Environmental Modelling & Software 47 (2013) 88–107

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Review

Data-derived soft-sensors for biological wastewater treatment plants: An overview

Henri Haimi[a,*], Michela Mulas[a], Francesco Corona[b], Riku Vahala[a]

[a] Department of Civil and Environmental Engineering, Aalto University, School of Engineering, P.O. Box 15200, FI-00076 Aalto, Finland
[b] Department of Information and Computer Science, Aalto University, School of Science, P.O. Box 15400, FI-00076 Aalto, Finland

A R T I C L E I N F O

Article history:
Received 21 September 2012
Received in revised form
8 May 2013
Accepted 17 May 2013
Available online

Keywords:
Water quality monitoring
Soft-sensors
Data-driven models
Wastewater treatment

A B S T R A C T

This paper surveys and discusses the application of data-derived soft-sensing techniques in biological wastewater treatment plants. Emphasis is given to an extensive overview of the current status and to the specific challenges and potential that allow for an effective application of these soft-sensors in full-scale scenarios. The soft-sensors presented in the case studies have been found to be effective and inexpensive technologies for extracting and modelling relevant process information directly from the process and laboratory data routinely acquired in biological wastewater treatment facilities. The extracted information is in the form of timely analysis of hard-to-measure primary process variables and process diagnostics that characterize the operation of the plants and their instrumentation. The information is invaluable for an effective utilization of advanced control and optimization strategies.

1. Introduction

During the recent decades, the increased awareness about the negative impact of eutrophication in the quality of water bodies (see, e.g., Ansari et al., 2010) and the advances in environmental technology have given rise to more stringent wastewater treatment requirements and regulations (Olsson et al., 2005; Olsson, 2012). In wastewater treatment plants (WWTPs), the tightening treatment regulations are leading towards the addition of new unit processes and towards the renewal of the existing ones. A typical example in municipal WWTPs is the update of the ammonia removal process towards total nitrogen removal; this is usually achieved through the conversion of the biological reactor from a single aerated tank to a sequence of anoxic and aerobic zones. The subsequent increase in operational and management investments, mostly associated to energy consumption and chemical dosing, stimulates modern WWTPs to face the challenges of maintaining and improving effluent quality, while guaranteeing efficient and safe operations and optimizing the costs. A major requirement for achieving these goals relies on the availability of real-time measurements of key or primary process indicators. These indicators are needed to efficiently monitor the operation of the plants in terms of influent and effluent quality, process and instrument performance and economic efficiency, with immediate implications for environmental compliance, safety, management planning and profitability. The real-time availability of primary indicators is invaluable for an effective utilization of advanced process control and optimization strategies in WWTPs.

The conventional approach to the monitoring problem in WWTPs relies upon on-line and off-line analysis of the primary variables. The primary variables are typically concentrations of ammonia, nitrates and total nitrogen, phosphates and total phosphorus, suspended solids, biochemical and chemical oxygen demand, as well as others process variables like the sludge blanket level. Such variables are hard-to-measure and their availability is often associated with expensive capital and maintenance costs, as well as being characterized by time-delayed responses that are often unsuitable for real-time monitoring. For instance, the organic compounds are still typically monitored by off-line laboratory measurements, of which the analysis of biochemical oxygen demand requires several days. Moreover, the harsh conditions in biological treatment processes such as the Activated Sludge Process (ASP) make reliable field measurements challenging. Already in an early survey where fifty wastewater treatment facilities in the USA were considered (Molvar et al., 1976) and in a publication on the state-of-the-art in wastewater treatment control (Olsson, 1977), the problems of on-line instrumentation were discussed. The typical problems included solids deposition, slime build-up and precipitation, which gave rise to poor performance and a frequent need for

* Corresponding author. Tel.: +358 50 407 4214.
E-mail addresses: henri.haimi@aalto.fi (H. Haimi), michela.mulas@aalto.fi (M. Mulas), francesco.corona@aalto.fi (F. Corona), riku.vahala@aalto.fi (R. Vahala).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.05.009

maintenance of the instrumentation. During the recent decades, considerable development in on-line instrumentation has taken place (see e.g., Vanrolleghem and Lee, 2003; Olsson, 2012). In spite of the recent advances, such as *in situ* nutrient sensors and luminescent dissolved oxygen sensors, instruments still tend to get fouled (Olsson, 2012). Nevertheless, trustworthy real-time analyses for many key variables is not there yet.

Due to the progress in measurement, automation and communication technologies, WWTPs are also becoming highly instrumented and many on-line easy-to-measure process variables are routinely acquired. The easy-to-measure or secondary variables are typically pressure, temperature, flow rate, and level measurements, as well as conductivity, turbidity, pH, and, perhaps dissolved oxygen. The secondary variables can be extensively used for characterizing the operational conditions of the unit processes and they offer an inexpensive opportunity to extract primary information useful to monitor both the processes and the instruments. For example, being the primary variables necessarily related to some of the secondary variables, their availability offers the opportunity to develop process models capable to reconstruct such a relationship and thus also to estimate the primary variables. Such models are at the core of virtual instruments often referred to as software- or soft-sensors; that is, computer programs that model the input information encoded in the secondary variables and output information related to the primary variables, in a similar fashion to their hardware counterparts (Kadlec et al., 2009). On the basis of their internal model, soft-sensors are often divided into two main classes, phenomenological and data-driven. Phenomenological soft-sensors are based on the first principle process models, whereas data-driven soft-sensors are built around process models derived from data. Hybrid soft-sensors combine these two modelling approaches.

In wastewater treatment, the most commonly used first principle models belong to the Activated Sludge Model (ASM) family (Henze et al., 2000) proposed by the IWA Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment. Moreover, the IWA Task Group on Good Modelling Practice has created a unified protocol for enhancing the quality of activated sludge modelling and dealing with uncertainties associated with the phenomenological approaches (Rieger et al., 2012). Because capable to describe both linear and nonlinear phenomena and to provide information on the internal states of the process, the detailed phenomenological modelling approach has proven to be efficient, for example, in wastewater treatment process design, renovation, employee training, optimization of the plant operation and understanding the system's behaviour and interactions of the components (Hauduc et al., 2009; Phillips et al., 2009). The ASM models have also been used in soft-sensor design, for instance by Spérandio and Queinnec (2004) and Grau et al. (2007). However, there are major challenges in using the first principle models for real-time applications. For example, characterizing the organic matter and determining the rate constants for the volatile fatty acid (VFA) uptake in wastewater is challenging, expensive and time-consuming and, yet, fundamental for successful model calibration (Dochain and Vanrolleghem, 2001; Petersen et al., 2003; Sin, 2004; Hauduc et al., 2011). Moreover, the high-dimensionality of detailed phenomenological models results in an enormous computational requirements and ill-conditioned problems due to the interaction between fast and slow dynamics (Dochain and Vanrolleghem, 2001).

The large amount of process data routinely measured and collected in modern WWTPs permits data-driven modelling as an interesting alternative for soft-sensor design. A data-derived soft-sensor is an input–output model, where the inputs usually consist of easy-to-measure secondary variables in the form of plant's signals. In the soft-sensor, the input information is modelled and the internal model is used to return output information associated with the hard-to-measure primary variables. Different families of models have been popular in designing the data-derived soft-sensors, which are commonly used for applications such as on-line prediction, process monitoring and process fault detection, and sensor monitoring and reconstruction (Kadlec, 2009). Today, data-driven soft-sensors are becoming more common in the wastewater treatment sector, even though they are still not as widespread as, for instance, in the process industry where soft-sensors are extensively exploited (see, e.g., Fortuna et al., 2007; Kadlec et al., 2009). Soft-sensors have also been used in the other environmental domains, where the open distributed architectures for sensor networks (Douglas et al., 2008) provide an increasing amount of the available data. For instance, environmental data has been applied for soft-sensors aiming at real-time anomaly detection in the meteorological signals (Hill et al., 2009; Hill and Minsker, 2010) and at prediction of the ammonia concentration in a river downstream the sewage and WWTP outlets (Masson et al., 1999). In the wastewater treatment facilities, the data-derived soft-sensor applications range from the proposals where a small number of variables are used for modelling (such as in Marsili-Libelli, 1990; Lumley, 2002; Puig et al., 2005; Äijälä and Lumley, 2006; Cecil and Kozlowska, 2010) to the studies where a larger number of measurements are processed together with a model (for instance, Teppola et al., 1999b; Rosen and Lennox, 2001; Aguado et al., 2007a; Lee et al., 2008; Corona et al., 2013), which in particular are in the scope of this review. Data-driven applications in wastewater treatment are included in earlier review papers, where their extent, however, is limited. Their limited amount is due to the main focus of these review publications being on the application of phenomenological modelling (Yoo et al., 2001; Gernaey et al., 2004; Banadda et al., 2011) or on the use of a specific data-driven modelling family (Khataee and Kasiri, 2011; Yetilmezsoy et al., 2011).

In this paper, we aim to provide an extensive overview of the applications of data-derived soft-sensors in wastewater treatment and to present a general guideline for the development of data-derived soft-sensors. The paper is organized as follows. Section 2 introduces briefly the biological wastewater treatment process types, which are used in the case studies, and characteristics of WWTP operation data. In Section 3, we give an overview of the practical steps to be undertaken in the design of data-derived soft-sensors. A review of publications focussing on the data-derived soft-sensor applications in wastewater treatment systems is given in Section 4. Next, Section 5 contains a discussion on our findings concerning the applications of the data-derived soft-sensors in WWTPs and the current status and future challenges of soft-sensing in the field of operation. A nomenclature of the terminology is provided in Table 1.

2. Wastewater treatment plants

Municipal wastewater treatment aims at reducing the amounts of nutrients, organic matter (determined as Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD) or Total Organic Carbon (TOC)) and Suspended Solids (SS) that influent wastewater contains. This is typically carried out by using several unit processes, including biological, chemical and physical treatment methods. The core of the treatment line is a biological reactor such in the ASP, where a high concentration of activated sludge consisting mainly of bacteria and protozoa is recycled in zones with different Dissolved Oxygen (DO) conditions, especially for nitrogen removal purposes. In most of the municipal WWTPs, primary and secondary clarifiers are applied for separation and thickening sludge. To maintain the microbiological population in the

Table 1
Nomenclature.

Abbreviation	Explanation	Abbreviation	Explanation
AdMSPCA	Adaptive Multiscale Principal Component Analysis	AD	Anaerobic Digestion
ANFIS	Adaptive Network-based Fuzzy Inference System	AF	Anaerobic Filter
ANN	Artificial Neural Network	AFBR	Anaerobic Fluidized Bed Reactor
APCA	Adaptive Principal Component Analysis	ASM	Activated Sludge Model
APLS	Adaptive Partial Least Squares	ASP	Activated Sludge Process
ARX	Auto-Regressive Exogenous	BOD	Biochemical Oxygen Demand
CPCA	Conventional Principal Component Analysis	BSM	Benchmark Simulation Model
CPCR	Conventional Principal Component Regression	COD	Chemical Oxygen Demand
CPLS	Conventional Partial Least Squares	DO	Dissolved Oxygen
DWT	Discrete Wavelet Transform	DSVI	Diluted Sludge Volume Index
EMPCA	Expectation-Maximization Principal Component Analysis	EBPR	Enhanced Biological Phosphorus Removal
FCM	Fuzzy C-Means	F/M	Food to Microorganisms ratio
FFNN	Feedforward Neural Network	LTP	Lagoon Treatment Plant
FPCR	Fuzzy Principal Component Regression	MBBR	Moving Bed Biofilm Reactor
FPLS	Fuzzy Partial Least Squares	MBR	Membrane Bioreactor
GK	Gustafsson-Kessel algorithm	MLSS	Mixed Liquor Suspended Solids
GLSR	Generalized Least Squares Regression	MLVSS	Mixed Liquor Volatile Suspended Solids
k-NN LLR	Local Linear Regression based on *k*-Nearest Neighbours	ORP	Oxidation Reduction Potential
KPLS	Kernel Partial Least Squares	SBFR	Submerged Biofilm Reactor
LAMDA	Learning Algorithm for Multivariable Data Analysis	SBR	Sequencing Batch Reactor
MLCA	Multilevel Component Analysis	SHARON	Single reactor system for High activity Ammonium Removal Over Nitrite
MLR	Multiple Linear Regression		
MPCA	Multiway Principal Component Analysis	SS	Suspended Solids
MPLS	Multiway Partial Least Squares	SSVI	Stirred Sludge Volume Index
MSPCA	Multi-Scale Principal Component Analysis	SVI	Sludge Volume Index
MSPLS	Multi-Scale Partial Least Squares	TKN	Total Kjeldahl Nitrogen
NNPLS	Neural Network Partial Least Squares	TN	Total Nitrogen
PARAFAC	Parallel Factor Analysis	TOC	Total Organic Carbon
PCA	Principal Component Analysis	TOD	Total Oxygen Demand
PCM	Possibilistic C-Means	TP	Total Phosphorus
PCR	Principal Component Regression	TSS	Total Suspended Solids
PLS	Partial Least Squares	UASB	Upflow Anaerobic Sludge Blanket (reactor)
QPLS	Quadratic Partial Least Squares	VFA	Volatile Fatty Acid
RAPLS	Robust Adaptive Partial Least Squares	WWTP	Wastewater Treatment Plant
RBFN	Radial Basis Function Network		
RF	Random Forest		
RPCA	Robust Principal Component Analysis		
RRBF	Repair Radial Basis Function Neural Network		
RTNN	Recurrent Trainable Neural Network		
SOM	Self-Organizing Map		
SSWN	State – Space Wavelet Neural Network		
TDNN	Time Delay Neural Network		
TSK	Takagi-Sugeno-Kang (model)		

