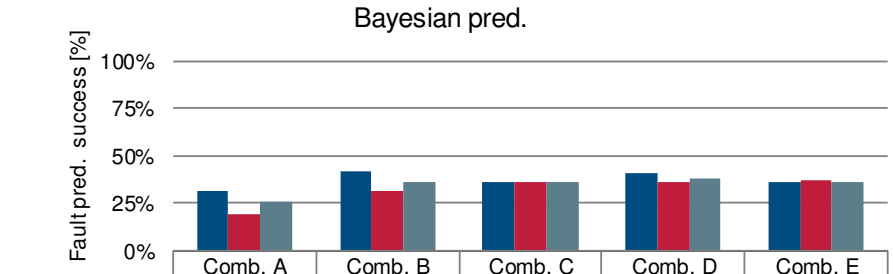
The performances of the method in the discrimination of the finer faults seem to be unaffected by the combination of parameters used. As can be seen in Figure 5.24, changing the combination of parameters the global prediction performances of the method do not change much, remaining around 60%.



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	59.6%	61.4%	58.0%	56.3%	58.0%
Anoxic phse	60.5%	61.2%	57.7%	56.8%	58.2%
Combined	60.1%	61.3%	57.8%	56.6%	58.1%

Figure 5.24 Global prediction success of the Bayesian method for the different combinations of parameters of the synthetic dataset

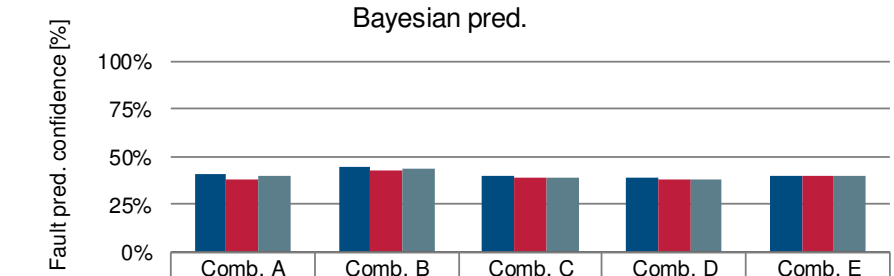
Some improvements on the prediction success of the fault events can instead be achieved using combinations with a higher number of diagnostic parameters but still the results are not satisfactory (Figure 5.25).



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	31.8%	41.8%	35.8%	40.6%	36.1%
Anoxic phse	19.2%	31.1%	36.0%	36.3%	36.9%
Combined	25.5%	36.4%	35.9%	38.4%	36.5%

Figure 5.25 Fault prediction success of the Bayesian method for the different combinations of parameters of the synthetic dataset

Also, the number of false alarms is high for all the combinations tested. Expressing its measure as its complementary, the degree of confidence in the faults predicted, i.e. the percentage of actual faults among all the predictions highlighted as fault, it can be observed that it hardly exceeds 40%, meaning that the prediction of the fault events cannot be trusted most of the times.



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	41.2%	45.2%	39.9%	39.0%	39.9%
Anoxic phse	38.0%	42.9%	39.1%	38.0%	39.8%
Combined	39.6%	44.0%	39.5%	38.5%	39.8%

Figure 5.26 Confidence rate on the fault predictions of the Bayesian method for the different combinations of parameters of the synthetic dataset

5.4.3 Results of the fault detection using the real-time mowing window principal component analysis method

The test of the RT-MWPCA described in 5.2.1 on the measurements dataset generated by the numerical model is conducted, after the initialization phase, using the partitioning of Table 5.14. With reference to the scheme of Figure 5.4, the algorithm is at first initialized for each process phase using a one-week-long set of fault-free observations in order to compute the initial reference models of the retained PCs. With the synthetic dataset, this is achieved using the parameters of the first 26 observations for each phase (i.e. in the one-week set of measurements were counted 26 aerobic phases and the same number of anoxic phases). Subsequently, the optimal level positions of the threshold levels, used to discriminate the normal phases from those containing the finer faults, are calibrated by minimizing eq.(5.2) using the data from the calibration test of

Table 5.14 and finally, the following 6 months of parameters are used to validate the method.

Table 5.21 shows the results of the optimization process of the statistical thresholds for the Combination A of the parameters, the same used for the operational plant data. Notice that in this case the number of components retained for the aerobic part is 3, one more than those used in the real measurements dataset. Moreover, the calibrated multiplying coefficient k_Q of the Q_{th} threshold for the anoxic phase is here smaller than one, suggesting that in order to achieve the best classification performance the method needs to adopt a more severe criterion for the violation of the Q statistics with respect to the theoretical threshold.

Table 5.21 Optimized thresholds of the RT-MWPCA method for the calibration set of Table 5.14 using the parameters of Combination A

		Aerobic phase	Anoxic phase
T^2	T^2_{lim}	9.874	7.089
	k_T	2.575	8.539
	T_{th}	25.423	60.540
Q	Q_{lim}	0.119	0.560
	k_Q	3.923	0.674
	Q_{th}	0.467	0.377
Total PCs		4	4
Retained PCs		3	2

The results of the calibration in terms of classifications are summarized in the confusion matrices of Figure 5.27.

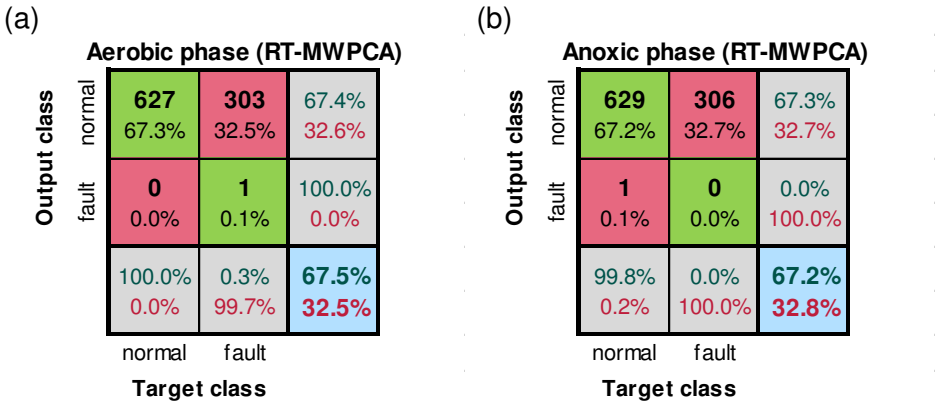


Figure 5.27 Confusion matrices describing the detection performances of the finer faults for the RT-MWPCA method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination A

The results obtained in calibration for both phases are similar. In both cases the methods fail to recognise the presence of faults (only one fault is detected, but it is correct just in the aerobic phase). Despite the overall classification success is greater than 60%, then, the RT-MWPCA method is once again unable to produce the required discrimination power.

As to the validation of the models obtained using the Combination A, we observe that the general classification efficiency is maintained but unfortunately so are the fault detection rates (Figure 5.28).

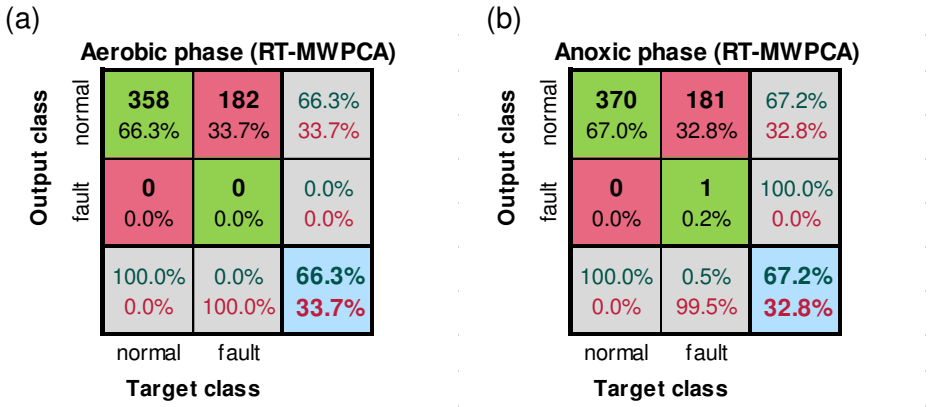


Figure 5.28 Confusion matrices describing the detection performances of the finer faults for the RT-MWPCA method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination A

The same conservative detection behaviour emerged when the method was applied to the synthetic dataset. Moreover, the detection performances of the RT-MWPCA method are only marginally affected by the choice of the combination of the diagnostic parameters. Using Combination E, for example, the performances marginally improve but are still far from being satisfactory (the results of the optimization of the diagnostic threshold position for the Combination E are reported in Table 5.22).

Table 5.22 Optimized thresholds of the RT-MWPCA method for the calibration set of Table 5.14 using the parameters of Combination E

		Aerobic phase	Anoxic phase
T^2	T^2_{lim}	15.981	15.981
	k_T	4.158	2.686
	T_{th}	66.458	42.939
Q	Q_{lim}	0.543	0.594
	k_Q	2.866	6.884
	Q_{th}	1.557	4.092
Total PCs		11	11
Retained PCs		5	5

Figure 5.29 show that even if the general classification error is, in this case too, around 67%, just a couple of finer faults are identified in the aerobic phases and the number of anomalies spotted in the anoxic one is only a small percentage of those observed.

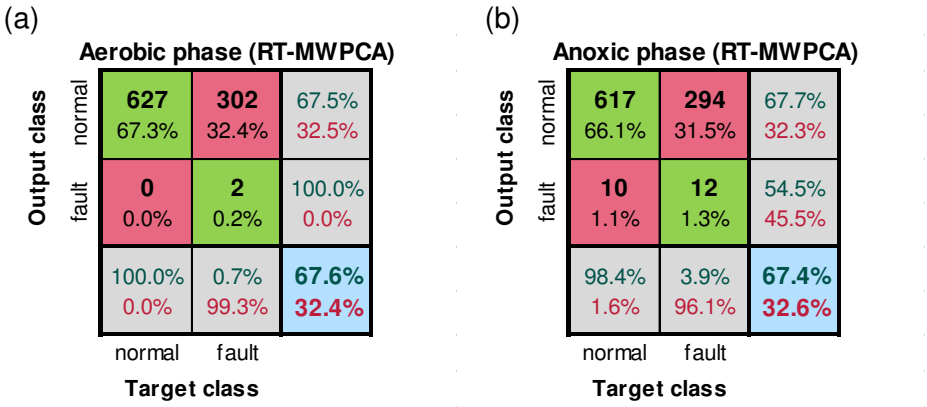


Figure 5.29 Confusion matrices describing the detection performances of the finer faults for the RT-MWPCA method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination E

Worse performances are then displayed for the validation set (Figure 5.30), where no anomalies are highlighted in the aerobic phase part of the algorithm and just a few in the anoxic phase, most of which, however, are false positives.

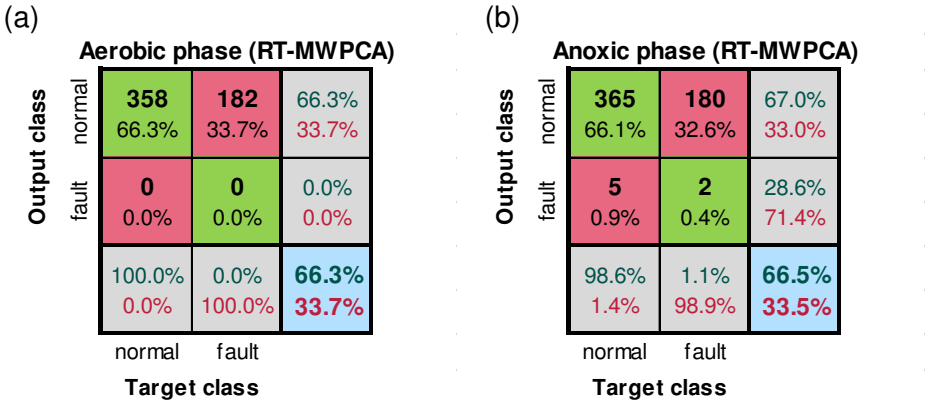


Figure 5.30 Confusion matrices describing the detection performances of the finer faults for the RT-MWPCA method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination E

Some more general considerations on the results achieved with the RT-MWPCA method for the different combination of parameters can be made by commenting Table 5.23 and Table 5.24, where the combined (aerobic + anoxic) results of the three performance indexes we have been focusing on (the global classification success rate, the fault identification rate and the detection reliability) are presented for all five the combinations of the diagnostic parameters tested.

Table 5.23 Global classification success, fault detection success and reliability of the detection of the RT-MWPCA method for the different combinations of parameters in the calibration set of Table 5.14

	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Global class. success [%]	67.3%	67.2%	67.2%	68.0%	67.5%
FD success [%]	0.1%	0.0%	0.0%	3.6%	2.3%
Detection confidence [%]	49.9%	0.0%	0.0%	73.4%	77.2%

Table 5.24 Global classification success, fault detection success and reliability of the detection of the RT-MWPCA method for the different combinations of parameters in the validation set of Table 5.14

	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Global class. success [%]	66.8%	66.8%	66.8%	66.1%	66.4%
FD success [%]	0.3%	0.3%	0.3%	0.3%	0.6%
Detection confidence [%]	50.5%	50.5%	50.5%	10.1%	14.5%

As anticipated, the influence of the particular choice of descriptors used for the instances is modest but still when more observed variables are available, as in Combination D and Combination E, this allow to have some detected faults at least in the calibration set. This is however no more true for the validation set. Regardless the combination of diagnostic features, in fact, the RT-MWPCA method seems unable to detect the anomalies and the reassuring 66% of the classification success here just actually reflects the proportion of normal phases on the whole dataset.

To better investigate the reasons of this poor outcome of the method let us present the performances of this when a different calibration/validation framework is used. With the partitioning of the dataset of synthetic measurements of Table 5.13, that is a one year-long subset of the data and using the first half of it for the training and the other half for the validation, the RT-MWPCA method shows quite different results. The calibration of the method using Combination A, for example, returns the results summarized in the confusion matrices of Figure 5.27.

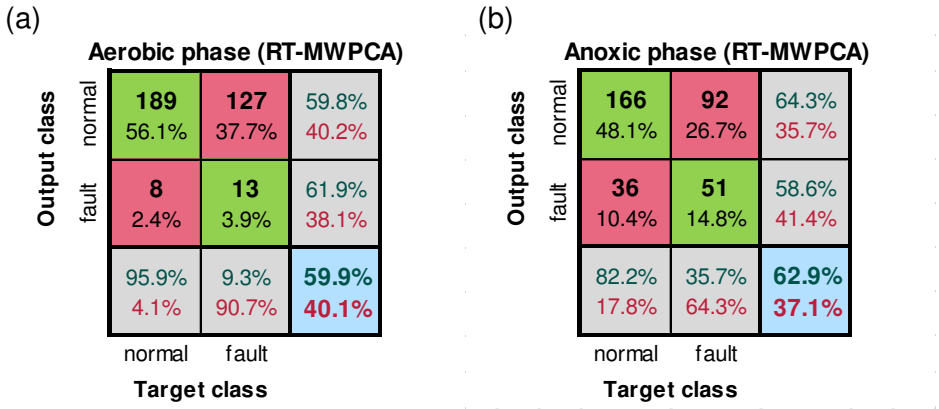


Figure 5.31 Confusion matrices describing the detection performances of the finer faults for the RT-MWPCA method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.13 using the parameters of Combination A

The position of the statistical thresholds determined in this case (T_{th} = 4.3 and Q_{th} = 0.86 for the aerobic part and T_{th} = 3.6 and Q_{th} = 0.78 for the anoxic) are way lower than those of Table 5.21, thus more frequently crossed by the instances. Therefore, the algorithms appear to be better able to detect the anomalies. In terms of global classification success nearly 60% of the phases are assigned to their correct class and one out of three of the finer faults observed is correctly identified in the anoxic phase. Moreover, the detection results reliable most of the times, at least in the calibration part,

since the number of false alarm is lower than the faults correctly identified. As for the validation of the models obtained using the Combination A we observe that the general classification efficiency is maintained but a substantial decrease of the accuracy of the fault detection is instead observed, and most of the anomalies detected in this section of the dataset are actually misclassified normal phases (false positives).

When the Combination E is used the RT-MWPCA methods trained on the calibration set of Table 5.13 display results that are on the same page of the results observed for the dataset of Table 5.14. The confusion matrices of Figure 5.29, reporting the performance of the method on the calibration set, show that even if the general classification success is in this case too, around 60%, not even one fault is identified in the aerobic phases and the number of anomalies spotted in the anoxic one is only a small percentage of those observed.

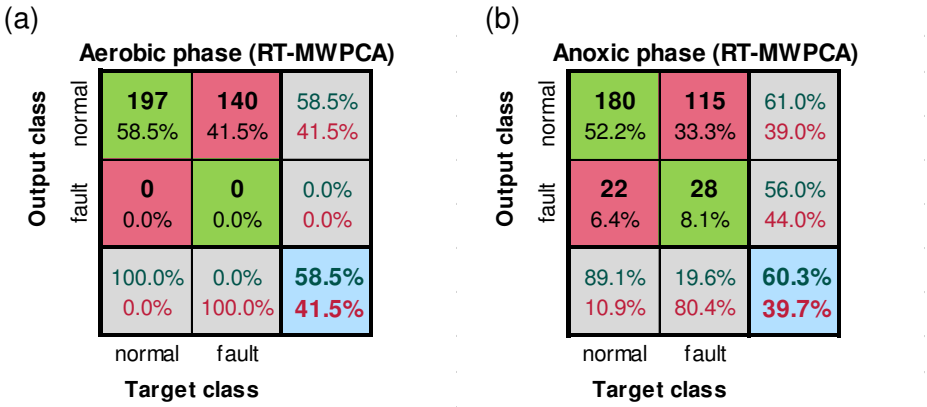


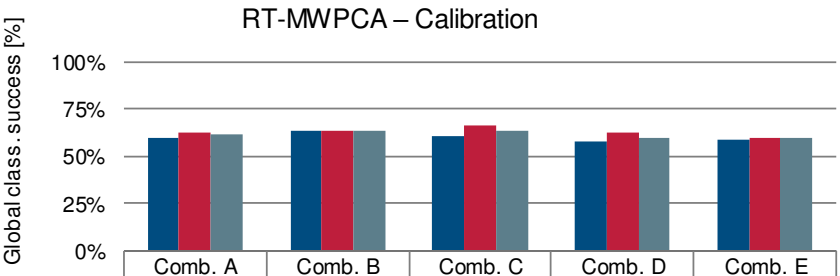
Figure 5.32 Confusion matrices describing the detection performances of the finer faults for the RT-MWPCA method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.13 using the parameters of Combination E

The same considerations apply to the validation of the models where the confidence in the detected faults during the anoxic phase is even worse.

Some more general considerations on the results achieved with the RT-MWPCA method for the different combination of parameters can be made by commenting Figure 5.33, Figure 5.34 and Figure 5.35, where the compared results of the three performance indexes we have been focusing on (the global classification success rate, the fault identification rate and the detection reliability) are presented.

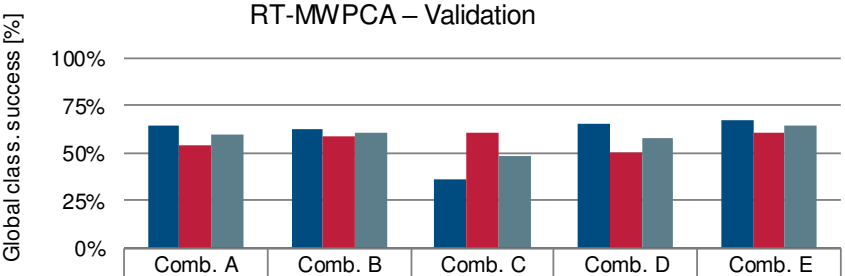
The classification efficiencies in Figure 5.33 show no particular influence of the diagnostic parameters set chosen, remaining around 60% also in validation. This is can be regarded as a positive aspect that proves the robustness of the general classification capabilities of the method even when a suboptimal set of observed variables is available.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	59.9%	63.8%	61.1%	58.2%	58.5%
Anoxic phse	62.9%	63.2%	66.1%	62.3%	60.3%
Combined	61.4%	63.5%	63.6%	60.3%	59.4%

(b)



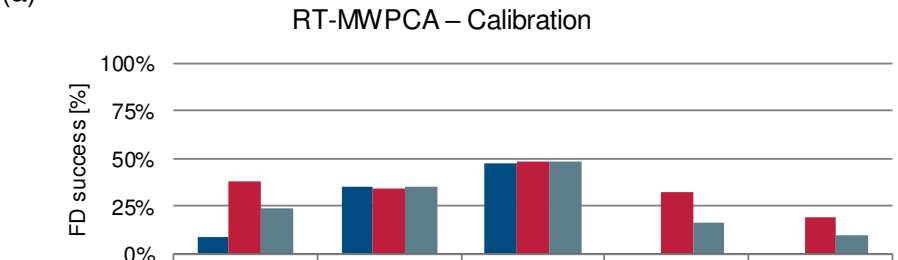
	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	64.7%	62.3%	35.8%	65.4%	67.9%
Anoxic phse	54.3%	58.8%	60.6%	50.9%	60.8%
Combined	59.5%	60.6%	48.1%	58.2%	64.4%

Figure 5.33 Global classification success of the RT-MWPCA method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

Conversely, regarding the ability to correctly identify and isolate the faulty state of the phases the RT-MWPCA demonstrates that if the thresholds are calibrated with the reduced dataset it can detect some anomaly but with some generally poor results (Figure 5.34). Varying the combination of the diagnostic features it is possible to achieve in the calibration phase almost the same identification success observed in 5.3.2 for the real measurements dataset, between 30% and 50% (Combination B and Combination C) which is however insufficient, especially since this index almost halves in the validation set. Moreover, as observed in Figure 5.29 and Figure 5.30, the use of Combination E do not produce the identification of any fault event in the aerobic phase, neither in the calibration set nor in the validation one. The same behaviour is displayed in calibration also by Combination D, but a small number of faults are anyway detected in validation.