bioreactor, the sludge from the secondary clarifiers is recirculated into the biological reactor. Phosphorus removal is typically achieved using chemical precipitation, but process configurations targeting for Enhanced Biological Phosphorus Removal (EBPR) also exist. In the EBPR process, polyphosphate-accumulating organisms are selectively enriched in the bacterial community and an additional anaerobic process stage and a more sophisticated sludge recirculation scheme are required. In addition, anaerobic digestion is a typical process for excess sludge treatment where the amount of organic matter in the sludge is considerably reduced and the biogas generated during the digestion process is often used for producing electricity and heat.

Also in many industrial sectors, biological wastewater treatment is widely applied. For instance in the pulp and paper industry, ASP is the treatment process most commonly used (see, e.g., Thompson et al., 2001; Pokhrel and Viraraghavan, 2004). Here, the whole biological reactor is aerated, being the main goal of the treatment to oxidize the high concentration of organic compounds in the wastewater. Clarifiers are applied in the treatment lines in a similar manner as in municipal facilities and, depending on the preceding industrial processes, unit operations such as cooling towers and equalization basins may be needed before the biological treatment. Furthermore, the neutralization of the wastewater is often required and nitrogen and phosphorus usually need to be dosed in the biological WWTPs of pulp and paper industry in order to maintain a sufficient balance between organic matter and the nutrients for the microbiological population. In the reviewed papers, ASP is also applied for the treatment of the wastewaters of other industrial sectors in the majority of the case studies.

In addition to the ASP, also other kinds of treatment processes have been used in the reviewed papers and they are briefly described. Sequencing Batch Reactors (SBRs) containing activated sludge are operated in a different number of phases such as filling, aeration, mixing, settling and decanting. The phases and their lengths are determined depending on the treatment targets. In the Membrane Bioreactors (MBRs), suspended solids are separated with membranes instead of secondary clarifiers as in the conventional ASPs. Considerably larger Mixed Liquor Suspended Solids (MLSS) concentrations can be used in the MBR due to getting rid of the operational constraints set by the clarifier units. Trickling filters are aerobic processes, where wastewater flows through permeable medium to which microorganisms are attached. The Moving Bed Biofilm Reactor (MBBR) technology is based on the carrier media that provides a large protected surface area for the biofilm attachment and that moves freely in the bioreactor where wastewater is treated in. The aerated Submerged Biofilm Reactors (SBFRs) are filled with porous support medium through which wastewater and air flow. Specifically, the support medium on which biofilm grows is maintained under the total immersion of the water flow in the SBFRs. In the Lagoon Treatment Plants (LTPs), earthen basins are used as reactors and typically they require a large surface area. Floating surface aerators are used to provide the required oxygen and mixing in the lagoons. In the Single reactor system for High activity Ammonium Removal Over Nitrite (SHARON) process, partial nitrification, i.e. oxidation of ammonium to nitrite, is established. Nitrate formation in the reactor is prevented by adjusting temperature, pH, and retention time and, thus, the aeration requirement of SHARON is smaller in comparison with a conventional complete nitrification process.

Several types of bioreactors operated in anaerobic or anoxic conditions have also been considered in the reviewed papers. Anaerobic Digestion (AD) is the process where sewage is treated, for instance, in septic tanks mainly for reducing the organic content of wastewater. In the Upflow Anaerobic Sludge Blanket (UASB) process, wastewater is introduced at the bottom of the reactor and it flows upwards through a sludge blanket composed of biologically formed granules or particles in the absence of oxygen. Anaerobic Filters (AFs) treat wastewater that is led through a bed of medium on which anaerobic bacteria grows. Typically, their operation targets on the removal of organic matter but, for example, in anaerobic post-treatment filters the goal may be reducing the nitrate content in wastewater in the presence of an external carbon source. In the Anaerobic Fluidized Bed Reactors (AFBRs), bacteria are immobilized on small fluidized medium though which wastewater flows and the process is operated in an up-flow mode in order to achieve fluidization.

The process data of a municipal WWTP has typical diurnal and seasonal trends, e.g., in influent flow rate, concentrations and temperature (Olsson et al., 2005). Also, differences in the wastewater composition between weekdays and weekends exist. In addition to time-related trends, municipal WWTPs are exposed to disturbances such as heavy rains and snow melt. In industrial WWTP data, rapid and abrupt changes in the influent temperature and concentrations are often caused by the permutations in the preceding industrial production process (Woodard & Curran, Inc., 2006). Moreover, depending on the specific field of industry, a large variety of pollutants are commonly regulated (see, e.g., Water Environment Federation, 2008) and, therefore, the analysis results for the case-specific pollutants are available in the data stored in the treatment plants.

Due to the demanding conditions in the biological WWTPs, the real-time measured process data usually contains missing and anomalous observations (Rosen, 2001; Olsson et al., 2005). The reasons for missing observations include instrumentation malfunctions and maintenance, faults in the data transmission and errors in the database. Typical wastewater treatment operation data also contains outliers, that is to say, inconsistent observations that deviate markedly from the regular operation of the process and its instrumentation. Additionally, drifting data can exist because of process drifts caused, for example, by weather changes or sensor drifts due to defective instruments. Under these circumstances, back-up and fault detection systems for the hardware instrumentation as well as real-time predictions of the process variables are beneficial for ensuring the efficient operation of WWTPs.

3. Soft-sensor development

A data-derived soft-sensor is conventionally described as an input–output process model. The model inputs consist of easy-to-measure secondary variables in the form of plant's signals and measurements and, sometimes, numerically encoded expert knowledge. The model outputs consist of information associated with the hard-to-measure primary variables. In the soft-sensor, the input and output process information is modelled empirically and the internal model is used to return the outputs when only the inputs are available.

The range of tasks that can be fulfilled by data-derived soft-sensors is broad and mainly dictated by the nature of the available input information, by the information that we are interested to output and the typology of the input–output model. The original and still most prominent application area is the *on-line prediction* of process variables that can be only measured either at low sampling rates or off-line. In this case, the inputs are those variables that are easy to measure and the outputs are estimates of the variables that are hard to measure; since the input–output relationship is encoded in the data used to calibrate the model, the soft-sensor model is used to reconstruct it and then to estimate the output variables when new inputs are available. At its core, this type of soft-sensors addresses a supervised learning problem in the form of regression or classification. The other typical application areas are related to *monitoring the state of the process* and to *monitoring the state of the instrumentation*. In this case, the inputs are again those variables that are easy to measure and the outputs are information on the operation of the process and the instruments, in the form of diagnostics and status characterization; since the output information is usually hidden in the data, the soft-sensor model is used to explore the data and then to estimate the outputs when new inputs are available. At its core, this typology of soft-sensors usually addresses an unsupervised learning problem in the form of dimensionality reduction or clustering.

This section overviews the practical steps to be undertaken in the design of data-derived soft-sensors. An overview is given in Fig. 1 and the main steps along with the most commonly used techniques are briefly discussed. The procedure is general and it does not refer to the design of specific types of soft-sensors or specific applications thereof. Moreover, it is not unequivocally standardized, though extensively adopted by both researchers and practitioners (Kadlec, 2009). The procedure consists of several independent steps (*data acquisition*, *data pre-processing*, *model design* and *model maintenance*) and it is to be understood as an iterative process where choices made in the design often need to be reconsidered before the soft-sensor is ready for deployment.

3.1. Data acquisition

In modern WWTPs, the historical process and laboratory data are routinely acquired and stored in the data acquisition system of the plant. The data can be easily retrieved and *data collection* and subsequent *data inspection* are the first steps in soft-sensor development. During the initial inspection, a preliminary exploration of the measurements is performed, in order to overview the prominent structures in the data (e.g., redundancies, taxonomies, functionalities and time-delays existing between variables and observations) and to identify the presence of obvious problems (e.g., locked variables, missing and drifting data and measurements outside the operating range of the instruments). In addition, periods of instrument calibrations and unit process maintenance are also annotated together with a selection of representative operations (e.g., steady states and periodicities, transients and seasonal disturbances). The inspection typically requires a large amount of manual work and expertise in the underlying processes, coupled with an extensive exploration of time series, scatter plots, and histograms of data.

3.2. Data pre-processing

Remarkable characteristics of the data acquired in wastewater treatment facilities are redundancy and possibly insignificance,

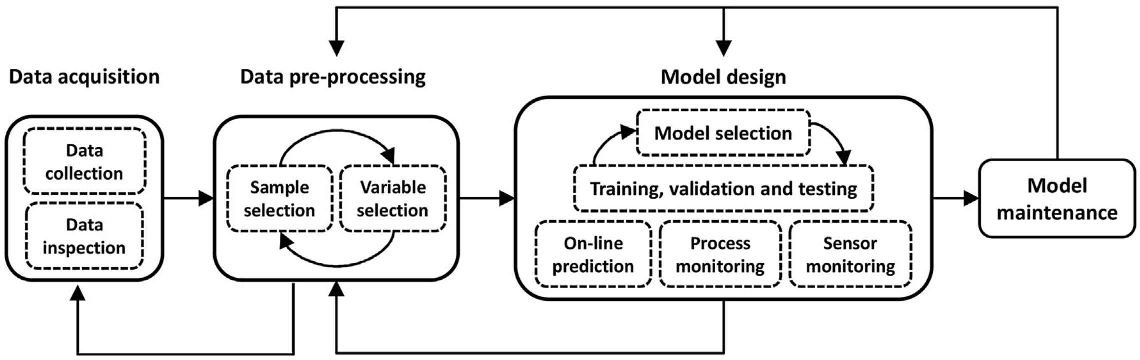

Fig. 1. Overview of the design steps for data-derived soft-sensors, adapted from Kadlec et al. (2009).

let alone the presence of other disturbances that corrupt the measurements. Very often, the amount and quality of the data together with their high-dimensionality can be a limiting factor for the development of the soft-sensors. Therefore, it is necessary to prepare the data before they are processed by the soft-sensor's model. Process understanding and *a priori* knowledge is required in this phase; the plant operators and experts can provide valuable information on the relevance of the variables and the observations. Such a knowledge can be supported and complemented by many statistical techniques for *variable selection* and *sample selection*. In the following, some of the most commonly used techniques for these tasks are briefly overviewed.

3.2.1. Variable selection

The choice of the input variables that influence the model output is a crucial stage. Variable selection consists of choosing those secondary variables that are most informative for the process being modelled, as well as those that provide the highest generalization ability. This step is fundamental because data models are built from a finite number of observations and having a model with too many inputs and hence too many parameters, may lead to overfitting and induce large computational times. In addition, describing a process in terms of a few selected variables allows to retain interpretability.

The most commonly used techniques for variable selection are often categorized as *filtering*, *wrapping* and *embedded* methods (Guyon and Elisseeff, 2003). Embedded methods perform the selection inside the soft-sensor's model and they will be overviewed in Section 3.3. Filters and wrappers select variables and subsets of variables by ranking them on the basis of their significance for the task. The latter methods rely on the conjunction of two independent elements:

- a *relevance criterion*, to score an input variable or a group of input variables according to its informative power;
- a *search procedure*, for finding among all the available variables the subset that optimizes the chosen criterion.

The criterion is often based either on statistical dependence or model accuracy measures. The simplest and most popular ways to measure the interactions between sets of variables are *correlation* and its multivariate extension obtained with Canonical Correlation Analysis (Hotelling, 1936). *Mutual information*, on the other hand, can be used to measure the total amount of linear and nonlinear information (Kraskov et al., 2004). Less conventional relevance criteria are *noise variance* (Liitiäinen et al., 2009) and the *generalization error* of a model (Kohavi and John, 1997). Given a criterion, the simplest strategy for variable selection consists of scoring all inputs and rank them accordingly, then, a threshold is set and only the variables that score more than the threshold are kept. *Variable ranking* is the simplest and the fastest of all the methods. *Exhaustive search*, conversely, is the most expensive procedure, for it performs the scoring and ranking for each possible subset of variables. A speed-complexity compromise is found in *incremental* and/or *decremental* methods; such methods add and/or remove one variable at each iteration of the search and their procedure is halted when there is no improvement in the criterion or, when a predefined number of variables is already selected.

In many situations the computational applicability of criterion-search schemes is limited by the too large number of inputs. An alternative approach to reduce the dimensionality of the problem consists of producing a small number of combinations of the original variables. New variables may be constructed on the basis of process knowledge (e.g., using balance calculations and averaging) or they can be derived using statistical projection methods for dimensionality reduction (Lee and Verleysen, 2007). Projection methods aim at reducing the data dimensionality by mapping the observations onto a low-dimensional subspace, in such a way that certain properties of the data are preserved as faithfully as possible. Linear multivariate techniques such as Principal Component Analysis (PCA, Jolliffe, 2002) and Partial Least Squares (PLS, Wold et al., 2001) are probably the most common methods for learning a low-dimensional representation of a set of data. PCA constructs new variables that are maximally independent in a linear sense, hence uncorrelated. PLS, on the other hand, constructs new variables that are maximally uncorrelated among themselves and also maximally linearly related with the output variable. The variables produced by projection methods do not necessarily retain an explicit connection with the original process variables and to regain interpretability on the process being modelled, they are often used in combination with contribution plots. Methods like PCA and PLS are most commonly used for continuous processes. In the case of batch processes, the multiway nature of the data is more appropriately exploited using consistent extensions of the aforementioned linear techniques (Smilde et al., 2004).

3.2.2. Sample selection

When analysing real data, it is not uncommon that some observations are different from the majority. Such observations are often called outliers. They may be due to data acquisition mistakes or they correspond to exceptional process circumstances and, in general, one can distinguish between two main types of outliers: i) *obvious outliers* are observations that violate physical or technological limitations and ii) *non-obvious outliers* are observations that do not violate any bound but still lay out of typical ranges. Sample selection consists of discarding or pinpointing outlying observations because not necessarily representative of normal operations and because their use may be detrimental for the performances of the soft-sensor model (Rosen et al., 2003). Alternatively, sample selection consists of choosing only those observations that are truly representative of the normal operation of the processes and instruments being modelled.

A detailed multi-phase procedure for obtaining high-quality WWTP data was recently proposed by Rieger et al. (2010) in a study that offers guidelines starting from the experimental design and the data collection phase. On already available data, multivariate statistical methods like the already mentioned PCA and PLS, coupled with an analysis of model residuals like the Hotelling's T^2 and the Squared Prediction Error (SPE), or Q-statistic, are frequently used (Robinson et al., 2005). Interestingly, however, also these classical models are sensible to outliers in the data and to overcome such a limitation, their robust extensions should be used instead (Rousseeuw and Hubert, 2011). Again, such approaches are most conveniently used for continuous processes, whereas multiway variants are more suitable in the case of batch processes (Hubert et al., 2012).

A conceptually different approach to sample selection can be seen in the application of *clustering* and *classification* methods (Hastie et al., 2009). Clustering and classification methods aim at reducing the amount of data by grouping the observations into subsets, either clusters or classes, consisting of similar observations. Similarity refers to the property where observations belonging to a group should be as much as possible similar to each other, but differ significantly from observations in other groups. Clustering is essentially an unsupervised learning problem, whereas classification is the analogous supervised problem based pre-labelled data. The most commonly used techniques for clustering are based on the K-means algorithm (Hartigan and Wong, 1979) and on Fuzzy C-means (FCM, Bezdek et al., 1984). Classification, on the other hand, is typically performed using methods based on Linear Discriminant Analysis (Fisher, 1936) and on Artificial Neural Networks (ANN, Haykin, 1999) and Support Vector Machines (Cristianini and Shawe-Taylor, 2000).

3.3. Model design

Model design is a critical step in soft-sensor development, for the structure of the model at the core defines the specific application task and it confirms the designer's assumptions on the problem being studied. Moreover, the selection of the model parameters determines the generalization abilities of the soft-sensor. However, a consistent approach to the task does not exist so far and, the model structure and its parameters are often selected in an ad hoc manner for each soft sensor (Kadlec et al., 2009). This is firstly due to the fact that model design depends on the task at hand and it is often subjected to developer's past experience and personal preference, as many designers strongly focus on the one approach in their field of expertise.

Despite the lack of a theoretically superior approach to model design, two main tasks can be recognized: i) *model structure selection* and ii) *model training, calibration and testing*. The common practice suggests to start with simple model types, assess their potential and performances and then gradually increase complexity, as long as significant improvements are observed. Furthermore, it is important that the models are not only accurate, but also simple and computationally efficient, interpretable and with low maintenance cost. While performing the task it is fundamental to optimize the model parameters in terms of generalization performances, through a validation of the results on independent data before testing the results. In the following the main family of model structures are overviewed and then the optimization of their parameters discussed using a standard method like cross-validation.

3.3.1. Model structure selection

3.3.1.1. Models for on-line prediction. Such models address the problem of reconstructing the functionality existing between the easy-to-measure inputs and the hard-to-measure process outputs. Usually, the inputs and the outputs take continuously varying real values and their relationships can be modelled as a regression problem. Less common is the case where the outputs take categorical values; in this case, their relationship can be modelled as a classification problem. In the following we will focus only on the most common regression models and we refer the reader interested on classification to the book by Duda et al. (2000).

The simplest regression techniques assume the existence of a linear input–output relationship and they fit a linear model to reconstruct it. Most of the commonly used techniques pertain to *multivariate statistics* (Anderson, 2003) and they provide an interpretable description of how the inputs affect the output. *Multiple linear regression* (MLR) is based on the ordinary least squares approximation of the linear model, which is simple and sometimes accurate. Accuracy and interpretability of MLR can be improved by shrinking the regression coefficients and possibly setting some of them to zero. This is achieved with linear regression models that also perform an *embedded variable selection* scheme (Hastie et al., 2009). The most commonly used *subset selection* methods are Best Subset Selection, Forward and Backward Stepwise Selection, Forward Stagewise Regression, whereas popular *shrinking* methods are Ridge regression, the LASSO and the LARS. In situations with a large number of inputs, multivariate statistical methods combining linear projection and linear regression can be used to reduce the dimensionality of the modelling problem, at the price of interpretability. The most commonly used in this category methods are Principal Component Regression (PCR) and PLS regression. Adaptive and recursive extensions of PCR and PLS (see e.g., Kadlec et al., 2011) can be used for capturing the dynamic nature of process data, whereas nonlinear kernel extensions like KPCA and KPLS (Rosipal and Trejo, 2001) can be used in the presence of nonlinearities.

Other nonlinear methods do not necessarily rely on any assumption on the input–output relationships. Nevertheless, they are widespread among researchers and practitioners. Methods based on supervised *artificial neural networks* (Haykin, 1999) and *neuro-fuzzy systems* (Fullér, 2000) are among the most popular ones. An ANN is a network of artificial neurons arranged in layers and connected to each other. The neurons nonlinearly transform the incoming signals using an activation function and then they distribute the result to the other neurons. The input–output relationship is encoded in the connection weights; the weights are adapted to minimize the error between the network outputs and the targets. In particular, Feedforward Neural Networks (FFNN, Haykin, 1999) have been popular in the wastewater treatment sector. Neuro-fuzzy systems combine the features of ANNs with the human-like reasoning style of fuzzy systems, aiming at complementary techniques and enhanced performance compared with the individual methodologies. Typically in neuro-fuzzy systems, the first layer corresponds to input variables, the middle layers encode fuzzy IF-THEN rules and the last layer corresponds to the output variables. The advantages of neuro-fuzzy systems include the ability of the ANN learning algorithms to learn both fuzzy sets and fuzzy rules, as well as the potential to use *a priori* knowledge. In particular, an Adaptive Network-based Fuzzy Inference System (ANFIS, Jang, 1993) has been popular among soft-sensor designers in the wastewater treatment sector.

3.3.1.2. Process and sensor monitoring. Such models address the problem of detecting, identifying and diagnosing normal and abnormal behaviours in the processes and in the field instruments, using the easy-to-measure process variables as inputs. The diagnostics comprise the hard-to-measure outputs and, usually, no prior information about them is available and it must be extracted from data using dimensionality reduction and clustering approaches. In the less common case where the output information is available, either as real or categorical values, the input–output relationship can be modelled as a regression or a classification problem, respectively. Again, we will mostly focus only on the most common models for dimensionality reduction and clustering.

In order to detect the occurrence of any variation having an exceptional or identifiable cause, univariate statistical control charts have been traditionally used to monitor a small number of process variables. Examining one variable at a time, as though they were independent, however, makes interpretation and diagnosis extremely difficult in environments where a large number of variables are continuously varying relative to one another. However, when the number of variables is large, one often finds that they are also highly dependent on one another and common methods for reducing the dimensionality of the problem are the already mentioned linear approaches based on PCA and PLS.

In the monitoring of the continuous processes and the hardware sensors, the conventional and adaptive PCA and PLS methods are popular for reducing the dimensionality of the variables of interest. The use of PCA methods enables process monitoring in a low-dimensional space defined by the principal components retained in the model at the price of losing the information encoded in the less significant principal components discarded. An established technique to isolate the variable(s) responsible for the detected anomalies is to study their contributions to the statistics considered for the analysis of model residuals. Another traditional monitoring approach is to use low-dimensional scatter plots defined by the most significant principal components, that include most of the information of the original variables, for observing the transitions in the process or in the relationships between the supervised sensors. When using PLS approaches for process monitoring, the output variables are usually hard, or impossible, to determine in real-time and they indicate the presence of anomalous situations. In wastewater treatment, such variables are, for instance, the indicators of the sludge settling properties determined by field and laboratory experiments.

In the batch process, an additional dimension to the data structure is addressed by the batch, the other dimensions representing time and the variables. The Multiway PCA (MPCA) and PLS (MPLS) are commonly used for dimensionality reduction when multiway data is considered, for instance in the cases of the batch processes (Smilde et al., 2004). The multiway extensions first unfold a three-dimensional data structure into a two-dimensional structure and, then, PCA or PLS is executed. These methodologies are often adopted, for instance, for monitoring the batch processes. Multilevel Component Analysis (MLCA, Timmerman, 2006) is an extension of PCA, which is useful if the variation in the data occurs on different levels simultaneously. For monitoring the batch processes, MLCA enables a separated interpretation of the transitions both within the batches and between the batches in low-dimensional subspaces.

The Kohonen Self-Organizing Map (SOM, Kohonen, 2001) is the most common unsupervised ANN method and it has also been used in many wastewater treatment applications. In the model training, the neurons in the hidden layer adapt themselves to the relationships within a set of input signals and the SOM outputs a low-dimensional (usually two-dimensional) representation of the patters encoded in the training data. In this representation, clusters corresponding to the characteristic features of the data are formed onto a topographic map that provides an interpretation of the input information.

Clustering methods (Everitt et al., 2011) are applied for monitoring using the process variables or new variables, created for instance by PCA approaches, as the model inputs. In the clustering analysis, the observations among the training data are grouped in the clusters based on their similarity. In the conventional clustering approaches based, for example, on the K-means algorithm, each observation belongs to one of the clusters. Instead, in the fuzzy clustering methods, each observation belongs to all the clusters to some extent, represented by their fuzzy memberships. When introducing unseen data, their discrete properties can be observed by monitoring their transition between the clusters defined in the model training step.