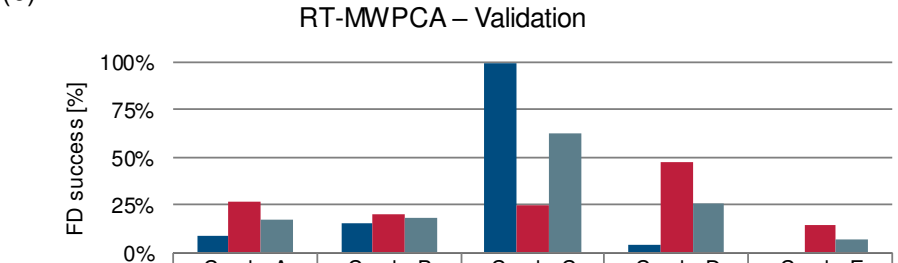
The alarms provided by this method (Figure 5.35) result generally as trustworthy as those provided by the Bayesian method. Despite, in fact, the reliability on the faults detected is generally acceptable (50-60%) in calibration, this confidence drops in the validation phase, hardly scoring more than 30% in some of the combinations tested.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	9.3%	35.7%	47.9%	0.0%	0.0%
Anoxic phse	37.7%	34.3%	49.0%	32.9%	19.6%
Combined	23.7%	35.0%	48.5%	16.6%	9.9%

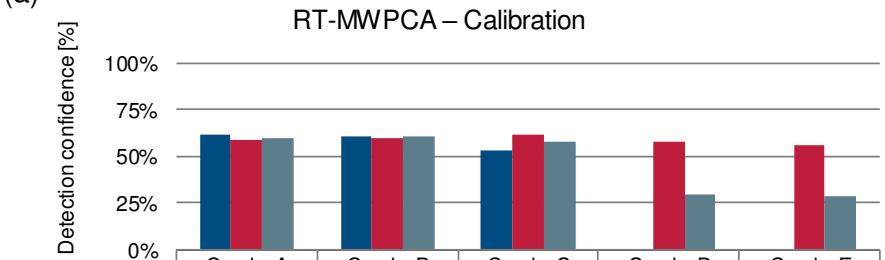
(b)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	8.4%	15.8%	99.5%	3.7%	0.0%
Anoxic phse	26.5%	20.0%	24.9%	47.6%	14.6%
Combined	17.4%	17.9%	62.4%	25.6%	7.3%

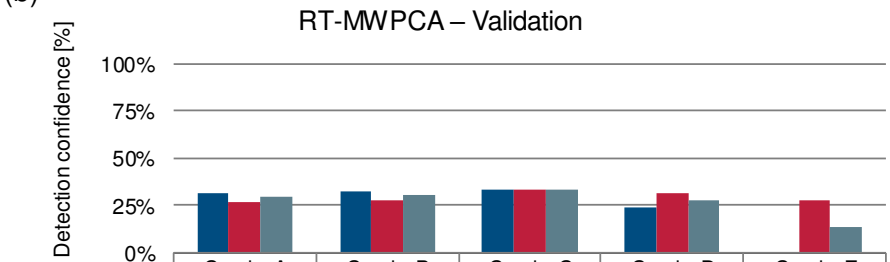
Figure 5.34 Fault identification success of the RT-MWPCA method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	61.9%	61.0%	53.6%	0.0%	0.0%
Anoxic phse	58.6%	59.8%	61.4%	58.0%	56.0%
Combined	60.2%	60.4%	57.5%	29.3%	28.3%

(b)

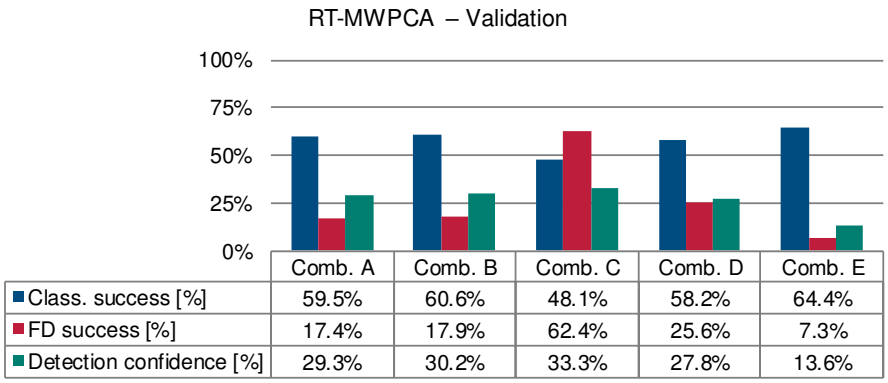


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	31.4%	32.2%	33.3%	24.1%	0.0%
Anoxic phse	27.1%	28.2%	33.3%	31.5%	27.3%
Combined	29.3%	30.2%	33.3%	27.8%	13.6%

Figure 5.35 Confidence rate on the faults identified by the RT-MWPCA method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

As a final consideration we can observe that from the comparison of the behaviour of RT-MWPCA method in the validation set of Table 5.13 under the different

combination of diagnostic parameters the configuration that seems to work better for this algorithm is provided by Combination C, where despite a higher global classification error most of the fault events are identified and the reliability on the alarms generated results a little higher than using other combinations.



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	59.5%	60.6%	48.1%	58.2%	64.4%
FD success [%]	17.4%	17.9%	62.4%	25.6%	7.3%
Detection confidence [%]	29.3%	30.2%	33.3%	27.8%	13.6%

Figure 5.36 Comparison of the performance indexes for the RT-MWPCA method under the different combinations of parameters in validation set of Table 5.13

Contrary to what was expected, from these comparisons emerge that the RT-MWPCA method seems to produce a better performance when the statistical thresholds used to discriminate the state of the phases are determined by minimizing the error on a shorter set of observations and using a reduced number of diagnostic parameters. In fact, as the combination of observed parameters gets more complex the method exhibits increasingly worse performances, ultimately resulting in no faults detected at all. Furthermore, the tests on the algorithm conducted on the dataset of Table 5.14 revealed that when a larger dataset is used to determine the optimal detection thresholds these are generally placed in a way to display an excessively conservative behaviour of the detection, so that almost all the instances tested are classified as normal.

5.4.4 Results of the fault detection using the simple binary tree

The same dataset partition of Table 5.14 has been also used to train and assess the performances of the classification trees of lower complexity (first row of Table 5.3) implemented as in paragraph 5.2.3. In the spirit of the framework of Figure 5.2 two classifiers have been trained, one for the anoxic process phases and one for the aerobic ones. Figure 5.37 shows that when the basic combination of diagnostic parameters is used (Combination A, the same available for the real plant data) the most important features, the roots of the tree, are associated with the nitrates: the nitrate concentration growth rate $s_{\text{aero},NOx-,\text{out}}$ in the aerobic phase (Figure 5.37.a) and the slope $s_{\text{anox},NOx-,\text{out}}$ in the anoxic (Figure 5.37.b). After the split of the root node in the aerobic phase all subsequent classifications of the state of the process are based on the concentration average of the ammonia, as in the anoxic phase.

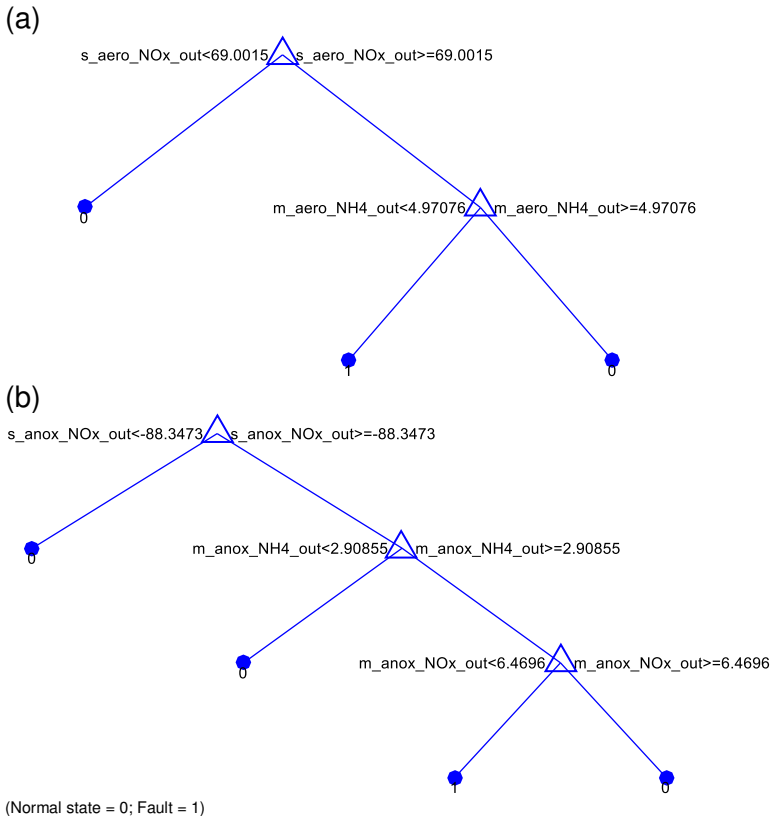


Figure 5.37 Structure of the classification trees of medium complexity trained on the calibration set described in Table 5.14 for the aerobic (a) and anoxic (b) phase using the Combination A of the diagnostic parameters

From the point of view of the performances achieved in the calibration phase using Combination A we observe (Figure 5.38) that also in this case the global classification error is above 60% for both the aerobic and anoxic process phase. A fair detection accuracy is obtained: 67% in the aerobic phase (Figure 5.38.a) while in the anoxic phase no false alarm is generated (Figure 5.38.b). The number of faults detected however is only a small part of those observed and the whole detection is off balanced in favour of the 'normal' class.

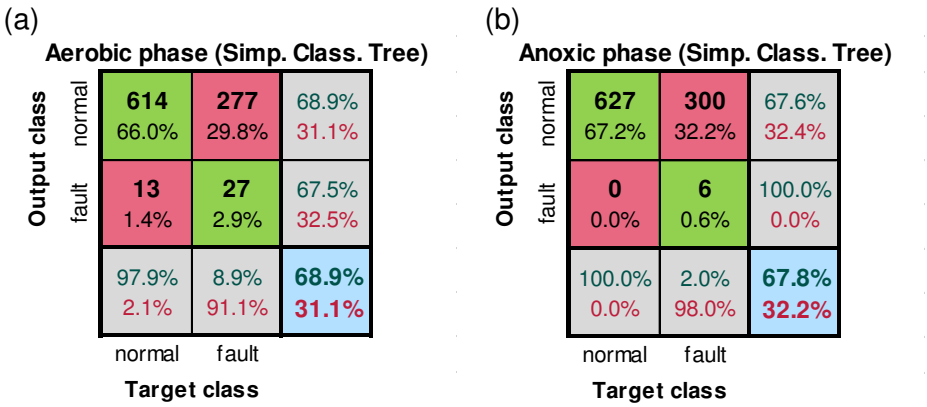


Figure 5.38 Confusion matrices describing detection performances of the finer faults for the simple binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination A

In validation the performances drop considerably (Figure 5.39). While, in fact, the global classification error is still around 66%, the number of faults identified is almost zero for the aerobic phase (it is exactly zero for the anoxic phase) and the accuracy of

the detection is scarce. In can be observed that also in the validation set the detection tends to classify most of the observations as well-behaved.

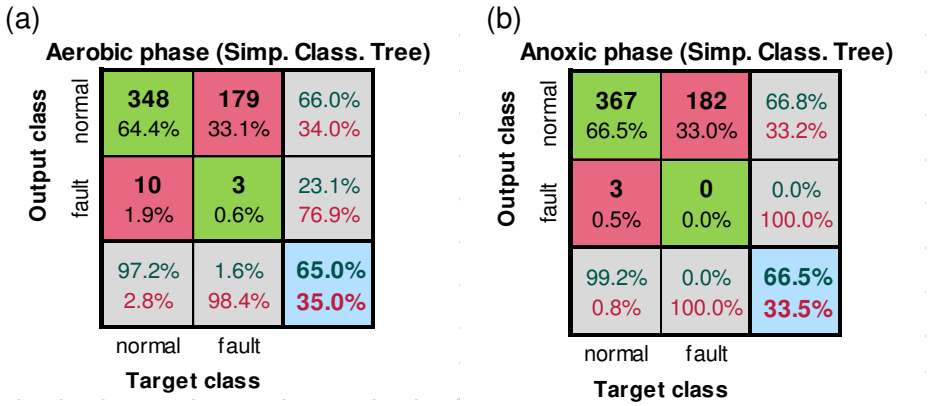


Figure 5.39 Confusion matrices describing detection performances of the finer faults for the simple binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination A

Training the classification trees on the dataset of Table 5.13, instead, the outcome in terms of shape of the tree and ranking of the descripting features is almost the same as in Figure 5.37 but they set of rules cut the features space in different points and the resulting performances are very different. Quite satisfying are in this case the detection rates, especially in the aerobic phase (Figure 5.40.a) where the 81% of the faults are identified. The algorithm trained on the anoxic phase data (Figure 5.40.b), instead, isolates a little less than half of the faults observed but its results are more reliable in 60% of the times.

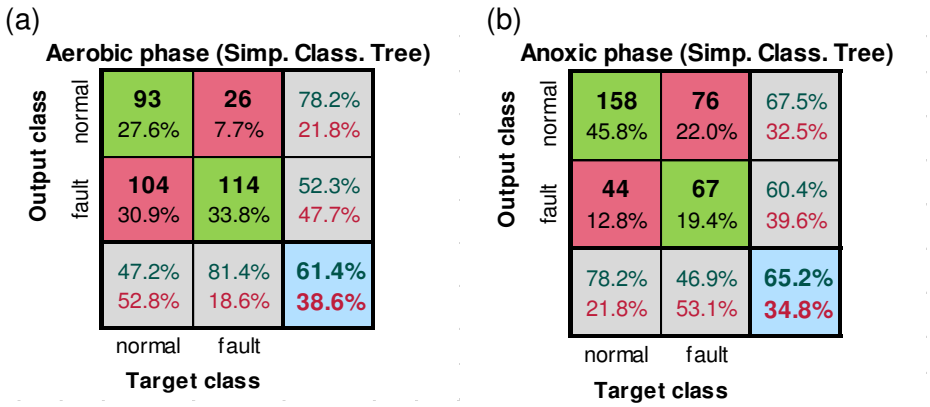


Figure 5.40 Confusion matrices describing detection performances of the finer faults for the simple binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.13 using the parameters of Combination A

As expected, in validation the performances drop (Figure 5.41), but less than with the previous case. While the global classification error, in fact, is around 50% and the accuracy of the detection drops to 30%, but the tree is able to identify more than a half of the observed anomalous phases.

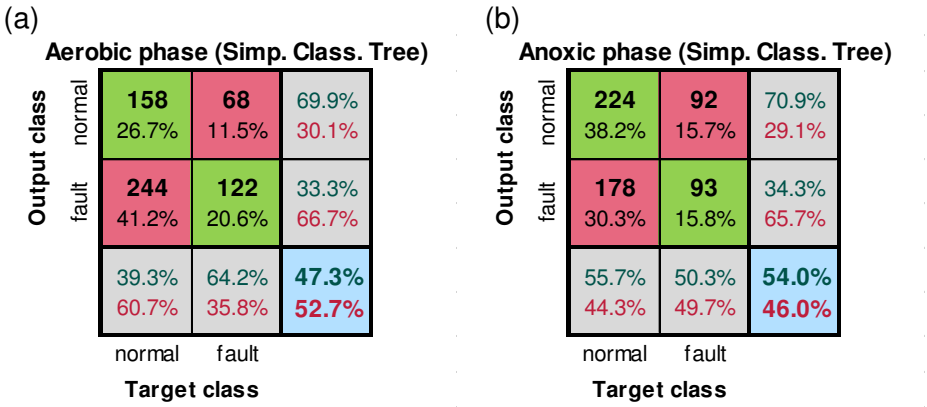


Figure 5.41 Confusion matrices describing detection performances of the finer faults for the simple binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.13 using the parameters of Combination A

Contrary to what was expected, in this case we observe that the use of a reduced number of training samples seems to generate classifiers that have more general classification capabilities.

Changing the combination of parameters and introducing some observed variables not available in the real measurements dataset, the performances of the binary tree and the relative importance of the diagnostic parameters change.

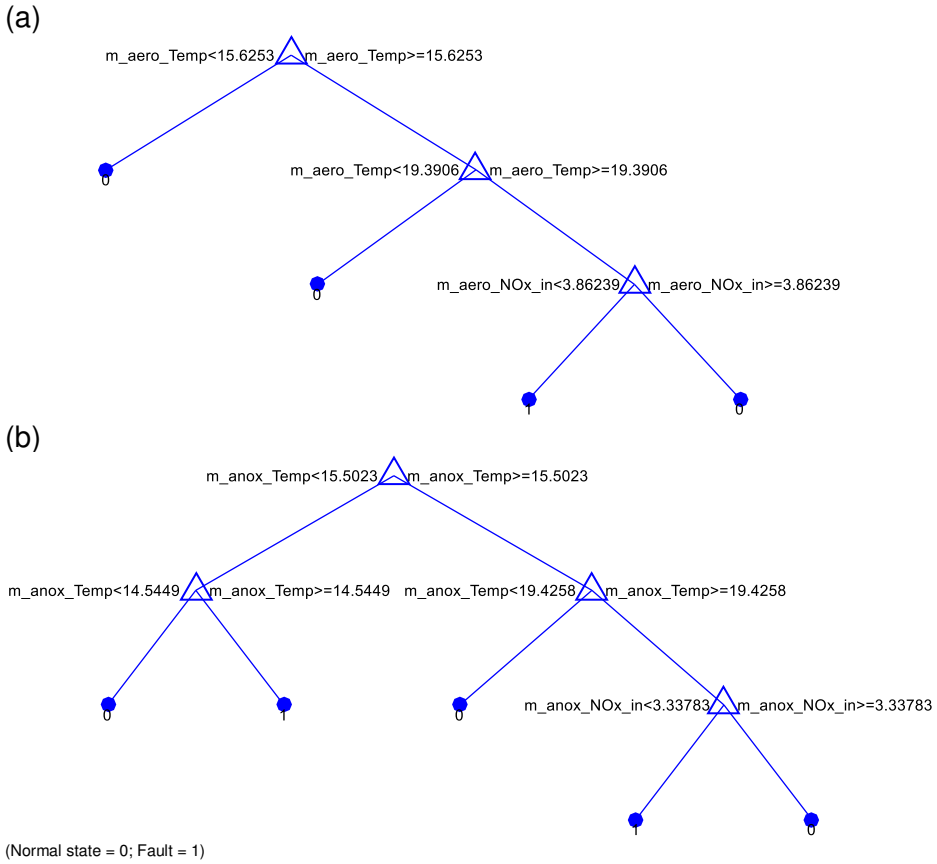


Figure 5.42 Structure of the classification trees of medium complexity trained on the calibration set described in Table 5.14 for the aerobic (a) and anoxic (b) phase using the Combination E of the diagnostic parameters

Using Combination E on the training set of Table 5.14, for example, it can be noticed that the most important parameter to discriminate the state of the process is now the

temperature (Figure 5.42). This is not surprising since all the kinetics of the biological processes, including nitrification and denitrification, are deeply influenced by the temperature, therefore enabling a consistency check on the average concentrations and most of all the nitrification and denitrification rates.

As for what concerns the performances of the calibrated models, with respect to the results shown in Figure 5.38 the new combination brings some improvements under all three of the aspects we are focusing on (Figure 5.43). A general increase can be observed in the global classification success rate, which reaches 69% of the tested aerobic phases and 71% of the anoxic ones. Moreover, the confidence in the detection is boosted, 66% in the aerobic (Figure 5.43.a) and even more in the anoxic phase (Figure 5.43.b).

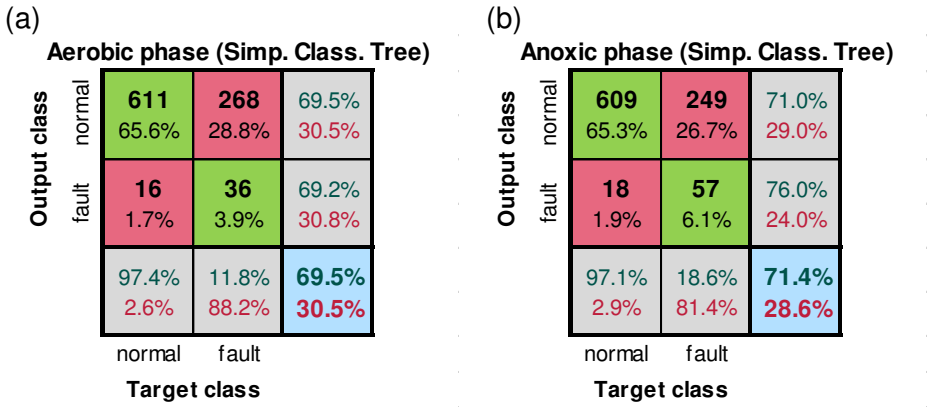


Figure 5.43 Confusion matrices describing detection performances of the finer faults for the simple binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination E

A general detection performance decrease can be observed in the validation set (Figure 5.44), where the classification results unreliable in the anoxic phase and has less than 50% accuracy in the anoxic case. Moreover, the number of detected faulty events is extremely small. In both the aerobic and the anoxic phase the trained classifiers show the same tendency observed with Combination A and exhibit a tendency to underestimate the anomalous phase of the tested samples.

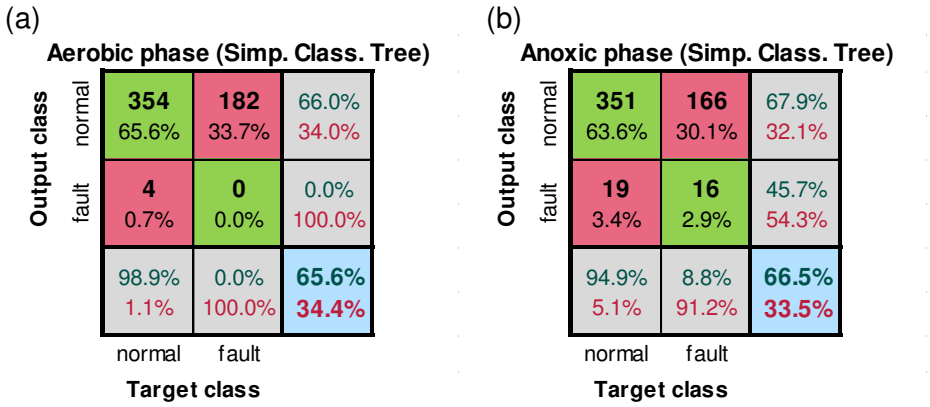
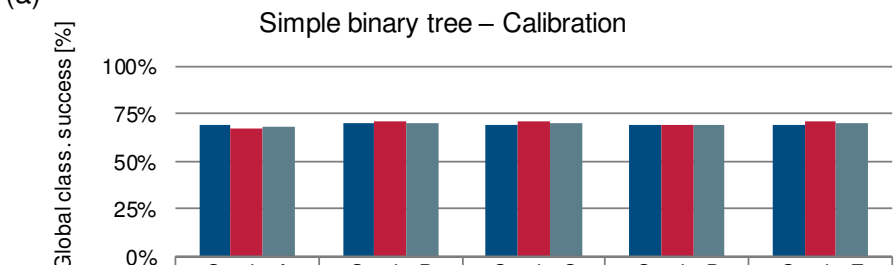


Figure 5.44 Confusion matrices describing detection performances of the finer faults for the simple binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination E

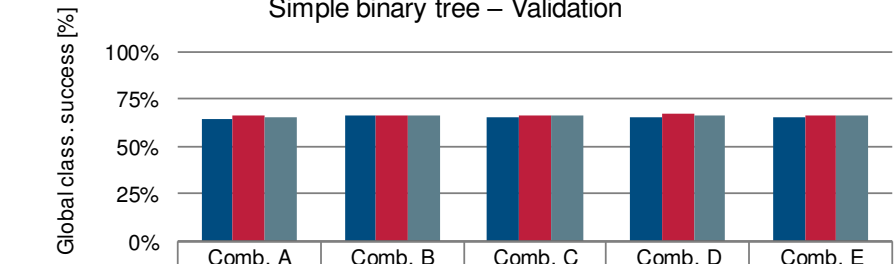
From the comparative analysis of the classification success rate for the different sets of parameters presented in Figure 5.45 in the dataset partitioning of Table 5.14 it is evident that a more populated set of observed variable has a small influence on the performances. The introduction of the average temperature value in Combination B to E improves the classification by no more than 2%.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	68.9%	70.2%	69.5%	69.5%	69.5%
Anoxic phse	67.8%	71.1%	71.4%	69.0%	71.4%
Combined	68.3%	70.7%	70.5%	69.2%	70.5%

(b)

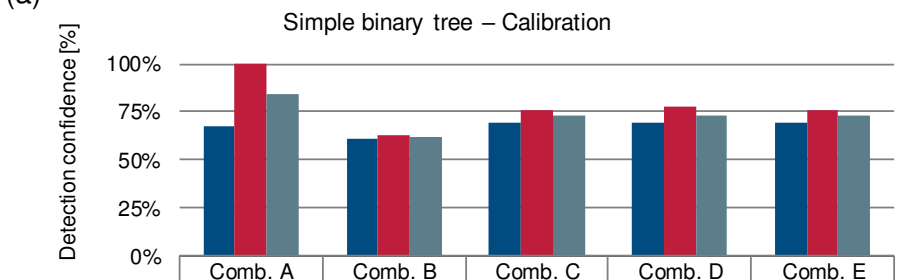


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	65.0%	66.7%	65.6%	65.6%	65.6%
Anoxic phse	66.5%	66.8%	66.5%	67.0%	66.5%
Combined	65.8%	66.8%	66.1%	66.3%	66.1%

Figure 5.45 Global classification success of the simple binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

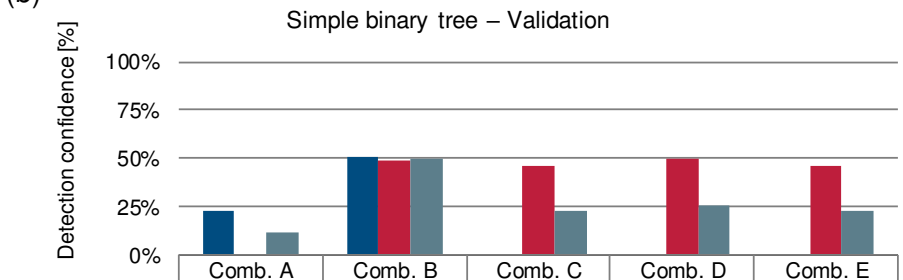
The detection accuracy benefits from a wider range of descriptors in the calibration phase (Figure 5.46.a), and in validation is just some percent point below 50%, at least in the anoxic phase (Figure 5.46.b). The result is however biased by the insufficient detection performances of the models generated for the aerobic part.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	67.5%	61.3%	69.2%	69.2%	69.2%
Anoxic phse	100.0%	62.7%	76.0%	77.4%	76.0%
Combined	83.8%	62.0%	72.6%	73.3%	72.6%

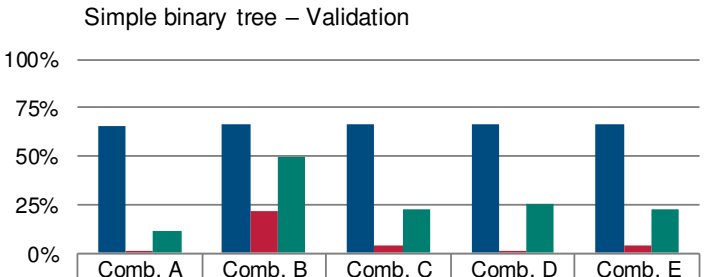
(b)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	23.1%	50.8%	0.0%	0.0%	0.0%
Anoxic phse	0.0%	48.8%	45.7%	50.0%	45.7%
Combined	11.4%	49.8%	23.1%	25.3%	23.1%

Figure 5.46 Confidence rate on the faults identified by the simple binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

As mentioned above, the trained classification trees tend to underestimate the faulty states and therefore a very small number of anomalies are identified. Figure 5.47 synthetizes the performance indexes of interest for all the combinations of diagnostic features tested using the partitioning of Table 5.14. It is clear that the so determined simple classification trees, given the generally poor detection results, cannot be a valuable candidate for the FD of the finer faults.



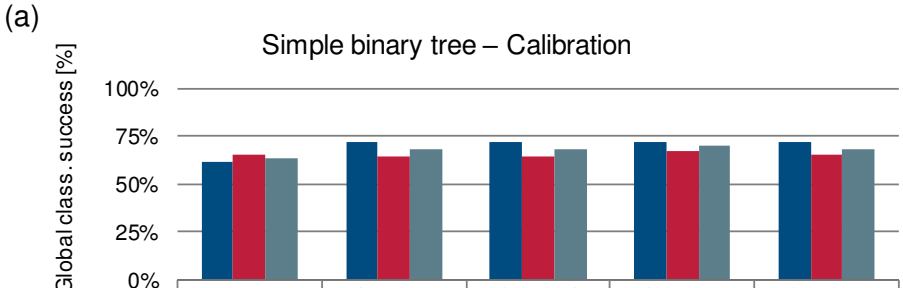
	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	65.8%	66.8%	66.1%	66.3%	66.1%
FD success [%]	0.8%	22.1%	4.4%	0.3%	4.4%
Detection confidence [%]	11.4%	49.8%	23.1%	25.3%	23.1%

Figure 5.47 Comparison of the performance indexes for the simple binary classification tree method using the different combinations of parameters in the validation set of Table 5.14

One can now wonder whether reducing the size of the training set, for example using the dataset partitioning of Table 5.13, and limiting the number of the observed variables, the performances could improve, as it is the case with the RT-MWPCA method. The

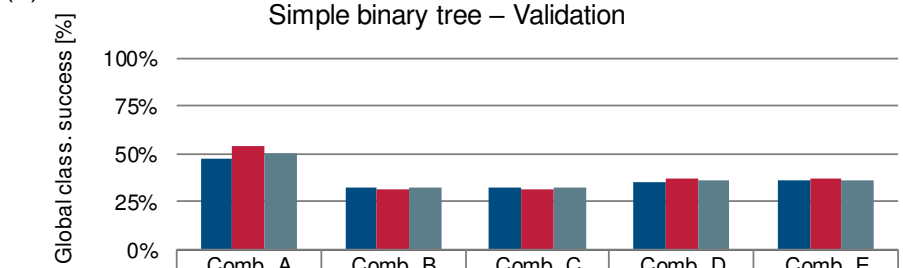
comparative analysis of the classification success rate in Figure 5.48 reveal that a larger set of observed variable leads to better results in the calibration of the models. The introduction of the average temperature value in Combination B to E seems however to produce a significant loss in the generalization capabilities of the trained classifiers, reflected in lower classification efficiencies (Figure 5.48.b).

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	61.4%	72.1%	72.1%	72.4%	72.4%
Anoxic phse	65.2%	64.9%	64.9%	67.8%	65.2%
Combined	63.3%	68.5%	68.5%	70.1%	68.8%

(b)

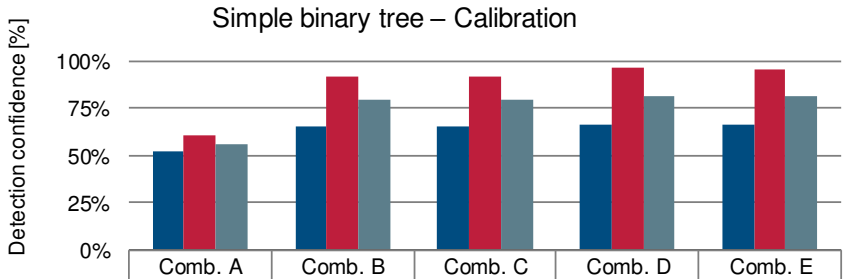


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	47.3%	32.4%	32.4%	35.3%	36.3%
Anoxic phse	54.0%	32.0%	32.0%	36.8%	36.8%
Combined	50.6%	32.2%	32.2%	36.0%	36.5%

Figure 5.48 Global classification success of the simple binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

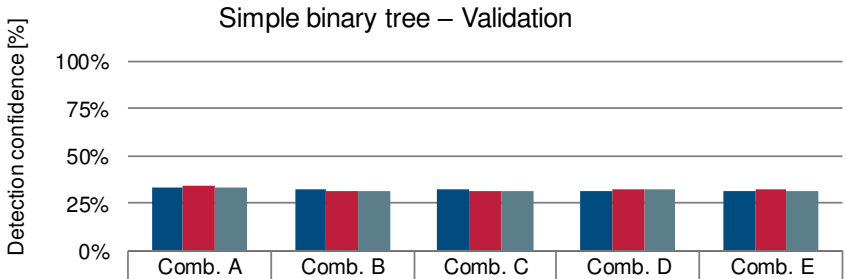
A similar behaviour is reflected in the detection accuracy rates. While the aerobic and anoxic classifiers produce quite high reliability levels on the calibration set (Figure 5.49.a), they seems to improve when a broader and more diversified spectrum of parameters is used, particularly for the anoxic processes, the same cannot be said for the validation set (Figure 5.46.b).

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	52.3%	66.0%	66.0%	66.4%	66.4%
Anoxic phse	60.4%	92.3%	92.3%	97.1%	96.0%
Combined	56.4%	79.3%	79.3%	81.9%	81.4%

(b)

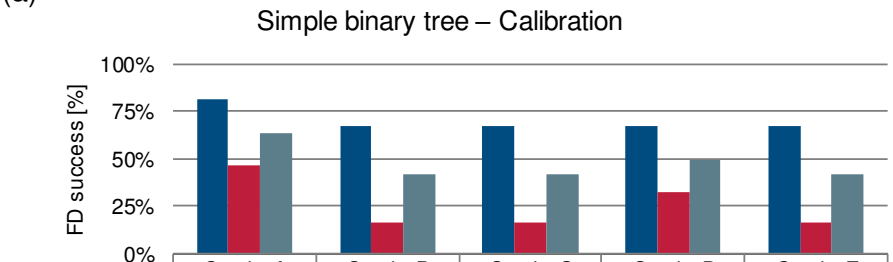


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	33.3%	32.2%	32.2%	32.0%	31.1%
Anoxic phse	34.3%	31.7%	31.7%	32.5%	32.6%
Combined	33.8%	32.0%	32.0%	32.2%	31.8%

Figure 5.49 Confidence rate on the faults identified by the simple binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

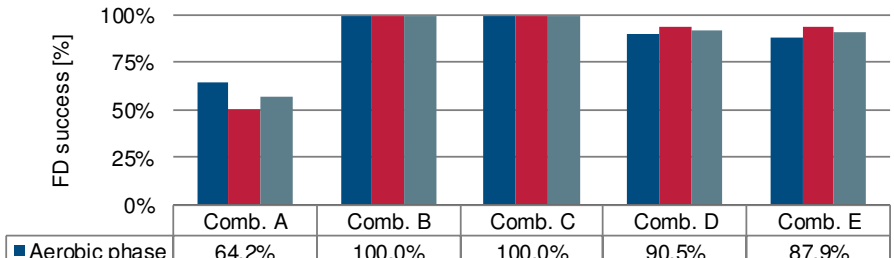
The effect of the different combinations of diagnostic parameters on the number of observed faults correctly identified is also very different in the calibration and in the validation set, but in an opposite fashion. While, in fact, in the calibration phase (Figure 5.50.a) the trained model for the aerobic processes is able to isolate most of the anomalous phases (but on the anoxic side the results are more disappointing), in the validation set most of the finer faults observed are identified by the method (Figure 5.50.b). However, this appears to be just a side effect of a fault-oriented detection, where the different conditions of the dataset used for the evaluation of the performances with respect to those of the calibration set are interpreted as indication of an anomalous behaviour. In both the aerobic and the anoxic case the classifiers show the tendency to overestimate the faulty state of the process and this is a probable shortcoming of having selected the temperature as a diagnostic variable in the calibration phase of the model. The climatic conditions in the validation phase are in fact different and the models seems to lack of the necessary generalization capability.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	81.4%	67.9%	67.9%	67.9%	67.9%
Anoxic phse	46.9%	16.8%	16.8%	32.1%	16.8%
Combined	63.9%	42.1%	42.1%	49.8%	42.1%

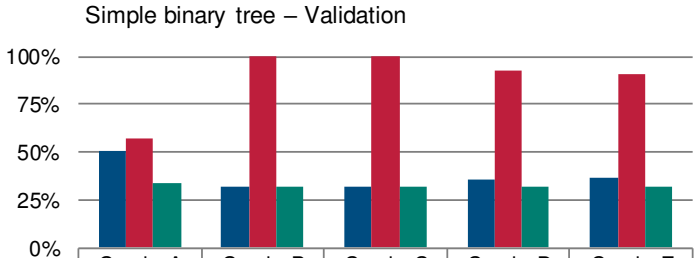
(b)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	64.2%	100.0%	100.0%	90.5%	87.9%
Anoxic phse	50.3%	100.0%	100.0%	94.1%	94.1%
Combined	57.3%	100.0%	100.0%	92.3%	91.0%

Figure 5.50 Fault identification success of the simple binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

As a consequence, the binary classification trees trained using a the shorter calibration set of Table 5.13 seems to only apparently benefit from the availability of a wider range of diagnostic features (Figure 5.51), since the increased information is exploited to create classifiers too much tailored on the specific features of the dataset they are trained on, lacking the necessary generalization capabilities. In this case too, but for opposite reasons than in Figure 5.47, it is not advisable to use these classifiers for the real-time FD due to the high number of false alarms and the relatively low global classification rate.



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	50.6%	32.2%	32.2%	36.0%	36.5%
FD success [%]	57.3%	100.0%	100.0%	92.3%	91.0%
Detection confidence [%]	33.8%	32.0%	32.0%	32.2%	31.8%

Figure 5.51 Comparison of the performance indexes for the simple binary classification tree method using the different combinations of parameters in the validation set of Table 5.13

5.4.5 Results of the fault detection using the medium binary tree

In order to assess whether the concerns emerged in the previous paragraph are related or not to the limitations imposed to the tree growth, the same tests have been conducted building for the same datasets partitioning some more complex classification trees (second row of Table 5.3).

From the point of view of the relative importance of the diagnostic variable we observe that for Combination A the preference that the simpler trees granted to the features associated to the nitrate it is here maintained: $s_{\text{aero,NO}x-\text{,out}}$ is still the root node for the classifier determined in the aerobic phase (Figure 5.37.a) and the slope $s_{\text{anox,NO}x-\text{,out}}$ provides the first splitting in the anoxic (Figure 5.37.b). This is a confirmation that a more detailed set of rules for the classification should not alter the general importance ranking of the antecedents used.

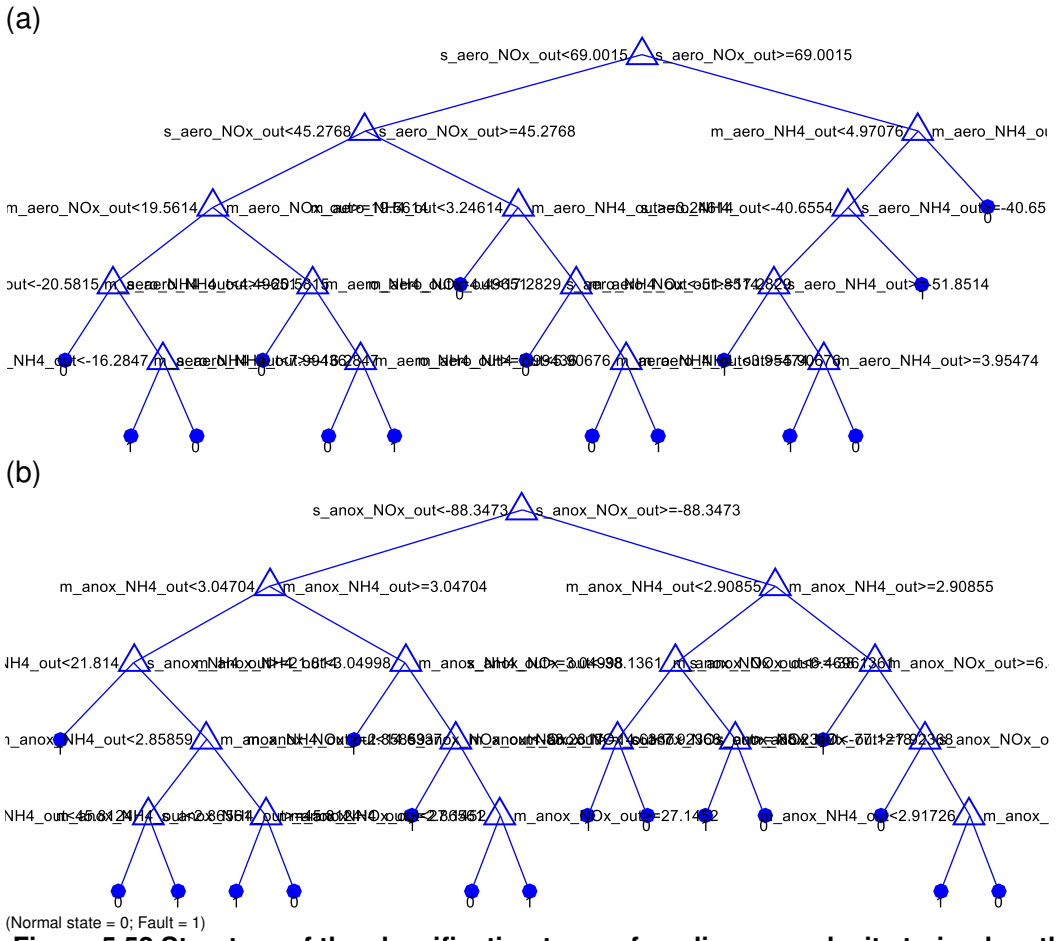


Figure 5.52 Structure of the classification trees of medium complexity trained on the calibration set described in Table 5.14 for the aerobic (a) and anoxic (b) phase using the Combination A of the diagnostic parameters

The models calibrated for each of the process phases using the training set described in Table 5.14 show better results than those in Figure 5.38. The increased level of sophistication of the detection process allows to train two classifiers that score above 70% of the classification success either in the aerobic (Figure 5.53.a) and in the anoxic

phase (Figure 5.53.b). Moreover, the general good classification rate comes together with an actual good discrimination capability, since despite a low detection rate index the reliability of the detection is quite high.