Since most of the basic data-derived techniques cannot deal with missing data directly, a strategy for their replacement has to be designed and implemented. A data imputation approach, which is primitive and not recommended but commonly applied, is to replace the missing values with the mean values of the affected variable. Another non-optimal approach is to skip the data samples consisting of variable or variables with the missing values. Also, imputing a missing value by a linear interpolation between the preceding and following existing values is a problematic approach especially when several sequential values are missing. More efficient approaches based on multivariate statistics of the data perform the reconstruction of the missing values from other variables of the affected samples (Walczak and Massart, 2001a,b). When the hardware sensor measurements studied, these approaches are considered as sensor reconstruction and this application area is closely connected with sensor monitoring. More strategies for dealing with the missing values are reported by Little and Rubin (2002).

3.3.2. *Model training, validation and testing*

Most of the model types discussed in this section are characterized by a number of basic parameters and a number of hyper-, or meta-, parameters that define their structure and optimize it in terms of its generalization performances. The basic parameters of the models are, for instance, the regression coefficients of linear regression methods, the connection weights of neural and neuro-fuzzy systems, the loading components in multivariate statistical methods like PCA and PLS, among the others. The meta-parameters are, on the other hand, the number of components to be retained in methods like PCA and PLS, the regularization parameter in linear shrinkage methods like LARS and LASSO, the number of neurons and layers in neural and neuro-fuzzy systems, the number of clusters, among the others. Before a model is able to operate on new unseen observations, it has to be trained to estimate its basic parameters and it has to be validated to optimize its meta-parameters. Model validation is a highly important step in soft-sensor development, in which the designer estimates how well the model will perform on new data.

Ideally, if enough data were available, the soft-sensor developer would set aside a validation set and use it to assess a model whose basic parameters are calibrated on a training set, for different values of its meta-parameters. After finding the optimal set of meta-parameters, the developer would then calibrate the model to ultimately set its basic parameters, using all the available learning data (i.e., both the training and the validation set). The resulting model is eventually assessed on an independent test set of data. However, it may be difficult to obtain a sufficient amount of historical data for the for learning the model according to the aforementioned procedure. In such a situation, the soft-sensor developer has to resort on error-estimation techniques, like the simple and widely used *cross-validation* (Hastie et al., 2009). K-fold cross-validation uses part of learning data to calibrate the model and a different part to validate it. The procedure consists of firstly splitting the data in K roughly equal-size parts, or folds, secondly to set aside the k-th part and calibrating the model to the other K-1 parts and, thirdly, to calculate a measure of model accuracy over the k-th part. After repeating the procedure for all the K parts, then accuracies are combined to give an average (over the folds) performance of the model, for a specific set of meta-parameters. The model whose meta-parameters have the best generalization accuracy is finally trained using all the learning data and assessed against a test set.

Although cross-validation is the most popular approach to optimize the generalization performances of data-derived soft-sensors, alternative techniques based on statistical resampling methods like *bagging* (Breiman, 1996) and *boosting* (Freund and Schapire, 1997) can be used for the task.

3.4. Model maintenance

After the successful design, it is not uncommon to observe a degradation of the performance of a data-derived soft-sensor. Such a degradation is often due to changes in process and instrumental characteristics or operating conditions. In wastewater treatment applications, the reason for this may be, e.g., variations in influent wastewater composition, temperature and flow rate, instruments recalibrations or operational changes inside the plant. To overcome such limitations, soft-sensors should be regularly maintained and updated as the system characteristics change but, their manual and repeated redesign should be avoided due to the heavy workload.

Most of the soft-sensors currently found in full-scale environments do not provide any automated mechanisms for their maintenance. To automatically cope with changes in process characteristics and operating conditions, a number of data-derived approaches have been however designed and are available for the developer (Kadlec et al., 2011). The majority of these approaches are inherently encoded in the adaptive and recursive versions of multivariate statistical methods like PCA and PLS. An approach related to the neuro-fuzzy methods also providing adaptation possibilities is local learning (Atkeson et al., 1997). An adaptive soft-sensor developed in this framework was published in Kadlec and Gabrys (2008). In addition, Fujiwara et al. (2009) and Zhu et al. (2011) discuss the development of maintenance-free soft-sensors for on-line prediction using local linear regression methods.

4. Soft-sensor applications in WWTPs

Data-derived methods can be used for (i) the on-line prediction of the primary process variables, (ii) process monitoring and process fault detection, and (iii) hardware-sensor monitoring and providing back-up for them, e.g., during the periods of sensor faults, maintenance and calibration. In this section, the reviewed case studies are organized according to the soft-sensor application categories.

In the beginning of each subsection representing the aforementioned application areas, we review a few representative studies without separating them based on the modelling methods. Next, the contributions belonging to the categories are further arranged considering the different families of modelling methods employed. Among the multivariate methods, case studies applying conventional methods are introduced first and, after that, studies concentrating on advanced techniques such as adaptive, nonlinear and multiway extensions are presented. As for the ANNs, we first introduce publications utilizing FFNNs as the most popular techniques used in WWTPs, then studies using other types of supervised ANNs and, finally, works employing unsupervised ANNs. Eventually, studies where neuro-fuzzy techniques and hybrid models are proposed for soft-sensor design are introduced. The modelling methods applied in the reviewed publications and their amounts are represented in Fig. 2.

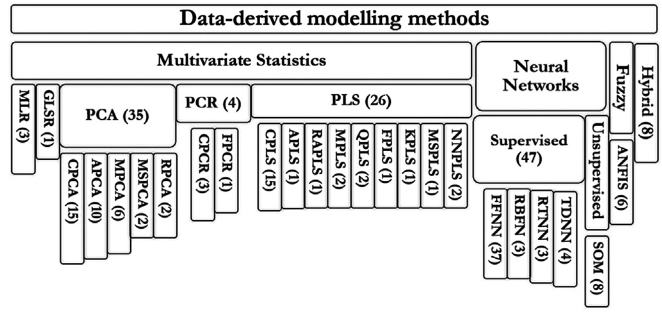

Fig. 2. Data-derived modelling methods used in the reviewed case studies.

4.1. On-line prediction

A common application for the data-derived soft-sensors in wastewater treatment systems involves predicting the primary process variables. The soft-sensor applications for prediction tasks in the reviewed publications are summarized in Table 2, where studies are arranged in the same order as they appear in the text. In addition, the main features of a few interesting conference publications have been included in the table (marked with an asterisk) even though they have not been reviewed in this paper.

Some of the most relevant works can be found in continuous WWTPs. In the earlier soft-sensor applications the SS, BOD and COD concentrations have been popular predicted outputs, whereas in more recent publications it has been more common to estimate nutrient concentrations. This indicates the progress in wastewater treatment technology, where modern-day municipal WWTPs are typically designed for nutrient removal and, therefore, reliable information on the nutrient concentrations in the process has become of interest to the plant operators. The secondary variables typically used as inputs in the case studies dealing with the continuous processes include flow rates, pH, temperature and DO, SS and nutrient concentrations measured in different locations of the process. As for the batch processes, the typical target has been to predict the nutrient concentration trends during the aerobic and anoxic phases using simple on-line measurements, e.g., DO, pH and Oxidation Reduction Potential (ORP) as the model inputs and, further, to employ the estimates for controlling the lengths of the phases. In both continuous and batch problems, the most commonly used modelling method for the reconstruction of the desired outputs has been ANNs, especially FFNNs, followed by multivariate statistical methods based on PLS.

A FFNN was applied for the estimation of NH_4–N and NO_3–N concentrations in a municipal ASP by Bongards (2001). Moreover, the predicted nutrient concentrations were successfully used as inputs for a fuzzy controller dosing high-strength sludge treatment reject water from a buffer tank into the wastewater treatment line. The proposed solution improved the treatment performance and helped in prevention of the high nitrogen concentration peaks in the effluent. Flow rate, conductivity, DO and nutrient measurements were used as inputs to predict future nutrient concentrations. Lee et al. (2008) developed soft-sensors for the prediction of COD, NH_4–N and PO_4–P concentrations in the effluent of small-scale municipal ASPs by using FFNN techniques. Simple on-line measurements such as turbidity, ORP, temperature, conductivity, pH and DO were used as inputs to a FFNN. The authors found that a significant increase in estimation accuracy could be, however, achieved by integrating the FFNN with an Auto-Regressive Exogenous (ARX) model providing time-lags. The results demonstrated the improved capability of the integrated methodology in capturing the nonlinearities and dynamics of the processes. The estimates

Table 2
Prediction applications of soft-sensors in the reviewed publications. M = municipal, I = industrial, L = laboratory-scale, P = pilot-scale, S = simulated, ✓ = practical implementation, * = conference publication.

Publication	Method(s)	Appl.	Process(es)	Predicted variable(s)
Bongards (2001)	FFNN	M ✓	ASP	NH_4–N, NO_3–N
Lee et al. (2008)	FFNN	M ✓	ASP	COD, NH_4–N, PO_4–P
Woo et al. (2009)	PLS, NNPLS, KPLS	I	ASP	COD, TN, cyanide
Aguado et al. (2006)	PCR, PLS, MPLS, FFNN	P	SBR	PO_4–P
Multivariate statistics				
Aarnio and Minkkinen (1986)	PLS	M	ASP	TP, COD, turbidity
Blom (1996)	PLS	M	WWTP	TP
Teppola et al. (1999b)	MLR, PCR, PLS	I	ASP	DSVI, COD
Lee et al. (2007)	PLS, APLS, RAPLS	I	AF	TOD, CH_4 production
Yoo et al. (2004)	PLS, FPLS, QPLS	S, I	BSM1, ASP	NH_4–N, NO_3–N, SVI, cyanide, COD
Dürrenmatt and Gujer (2012)	GLSR, FFNN, SOM, RF	M	ASP	COD, NH_4–N
Corona et al. (2013)	MLR, PLS, *k*-NN LLR	M	Post-filters	NO_3–N
Neural networks				
Capodaglio et al. (1991)	FFNN	M	ASP	SVI
Tyagi and Du (1992)	FFNN	M	ASP	Inhibition by heavy metals
Boger (1992)	FFNN	M ✓	ASP	NH_4–N
Pu and Hung (1995)	FFNN, MLR	M	Biofilter	BOD, SS
Häck and Köhne (1996)	FFNN	M	ASP	COD, NH_4–N, NO_3–N
Hamed et al. (2004)	FFNN	M	ASP	BOD, SS
Bisschops et al. (2006)	FFNN	I, S	ASP, ASM1	Exogenous respiration end-point
Pai et al. (2007)	Grey models, FFNN	I	ASP	SS, COD
Pai et al. (2008)	Grey models, FFNN	I	ASP	SS, COD, pH
Poutiainen et al. (2010)	FFNN	M	ASP	TSS
Fu and Poch (1995)	FFNN	M	ASP	COD
Hamoda et al. (1999)	FFNN	M	ASP	BOD, TSS
Hanbay et al. (2007)	FFNN	M	WWTP	COD
Machón et al. (2007)	FFNN	L	ASP	NH_4–N
Mjalli et al. (2007)	FFNN	M	ASP	BOD, COD, SS
Moral et al. (2008)	FFNN	M, S	ASP, ASM1	MLVSS, COD
Çinar et al. (2006)	FFNN	I, P	MBR	COD, NH_4–N, NO_3–N, TP
Aguado et al. (2009)	FFNN	L	SBR	PO_4–P
Delnavaz et al. (2010)	FFNN	P	MBBR	COD, aniline
Ráduly et al. (2007)	FFNN	S	BSM1	COD, SS, NH_4–N, BOD, TN, TKN
Caraman et al. (2007)	FFNN	S	ASP	Substrate, MLSS, DO
Clara (2008)	FFNN	S	BSM1	TN, TKN, TSS, BOD, COD
Belanche et al. (1999)	TDNN	M	ASP	BOD, COD
Belanche et al. (2000)	TDNN	M	ASP	SS
Dellana and West (2009)	ARIMA, TDNN	S, M	Artificial data, ASP	BOD, SS, TN, TP
Luccarini et al. (2002)	Elman NN	L	SBR	NH_4–N, NO_3–N, PO_4–P
Borowa et al. (2007)	SSWN	M	ASP	Flow rate, NO_3–N, NH_4–N, DO
Qiao et al. (2011)	RRBF	M	ASP	COD
Grieu et al. (2006)	SOM + FFNN	M	ASP	COD, NH_4–N, SS
Rustum and Adeloye (2007)	SOM, MLR, FFNN	M	ASP	Sludge Age, F/M, BOD, SS
Rustum et al. (2008)	SOM	M	ASP	BOD
Zhao and Chai (2005)*	EMPCA + TDNN	M	ASP	BOD
Lee et al. (2011)*	FFNN	M	ASP	BOD, COD, SS, TN
Fuzzy				
Tay and Zhang (1999)	ANFIS	I, L	AFBR, UASB	CH_4, TOC, VFA, propionic acid, acetic acid, butyrate acid, COD
Tay and Zhang (2000)	ANFIS	L	UASB, AFBR, AF	CH_4 production, TOC, VFA
Civelekoglu et al. (2007)	PCA + ANFIS	I	ASP	COD, NH_4–N, TN
Fernandez et al. (2009)	Fuzzy ANN	M	ASP	Flow rate
Huang et al. (2010)	MPCA, FFNN, ANFIS	L	SBR	NH_4–N, NO_3–N, PO_4–P
Wan et al. (2011)	ANFIS, ANN	I	SBFR	SS, COD
Pai et al. (2011)	ANFIS, FFNN	I	ASP	SS, COD, pH
Hybrid				
Côté et al. (1995)	ASM1 + FFNN	M	ASP	DO, SS, COD, NH_4–N, SS
Cohen et al. (1997)	FFNN + fuzzy logic	M, I ✓	SBR	COD, TN
Choi and Park (2001)	PCA + FFNN	I	ASP	TKN
Lee et al. (2002)	FFNN, ASM1 + FFNN	I	ASP	MLSS, SS, COD, cyanide
Lee et al. (2005)	FFNN, RBFN, PLS, QPLS, NNPLS	I	ASP	MLSS, SS, COD, cyanide
Hong et al. (2007)	MPCA + FFNN	L	SBR	NH_4–N, NO_3–N, PO_4–P
Kim et al. (2009)	MPCA + FFNN	M	ASP	COD, TN, TP
Rustum (2009)	SOM + ANFIS	M	ASP	BOD, SS