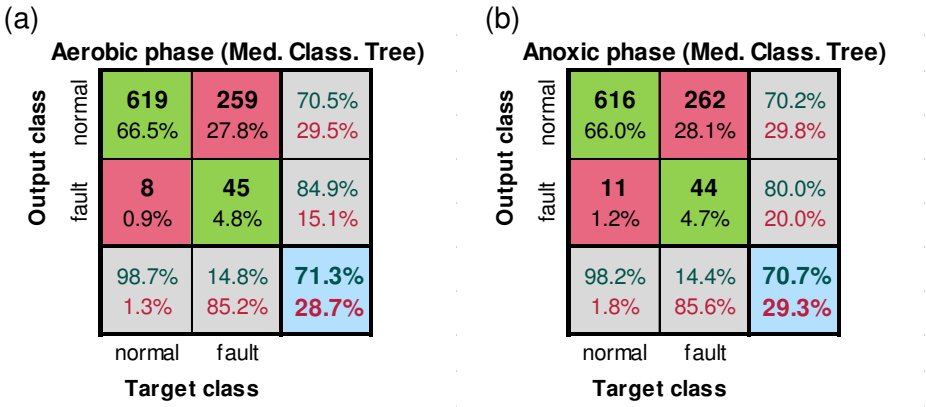


Figure 5.53 Confusion matrices describing detection performances of the finer faults for the medium binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination A

The results in the validation set (Figure 5.54) display the same behaviour observed for the simpler classification trees with this partitioning of the dataset. The classifiers tend to label the vast majority of the tested instances as 'normal', leading to a scarce identification rate of the anomalies. The accuracy of the detection also drops in the validation, touching 40% for the aerobic phase and 22% for the anoxic.

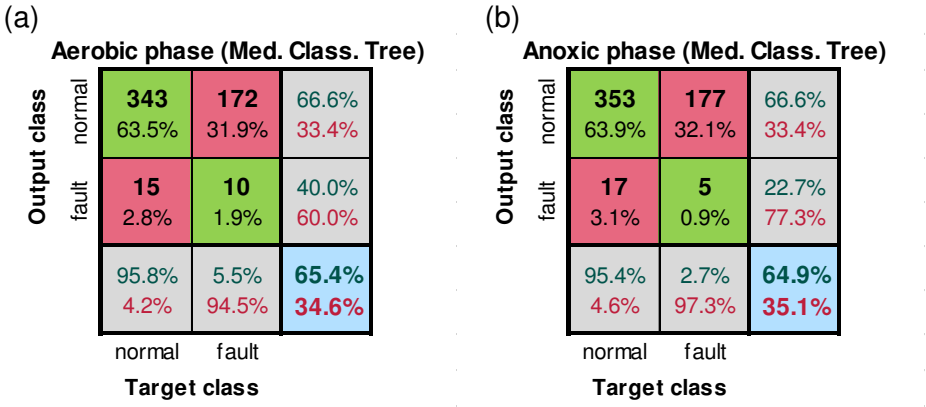


Figure 5.54 Confusion matrices describing detection performances of the finer faults for the medium binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination A

The same considerations made for the Combination A apply also when the Combination E of the diagnostic parameters is used to train the trees. Both the classifiers calibrated on the data from the aerobic and anoxic phases (Figure 5.55), in fact, consider the temperature as most important parameter, followed by the average DO concentration. The higher number of branches and nodes brings as a consequence that all the other parameters now play some role.

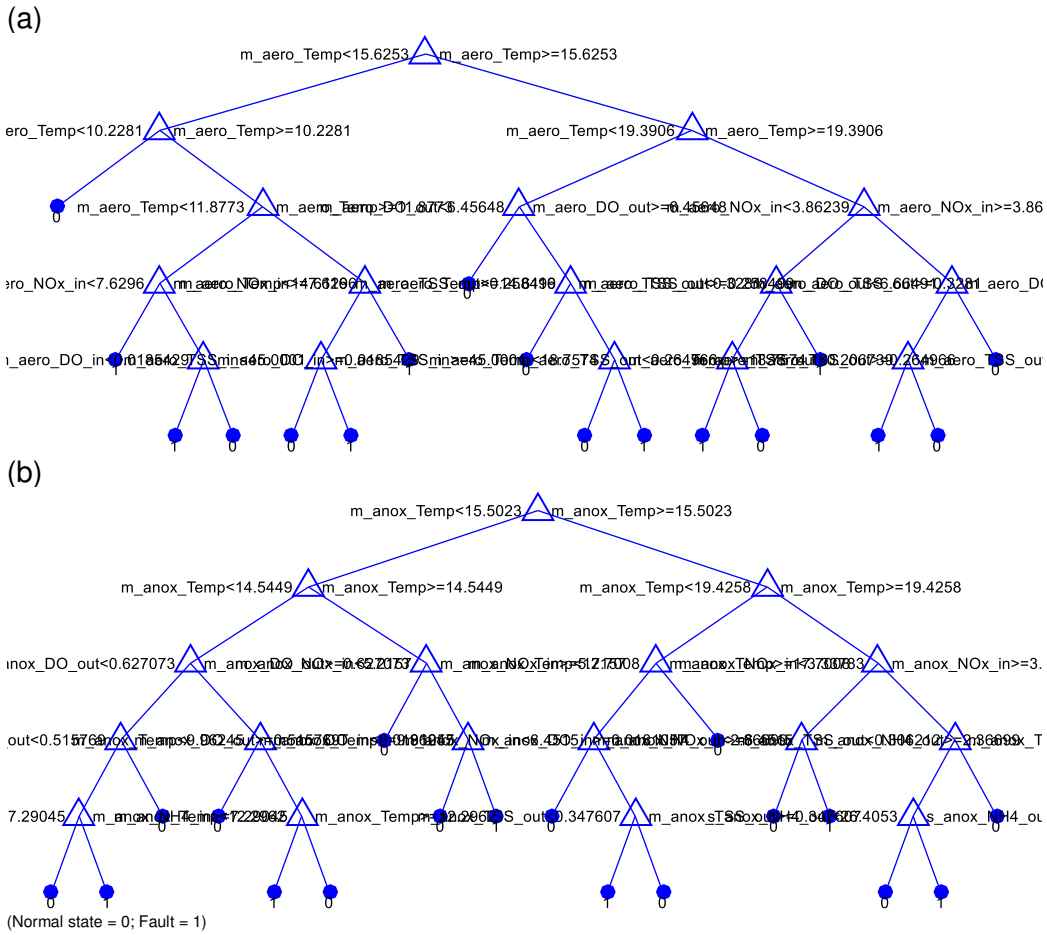


Figure 5.55 Structure of the classification trees of medium complexity trained on the calibration set described in Table 5.14 for the aerobic (a) and anoxic (b) phase using the Combination E of the diagnostic parameters

Good classification results are displayed by the trained trees on the calibration set for both the process conditions, with an extremely limited number of wrong assignments and an acceptable fault identification efficiency (Figure 5.56).

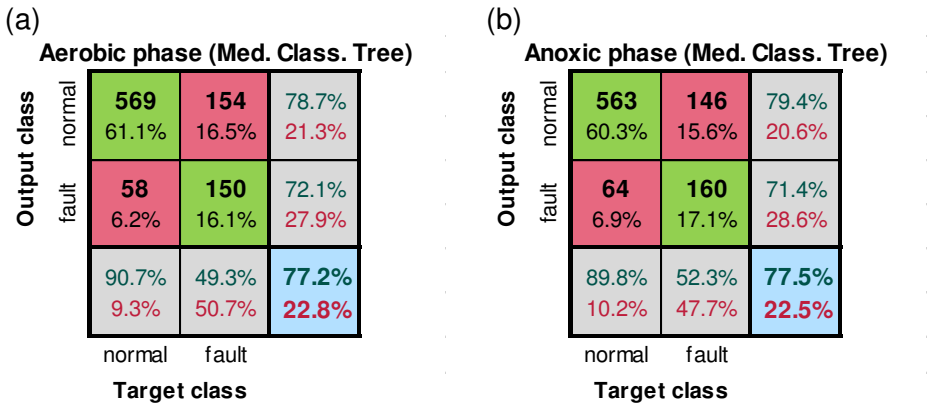


Figure 5.56 Confusion matrices describing detection performances of the finer faults for the medium binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination E

These models, however, show the tendency of overlooking the anomaly of the tested observations. Both in the calibration and in the validation set the number of instances assigned with the 'normal' class are the majority and only ¼ of them are classified as faults. The accuracy of the detection in the validation set is however higher than in the previous cases, being 53% for the aerobic phase and 40% for the anoxic.

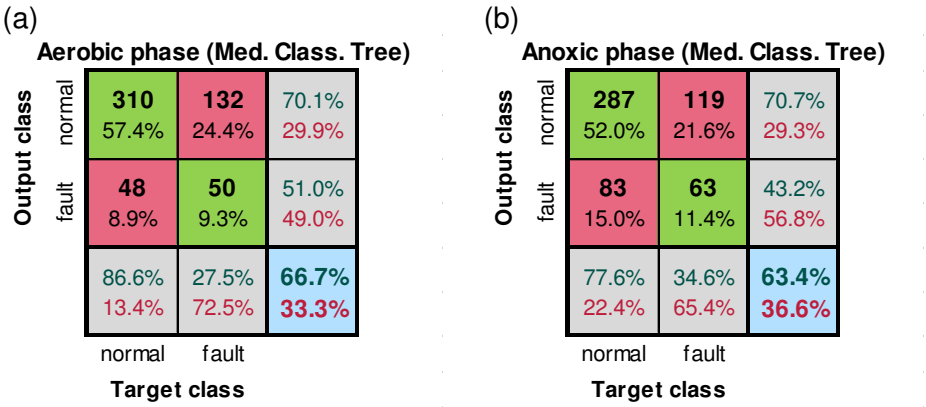


Figure 5.57 Confusion matrices describing detection performances of the finer faults for the medium binary classification tree method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination E

The comparative analysis of the classification success rates for the different combinations of diagnostic parameters presented in Figure 5.58 is almost identical to the one reported in Figure 5.45. In terms of global classification efficiency, the results on the calibration set are slightly better as the number of variables composing the dataset increases and this behaviour is basically maintained in the validation phase.

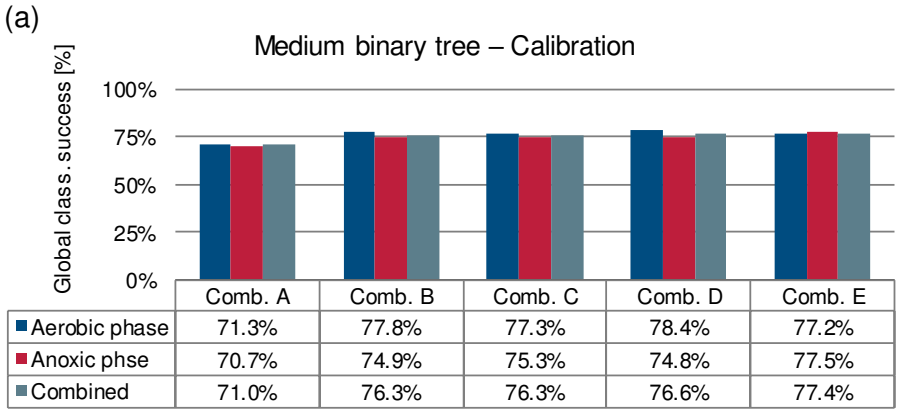
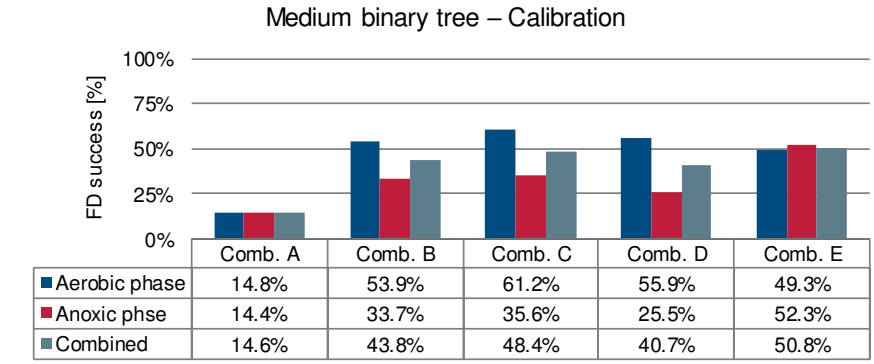


Figure 5.58 Global classification success of the medium binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

As already observed for the Combination A and Combination E, also for the other combinations of diagnostic variables the classifiers trained on the calibration set of Table 5.14 favour the 'normal' class. This naturally reflects on the generally poor rate of faults identified both in the calibration (Figure 5.59.a) and in the validation (Figure 5.59.b) phase. It is evident however that the increased complexity of the tree benefits

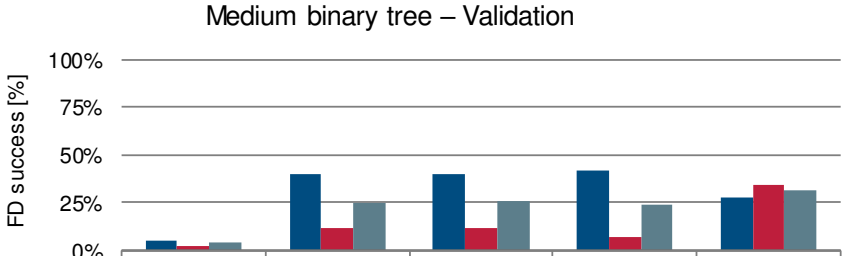
from an equally increased complexity in the number of antecedents used to represent the observations.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	14.8%	53.9%	61.2%	55.9%	49.3%
Anoxic phse	14.4%	33.7%	35.6%	25.5%	52.3%
Combined	14.6%	43.8%	48.4%	40.7%	50.8%

(b)

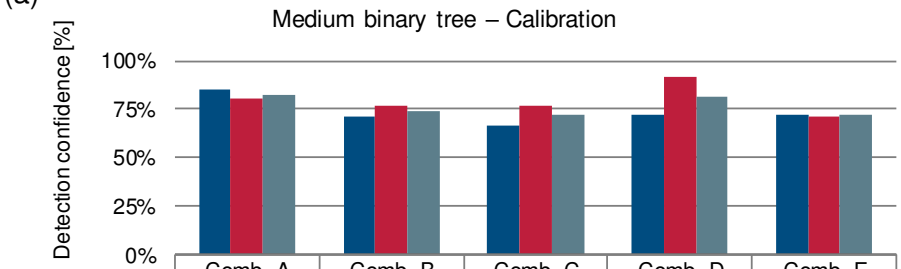


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	5.5%	39.6%	40.1%	41.8%	27.5%
Anoxic phse	2.7%	11.5%	12.1%	6.6%	34.6%
Combined	4.1%	25.4%	25.9%	24.0%	31.1%

Figure 5.59 Fault identification success of the medium binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

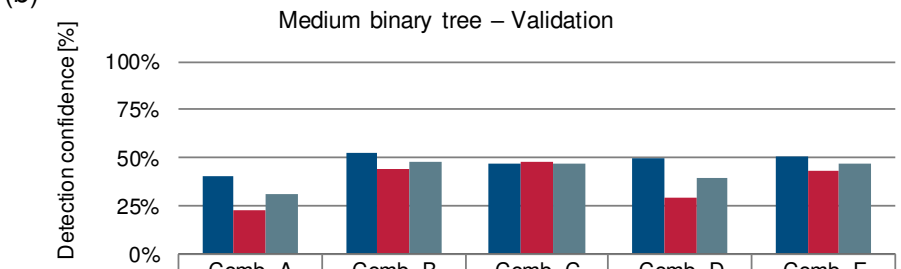
The same correspondence between the dimensionality of the descriptors and the FD performance, can be observed in the assessment of the detection accuracy of the faults (Figure 5.60). In fact, the increase in the diversity of the diagnostic parameters' combinations used is proportional to the reliability of the detection, and the combinations which include most of the observed parameters are the same that score higher both in the calibration and in the validation set.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	84.9%	71.0%	66.7%	71.7%	72.1%
Anoxic phse	80.0%	76.9%	76.8%	91.8%	71.4%
Combined	82.4%	74.0%	71.8%	81.8%	71.7%

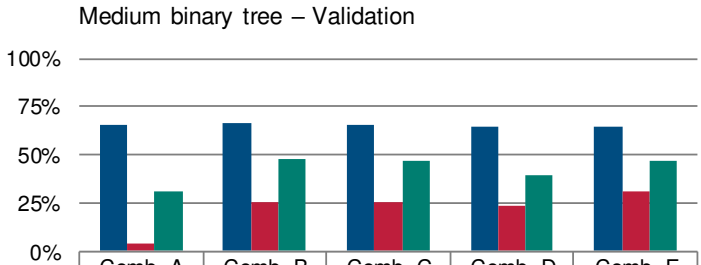
(b)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	40.0%	52.6%	46.5%	49.7%	51.0%
Anoxic phse	22.7%	43.8%	47.8%	29.3%	43.2%
Combined	31.3%	48.2%	47.2%	39.4%	47.1%

Figure 5.60 Confidence rate on the faults identified by the medium binary classification tree method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

From the histograms in Figure 5.61, which summarize the performances of the classification tree emerges that despite the low identification rate the Combination E is the one that provides better results, also in terms of general classification capability and in accuracy of the finer faults detected.

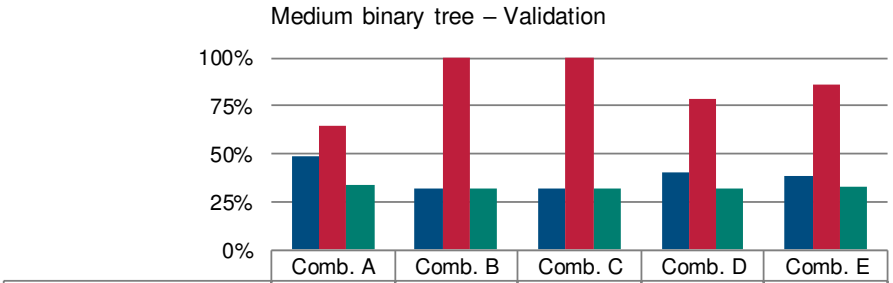


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	65.1%	66.7%	65.5%	65.0%	65.0%
FD success [%]	4.1%	25.4%	25.9%	24.0%	31.1%
Detection confidence [%]	31.3%	48.2%	47.2%	39.4%	47.1%

Figure 5.61 Comparison of the performance indexes for the medium binary classification tree method under the different combinations of parameters in the validation set of Table 5.14

Testing the method on the reduced training and validation sets of Table 5.13 as before does not bring any improvement in the detection. The models present a slightly different feature ranking than the ones in Figure 5.55, having the average concentration of suspended solids as second most important parameter instead of the DO, but also in this case the classifiers result too much tailored on the calibration data, off-balancing the detection in favour of the 'fault' class when tested with a different dataset. This is

particularly evident from the results summarized in Figure 5.62. By relaxing the limitations on the number of rules used to describe the classification process, the method produces classifiers that are extremely accurate in sorting the elements of dataset with close similarities to the ones used for the training. However, when tested on datasets with different characteristics, the models calibrated using the dataset partitioning of Table 5.13 reveal poor generalization capabilities and are more likely to classify the instances in the 'fault' category than in the 'normal' one.



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	49.2%	32.0%	32.2%	40.7%	38.9%
FD success [%]	64.5%	100.0%	100.0%	79.0%	86.2%
Detection confidence [%]	34.1%	31.8%	31.9%	32.3%	32.6%

Figure 5.62 Comparison of the performance indexes for the medium binary classification tree method under the different combinations of parameters in the validation set of Table 5.13

This more refined category of decision trees seems to be better suited for the detection of the finer faults, provided it can be trained with sufficiently well-described observation and the training set includes examples that encompass as much as possible of the variability of the process.

5.4.6 Results of the fault detection using the linear support vector machine method

The dataset defined in Table 5.14 has also been tested for the presence of finer faults using the SVM methods described in Chapter 4. The first test considered the performances of the SVM method as a linear classifier, i.e. under the assumptions that the 'fault' and 'normal' class are linearly separable in their features space. Extracting from the calibration set the Combination A of diagnostic parameters the optimal separation hyperplanes found for the aerobic and the anoxic phases have the characteristics shown in Table 5.25.

Table 5.25 Characteristics of the linear SVM model trained on the calibration set of Table 5.14 for the Combination A of the diagnostic parameters

Property	Aerobic phase	Anoxic phase
Number of training instances (N)	931	933
Number of support vectors (N_s)	682	681
Kernel	linear	linear
Bias (w_0)	-	-
$\boldsymbol{w} = [m_{\cdot,NH4+,out}$	5.226e–06	1.917e–06
$s_{\cdot,NH4+,out}$	3.231e–06	3.638e–06
$m_{\cdot,NOx-,out}$	–1.571e–05	–2.120e–05
$s_{\cdot,NOx-,out}]$	2.773e–05	5.818e–06

Using just the four parameters of Combination A the method is not able to determine a satisfying decision surface. The optimal separation planes found for the aerobic and anoxic phases, in fact, do not lead to the detection of any fault event, not even in the calibration set, as it is shown in Figure 5.63.

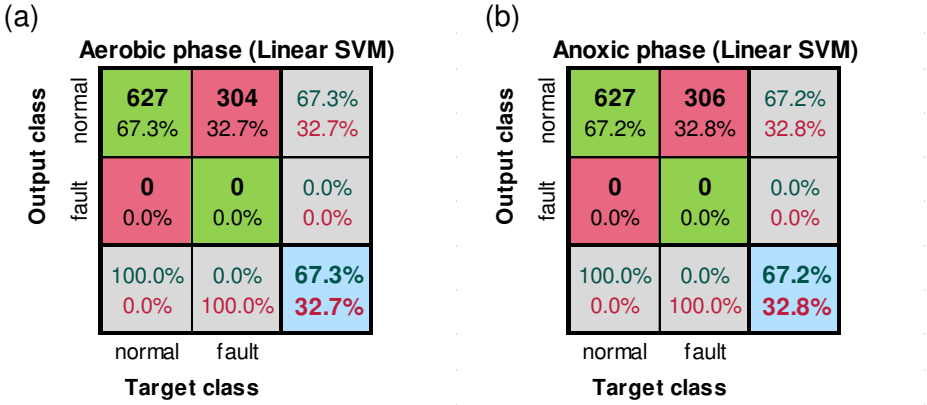


Figure 5.63 Confusion matrices describing detection performances of the finer faults for the linear SVM method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination A

The same behaviour is displayed by the confusion matrices presenting the performances in the validation set in Figure 5.64. Here too the calibrated linear classification models are unable to identify any anomaly neither in the aerobic nor in the anoxic phase.

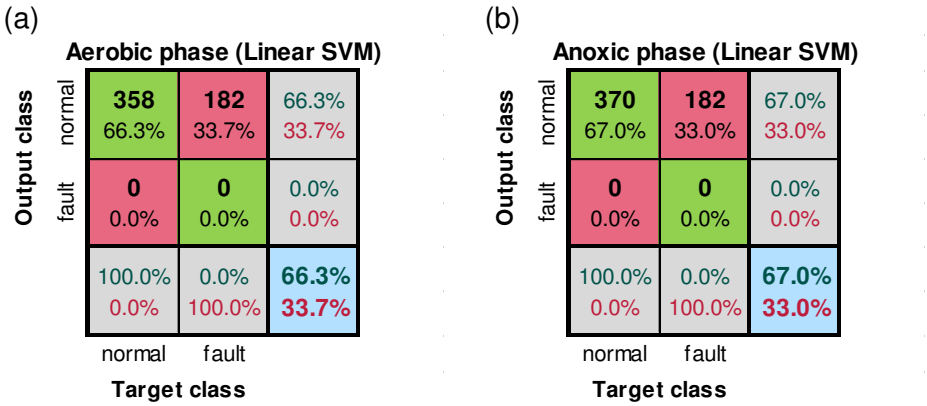


Figure 5.64 Confusion matrices describing detection performances of the finer faults for the linear SVM method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination A

The same result is obtained when a wider range of variables is used. Using the Combination E the classifiers for each process phase have the characteristics reported in Table 5.26.

Table 5.26 Characteristics of the linear SVM model trained on the calibration set of Table 5.14 for the Combination E of the diagnostic parameters

Property	Aerobic phase	Anoxic phase
Number of training instances (N)	931	933
Number of support vectors (N_s)	669	621
Kernel	linear	linear
Bias (w_0)	–1.01	–0.993
$\boldsymbol{w} = [m_{\cdot,NH4+,in}$	0.001	0.0541
$m_{\cdot,NH4+,out}$	0.001	0.0138
$s_{\cdot,NH4+,out}$	–0.001	–0.042
$m_{\cdot,NOx-,in}$	0.001	0.005
$m_{\cdot,NOx-,out}$	–0.002	–0.0128
$s_{\cdot,NOx-,out}$	–0.001	–0.010
$m_{\cdot,TSS,in}$	0.000	–0.015
$m_{\cdot,TSS,out}$	0.001	0.097
$m_{\cdot,DO,in}$	–8.999e-05	–0.001
$m_{\cdot,DO,out}$	–4.782e–06	–0.006
$m_{\cdot,Temp}]$	0.000	0.004

This time too the decision surfaces are unable to discriminate the anomalies in the aerobic phase (Figure 5.65a) and just two faults are detected in the anoxic phase.

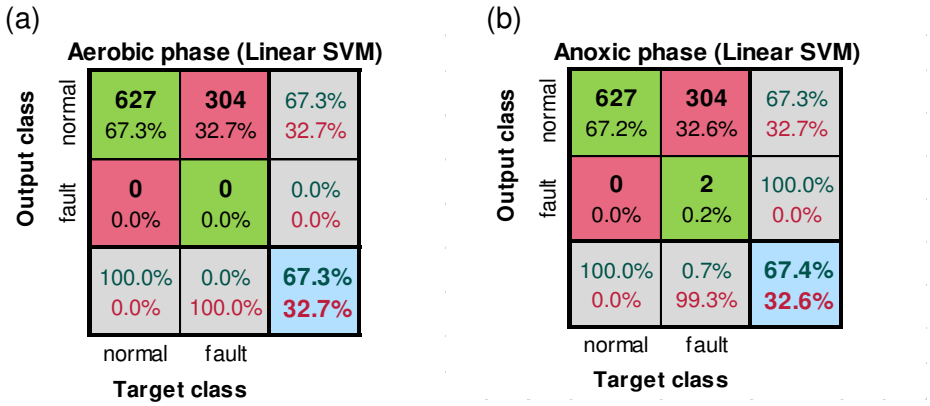


Figure 5.65 Confusion matrices describing detection performances of the finer faults for the linear SVM method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination E

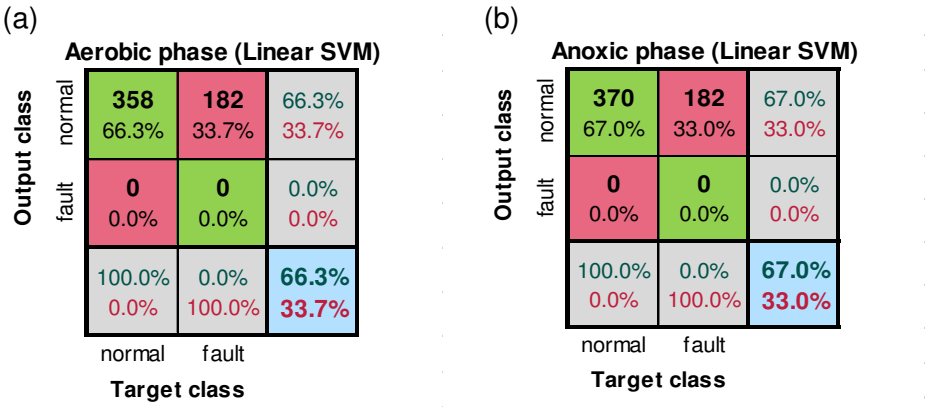
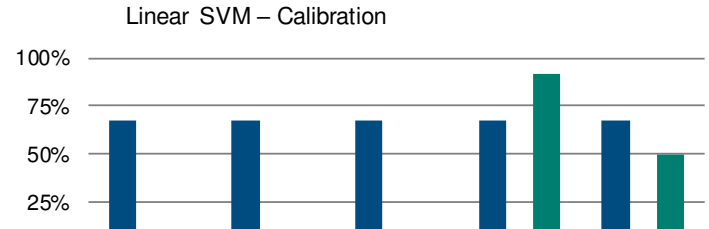


Figure 5.66 Confusion matrices describing detection performances of the finer faults for the linear SVM method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination E

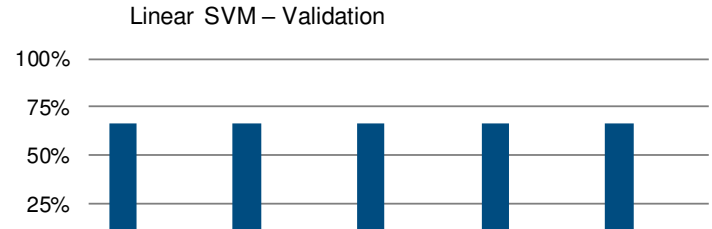
Considering the results just shown, it makes little sense to compare the performances of the method when varying the combination of descriptors by focusing on the performance indexes of interest separately as we did for the other methods. Figure 5.67 compares the indexes describing the performances of the combined aerobic and anoxic classifiers both on the calibration and on the validation phase. Of course, the high percentage global classification rates in light of the absence of instances assigned to the 'fault' class here merely assume the meaning of the frequencies of occurrence of the phases containing correct measurements in the whole dataset. Using a higher number of descriptors the method is able to recognise some faults but just in the calibration dataset.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	67.2%	67.2%	67.2%	67.7%	67.4%
FD success [%]	0.0%	0.0%	0.0%	1.6%	0.4%
Detection confidence [%]	0.0%	0.0%	0.0%	91.7%	50.1%

(b)

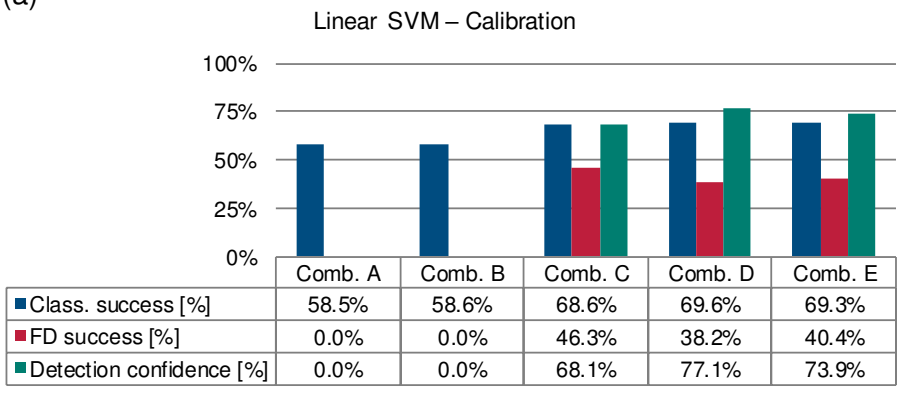


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	66.7%	66.7%	66.7%	66.3%	66.7%
FD success [%]	0.0%	0.0%	0.0%	0.0%	0.0%
Detection confidence [%]	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 5.67 Comparison of the performance indexes for the linear SVM method under the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

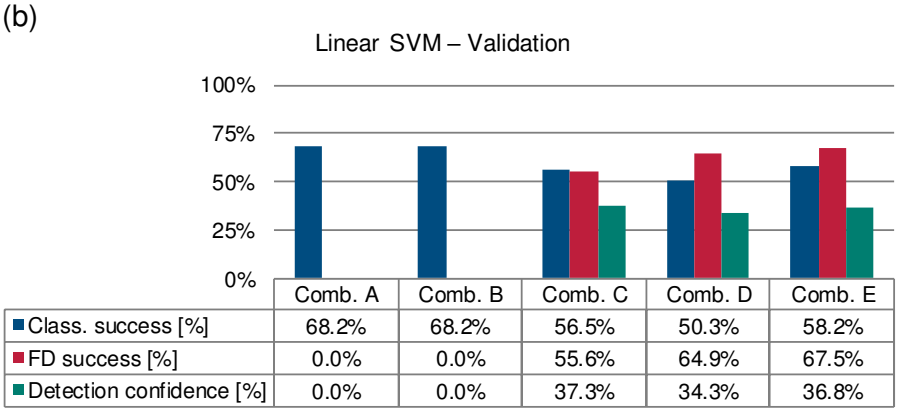
Better results can be obtained when the linear classifier is trained on the calibration dataset of Table 5.13 and a high number of features is available. From the comparison of the histograms of Figure 5.68 with those of Figure 5.67 is evident that the models trained with just the 6 months dataset are able to detect, except for Combination A and Combination B, a fair number of faults even though the accuracy of the detection, particularly high in calibration, drops below 40% in the validation set.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	58.5%	58.6%	68.6%	69.6%	69.3%
FD success [%]	0.0%	0.0%	46.3%	38.2%	40.4%
Detection confidence [%]	0.0%	0.0%	68.1%	77.1%	73.9%

(b)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	68.2%	68.2%	56.5%	50.3%	58.2%
FD success [%]	0.0%	0.0%	55.6%	64.9%	67.5%
Detection confidence [%]	0.0%	0.0%	37.3%	34.3%	36.8%

Figure 5.68 Comparison of the performance indexes for the linear SVM method under the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

It is then clear that the assumption of linearly separable classes, although it was expected, is not suited for this problem and a separation surface with good discrimination properties and generality cannot be found in the feature space.

5.4.7 Results of the fault detection using the cubic kernel support vector machine

The search for the optimal separation phase suggested to expand the feature space with a polynomial of order 3 as kernel function. The characteristics of the SVM classification models trained using the Combination A of diagnostic features on the calibration set defined in Table 5.14 are reported in Table 5.27.

Table 5.27 Characteristics of the cubic kernel SVM model trained on the calibration set of Table 5.14 for the Combination A of the diagnostic parameters

Property	Aerobic phase	Anoxic phase
Number of training instances (N)	931	933
Number of support vectors (N_s)	891	882
Kernel	Polynomial (cubic)	Polynomial (cubic)r
Bias (w_0)	–2.570	–2.715

The FD performance on the calibration set, visualised by the confusion matrices of Figure 5.69, shows that the method can identify a relatively small number of faults with little more than 40% of accuracy, which is a better result compared to the linear kernel, but still far from satisfactory. Moreover, the same conservative behaviour displayed by the linear SVM models trained on the same dataset can be observed also in this nonlinear setting, assigning more faulty instances to the 'normal' class than to the other.

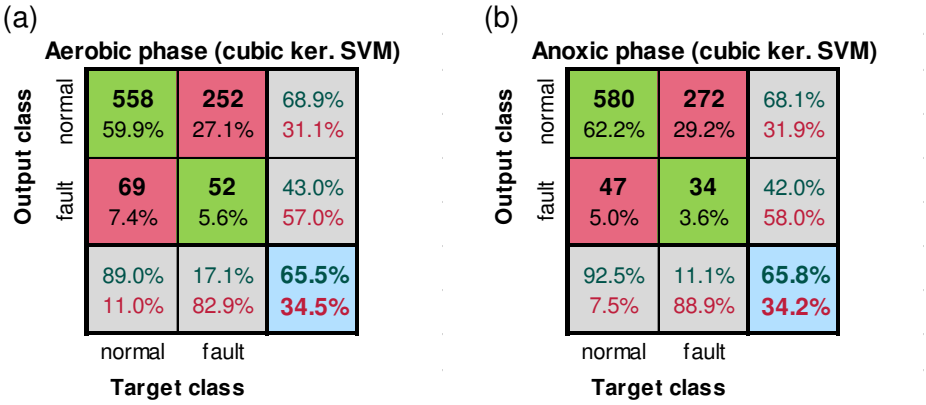


Figure 5.69 Confusion matrices describing detection performances of the finer faults for the cubic kernel SVM method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination A

The same considerations apply to the validation of the models (Figure 5.70): both the model for the aerobic and the one for the anoxic phase tend to classify the faulty instances as "normal", reducing the percentage of detected faults.

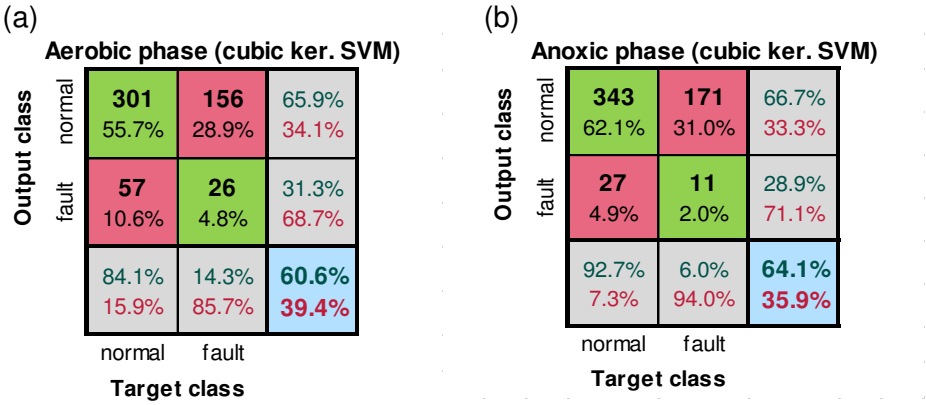


Figure 5.70 Confusion matrices describing detection performances of the finer faults for the cubic kernel SVM method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination A

The effect of a different combination of the descriptors used to represent the instances produces better results, especially in the calibration set. For example using Combination

E the optimal separation hyperplane can be determined using a lower number of support vectors (Table 5.28) and the results obtained are better than in Figure 5.70.

Table 5.28 Characteristics of the cubic kernel SVM model trained on the calibration set of Table 5.14 for the Combination E of the diagnostic parameters

Property	Aerobic phase	Anoxic phase
Number of training instances (N)	931	933
Number of support vectors (N_s)	505	509
Kernel	Polynomial (cubic)	Polynomial (cubic)r
Bias (w_0)	–0.532	–1.334

Figure 5.71 reports the results obtained on the calibration set and they are quite good. Most of the observed anomalies are detected and the accuracy exceeds 85%. The general classification efficiency is also of the same entity.

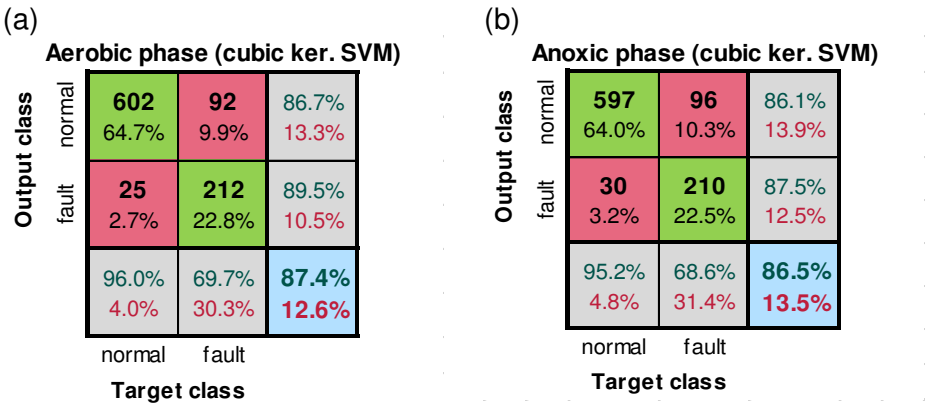


Figure 5.71 Confusion matrices describing detection performances of the finer faults for the cubic kernel SVM method in both the aerobic (a) and the anoxic phases (b) of the calibration set of Table 5.14 using the parameters of Combination E

The results on the validation set, however, show a lower rate of detected faults compared to the real ones and a lower accuracy of their identification (Figure 5.72). However, the performance of the models as general classifiers are still satisfying as they score over 60% in both the anoxic and the aerobic phase.