were considered sufficient alternatives for the expensive instruments for measuring nutrient concentrations. Woo et al. (2009) applied different PLS approaches for the real-time prediction of three outputs, the effluent COD, Total Nitrogen (TN) and total cyanides concentrations, in an ASP treating cokes wastewater. The authors used 27 variables measured in influent, anoxic tank, aerobic tank, secondary clarifier and effluent as model inputs, some of them being time-lagged. While the conventional PLS model was not capable of modelling the process satisfactorily, a Neural Network PLS (NNPLS, Qin and McAvoy, 1992) achieved improved estimation accuracies. The best prediction performance was, however, obtained by a KPLS model. Interestingly, the models were further compared in terms of their complexity; for the task, the authors used the Bayesian Information Criterion which confirmed the ability of KPLS to outperform both linear and nonlinear PLS when considering both accuracy and complexity.

As for the batch processes, a study by Aguado et al. (2006) compared the performances of several approaches based on PCR, PLS and FFNNs for predicting the PO_4–P concentration profile in a pilot-scale SBR treating synthetic sewage and targeting for the EBPR. The authors found the batch-wise unfolding MPLS models to outperform the other techniques used for the estimation task. Moreover, the proposed MPLS model used only on-line measurements of pH, conductivity, ORP, DO and temperature as inputs, whereas some of the other models also used time-lagged inputs.

4.1.1. Multivariate statistical models

Multivariate statistics, in particular those based on PLS, are one of the typical techniques used as soft-sensors for prediction tasks. In an early application in a municipal ASP, conventional PLS was used for estimating Total Phosphorus (TP) and COD concentrations and turbidity in the effluent (Aarnio and Minkkinen, 1986). The authors found the methodology to be feasible for recognizing the reasons for sludge bulking episodes which were responsible for poor effluent quality. Blom (1996) designed a PLS model for estimating influent TP concentrations in a municipal WWTP. He considered the prediction accuracy of a model based on daily laboratory analyses quite acceptable. Adaptive extensions of multivariate methods have also been proposed for prediction purposes. In a study by Teppola et al. (1999b), MLR, PCR and PLS models combined with a Kalman filter providing time-lags were presented for the estimation of the Diluted Sludge Volume Index (DSVI) and COD reduction in an ASP of a paper mill. The static models were not successful in approximating the output variables, whereas the use of the Kalman filter remarkably improved the predictions of the DSVI. In the case of COD estimation, there were no notable improvements obtained by using the Kalman filter. Lee et al. (2007) proposed Robust Adaptive PLS (RAPLS) for prediction of the Total Oxygen Demand (TOD) in the effluent and the production rate of methane gas in an AF process treating industrial wastewater. The authors indicated that the estimates could be also used for detecting abnormal process operations.

Yoo et al. (2004) applied conventional PLS, Quadratic PLS (QPLS, Baffi et al., 1999) and Fuzzy PLS (FPLS, Bang et al., 2003) models for two soft-sensor case studies. In the case where the IWA/COST Benchmark Simulation Model BSM1 (Copp, 2002) was used as a test platform and effluent NH_4–N and NO_3–N concentrations were considered as the output variables, it was indicated that FPLS had the best prediction performance. In a full-scale industrial ASP application, the researchers used the Sludge Volume Index (SVI) and the reductions of cyanide and COD in the process as the estimated outputs. Here, PLS and QPLS models showed a slightly better prediction performance than FPLS models. However, the prediction accuracy of SVI was poor using any of the methods. Dürrenmatt and Gujer (2012) compared different data-driven modelling approaches for approximating COD concentration in the primary clarifier effluent and NH_4–N concentration in the bioreactor of a municipal ASP. The authors found Generalized Least Squares Regression (GLSR, Kariya and Kurata, 2004) estimates to be less accurate than the ANN estimates, but on the other hand, they considered the transparency of the GLSR models to be a significant advantage compared with the opaqueness of the ANNs. Therefore, they considered the interpretability to justify the selection of the GLSR approach for the soft-sensor development. In a recent publication, Corona et al. (2013) designed an array of soft-sensors for estimating the NO_3–N concentrations in the denitrifying post-filtration unit in a municipal WWTP. The authors found only minor improvements in prediction accuracies when linear MLR and PLS models were replaced by nonlinear Local Linear Regression based on *k*-Nearest Neighbours (*k*-NN LLR, Stone, 1977), at the price of slightly heavier computational cost. The estimation accuracies of the models were reported being comparable with the resolution of the hardware sensors.

4.1.2. Artificial neural network models

As already pointed out, ANN techniques have been popular in soft-sensor for prediction of process variables in the biological WWTPs.

Most commonly the applications in this research area have considered the use of **FFNNs in full-scale ASPs**. In an early study, Capodaglio et al. (1991) developed a FFNN for estimating SVI in a municipal ASP targeting at monitoring bulking sludge episodes. Tyagi and Du (1992) proposed to use a FFNN for the prediction of the inhibitory effect of heavy metals on activated sludge. The authors used the ratios of the maximum specific growth-rate in the presence and absence of heavy metals as model outputs to be estimated. In their conclusion, the authors found the estimates to be close to the experimental results. At about the same time, Boger (1992) suggested a soft-sensor based on a FFNN for predicting the average weekly concentration of effluent NH_4–N in a municipal ASP and diagnosed the process variables associated with the high NH_4–N concentrations. A study on a FFNN-based soft-sensor for estimating the effluent BOD and SS concentrations in a trickling filter for municipal wastewater treatment was published by Pu and Hung (1995). However, the prediction errors indicated that the performance of the models was not adequate, because of the long interval between the measurements available for training the model. Häck and Köhne (1996) used FFNN models for estimating COD and NH_4–N concentrations in the influent wastewater and NH_4–N and NO_3–N concentrations in the aeration basin of a municipal ASP. They found the ANN estimates reliable but only for a limited period of time, yet still considerably better when compared with the least square estimates.

More recently, Hamed et al. (2004) presented a FFNN for the estimation of BOD and SS concentrations in the effluent of a municipal ASP in a study where they used the daily records of BOD and SS concentrations in several process stages as model inputs. They found the prediction accuracy fairly low, especially for the SS concentration, apparently due to the insufficient amount and quality of the available process data. In their study, Bisschops et al. (2006) investigated the use of a FFNN for predicting the end-point of exogenous respiration, which indicates complete oxidation of biodegradable matter, in an activated sludge system. Interestingly, they used recorded industrial wastewater respirograms and artificial respirograms simulated based on the ASM1 model (Henze et al., 2000) for training and testing the models. The amount of the available data was too limited for considering a practical implementation of the model. The researchers speculated that repeating the study with a larger number of respirograms would highlight the true potential of FFNNs for determining the treatability of wastewater in real-time.

Grey modelling and FFNN techniques were applied for the prediction of SS and COD concentrations in the effluent of a continuous sequencing batch ASP treating hospital wastewater (Pai et al., 2007) and, of a conventional industrial ASP (Pai et al., 2008) where also effluent pH was estimated. In these studies, the authors found the grey models more promising for the prediction tasks compared with the FFNN models, especially when the operation data available was not extensive. In their research, Poutiainen et al. (2010) employed FFNN for estimating the next day's effluent Total Suspended Solids (TSS) load in a municipal ASP. However, the prediction accuracy suffered from an inadequate amount of available process data, which did not contain all the seasonal patters. In addition, the effluent concentrations of nutrients, organic matter or SS in full-scale ASPs have been predicted using FFNN methods in publications by Fu and Poch (1995), Hamoda et al. (1999), Hanbay et al. (2007), Machón et al. (2007), Mjalli et al. (2007) and Moral et al. (2008).

Case studies targeting at prediction of the variables using **FFNNs in laboratory and pilot-scale processes** have also been presented. The effluent concentrations of a pilot-scale MBR treating industrial wastewater were predicted by using a cascade-forward ANN model by Çinar et al. (2006). The authors reported the prediction accuracy of the model to be successful for NH_4–N and NO_3–N concentrations, whereas the results for COD and TP prediction were considered only satisfactory. The reasons for achieving only satisfactory accuracies obviously include the scarcity of the data used in the work: only 20 daily observations for model training and 10 observations for testing. In a publication by Aguado et al. (2009), FFNN models were employed for the prediction of the PO_4–P concentration trends in a laboratory-scale SBR aiming at EBPR. Particularly, they showed the real-time prediction to enable the optimization of the lengths of the phases in the SBR instead of the fixed time schedule basis typically applied for the purpose. Delnavaz et al. (2010) designed a FFNN-based soft-sensor for predicting the COD reduction and effluent aniline concentration of a pilot-scale MBBR fed with synthetic wastewater of various strengths and operated with different retention times. They found the estimation results to be satisfactory and ANN models to provide robust tools since the prediction errors in the experiments varied only slightly.

Soft-sensors aiming at on-line prediction have been investigated by employing **FFNNs in simulated treatment processes**. In an interesting study by Ráduly et al. (2007), the prediction performance and simulation time of a FFNN model were compared with a phenomenological WWTP model (ASM3, Henze et al., 2000) where the preliminary influent data of the BSM2 platform (Jeppsson et al., 2006) extended with a disturbance scenario generator was applied as the influent module. The results showed that the FFNN model was capable of providing good prediction accuracies for effluent NH_4–N, BOD and SS concentrations, whereas the estimations of COD and TN concentrations were less satisfactory. The authors also found that FFNN models allow for a considerably faster WWTP performance evaluation compared with the ASM3 model, including the time needed for training the models. Caraman et al. (2007) proposed a FFNN model for the estimation of variables and to be used as internal model for a predictive controller in a simulated ASP. The tests with a simplified ASP model indicated that the predicted substrate and DO concentrations in the ASP were accurate, but there was a noticeable shift between the estimated and measured biomass concentrations in the bioreactor. In addition, Clara (2008) presented FFNN-based models for approximating the effluent TN and Total Kjeldahl Nitrogen (TKN) concentrations using the BSM1 as the test platform. Interestingly, he applied successfully general algorithms and fuzzy systems for variable selection and found FFNN a confident tool especially for estimating the TN behaviour.