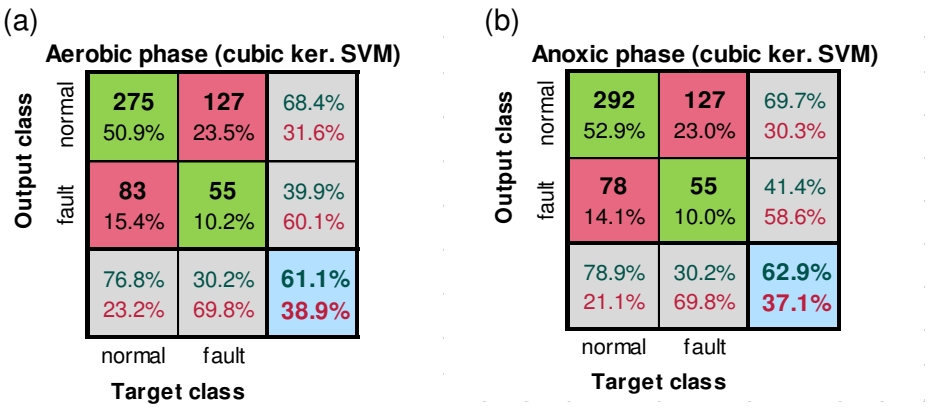
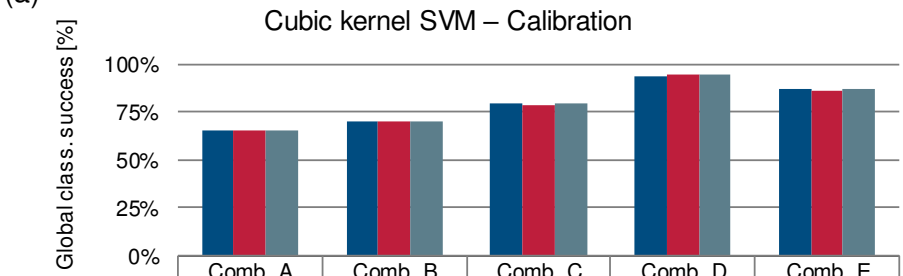


Figure 5.72 Confusion matrices describing detection performances of the finer faults for the cubic kernel SVM method in both the aerobic (a) and the anoxic phases (b) of the validation set of Table 5.14 using the parameters of Combination E

Of all the method tested, the kernel SVM is the one that better responds to the increment of available information provided by the use of a larger number of parameters by delivering higher performing models in the calibration set. As shown in Figure 5.73.a

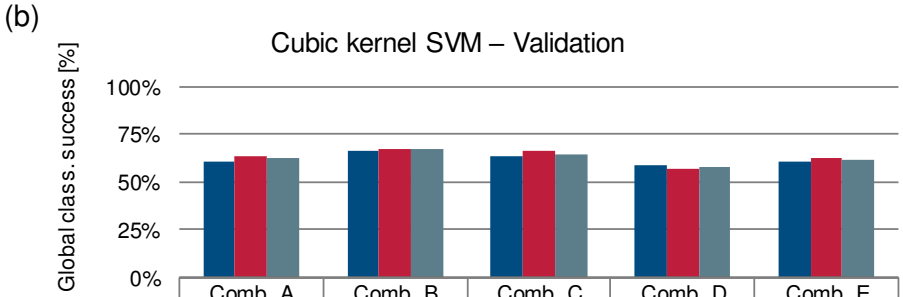
the global classification success is higher when more comprehensive descriptors sets, as Combination D and Combination E, are used. At the same times this does not translate in overfitted classifiers as in the validation sets the global classification efficiency does not drop below 50% in any of the tested combinations.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	65.5%	70.0%	80.0%	94.0%	87.4%
Anoxic phse	65.8%	70.2%	79.1%	95.2%	86.5%
Combined	65.7%	70.1%	79.5%	94.6%	86.9%

(b)

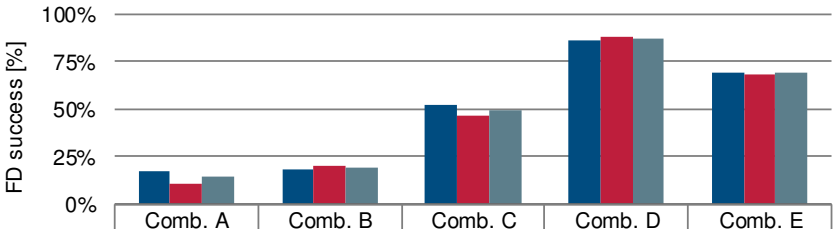


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	60.6%	66.9%	63.7%	58.5%	61.1%
Anoxic phse	64.1%	67.6%	66.3%	56.9%	62.9%
Combined	62.4%	67.3%	65.0%	57.7%	62.0%

Figure 5.73 Global classification success of the cubic kernel SVM method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

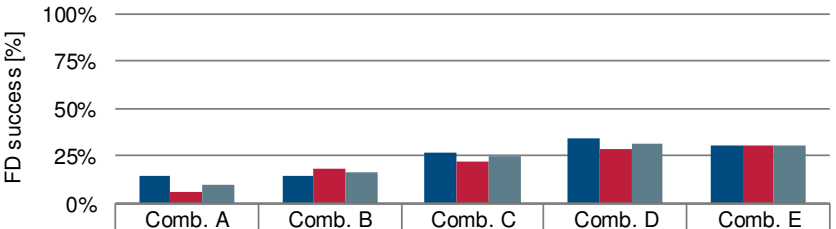
As shown in Figure 5.74, the number of detected faults is another index deeply influenced by the choice of parameters, even more than the global classification success rate. In fact, the improvement provided by the use of larger number of descriptors in both the calibration and the validation sets is considerable if compared to the results obtained with the basic configuration of features of the Combination A. None of the tested combinations, however, is able to achieve in the validation set a detection rate similar to the one obtained in the calibration, and even if with Combination D the classifiers detect over the 80% of the faults, in validation this value decreases to 30%.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	17.1%	18.4%	52.3%	86.5%	69.7%
Anoxic phse	11.1%	20.3%	46.4%	88.2%	68.6%
Combined	14.1%	19.4%	49.3%	87.4%	69.1%

(b)

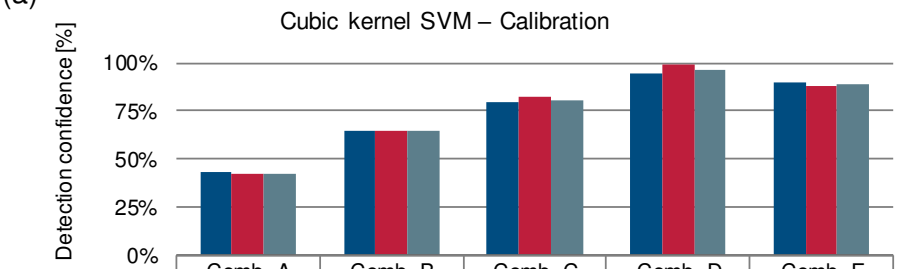


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	14.3%	14.8%	26.9%	34.1%	30.2%
Anoxic phse	6.0%	18.1%	22.5%	29.1%	30.2%
Combined	10.1%	16.5%	24.7%	31.6%	30.2%

Figure 5.74 Fault identification success of the cubic kernel SVM method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

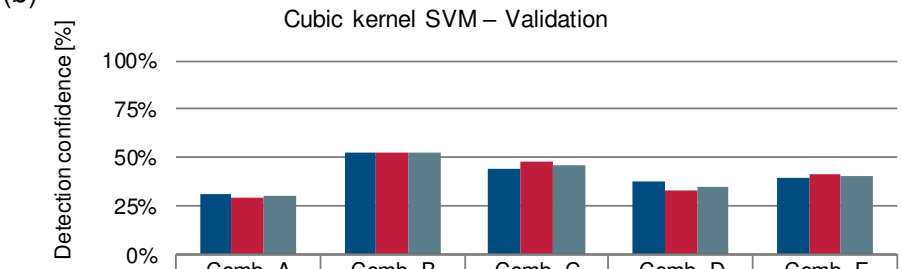
Regarding the accuracy of the detection of the anomalous events in the calibration dataset, this is also linked to the combination of diagnostic parameters (Figure 5.75). For Combination C, Combination D and Combination E, in fact the optimized classifiers return a very low number of false positives. In the validation set, instead, the accuracy, although higher than in other methods exceeds the 50% threshold only for the Combination B.

(a)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	43.0%	64.4%	79.5%	94.6%	89.5%
Anoxic phse	42.0%	64.6%	82.1%	98.8%	87.5%
Combined	42.5%	64.5%	80.8%	96.7%	88.5%

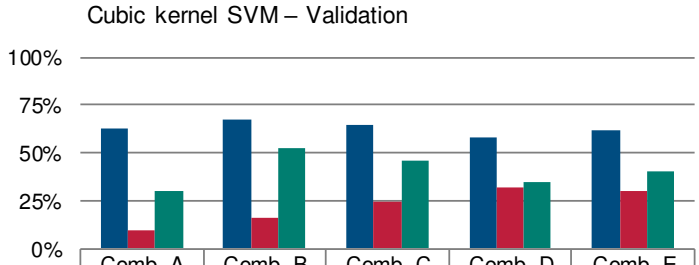
(b)



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Aerobic phase	31.3%	52.9%	43.8%	37.3%	39.9%
Anoxic phse	28.9%	52.4%	47.7%	32.7%	41.4%
Combined	30.1%	52.6%	45.8%	35.0%	40.7%

Figure 5.75 Confidence rate on the faults identified by the cubic kernel SVM method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

The aggregated results of the three performance indexes reported in Figure 5.76 are quite similar to those obtained with the medium classification tree and shown in Figure 5.61. The high performance as a classifier also in the validation set and the accuracy, generally higher with respect to the other trained methods, make this method a good and fairly reliable fault detection tool, especially if a high number of observed variables is available.

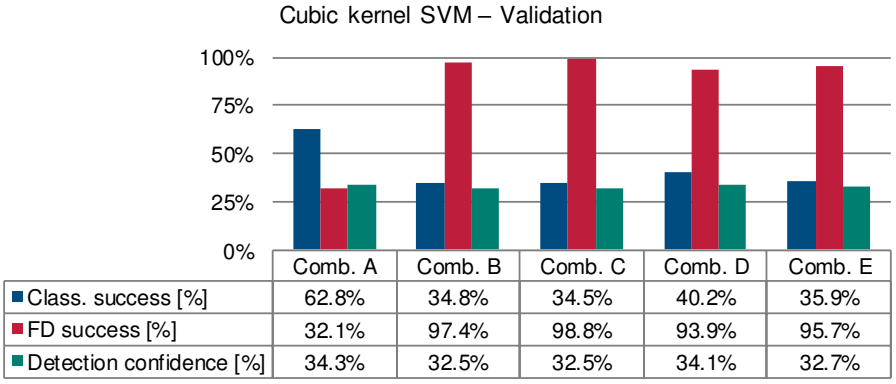


	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	62.4%	67.3%	65.0%	57.7%	62.0%
FD success [%]	10.1%	16.5%	24.7%	31.6%	30.2%
Detection confidence [%]	30.1%	52.6%	45.8%	35.0%	40.7%

Figure 5.76 Comparison of the performance indexes for the cubic kernel SVM method under the different combinations of parameters in the validation set of Table 5.14

The learning characteristics of the kernel SVM method are such that it is recommended that the training dataset contain a large number of diversified examples covering almost all the possible variations of the observed variables. The classifiers, in fact, tend to perform extremely well on the calibration set and therefore when they are

used on datasets with different characteristics it is possible that most of the instances are misclassified. We tested for example the method using the reduced calibration set of Table 5.13 achieving very good results in all the performance indexes, even higher than those reported for the calibration in Figure 5.73, Figure 5.74 and Figure 5.75. The results for the validation set shown in Figure 5.77, however, reveal the poor generalization capability of the trained classifiers. The observed high detection success rate, compared to the other indexes, must be seen as an indication that most of the samples of the validation dataset are interpreted as anomalies, since they are much different from the example the models have been trained on. Acceptable performances as general classifier are only displayed by the most basic combinations of features.



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Class. success [%]	62.8%	34.8%	34.5%	40.2%	35.9%
FD success [%]	32.1%	97.4%	98.8%	93.9%	95.7%
Detection confidence [%]	34.3%	32.5%	32.5%	34.1%	32.7%

Figure 5.77 Comparison of the performance indexes for the cubic kernel SVM method under the different combinations of parameters in the validation set of Table 5.13

Thus, if a proper and diversified training set is available the kernel SVM method can be trained using and a sufficient number or of diagnostic features and it will be a good fault detection tool, due to the relatively good degree of confidence that can be put on its detection. If, however, the dataset available is small or the example contained only represent a part of the full variability of the process, it is preferable to use a reduced number of descriptors for the instances otherwise the risk is to overfit the training set.

5.5 Comparative assessment of the of the methods

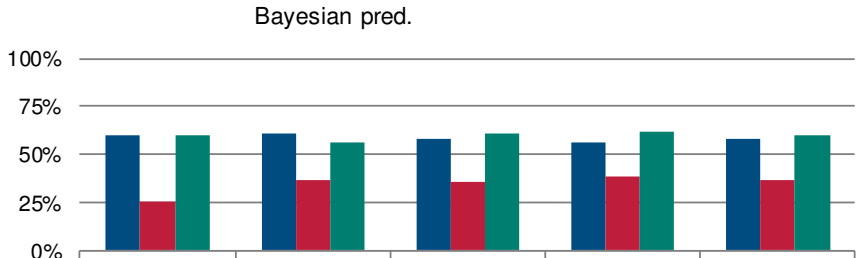
In this paragraph we provide a comparative assessment of the performances obtained by the different methods used to detect the presence of the finer faults and tested with the five combinations of diagnostic parameters. Finally, some conclusions on their reliability as FD tools is drawn.

Since the PCA-based method, the classification trees and the SVM methods tested basically share the same calibration/validation settings, in order to facilitate the comparison for each dataset partition their performances will be evaluated together.

Regarding the Bayesian method, instead, it is not possible to exactly separate the calibration and the validation part on the basis of the dataset partitioning, because in the real-time-like setting implemented each observation is first used to predict the state based on the knowledge acquired by the previous observations (so it is used to validate the current model) then contributes to update the predictive model used in the following iteration (thus becoming part of the training). For these reasons, the results for the Bayesian method will be assessed separately.

5.5.1 The naïve Bayesian predictor

The naïve Bayesian method, implemented as described in 5.2.2, has been tested on the 1 year-long dataset described in Table 5.15, which is basically the same as the calibration set of Table 5.14 used to train the other methods. The performance indexes reported in Figure 5.78, displaying the performances achieved by the method at the end of the dataset under the different combinations of parameters used to represent the observations, reveal that the method is quite robust with respect to the specific combination adopted. All the three performance indexes monitored are only marginally affected by it: with respect to the case in which the state is only described in terms of the average concentrations and the slopes of the nitrogen and the ammonia, the number of faults detected by the Bayesian predictor slightly increases as more parameters are introduced. Unfortunately, none of the combinations tested allow to identify more than 38% of the faults observed. The general prediction capabilities and the reliability of the prediction are conversely almost unaffected by the descriptors selection and the success rate is around 60% for each tested combination.



	Comb. A	Comb. B	Comb. C	Comb. D	Comb. E
Pred. success [%]	60.1%	61.3%	57.8%	56.6%	58.1%
Fault pred. success [%]	25.5%	36.4%	35.9%	38.4%	36.5%
Pred. confidence [%]	60.4%	56.0%	60.5%	61.5%	60.2%

Figure 5.78 Comparison of the performance indexes for the Bayesian method using the different combinations of parameters for the dataset of Table 5.15

The statistics on the synthetic dataset are not as good as those shown in Figure 4.17 for the real measurements and are generally worse than those achieved in the calibration phase with most of the other methods, especially regarding the number of identified finer faults, however the robustness of the prediction accuracy with respect to the

variables used to describe the process, together with the capability of improving the prediction at each step, make this method a valuable FD instrument.

5.5.2 Comparison of the results for the different classifiers

A first general assessment of the goodness of the tested methods can be achieved by comparing their global classification success rate, i.e. the portion of instances correctly assigned to their right class. This is a complementary measure of the global classification error, the cost minimized by the training of the models. We recall from Table 5.14 that of the 1864 instances composing the combined aerobic/anoxic training dataset most of them (~67%) represent normal phases and only 610 contain a finer fault. Then any method that assigns the instances only to the normal class would still have the 67% of global classification efficiency. As seen in the previous paragraphs that is what happened with the RT-MWPCA method using Combination B and Combination C and with the linear SVM method in Combination A. Combination B and Combination C, where the models, even in calibration phase, failed to find an optimal solution that allow them to discriminate the samples in two different classes instead of one. As can be observed from the histograms shown in Figure 5.79.a, the only calibrated models able to reach a higher classification efficiency are the medium size classification tree and the polynomial kernel SVM (and the simple tree, but the improvement is marginal) but only when a detailed set of descriptors for the instances is provided.

The global classification success in the validation set (Figure 5.79.b) is instead very similar for all the methods and the combinations tested, generally scoring more than 60%.

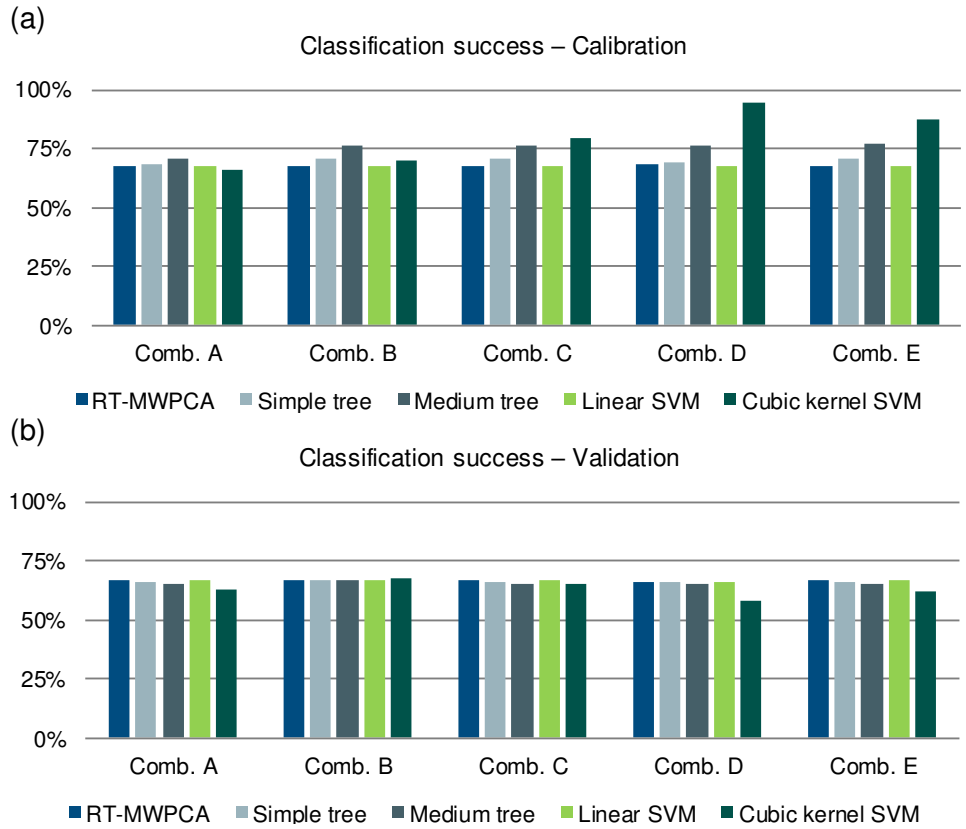


Figure 5.79 Global classification success of the cubic kernel SVM method for the five combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

Figure 5.80 shows the global classification success in case the methods are trained and tested on smaller and less diversified datasets as those of Table 5.13. High classification rates can be achieved in calibration also in this case, especially using Combination D and Combination E, but the drop of the performance in the validation set is dramatic. Except for the RT-MWPCA and the linear SVM methods, all the others produce a general classification error higher than 50%.

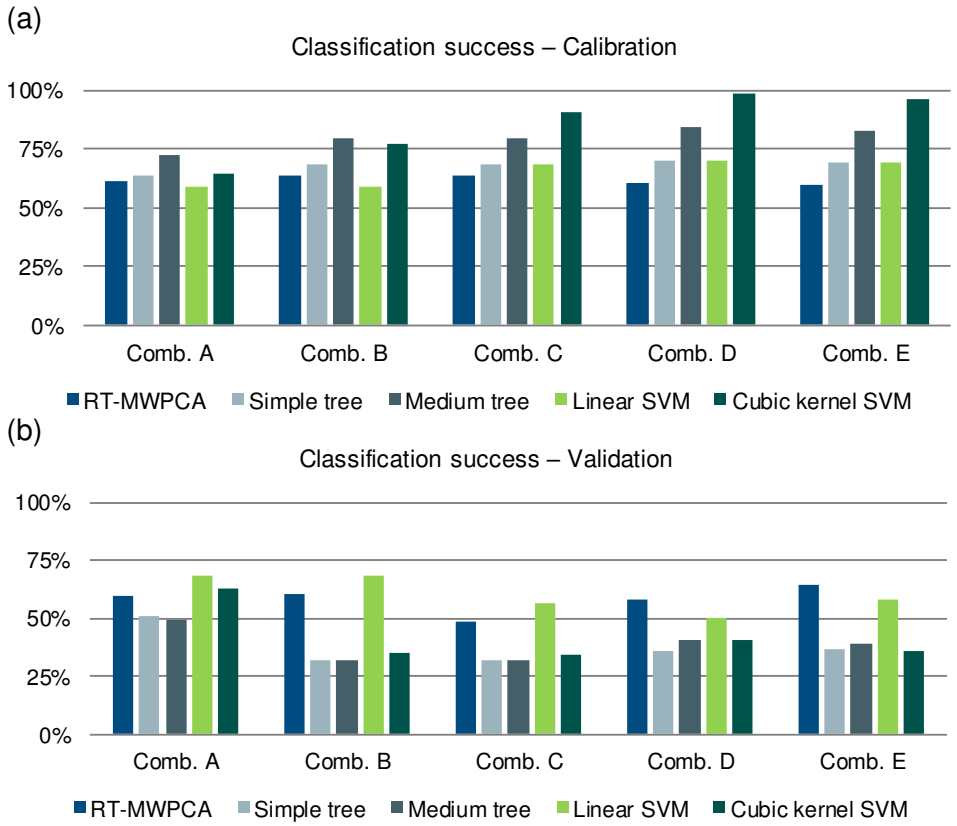


Figure 5.80 Global classification success of the cubic kernel SVM method for the five combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

The general good results observed in terms of global classification success, though, must be weighed considering of the number of faults detected by the method and the accuracy that can be reached in their detection. As displayed in Figure 5.81 the RT-MWPCA or the linear SVM method are able to detect only a very small part (if any) of the observed anomalies. Such poor results are surprising, especially for the calibration of the RT-MWPCA method, since when it was trained on the real measurements successful detection was achieved in more than 30% of the faults and with a fairly limited number of false alarms (Figure 4.13). Also, the method trained with the dataset of Table 5.13 showed better detection rates (Figure 5.34). The only methods presenting good identification performances in the training set are the classification tree of medium complexity (but only using Combination C and Combination E it is able to identify almost half of the anomalies), and the cubic kernel SVM, which with Combination C, Combination D and Combination E detects more than 50% of the faults observed in the training set. Similar, but lower, detection performances are observed in the validation set, where both methods score between 20% and 30% using the more refined combinations of the diagnostic parameters.