Apart from FFNNs, also **other types of supervised ANNs** have been used for estimation purposes. In a study by Belanche et al. (1999), the effluent BOD and COD concentrations of a municipal ASP were estimated and the fuzzy heterogeneous Time Delay Neural Network (TDNN, de Vries and Principe, 1991) was shown to outperform the classical TDNN in the investigation. In another application in the same WWTP, Belanche et al. (2000) reported that a hybrid TDNN model provided the best prediction performance for the effluent SS concentration out of the three ANN approaches and the *k*-NN algorithm tested. The proposed methodology was also able to handle missing values in data. Dellana and West (2009) investigated the performances of the TDNN and linear ARIMA models for predicting the wastewater effluent BOD, SS, TN and TP concentrations in a work where they used both artificial and real-world WWTP data sets. The authors concluded that the TDNN models were more accurate in the cases concerning all the real-world process data, but in half of the artificial data cases, the ARIMA models were found more precise. This outcome highlights the need for a careful consideration in the model structure selection taking into account the nature of the data. In another study, Luccarini et al. (2002) proposed soft-sensors for the estimation of the NH_4–N, NO_3–N and PO_4–P concentration trends in a laboratory-scale SBR treating synthetic wastewater by using the Elman NN methodology (Elman, 1990). They reported that the soft-sensors had the potential to be used for controlling the lengths of the anoxic and aerobic phases in SBRs. Borowa et al. (2007) presented a soft-sensor for the prediction of the effluent flow rate, NO_3–N and NH_4–N concentrations and DO concentration in the aerobic zone of a modelled ASP using State - Space Wavelet Neural Networks (SSWN, Zamarrenoa and Vega, 1998) and they demonstrated that the method was able to approximate the variables. Recently, Qiao et al. (2011) proposed Repair Radial Basis Function Neural Network (RRBF) for estimating the effluent COD concentrations in municipal WWTPs, resulting in acceptable accuracy.

As for the **unsupervised ANN methods**, Grieu et al. (2006) proposed the on-line prediction of influent and effluent COD, NH_4–N and SS concentrations by means of SOM and FFNN for a municipal WWTP. In particular, SOM was used satisfactorily for recognizing correlations between the process variables and for choosing the input variables for the FFNN models. However, the prediction accuracies were considered only satisfactory. This is possibly due to the nature of the acquired data consisting of daily average measurements for a four-year period. Rustum and Adeloye (2007) proposed SOM estimates for replacing outliers and missing values in the process data of a municipal ASP. The results indicated that the performance of the SOM approach was superior for the considered task compared with FFNN and MLR models. The SOM model also did particularly well in predicting the Sludge Age and Food to Microorganism Ratio (F/M). Moreover, Rustum et al. (2008) used the SOM technique successfully in a soft-sensor for approximating the influent BOD concentration of three municipal WWTPs.

4.1.3. Neuro-fuzzy systems

Neuro-fuzzy systems, especially ANFIS models, have been applied for prediction tasks in various wastewater treatment processes. In a detailed study by Tay and Zhang (1999), ANFIS models were used for estimation in a laboratory-scale UASB reactor and a full-scale AFBR. In the case of the UASB reactor, the neuro-fuzzy technique was successfully applied for predicting the volumetric methane production and the effluent Total Organic Carbon (TOC) concentration. In the AFBR application, a model with six output variables provided satisfactory estimates given the limited amount of available process data. Tay and Zhang (2000) used the ANFIS methodology successfully also for prediction of the volumetric methane production, the effluent TOC concentration and the total VFA concentration in three laboratory-scale wastewater treatment processes subjected to shock tests. In the work by Civelekoglu et al. (2007), ANFIS models were used for estimating the effluent COD, NH_4–N and TN concentrations of an ASP treating the wastewater of a sugar factory. They demonstrated that the estimates fit well with the measured values. In addition, PCA was applied as a data preprocessing method which actually decreased significantly the prediction errors. Fernandez et al. (2009) developed a neuro-fuzzy system to predict the daily influent flow rates in two municipal ASPs, and the model validation showed a strong agreement between the estimates and the measured data. In their case study, Huang et al. (2010) suggested an approach utilizing the ANFIS technique for the estimation of NH_4–N, NO_3–N and PO_4–P concentrations in a laboratory-scale SBR. The data were first analysed with the MPCA methodology in order to select adequate training data sets and both normal and abnormal testing data sets. The model testing with normal data indicated that the predicted values were able to follow the measured values well, whereas the use of abnormal data resulted in poor prediction ability, which was expected since the training data only spanned the normal operating range. Wan et al. (2011) presented a soft-sensor application where they used an ANFIS model for estimating the effluent COD and SS concentrations in an aerated SBFR treating paper industry wastewater. The authors also applied PCA successfully for the selection of the input variables for the ANFIS models. The ANFIS methodology has also been recently used in a proposal for the soft-sensors in industrial wastewater treatment by Pai et al. (2011).

4.1.4. Hybrid models

Hybrid models combining different modelling approaches on the system level provide interesting alternatives for soft-sensor design. In an early study, Côté et al. (1995) used hybrid models consisting of a modified ASM1 model and FFNN models for predicting the concentrations of DO in the bioreactor, of SS, COD and NH_4–N in the effluent and of SS in the return sludge with good accuracy in a municipal ASP. A study by Cohen et al. (1997) applied a combination of FFNN and fuzzy methods to modelling process variables in a SBR treating municipal, dairy and beef processing wastewaters. The models did not provide accurate predictions, but nevertheless they were found to be useful for indicating the general process trends. In their research, Choi and Park (2001) proposed a hybrid system combining PCA and FFNN techniques for estimating the influent TKN concentration of an industrial ASP. Particularly, they found that the hybrid technique reduced the overfitting problems of the FFNN models. Lee et al. (2002) estimated MLSS, SS, COD and cyanide concentrations in a full-scale ASP treating cokes wastewater using several modelling techniques. In their conclusion, the authors suggested that a parallel hybrid technique combining phenomenological and FFNN models provided the best prediction performance. In another work aiming at approximating the process parameters of an ASP for cokes wastewater treatment, Lee et al. (2005) combined phenomenological model based on ASM1 and process knowledge with a number of PLS and ANN methods. As a result, they indicated a hybrid NNPLS model to achieve the best prediction performance, also providing an ability to detect and isolate process faults. In a soft-sensor proposal by Hong et al. (2007), the real-time estimation of NH_4–N, NO_3–N and PO_4–P concentrations in a laboratory-scale SBR was implemented by a combination of MPCA and FFNN with ARX structuring. The model with simple on-line measurements as the input data showed an excellent prediction capability with normal validation sets, whereas the prediction ability was not satisfactory with abnormal data sets employed. Kim et al. (2009) integrated MPCA and FFNN methodologies for soft-sensor modelling of the effluent COD, TN and TP concentrations in a municipal ASP. Specifically, they utilized MPCA for extracting information from the one-day process data batches to be used for the input variable selection for the FFNN model. The testing results showed that the predictions were not accurate, but they could follow the trends of the measured data. Presumably, more frequent data sets would have helped in catching the daily variations in the process and, thus, in increasing the estimation performance. In their work, Rustum (2009) investigated a combination of SOM and ANFIS methods for estimating effluent BOD and SS concentrations in a municipal ASP. The SOM model was shown to be able to extract features from noisy data and eliminate the effect of the missing values that the ANFIS model was not capable of dealing with.

4.2. Process monitoring and process fault detection

Various methodologies have been employed in the publications concentrating on the process monitoring and process fault detection in WWTPs. The reviewed studies targeting for monitoring the treatment processes are summarized in Table 3, where the modelling methods, plant types and treatment processes are presented. Also, the main features of a few conference publications have been included in the table (marked with an asterisk), but those have not been reviewed in this paper.

Many of the early soft-sensor applications in wastewater treatment considered monitoring bulking sludge episodes, which diminish the effluent wastewater quality and disturb the operation of the ASP. For instance monitoring SVI or its variants, that describe sludge settling properties and indicate presence of the bulking sludge, have been the goal of several studies. Taking into account the fact that the bulking sludge episodes do not usually appear rapidly, it is reasonable that their detection was among the first

Table 3
Process monitoring and process fault detection applications of soft-sensors in the reviewed publications. M = municipal, I = industrial, L = laboratory-scale, P = pilot-scale, S = simulated, ✓ = practical implementation, * = conference publication.

Publication	Method(s)	Appl.	Process
Rosen and Lennox (2001)	PCA, APCA, MSPCA	M	ASP
Lee et al. (2008)	PCA, APCA	M ✓	ASP
Multivariate statistics			
Rosen and Olsson (1998)	PCA, PLS	M	ASP
Teppola et al. (1999a)	PCA + FCM	I	ASP
Tomita et al. (2002)	PCA	S	ASM1
Olsson et al. (2003)	PCA	M	ASP
Miettinen et al. (2004)	PCA, PARAFAC	M	LTP
Moon et al. (2009)	PCA + K-means	M	ASP
Teppola et al. (1997)	PLS + autocorrelation function	I	ASP
Teppola et al. (1998)	PLS + FCM	I	ASP
Mujunen et al. (1998)	PLS	I	ASP
Teppola and Minkkinen (1999)	PLS + FCM/PCM	I	ASP
Rosen and Yuan (2001)	APCA + FCM	S	BSM1
Rosen et al. (2002)	APCA + FCM	S	BSM1
Lennox and Rosen (2002)	APCA, AdMSPCA	M	ASP
Rosen et al. (2003)	PCA, APCA	M	ASP
Yoo et al. (2003)	FPCR, TSK fuzzy regression, PCR	I	ASP
Aguado and Rosen (2008)	PCA, APCA, MPCA	S	BSM1_LT
Aguado et al. (2007b)	MPLS	L	SBR
Aguado et al. (2007a)	MPCA	L	SBR
Villez et al. (2008)	MPCA + LAMDA	P	SBR
Villez et al. (2010)	Hotelling's T^2	P	SBR
Ruiz et al. (2006)*	MPCA	P	SBR
Zhao et al. (2006)*	EMPCA	S	BSM1
Neural networks			
Fuente and Vega (1999)	FFNN	M	ASP
West and Mangiameli (2000)	FFNN, RBFN, Fuzzy ARTMAP	M	ASP
Baruch et al. (2005)	RTNN	S	ASP
Baruch and Mariaca-Gaspar (2009)	Dynamic RTNN	S	ASP
Brault et al. (2010)	FFNN	I	ASP
Hong et al. (2003)	SOM	M	ASP
Çinar (2005)	SOM	M	ASP
González and López García (2006)	SOM	I ✓	SBR
Heikkinen et al. (2011)	SOM	I	ASP
Fuzzy			
Marsili-Libelli and Müller (1996)	FCM	P	AD
Marsili-Libelli (2006)	DWT + GK	L	SBR

process fault monitoring interests, as the data-derived models were substantially dependent on the laboratory measurements. Together with the increase of the real-time measurements in WWTPs, the proposals for process monitoring have focused on a wider range of anomalies, which appear more abruptly. In addition, the monitoring of the states in SBRs has been a popular research objective, particularly, aiming at optimized control on the lengths of the phases. The usual secondary variables used for monitoring tasks are flow rates, pH, temperature and DO, SS and nutrient concentrations measured in different locations of the process. Most typically, multivariate methods and the SOM combined with clustering algorithms have been applied for process monitoring purposes.

The actual pioneering work on the applications of multivariate methods in soft-sensor design in the wastewater treatment sector has been done for monitoring ASPs in the pulp and paper industry by Mujunen (1999) and Teppola (1999) and for monitoring and control in municipal ASPs by Rosen (2001). The multivariate methods have been employed for soft-sensor development in these works and especially Rosen (2001) developed advanced techniques that were shown to be more adequate than the conventional techniques for the field of operation. For instance, Rosen and Lennox (2001) tested Adaptive PCA (APCA) and Multiscale PCA (MSPCA) for monitoring a municipal ASP. As model inputs, the authors used 14 variables including flow rate, pH, ammonia, temperature and air-valve positions at several points in the plant. The researchers concluded that both techniques had potential to overcome the difficulties associated with the changing process conditions. The MSPCA method separating the data in different time scales is more complex than APCA, but it was shown to provide more information about the process disturbances. As an example of a practical implementation, Lee et al. (2008) applied APCA based on the moving window technique for the real-time remote monitoring of small-scale municipal ASPs. They noticed that the APCA models overcame the problem of evolving dynamics and reduced the number of false alarms significantly. The models were used to detect abnormal process behaviour and to provide the plant operators with an early warning when such situations were identified.