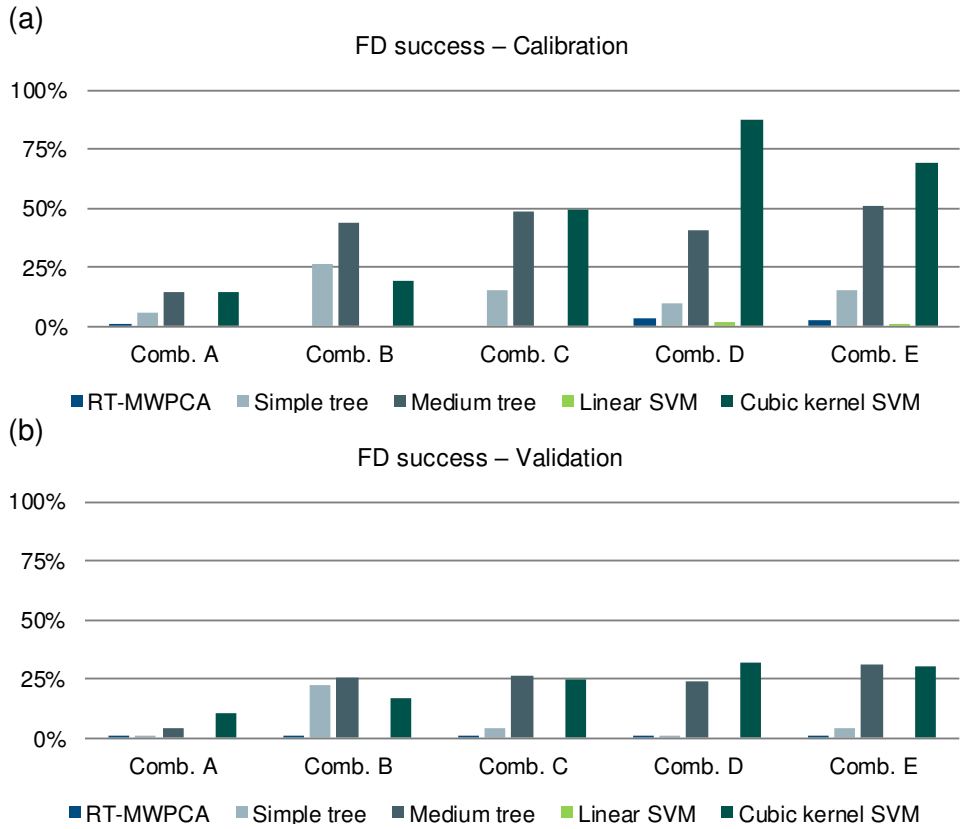


Figure 5.81 Fault identification success of the cubic kernel SVM method for the five combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

Higher rates of detection, both in training and in validation, are observed when the models are calibrated on a less diversified dataset as the one in Table 5.13. In this case, however, the higher number of faults identified in the validation set (see Figure 5.82) is biased by a generally off-balanced detection, where the trained models tend to assign to the 'fault' category the instances of the validation set since they are much different from the ones they have been trained on.

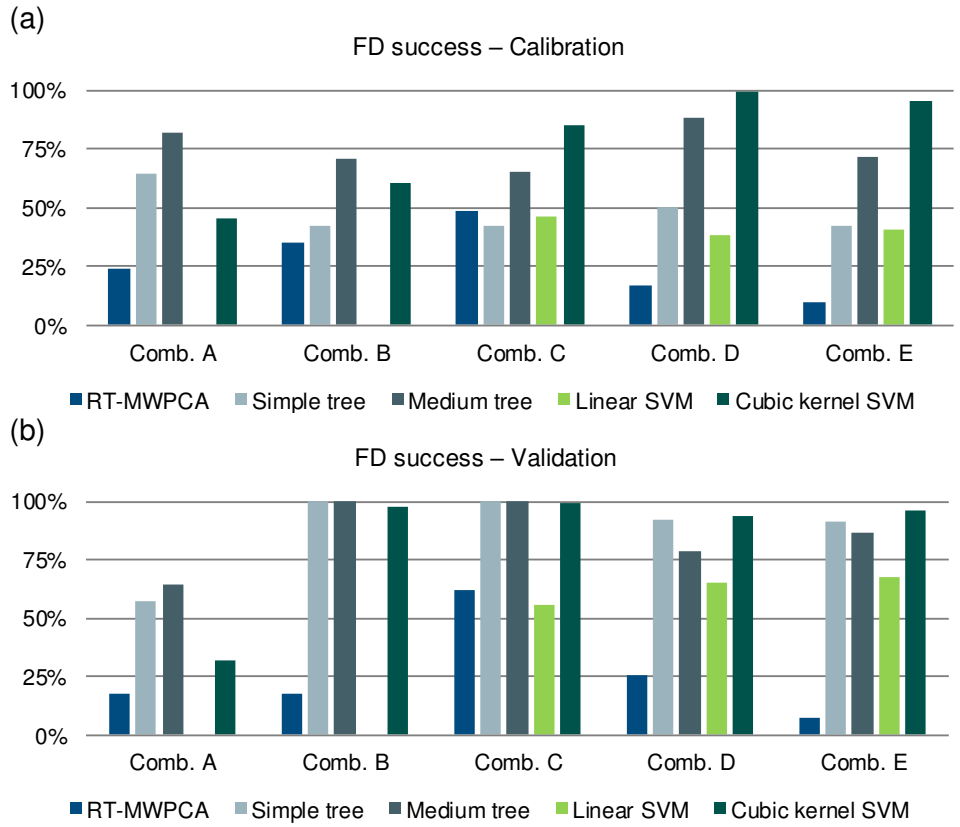


Figure 5.82 Fault identification success of the cubic kernel SVM method for the five combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

The accuracy of the detection is a key aspect to consider when choosing the most appropriate method to implement. The results in terms of confidence in the detection of fault events observed for the dataset partitioning of Table 5.14 are shown in Figure 5.83.

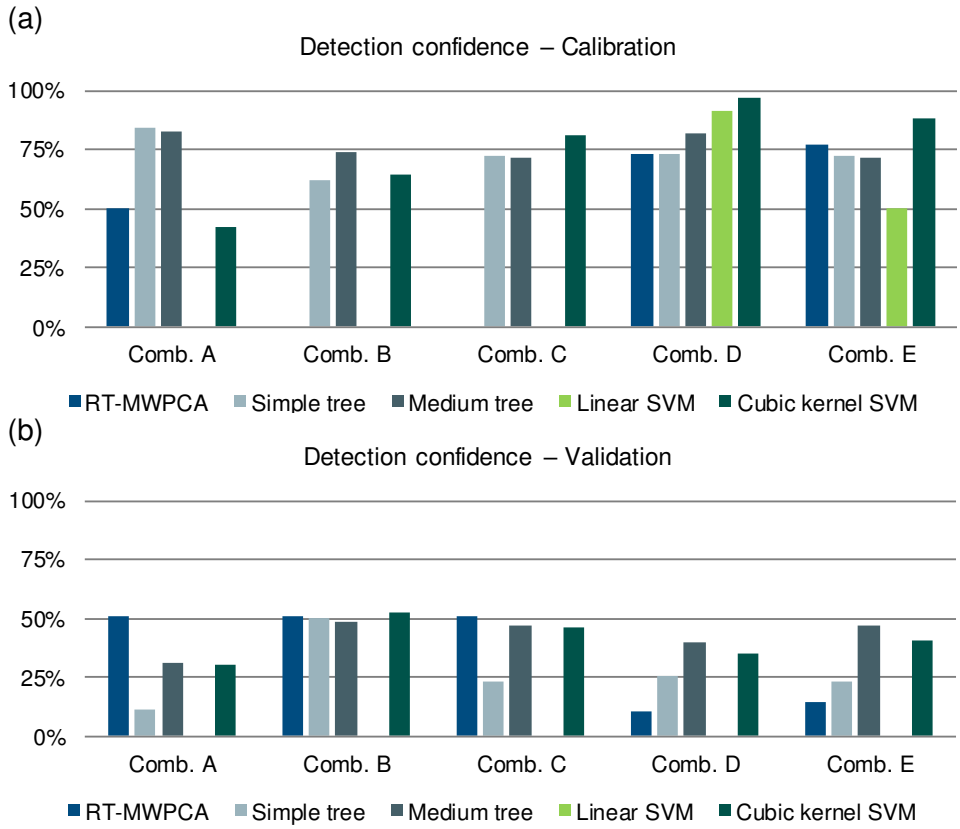


Figure 5.83 Confidence rate on the faults identified by the cubic kernel SVM method for the five combinations of parameters in the calibration (a) and validation (b) set of Table 5.14

It can be observed that the reliability of the fault detection in the calibration set is above 50% for most of the methods tested and the wider is the set of descriptors, the more reliable is the detection. This is especially true for the medium classification tree and the polynomial kernel SVM method. The latter, for example, is able to achieve more than 80% of accuracy when Combination C, D or E are used. On the validation set the performances are quite different and the confidence that the detected faults are actual anomalies generally becomes less than half of the training (Figure 5.83.b). The RT-MWPCA method is the one showing the higher accuracy (little more than 50%) when a reduced number of variables are used (Combination A, B, C) but the performance with the more complex set of descriptors is worse. The classification tree of medium complexity and the cubic kernel SVM methods instead have similar performances on the same combinations, while the medium tree performs slightly better with a high number of descriptors and the SVM achieves the highest accuracy of all the other methods and combinations when Combination B is selected. The usual comparison with the performances achieved using the dataset partitioning described in Table 5.13 displays that in the calibration phase all the methods have a higher degree of reliability (Figure 5.84.a) but in the validation phase none of the methods tested is able to achieve an accuracy higher than 30%, revealing the lack of generalization capability of the models trained with a less diversified dataset.

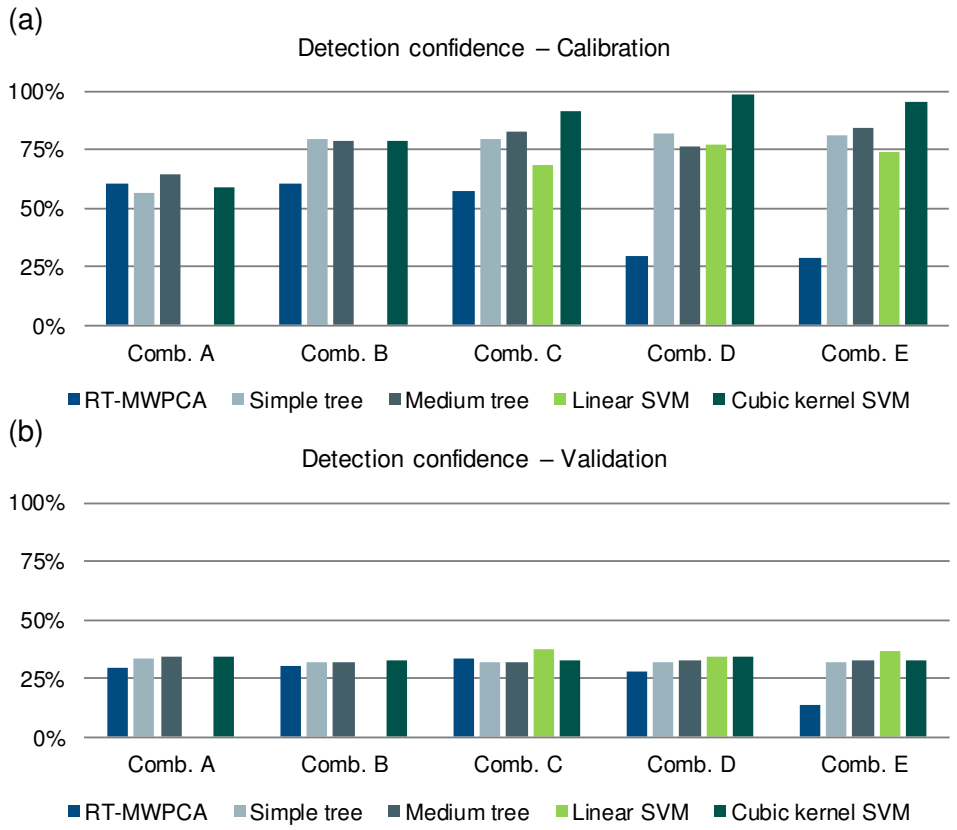


Figure 5.84 Confidence rate on the faults identified by the cubic kernel SVM method for the different combinations of parameters in the calibration (a) and validation (b) set of Table 5.13

Chapter 6 Conclusions

The fault detection algorithms investigated in this research have been conceived to detect the possible faults of the instruments measuring the nitrate and ammonia concentrations in alternate cycles wastewater treatment plants. In this process configuration the nitrogen removal is obtained through alternate oxic and anoxic cycles by switching on and off the aeration in the oxidation tank. In this way the ammonium-N is oxidized to nitrate-N in the aerobic phase, while in the anoxic phase this is reduced to dinitrogen, provided that enough organic carbon is available. The fault detection (FD) algorithms are designed to operate as a real-time tool operating under the supervision of the plant supervisory control and data acquisition (SCADA) system. The future role of this component is to enable the plant manager to adapt the control strategies to compensate for the possible loss in quality of the treatment or cost efficiency and enact the necessary countermeasures, whenever a sensor fault is detected. The choice of the optimal remedial action according to the different anomalies in the measurements quality is however beyond the scope of this research.

The FD procedure proposed in this study is composed of two hierarchically organized phases: a preliminary screening of gross failures and a further, more refined search of finer faults. The first part consists of some low-level controls performed on the raw signals, in order to detect gross malfunctions, like signals data interruptions, anomalous constant measurements and irregular duration of the phases. This preliminary screening is performed as soon as the measurements are acquired and provide an immediate alert to the plant manager. In the second phase, more sophisticated detection methods are implemented to identify finer faults that cannot be detected by the simple screening of the first phase. This class of anomalies may be caused by the presence of faults that cause the measurements to diverge from their true value. The progressive drift introduced by these malfunctions cannot be easily discriminated from the natural variability of the process by analysing one sample at the time. Considering the particular configuration of the treatment process based on alternate cycles, the second phase was split in two parts: one for the aerobic phase and one for the anoxic phase.

A preliminary operation is however the parametrization of the fault, as described in Chapter 3, based on the choice of a number of process variables, selected as fault indicators. The influence of the process variables on the performance of the FD algorithms have been assessed in Chapter 5.

In order to quantify the relationships among the variables for each investigated phase the portions of signals of the observed quantities are parametrized and a number

diagnostic features are extracted. It this context, it is clear that the response to a finer fault can only be obtained when the phase ends and its parameters have been extracted. Moreover, it should be underlined that although sharing the same principles, the FD methods applied to the different process phases have been trained separately minimizing the global classification error (number of misclassified instances over the whole dataset).

The algorithms have been tested in Chapter 5 using different measurement datasets and applying different parametrizations to the signals. The operational data used in the present work were provided by the plant managing society of the municipal treatment facility of Mantua, in Italy, and consisted of a nearly nine month-long set of measurements of ammonia and nitrates concentrations measured by the ion-specific probe installed in one of the four reactors composing the WWTP. This dataset, more extensively discussed in section 3.2, is composed by the concentrations values of the nitrates and ammonia in the aerobic reactor (and by the activation/deactivation Boolean signal of the aeration), thus four parameters are extracted: the ammonia and nitrate concentrations rate of change and their average values over each phase.

Due to the fairly limited information that could be extracted from this set of measurements and a relatively poor characterization of the fault events occurred, a further set of synthetic data was then generated using the numerical AC model described in section 3.3. Based on the standard Benchmark simulation protocol and an ASM3-like kinetics, this numerical model included an improved nitrogen kinetics including the seasonal temperature variation. To this end the model has been fed with the synthetic input data of the BSM_LT protocol. Detailed sensor models were also included, so that the occurrence of finer faults could be totally controlled, reproducing the combined effect of the main anomalies observed in the operational data: measurements noise, drift, and fouling. The occurrence of the ‘gross faults’, instead, was not simulated. The numerical model provided a broader spectrum of possible observed variables but, since it was intended as a surrogate of the real plant, we chose to use only those related to the nitrification/denitrification process and having an actual physical meaning, i.e. the concentrations of suspended solids, dissolved oxygen, the temperature, in addition to the concentration of nitrate and ammonia. This extended choice expanded therefore the set of features that can be extracted for each phase for FD purposes. Five combinations of diagnostic parameters, one of which containing the same four features available from the real plant dataset, have been used to parametrize the synthetic dataset. These combinations are intended to reflect the possible different availability of observed variable, and therefore of parameters used to describe the processes.

The performance of the algorithms has then been assessed by comparing the anomalies detected with those actually observed, either in case of real operational data, or produced by the numerical model, and using the different combinations of diagnostic features available. The FD problem has been treated either in terms of classification problem, testing different algorithms based on classification techniques such as binary trees, support vector machines (SVM) and principal component analysis (PCA), or as a forecasting problem, using an approach based on the Bayesian theory to predict the faulty or normal state of the process based on the previous records.

In the PCA-based algorithm, the classification method was improved by redefining new variable thresholds based on the classical statistical indicators, like the Hotelling's T^2, and a moving window of fixed length to keep track of the normal variability of WWTP operation due to seasonality or varying loading conditions in a real-time setting. In this context, the moving window is used to overcome the limitations of the static PCA method by defining an adaptive reference model of the good behaved observations. In the real-time-moving window PCA method (RT-MWPCA) the parameters extracted are projected onto this reference space and the Hotelling's T^2 and Q statistics are used to discriminate the phases with an irregular behaviour: if the tested parameters produce scores that are greater than the newly defined thresholds in both statistics the corresponding phase is reported as fault, otherwise it is considered "normal" and its parameters are used to update de reference space.

The Bayesian method also is intrinsically an adaptive one. Here the occurrence of a fault in the tested phase is predicted at each iteration based on the previous history of the system. In other words, by means of the Bayes' theorem the conditional probability of experiencing, or not, a fault given the current parameters is estimated at each iteration, and this is a function of which value of parameters were associated to the fault events in the previously tested sample and which to the normal state. Thus, contrary to the RT-MWPCA method, here all the previous history contributes to the definition of the fault model and not just the finite number of samples included in the last window. Moreover, the Bayesian method actually consists of two predictors, one assessing the probability of having a fault and the other the probability of not having one, and the outcome is assigned to the most probable of the two.

As for the classification trees, we decided to test two versions of the method. First, by imposing a strong limitation to the number of nodes, a simple tree was tested. Then this limitation has been relaxed letting the method to learn an increased number of rules.

Also for the SVM two different approaches have been tested. In one, under the strong assumption of linearly separable classes a simple linear kernel SVM has been trained,

to separate the normal and faulty instances on the space spanned by the features. The other approach consisted in searching the optimal separation surface in a more complex space defined by a polynomial transformation of order three of the features.

The tests on the nine-month time series of the real plant measurements revealed excellent performances for the pre-processing screenings that are able to identify the majority of the malfunctions observed in the analysed period. Only a small part (70) of the anomalies occurred in the analysed period belong to the 'finer fault' category and their detection has been tested using both the Bayesian predictor and RT-MWPCA method, although for the latter it has only been possible to perform the calibration, due to the limited quantity of faults observed. The performances achieved by the two methods are very similar, both in terms of general classification success (number of correctly assigned instances on the total number of observations) and accuracy of the prediction limited to the finer faults (share of correctly identified faults on the whole number of instances classified as fault by the method). The number of faults identified/predicted by the Bayesian method, however, is higher and it succeeds in identifying most of the observed anomalies.

A more than one year and a half-long set of synthetic measurements has been generated by a numerical model based on the ASM3 kinetics, from which several differing sets of process variables were used to derive five possible combinations of diagnostic parameters. It contains a small number of 'gross fault' which are promptly and effectively detected by the preliminary checks devised. For the assessment of the detection of finer faults two different datasets were used, the first consisting in a one year-long calibration set and an almost nine-month long validation set, the other obtained by splitting one year of data into a six-month training set and an equally long validation set. Each dataset partitioning has been tested under all the five possible combinations of the diagnostic parameters. We chose to use these two dataset partitioning strategies in order to better investigate the responses of the tested methods either in case a limited and poorly diversified dataset is available (the six-month long calibration set) and in the opposite case in which it is possible to use a longer set of measurements, encompassing all the seasonal variability of the variables. This partitioning does not apply of course to the Bayesian predictor since for this method the calibration and validation phase intermittently switch at each iteration. The resulting performances of the Bayesian method tested on one year of synthetic data are essentially independent from the particular choice of descriptors used and, despite the number of observed faults does not exceed 30% for any of the combinations, both the global classification success rate and the reliability of the fault detection are around 60%. These

results are not as high as those achieved in the real plant dataset but still the robustness of the prediction accuracy displayed varying the variables used to describe the process, combined with the possibility to improve the prediction model at each step, make this method a valuable FD instrument also in this dataset.

The other methods provided mixed results. Overall, the calibrated methods produced high performing classifiers for the training dataset, with high performance indexes for almost all the combinations of diagnostic parameters. The only exception is represented by the linear SVM classifiers, which was unable to detect any fault for most of the combinations. Conversely the cubic kernel SVM and the more complex classification tree are the methods that scored higher, especially when the observations included a high number of parameters. It is interesting to notice that the calibration results on the six-months training set are usually better but this can be explained considering that the data are less diversified, therefore more easily separable.

Opposite results are obtained in validation. The methods that are trained on the short-period calibration set suffer from the scarce variability of the data and display poor generalization capabilities. Except for the RT-MWPCA method, which in calibration had the worst performances, all other classification techniques tend to label as faulty most of the instances of the validation, which makes sense since they have been trained on examples having very different characteristics. The lack of generalization is especially confirmed by the high misclassification rate achieved with the short parametrizations and by the generally low accuracy of the fault detections. Conversely, the classifiers trained on the longer dataset, when tested on the instances from the validation set, report comparatively higher scores in terms of accuracy and general classification success, despite the detection performances is between 20% and 30% for the most performing methods (which are also in this case the medium classification tree and the cubic kernel SVM, but only in case more populated combinations of diagnostic features can be extracted from the signals). The RT-MWPCA, among all methods, is the one showing the higher accuracy when a reduced number of variables are used but the performances with the more complex set of descriptors are worse. The classification tree of medium complexity and the cubic kernel SVM methods, instead, have similar accuracy if tested on the same combinations, with the medium tree performing slightly better with a high number of descriptors but the number of faults identified by the SVM in this case is higher.