4.2.1. Multivariate statistical models

Among multivariate statistics, the conventional PCA technique has been used for process state identification especially in earlier applications. Rosen and Olsson (1998) applied the PCA score plots for monitoring operational states in a full-scale municipal ASP. First, they built a PCA model using five input variables and samples corresponding to normal operating conditions. After that, deviations from the normal process behaviour were monitored using a two-dimensional score plot. Teppola et al. (1999a) combined PCA and FCM clustering algorithms for monitoring and visualizing process states and seasonal fluctuations in an ASP treating wastewater from a paper mill. The researchers were able, for instance, to detect and isolate an excess addition of phosphorus compounds with the proposed methodology. In their investigation, Tomita et al. (2002) proposed the PCA method for the analysis and disturbance detection in a modelled ASP based on the ASM1 protocol and they found three groups of process variables that characterized the system behaviour. Olsson et al. (2003) used PCA for monitoring the operational states in a municipal ASP as one example of using information technology for decision support purposes. Specifically, four process states were identified based on the training data and a score plot was applied for monitoring the shifting of the new data between the clusters representing the different operational states. A study by Miettinen et al. (2004), applied PCA and Parallel Factor Analysis (PARAFAC, Bro, 1997) for monitoring a multistage biological LTP for municipal wastewater. Their results established that the operational states and the ponds, where particular reactions occurred, could be identified using both of the methods. Moon et al. (2009) presented a methodology for the identification of the process states in a municipal ASP by means of PCA and K-means clustering. They identified five operational groups and showed that the operational map could provide visual information on the dynamic trends of the process states.

The conventional PLS combined with an auto-correlation function was used for modelling DSVI and the COD, nitrogen and phosphorus reductions in the ASP for the treatment of paper mill wastewater in order to detect the various process shifts by Teppola et al. (1997). Although the model was able to estimate DVSI well, it did not provide good estimates for the reductions, which was due to data obtained from single daily samples or daily averages of the continuous measurements. Nevertheless, the authors concluded that almost in every case, the disturbance was successfully isolated and the reasons for it were effectively analysed. In another soft-sensor application, Teppola et al. (1998) combined a PLS with a FCM clustering aiming at novel monitoring tools for an ASP of a paper mill. They applied the methodology successfully for monitoring DSVI and, for instance, recognized seasonal variations in the process. A study by Mujunen et al. (1998) employed PLS for monitoring SVI and DVSI, and the COD, TN and TP concentrations and their reductions in three ASPs treating wastewater from the pulp and paper industry. The authors concluded that a higher sampling frequency or new input variables would have been needed for modelling the peak values successfully. Rosen and Olsson (1998) demonstrated the use of PLS for monitoring the operational states of a municipal WWTP when considering the effluent turbidity as the model output. In their work, the process variables associated to the disturbances were isolated by using contribution plots. Teppola and Minkkinen (1999) employed combinations of PLS and clustering methods for monitoring the ASP of a paper mill. Particularly, they used the latent variables of PLS analysis as inputs for the FCM and Possibilistic C-means (PCM, Krishnapuram and Keller, 1993) clustering algorithms.

Adaptive extensions of multivariate techniques have been proposed for WWTP monitoring in a number of publications. The work of Rosen and Yuan (2001) concentrated on APCA and FCM clustering methods combined for monitoring and control set-point definition purposes in a case study where a simulated step-feed ASP was used. In particular, they used the influent data of a preliminary version of BSM1 modified with an extreme NH_4–N load disturbance for the simulation. Supervisory control strategies were then successfully used under normal conditions, storm with sewer flush-out, storm, rain and extreme NH_4–N load. Later, Rosen et al. (2002) utilized the same process state estimation approach in a proposal for a predictive supervisory controller during extreme events using a modified BSM1 as test bench. Adaptive Multiscale PCA (AdMSPCA) extension was successfully applied for monitoring a municipal ASP by Lennox and Rosen (2002). When compared with the APCA, the authors observed AdMSPCA to have the ability to adapt to a much broader range of changes. A publication by Rosen et al. (2003) focused on the challenges of multivariate monitoring in municipal wastewater treatment. The authors did not find the PCA suitable for the monitoring task, but an adaptive scaling of the its parameters improved the results considerably. A study by Yoo et al. (2003) proposed Fuzzy PCR (FPCR) as an adaptive monitoring tool for an industrial ASP. The methodology was demonstrated to be able to distinguish between a large process change and a simple fault or a short disturbance. A study by Aguado and Rosen (2008) investigated successfully APCA and FCM for the efficient monitoring of the operational states and for tracing the most likely causes of the disturbances in municipal ASPs. The influent data of the BSM1_LT platform (Rosen et al., 2004), including a realistic set of process disturbances and failures, was used in the research.

Multiway methods have been popular in particular in the SBR case studies. Aguado et al. (2007b) used a MPLS to correlate several on-line variables with phosphorus removal efficiency in a laboratory-scale SBR aiming at EBPR. They identified conductivity as a suitable variable for detecting the process upsets associated with a negative effect on the phosphorus removal efficiency. A study by Aguado et al. (2007a) proposed MPCA models for fault monitoring in a laboratory-scale SBR operated for an EBPR purpose. The authors found the methodology to be straightforward and very consistent in the diagnosis of the detected process abnormalities. Villez et al. (2008) applied a combination of MPCA and Learning Algorithm for Multivariable Data Analysis (LAMDA, Aguilar-Martin and López de Mántaras, 1982) clustering techniques for an analysis of a pilot-scale SBR treating synthetic sewage. They showed the combined methodology to provide an efficient and robust tool for screening and interpreting the data from a batch process.

Another approach presented by Villez et al. (2010) employed the Hotelling's T^2 statistic for detection of the operational states in a pilot-scale SBR aiming at EBPR. They used the information obtained from the model with five input variables for the control of the process which reduced the length of the aerated phase by 41%. The authors also conclude that the proposed methodology is general in nature and, thus, not limited only for the reported application.

4.2.2. Artificial neural network models

As for the ANN methods, Fuente and Vega (1999) used the frequency content of the fault-indicating signals and a FFNN for process fault detection in a municipal ASP. In particular, faults in aeration turbines were monitored and the results demonstrated the reliability of the proposed method. West and Mangiameli (2000) compared the performances of different statistical models and ANN architectures for monitoring a municipal ASP. As a result, they found Radial Basis Function Network (RBFN, Haykin, 1999) to be the most competent method for detecting normal and abnormal process conditions. In their investigation, Baruch et al. (2005) used a nonlinear Recurrent Trainable Neural Network (RTNN, Baruch et al., 2001) for real-time state estimation and prediction of the biomass concentration in the recycle sludge flow of a simulated ASP. A RTNN was also employed for systems identification and state estimation for a simulated ASP by Baruch and Mariaca-Gaspar (2009). A work of Brault et al. (2010) presented a FFNN-based soft-sensor for monitoring the settling trends of the sludge, in particular the Stirred Sludge Volume Index (SSVI), in an ASP treating wastewater from a pulp and paper industry. The input variables included non-conventional parameters, such as an adenosine triphosphate (ATP) concentration, which was found to improve the model performance by providing with an early signal for sludge bulking.

The operating conditions and relationships between the process variables were monitored in a municipal ASP with the SOM by Hong et al. (2003). They identified five operational clusters and extracted information on the reasons for a poor performance from the properties of the clusters. Çinar (2005) used the SOM for analysing the causes for the high effluent concentrations in a municipal ASP. He established that low pH values were responsible for the high effluent BOD and SS concentrations, and a high solids retention time for an increased effluent fecal coliform concentration. A study by González and López García (2006) proposed a combination of the SOM and K-means clustering algorithms for monitoring operational states, especially the end-point of the aerobic phase, in a SBR treating coke wastewater. Also Heikkinen et al. (2011) applied the same combination of methods to detect the process states in an industrial ASP. As a result, they identified four operational clusters corresponding to summertime, abnormal situations such as stoppages, wintertime, and unstable operational states such as bulking and foaming.

4.2.3. Fuzzy systems

Fuzzy methods have also been proposed for process monitoring in wastewater treatment. Marsili-Libelli and Müller (1996) used FCM for monitoring operational states in a pilot-scale AD process that was used in parallel with the main anaerobic digestion stage located before the aerobic bioreactors in a municipal WWTP. The methodology was successful in recognizing the process shifting between operational states such as normal, toxic, overload and inhibition situations. The authors used biogas production and hydrogen concentration in off-gas as input variables and the data were selected from a collection of shock experiments. In another paper, Marsili-Libelli (2006) presented a pattern recognition system in a laboratory-scale SBR based on fuzzy clustering of the derivatives of three process signals that were denoised by using Discrete Wavelet Transform (DWT, Strang, 1994). In particular, he employed Gustafsson-Kessel fuzzy clustering algorithm (GK, Gustafson and Kessel, 1979) and the information provided by the procedure was used for the control of the durations of the phases in the process.

4.3. Sensor monitoring

Another task for soft-sensors in WWTPs is the validation, fault detection and diagnosis of the hardware instrumentation. The sensor monitoring applications in the reviewed publications and a few conference publications that are not reviewed in this paper (marked with an asterisk) are summarized in Table 4, where the modelling methods, plant types and treatment processes used in the case studies are collected.

Typically, advanced multivariate approaches have been applied to the identification of reasons for sensor faults, such as bias, drift, complete failure and precision degradation. Most commonly, monitoring of the NH_4–N and NO_3–N sensors in ASPs have been studied in the reviewed publications. The scarcity of the reviewed papers in this application area and the fact that all of them are relatively recent suggests that sensor monitoring in wastewater treatment facilities is still an emerging research field. In addition, data from a full-scale WWTP was used only in one of the publications whereas the majority of them considered simulated processes.

Lee et al. (2004) proposed a PCA-based sensor fault detection and isolation application using BSM1 as a test environment. They used a time-lagged APCA model to identify faults in seven sensors and tested two case studies: the precision degradation of an influent NH_4–N sensor and the bias in an influent flow rate sensor.

Table 4
Sensor monitoring applications of soft-sensors in the reviewed publications. M = municipal, I = industrial, L = laboratory-scale, S = simulated, * = conference publication.

Publication	Method(s)	Appl.	Process
Multivariate statistics			
Lee et al. (2004)	PCA, APCA	S	BSM1
Lee et al. (2006)	PCA, APCA	S	BSM1
Yoo et al. (2008)	PCA	L	SHARON
Baggiani and Marsili-Libelli (2009)	APCA	M	ASP
Lee et al. (2009)	PLS, MSPLS	I	AF
Corominas et al. (2009)*	PCA, MPCA	S	BSM1_LT
Tharrault et al. (2009)*	RPCA	I	ASP
Neural networks			
Caccavale et al. (2010)	FFNN	S	ASM1

In both of the cases, the APCA model detected and isolated faulty sensors clearly and consistently. An investigation by Lee et al. (2006) presented a Sensor Quality Index in a publication where APCA was applied for the detection of sensor faults in the BSM1 platform. Moreover, the studied fault scenarios included an influent NH_4–N sensor corrupted by a drifting fault, an influent flow rate sensor corrupted by a bias fault and a NO_3–N sensor in the ASP corrupted by a precision degradation fault. The authors indicated the proposed approach to perform well for sensor fault detection and to identify the faulty sensors efficiently in the dynamic process. However, a limitation of the presented APCA technique concerned its inability to identify the faulty sensors that cause process transitions, i.e. the situations when a faulty instrument is connected to a control loop. Yoo et al. (2008) investigated a PCA technique combined with fault indices for sensor fault identification and reconstruction. In particular, they successfully tested two realistic fault scenarios related to a DO sensor in a case study dealing with a laboratory-scale SHARON process used for the treatment of wastewater containing high concentrations of nitrogen. In a work by Baggiani and Marsili-Libelli (2009), APCA with differing moving window lengths were proposed for real-time fault detection and isolation in NH_4–N and NO_3–N sensors in an ASP treating municipal sewage and septic tank discharges. The faults were classified into three categories: sensor faults, spikes and process anomalies. The short-window model was capable of detecting all the faults, while the detection percentage of the long-window model decreased to 84%. In addition, the sensors responsible for the faults were isolated by using the conventional contribution plots. Lee et al. (2009) proposed a Multi-Scale PLS (MSPLS) algorithm combining PLS and wavelets for sensor fault detection in a biological AF process treating wastewater from a petrochemical industry. They found the fault detection ability of the MSPLS approach to be good and the methodology to properly diagnose the detected sensor faults and to provide scale-level information about the fault characteristics.