These considerations lead to the conclusion that the optimal choice of an FD method depends on the characteristics of the training dataset and the number of observed variables available. If the historical data series used to train the models cover a short

time period and it is known to be only partially representative of the whole seasonal variability of the investigated process then the RT-MWPCA is preferable, since thanks to its built-in adaptive framework it is more able to keep track of the changing loading and environmental conditions than the other methods. On the contrary, when an exhaustive amount of data is available and the number of monitored variables is large enough to allow fora higher number of diagnostic features, the nonlinear SVM method prevails as it exploits the increased information available and as a consequence has better generalization capability. However, despite a low number of detected faults, the Bayesian predictor is the tested method that returned the higher accuracy of the detection and a substantial independence of its performance from the parametrization chosen.

6.1 Key aspects of this research

The malfunction of the measuring instruments can introduce disturbances in the signals that influence the control of the process, possibly deteriorating the efficiency of the treatment. A complete real-time fault detection algorithm is then provided, with characteristics specifically designed for the application to a wastewater treatment plant employing the alternate aeration process configuration to improve the nitrogen removal efficiency. The algorithm takes into consideration the various malfunctions that can affect the monitored signals, dealing with the detection of the most gross malfunctions testing each measurement as soon as it is acquired and on a larger time scale investigating the presence of the finer faults in each process phase as soon as it is concluded. In this context, a parametrization of the portions of signals in each process phase is performed to extract meaningful diagnostic parameters.

Several fault detection algorithms were devised and tested either using the measurements of a medium-sized municipal plant in the north of Italy and a more exhaustive synthetic dataset generated by a numerical model of an alternated cycles plant based on the benchmark simulation model protocol. The latter included fairly detailed models of the sensors and their disturbances. Mainly developed in the Matlab environment, at the earlier stage of the research the algorithm was ported in LabView for its application on the treatment plant in Mantua from which the field dataset was obtained.

Using different parametrizations for the two the process phases and different partitioning of the available dataset, the feasibility of the different methods was investigated, with special attention to the detection of the finer faults. The approaches to the fault detection task tested in this work on one hand treated the problem in terms of a classification problem, comparing a method based on a technique traditionally

employed for the fault detection in WWTPs such as principal component analysis (PCA) with other classification techniques less popular in this field such as binary trees and support vector machines (SVM), on the other hand the fault detection was tackled as a forecasting problem, using an approach based on the Bayes' theory to predict the faulty or normal state of the process based on the previous records.

The main findings and contributions of this research are listed below:

- The main contribution of this thesis lies in the proposal of a framework for the fault detection on wastewater treatment plants where the problem of the detectability of the gross and most evident faults, such as spikes and constant signals, is separated from the one related to more serious and persistent malfunctions of the measurement instruments. This division was motivated by the fact that the first class of faults is generally easier to promptly detect using simpler screening tools, while the second class of anomalies, whose occurrence is more likely to produce long-lasting effect on the quality of the treatment process, is more difficult to identify in a dynamic system, especially in an AC wastewater treatment plant, therefore more sophisticated monitoring methods are required;
- The results showed a very high performance of the methods investigating the presence of the first class of faults, which were able to identify the majority of the observed anomalies;
- As for the detection of the finer faults, the PCA-based algorithm showed a good compromise between fault detection performance and reliability of the identification when trained with a relatively short and moderately diversified dataset and when a small number of observed variable is provided. Conversely, when a larger set of instances covering all the seasonal variability of the process and a relatively higher number of signals are available for the training, the use of a nonlinear support vector machine emerged as the most performing choice;
- On the other hand, the fault detection provided by the Bayesian predictor, although generally identifying a lower number of fault events, is the method that provides the highest accuracy for the two datasets and under all the different combinations of diagnostic parameters.

6.2 Outlook

Several insights for further research are provided by the results and the experience gained during this work. The different machine learning techniques applied to the identification of the finer faults under different conditions of data availability and

parametrization strategies revealed that none of them, except perhaps the Bayesian predictor, can be considered the optimal method in absolute terms. An interesting further development could be represented by the testing of a FD algorithm involving different and independently trained methods working at the same time and each reporting their estimation on the state of the system. A polling system could then determine with some criteria the final classification of the tested phase.

From the operational point of view an extensive testing of the methods devised on one or more real treatment plants is required, together with the development of a framework for the definition of the paradigm of the fault events. Especially in the historical series used for the initial training of the algorithms this aspect was in fact one of the main reasons of the poor results achieved using the real measurements dataset: no laboratory concentration measurements were available to provide a comparison with the measurement values and often from the expert judgement used to identify the fault events it was not possible to distinguish the anomalies associated to an instrument fault from those related to the input concentrations.

Particularly interesting are the possible research topics arisen from the Bayesian approach. In fact, the use of a probabilistic approach and the quantification of the uncertainties in the WWTPs models, either regarding the inputs or the constituent parameters, seems a promising field that could lead to a new class of better performing fault detection algorithms.

Bibliography

Abdul-Talib, S., Hvitved-Jacobsen, T., Vollertsen, J., and Ujang, Z., 2002. Half saturation constants for nitrate and nitrite by in-sewer anoxic transformations of wastewater organic matter. *Water Science and Technology*, **46**(9), 185–192.

Abeling, U. and Seyfried, C. F., 1992. Anaerobic-aerobic treatment of high-strength ammonium wastewater – nitrogen removal via nitrite. *Water Science and Technology*, **26**(5–6), 1007–1015.

Aguado, D. and Rosén, C., 2008. Multivariate statistical monitoring of continuous wastewater treatment plants. *Engineering Applications of Artificial Intelligence*, **21**(7), 1080–1091.

Alex, J., Benedetti, L., Copp, J. B., Gernaey, K. V., Jeppsson, U., Nopens, I., Pons, M.-N., Rieger, L. P., Rosén, C., Steyer, J.-P., Vanrolleghem, P. A., and Winkler, S., 2008. *Benchmark Simulation Model no.1 (BSM1)*, Lund, Sweden, pp. 1-62.

Alex, J., Rieger, L. P., Winkler, S., and Siegrist, H., 2003. Progress in sensor technology – progress in process control? Part II: results from a simulation benchmark study. *Water science and technology*, **47**(2), 113–120.

Alferes, J., Tik, S., Copp, J. B., and Vanrolleghem, P. A., 2013. Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection. *Water science and technology*, **68**(5), 1022–1030.

Alleman, J. E., 1984. Elevated nitrite occurrence in treatment systems. *Water Science & Technology*, **17**, 409–419.

Baggiani, F. and Marsili-Libelli, S., 2009. Real-time fault detection and isolation in biological wastewater treatment plants. *Water Science & Technology*, **60**(11), 2949–2961.

Battistoni, P., Boccadoro, R., Bolzonella, D., and Marinelli, M., 2003. An alternate oxic-anoxic process automatically controlled. Theory and practice in a real treatment plant network. *Water Science and Technology*, **48**(11–12), 337–344.

Battistoni, P., Fatone, F., Bolzonella, D., and Pavan, P., 2006. Full scale application of the coupled alternate cycles-membrane bioreactor (AC-MBR) process for wastewater reclamation and reuse. In: *IWA world water Congress and Exibition*. Beijing, China,

Battistoni, P., Fatone, F., Passacantando, D., and Bolzonella, D., 2007. Application of food waste disposers and alternate cycles process in small-decentralized towns: a case study. *Water research*, **41**(4), 893–903.

Cola, E., Battistoni, P., Fatone, F., and Pavan, P., 2007. A methodology to upgrade

municipal WWTPs by alternating system. In: *8th International Conference on Chemical and Process Engineering*. Ischia, Italy, pp. 2–7.

Copp, J. B. (ed.), 2002. *The COST simulation benchmark: description and simulator manual*, Office for official publications of the European Community, Luxembourg, LB, pp. 154.

Corominas, L., Villez, K., Aguado, D., Rieger, L. P., Rosén, C., and Vanrolleghem, P. A., 2011. Performance evaluation of fault detection methods for wastewater treatment processes. *Biotechnology and bioengineering*, **108**(2), 333–344.

Dairi, S., Mrad, D., Bensoltane, M., Djebbar, Y., and Abida, H., 2010. Optimal operation of alternating activated sludge processes of the municipal wastewater treatment plant case of Souk-Ahras (Algeria). In: *14th International Water Technology Conference*. Cairo, Egypt, pp. 431–443.

Dunteman, G. H., 1989. *Principal Component Analysis*, SAGE Publ. Inc., Newbury Park, CA, U.S.A., pp. 487.

Fatone, F., Battistoni, P., Bolzonella, D., Pavan, P., and Cecchi, F., 2008. Long-term experience with an automatic process control for nitrogen removal in membrane bioreactors. *Desalination*, **227**(1–3), 72–84.

Fuente, M. J., Garcia-Alvarez, D., Sainz-Palmero, G. I., and Vega, P., 2012. Fault detection in a wastewater treatment plant based on neural networks and PCA. In: *20th Mediterranean Conference on Control & Automation (MED)*. Barcelona, Spain, pp. 758–763.

Garcia-Alvarez, D., Fuente, M. J., and Palacin, L. G., 2011. Monitoring and fault detection in a reverse osmosis plant using principal component analysis. In: *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. Orlando, FL, U.S.A., pp. 3044–3049.

Garcia-Alvarez, D., Fuente, M. J., and Sainz-Palmero, G. I., 2012. Fault detection and isolation in transient states using principal component analysis. *Journal of Process Control*, **22**(3), 551–563.

Garcia-Alvarez, D., Fuente, M. J., Vega, P., and Sainz-Palmero, G. I., 2009. Fault detection and diagnosis using multivariate statistical techniques in a wastewater treatment plant. In: *7th IFAC International Symposium on Advanced Control of Chemical Processes*. Koç, Turkey, pp. 286–296.

Giusti, E., Marsili-Libelli, S., and Spagni, A., 2011. Modelling microbial population dynamics in nitritation processes. *Environmental Modelling and Software*, **26**(7), 938–949.

Goode, P. V. and Chow, M.-Y., 1994. A hybrid fuzzy/neural system used to extract

heuristic knowledge from a fault detection problem. In: *IEEE 3rd International Fuzzy Systems Conference*. Orlando, FL, U.S.A., pp. 1731–1736.

de Gracia, M., Larrea, A., and Fabiyi, M., 2013. Optimization of the Alternate Cycling Process for Nutrient Removal in Industrial Wastewater Treatment Plants–Full Scale Study. In: *Proceedings of 86th Annual Water Environment Federation Technical Exposition & Conference*. Chicago, IL, U.S.A, pp. 286–296.

Guglielmi, G. and Andreottola, G., 2011. Alternate anoxic/aerobic operation for nitrogen removal in a membrane bioreactor for municipal wastewater treatment. *Water Science & Technology*, **64**(8), 1730–1735.

Gujer, W., Henze, M., Mino, T., and van Loosdrecht, M. C. M., 2000. Activated Sludge Model no.3. *Water Science & Technology*, **39**(1), 189–193.

Héduit, A., Quinio, I., Stadmuller, D., and Thévenot, D. R., 1996. Modified platinume electrodes: electrochemical characteristics and behaviour in activated sludge. *Water Science & Technology*, **34**(3–4), 143–150.

Henze, M., 2008. *Biological wastewater treatment: principles modelling and design* (M. Henze, M. C. M. van Loosdrecht, G. A. Ekama, and D. Brdjanovic, eds.), IWA Publishing, London, UK, pp. 518.

Henze, M., Grady jr, C. P. L., Gujer, W., Marais, G. v. R., and Matsuo, T., 1987. *Activated Sludge Model no.1*. IAWPRC Scientific and Technical Report no.1. IAWPRC, London, UK.

Henze, M., Gujer, W., Mino, T., and van Loosdrecht, M. C. M., 2000. *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*. Scientific and Technical Report no.9. IWA Publishing, London, UK, pp. 121.

Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M. C., and Marais, G. v. R., 1995. *Activated Sludge Model no.2*. IAWQ Scientific and Technical Report no.3. IWAQ, London, UK.

Hotelling, H., 1947. Multivariate quality control illustrated by the testing of sample bombsights. in: O. Eisenhart (ed.), *Selected techniques of statistical analysis*. McGraw-Hill, New York, NY, U.S.A.

Iacopozzi, I., Innocenti, V., Marsili-Libelli, S., and Giusti, E., 2007. A modified Activated Sludge Model no.3 (ASM3) with two-step nitrification-denitrification. *Environmental Modelling & Software*, **22**(6), 847–861.

Jia, M.-X., Xu, H.-Y., Liu, X.-F., and Wang, N., 2012. The optimization of the kind and parameters of kernel function in KPCA for process monitoring. *Computers & Chemical Engineering*, **46**, 94–104.

Jolliffe, I. T., 2002. *Principal Component Analysis*, John Wiley & Sons Ltd, New York,

NY, U.S.A., pp. 487.

Ku, W., Storer, R. H., and Georgakis, C., 1995. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **30**(1), 179–196.

Lee, D. S. and Vanrolleghem, P. A., 2003. Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis. *Biotechnology and Bioengineering*, **82**(4), 489–497.

Lee, J.-M., Yoo, C.-K., Choi, S. W., Vanrolleghem, P. A., and Lee, I.-B., 2004. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, **59**(1), 223–234.

Lennox, J. A. and Rosén, C., 2002. Adaptive multiscale principal components analysis for online monitoring of wastewater treatment. *Water Science & Technology*, **45**(4–5), 227--235.

Li, W., Yue, H. H., Valle-Cervantes, S., and Qin, S. J., 2000. Recursive PCA for adaptive process monitoring. *Journal of Process Control*, **10**(5), 471–486.

Liu, Y.-J., Chen, T., and Yao, Y., 2014. Nonlinear process monitoring and fault isolation using extended maximum variance unfolding. *Journal of Process Control*, **24**(6), 880–891.

Marsili-Libelli, S., El Basri, E., Plotegher, C., and Giusti, E., 2012. Fuzzy classification trees as environmental indicators. In: *6th Biennial Meeting of iEMSs*. Liepzig, Germany, pp. 2413–2420.

Marsili-Libelli, S., Ratini, P., Spagni, A., and Bortone, G., 2001. Implementation, study and calibration of a modified ASM2d for the simulation of SBR processes. *Water Science and Technology*, **43**(3), 69–76.

Martín de la Vega, P. T., Jaramillo, M. A., and Martínez de Salazar, E., 2013. Upgrading the biological nutrient removal process in decentralized WWTPs based on the intelligent control of alternating aeration cycles. *Chemical Engineering Journal*, **232**, 213–220.

Mason, R. L., Chou, Y.-M., and Young, J. C., 2001. Applying Hotelling's T2 statistic to batch processes. *Journal of Quality Technology*, **33**(4), 466–479.

Metcalf & Eddy Inc., Tchobanoglous, G., Burton, F. L., and Stensel, H. D., 2004. *Wastewater engineering: treatment, disposal, and reuse*, McGraw-Hill, New York, NY, U.S.A., pp. 1819.

Meyer, D., Leisch, F., and Hornik, K., 2003. The support vector machine under test. *Neurocomputing*, **55**(1–2), 169–186.

Nagy-Kiss, A. M. and Schutz, G., 2013. Estimation and diagnosis using multi-models

with application to a wastewater treatment plant. *Journal of Process Control*, **23**(10), 1528–1544.

Nardelli, P., Gatti, G., Eusebi, A. L., Battistoni, P., and Cecchi, F., 2009. Full-scale application of the alternating oxic/anoxic process: an averview. *Industrial & Engineering Chemistry Research*, **48**(7), 3526–3532.

Olaru, C. and Wehenkel, L., 2003. A complete fuzzy decision tree technique. *Fuzzy Sets and Systems*, **138**(2), 221–254.

Olsson, G., Nielsen, M. K., Yuan, Z., Lynggaard-Jensen, A., and Steyer, J.-P., 2005. *Instrumentation, control and automation in wastewater systems*, IWA Publishing, London, UK, pp. 246.

Pambrun, V., Paul, E., and Spérandio, M., 2008. Control and modelling of partial nitrification of effluents with high ammonia concentrations in sequencing batch reactor. *Chemical Engineering and Processing*, **47**(3), 323–329.

Parzen, E., 1962. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, **33**(3), 1065–1076.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, **2**(6), 559–572.

Quinlan, J. R., 1990. Decision trees and decision-making. *IEEE Transactions on Systems, Man, and Cybernetics*, **20**(2), 339–346.

Quinlan, J. R., 1987. Decision trees as probabilistic classifiers. In: *Proceedings of the Fourth International Machine Learning Workshop*. San Mateo, CA, U.S.A., pp. 31–37.

Rieger, L. P., Alex, J., Winkler, S., Boehler, M., Thomann, M., and Siegrist, H., 2003. Progress in sensor technology – progress in process control? Part I: sensor property investigation and classification. *Water Science & Technology*, **47**(2), 103–112.

Rosén, C., Jeppsson, U., and Vanrolleghem, P. A., 2004. Towards a common benchmark for long-term process control and monitoring performance evaluation. *Water Science & Technology*, **50**(11), 41–49.

Rosén, C. and Lennox, J. A., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. *Water Research*, **35**(14), 3402–3410.

Rosén, C. and Olsson, G., 1998. Disturbance detection in wastewater treatment plants. *Water Science & Technology*, **37**(12), 197–205.

Ruiz, G., Jeison, D., and Chamy, R., 2003. Nitrification with high nitrite accumulation for the treatment of wastewater with high ammonia concentration. *Water research*, **37**(6), 1371–1377.

Ruiz Ordóñez, M. L., Sin, G., Berjaga, X., Colprim, J., Puig, S., and Colomer, J., 2011.

Multivariate Principal Component Analysis and Case-Based Reasoning for monitoring, fault detection and diagnosis in a WWTP. *Water Science & Technology*, **64**(8), 1661.

Santinelli, M., Eusebi, A. L., Cola, E., and Battistoni, P., 2011. A Hybrid Denitrification–Alternate Cycles Reactor To Enhance the Nitrogen Biological Removal in a Real Wastewater Treatment Plant. *Industrial & Engineering Chemistry Research*, **50**(24), 13947–13953.

Schölkopf, B., Smola, A., and Müller, K.-R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, **10**(5), 1299–1319.

Schraa, O., Tole, B., and Copp, J. B., 2006. Fault detection for control of wastewater treatment plants. *Water Science & Technology*, **53**(4–5), 375–382.

Shlens, J., 2014. *A tutorial on principal component analysis*, Mountain View, CA, USA.

Sin, G., Kaelin, D., Kampschreur, M. J., Takács, I., Wett, B., Gernaey, K. V., Rieger, L. P., Siegrist, H., and van Loosdrecht, M. C. M., 2008. Modelling nitrite in wastewater treatment systems: a discussion of different modelling concepts. *Water Science and Technology*, **58**(6), 1155–1171.

Suárez, A. and Lutsko, J. F., 1999. Globally optimal fuzzy decision trees for classification and regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(12), 1297–1311.

Takács, I., Patry, G. G., and Nolasco, D., 1991. A dynamic model of the clarification thickening process. *Water Research*, **25**(10), 1263–1271.

Theodoridis, S. and Koutroumbas, K., 2009. *Pattern Recognition*, Academic Press, London, UK, pp. 961.

Tong, C., Palazoglu, A., and Yan, X.-F., 2013. An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding. *Journal of Process Control*, **23**(10), 1497–1507.

Turk, O. and Mavinic, D. S., 1989. Maintaining nitrite build-up in a system acclimated to free ammonia. *Water Research*, **23**(11), 1383–1388.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K., 2003. A review of process fault detection and diagnosis. *Computers & Chemical Engineering*, **27**(3), 293–346.

Villez, K., 2007. *Multivariate and qualitative data-analysis for monitoring, diagnosis and control of sequencing batch reactors for wastewater treatment*. Ph.D. Thesis. Universiteit Gent, Ghent, Belgium, pp. 370.

Villez, K., Rosén, C., Anctil, F., Duchesne, C., and Vanrolleghem, P. A., 2013. Qualitative Representation of Trends (QRT): Extended method for identification

of consecutive inflection points. *Computers & Chemical Engineering*, **48**, 187–199.

Villez, K., Ruiz Ordóñez, M. L., Sin, G., Colomer, J., Rosén, C., and Vanrolleghem, P. A., 2008. Combining multiway Principal Component Analysis (MPCA) and clustering for efficient data mining of historical data sets of SBR processes. *Water Science and Technology*, **57**(10), 1659–1666.

Wareham, D. G., Hall, K. J., and Mavinic, D. S., 1993. Real-time control of wastewater treatment systems using ORP. *Water Science and Technology*, **28**(11–12), 273–282.

Xu, F., Puig, V., Ocampo-Martinez, C., Stoican, F., and Olaru, S., 2014. Actuator-fault detection and isolation based on set-theoretic approaches. *Journal of Process Control*, **24**(6), 947–956.

Yao, M., Wang, H., and Xu, W., 2014. Batch process monitoring based on functional data analysis and support vector data description. *Journal of Process Control*, **24**(7), 1085–1097.

Yuan, Y.-F. and Shaw, M. J., 1995. Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, **69**(2), 125–139.