In an ANN-based application, Caccavale et al. (2010) employed a FFNN for the detection and diagnosis of NH_4–N and NO_3–N sensor faults in a simulated alternated aeration ASP. They showed that the faulty sensors were correctly isolated and information on the typology of the fault could be obtained by analysing the shape of the corresponding residuals between the predicted outputs and sensor measurements.

5. Discussion

5.1. Applications in the wastewater treatment facilities

Recently, the majority of the soft-sensor applications in wastewater treatment have been associated with prediction tasks. This can be observed in Fig. 3, where the shares of the on-line prediction, process monitoring and sensor monitoring applications in the reviewed publications have been divided into the studies published in 2005 or earlier and the studies published after 2005. The need for reliable on-line predictions indicates the importance of efficient process monitoring and control in WWTPs under strict treatment regulations. The variables describing the content of organic matter that are challenging to measure in real-time have been the most common choices for the predicted model outputs in the reviewed publications. Also, in an extensive number of soft-sensor applications, nitrogen compounds have been selected for the estimated output variables. This is the case especially in more recent studies, which shows the trend that present-day municipal treatment processes are being designed, in particular, for nutrient removal. In addition, data-derived models for sensor fault monitoring have been proposed in recent publications. However, the rather small number of existing publications suggests that the research area is relatively new in wastewater treatment.

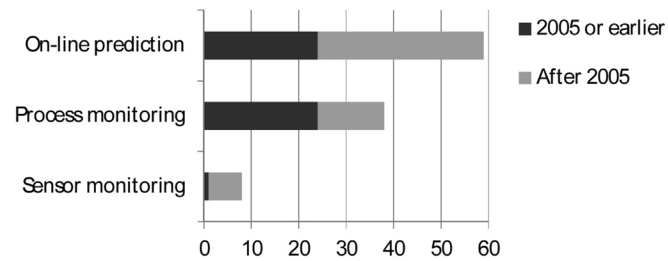

Fig. 3. The soft-sensor applications in the reviewed publications.

As for the process types in the case studies shown in Fig. 4a, municipal and industrial WWTPs corresponded for 38% and 25%, respectively. In addition, pilot- and laboratory-scale processes were used in 18% of investigations, most of them being SBRs. The share of the simulated processes in the reviewed papers papers was 19%, where the BSM platforms were the most popular virtual test environments. According to the data descriptions in the studies, it appears that in the municipal WWTPs, a larger amount of process variables measured on-line are available for modelling purposes compared with the industrial WWTPs. Even though biological treatment processes are common in both municipal and industrial plants, it is obvious that the differences in the composition of influent wastewater, especially regarding the nutrient concentrations, give rise to more complex treatment processes and, hence, more versatile monitoring and control requirements in the municipal facilities. In spite of that, interesting data-derived soft-sensor applications have been proposed also for industrial wastewater treatment for a wide range of tasks.

The useful soft-sensor applications ease with the operation of the treatment processes, for instance, by providing beneficial monitoring tools and helping in reducing the operational costs or reaching the treatment requirements. The desired monitoring tools often depend on the type of the process. For example, considering SBRs, they typically relate to the recognition of the optimal lengths of the phases, while in ASPs they may concern the detection of abnormal process states. As for the cost reduction, the on-line variable estimates used for supporting the process control may give advantages, e.g., in avoiding excessive aeration or chemical dosing. In the modern-day plants, soft-sensors for such purposes typically aim at the prediction of nutrient concentrations in various locations of the treatment line, or at monitoring the hardware instruments measuring the nutrient contents. On the other hand, the

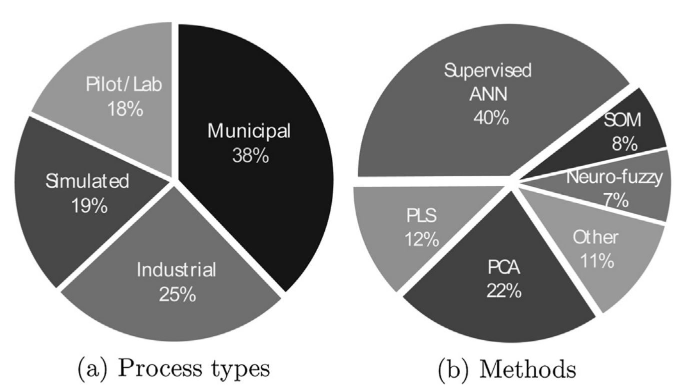

Fig. 4. The wastewater treatment process types (a) and the methods for soft-sensor design (b) in the reviewed publications.

treatment requirements vary depending, e.g., on the local legislation, the size of the plant and the characteristics of the wastewater, which affect case-specifically the types of useful soft-sensors. For instance, the shifting of the typical treatment requirements from organic matter to nutrients in the municipal sector during the recent decades has also made soft-sensor estimates of nutrient concentrations more preferable. In the industrial sector, the removal of compounds specific to the field of industry is regulated by the authorities. Therefore, the estimates of such specific compounds has been of special interest in the soft-sensor design and shown to be useful for the plant operation. The appropriate modelling methodologies for soft-sensor design are typically case-specific and they depend on the problem and process considered as well as the available process data.

5.2. Modelling approaches used for the soft-sensor design

Different multivariate and ANN methodologies have typically been used for the data-driven soft-sensor development in wastewater treatment, with the supervised ANN techniques having been the most popular. The shares of the data-derived methods applied for soft-sensor design in the reviewed publications are presented in Fig. 4b. The supervised ANN techniques and the unsupervised SOM techniques were used in 40% and 8% of the reviewed studies, respectively. Multivariate methods were applied for soft-sensor design in 36% of the publications, divided in particular in PLS (12%), PCA (22%) and PCR (2%) techniques. Moreover, neuro-fuzzy methods were used in 7% of the publications.

The most applied methods used for different tasks in the continuous and batch processes in the studies conducted after the year 2005 have been collected in Table 5. Even though the popularity of the modelling methods among the research published during the recent years does not mean that the considered methods would be superior compared with the others methods, it gives a hint for which tasks and process types these methods have been found to be adequate. The FFNNs have been by far the most popular methods proposed for the estimation tasks, corresponding for 58% and 50% of the recent studies concerning the continuous and batch processes, respectively. Other popular methods for the prediction tasks in the publications dealing with the continuous processes have been ANFIS and various PLS approaches. Different PCA techniques have been applied in 57% of the studies targeting for process monitoring in the continuous treatment process after 2005. Among them, the APCA variants have been the most popular. As for the monitoring of the batch processes, the MPCA approach has been employed in 60% of the recent papers. In the limited number of papers where sensor monitoring has been studied, the researchers have used various PCA methods in the continuous processes in 71% of the works reported after the year 2005. None of the reviewed publications targeted for sensor monitoring in SBRs.

Due to the dynamic and nonlinear nature of the wastewater characteristics and treatment process conditions, the conventional linear multivariate techniques have often been found to be unsatisfying methods for soft-sensor development in the wastewater sector as such. Therefore, a number of adaptive and nonlinear PCA and PLS extensions have been proposed and shown to be more feasible for soft-sensor design in the reviewed papers. In particular, adaptive PCA and PLS approaches based on, e.g., moving window or recursive techniques have been demonstrated to overcome the difficulties associated with the changing process conditions, however, with an increased computational cost. An additional challenge in the adaptive methods based on the moving window techniques is the choice of an adequate time window length. As for the nonlinear extensions, researchers have established, for instance, FPLS and KPLS to be adequate for prediction tasks in biological WWTPs. On the other hand, in all the case studies the performance of nonlinear methods has not been shown superior to the simpler linear methods. In those cases, employing the simpler methods is suggested due to their lighter computational burden. The researchers have also indicated the multiway extensions of PCA and PLS useful, in particular, for the monitoring and analysis of SBRs, and found the multiscale approaches to extract the features of treatment processes in different time-scales. In addition, the conventional and multiway PCA methods have been popular preprocessing techniques applied in 10% of the ANN-based soft-sensors proposed in the studies, which indicates the potential of the PCA techniques in the compression of information.

In the publications where the performances of multivariate and supervised ANN methods have been compared for prediction tasks in WWTPs, usually the ANN methods have been found to be more feasible. Especially FFNNs with various architectures and back-propagation training algorithms have been popular among the soft-sensor developers. Even though in most of the case studies the FFNN estimates in the biological plants have been shown accurate, some drawbacks of the methodology have been notified. Those include the extent of the training data required and the challenges in simulating outputs outside the range of the training data. For instance, the FFNN estimates resulting only in satisfactory accuracies due to an insufficient amount of training data has been reported in several publications. Moreover, the structure of the ANNs is not easily interpretable and, therefore, they are not very useful for learning interactions between the process variables. Another challenge with FFNN modelling is defining the topology of the network, which is often done based on trial and error. In addition to the supervised ANNs, the researchers have applied unsupervised SOM models successfully for monitoring, analysis and preprocessing tasks in a number of publications.

In several case studies, neuro-fuzzy models such as ANFIS have outperformed the supervised ANN models when tested for prediction accuracy in the dynamic conditions of WWTPs. Also, the structure of ANFIS may be easier to interpret than, e.g., the structure of FFNNs due to the IF-THEN rules, but a drawback of the ANFIS methodology is a strong computational power required. Hybrid models, where different modelling methods are fused on the system level, have been shown to be efficient tools for soft-sensor tasks in many of the reviewed publications. Particularly, the hybrid techniques have been indicated to enhance the strengths of the individual modelling methods and to overcome their limitations. The features, pros and cons of the basic data-derived methods used for soft-sensor design in the reviewed publications are summarized in Table 6.

Table 5
The most popular method employed for different tasks in the continuous and batch processes in the reviewed papers published after 2005.

Task	Continuous	Batch
On-line prediction	FFNN, ANFIS, PLS variants	FFNN
Process monitoring	APCA, other PCA variants	MPCA
Sensor monitoring	PCA variants	No papers aiming at this task

5.3. Factors limiting the use of soft-sensors and needs for the future research

Even though full scale WWTP data has been used in most of the case studies, only a small minority of the publications reported a practical implementation of the proposed methods (see Tables 2–4). This implies that many of the studies have been done mainly for academic purposes. In other words, there is still clearly a gap to

Table 6
The features, pros and cons of the basic data-derived methods used for soft-sensor design in the reviewed publications.

Method	Features	Pros	Cons
Multivariate statistics	Dimensionality reduction, removing multi-collinearity, describes relationships between variables	Potential in pre-processing, data inspection, simplicity, compression of information, classification, regression, interpretability	Linearity, stationarity, assumption of normal distribution, challenges with selecting number of PCs/LVs
Artificial neural networks	Supervised or unsupervised learning, number of layers and neurons paradigm, different training algorithms	Adaptivity, nonlinearity, regression, classification	Challenges in choosing network topology, risks of over-fitting, local minimum, interpretability of structure
Neuro-fuzzy systems	Combination of human-like reasoning and ANNs, different inference systems and configurations	Nonlinearity, adaptivity, regression, IF-THEN rules easy to interpret	Challenges in choosing network topology, computational heaviness, challenges with noisy data

be narrowed between the motivations of academia and the needs of the people working in the plants. This problematic is discussed, e.g., in a recent subjective review by Olsson (2012). It is likely that especially in the smaller plants where the number of the technically educated employees is lesser, a limiting factor for soft-sensor implementations is the insecurity caused by unfamiliar solutions. In such cases, it may feel more confident for the operators to rely on the conventional hardware sensors and their maintenance provided by the instrument suppliers when available. Instead, if the plant operators have more technical competence and they are research-orientated, it is more likely that they are open-minded for novel techniques such as soft-sensors. In the best situations, the operators are actually involved in the soft-sensor development, they have recognized problems that can potentially be solved by soft-sensing and, therefore, they are willing to test and implement these techniques (see e.g., Äijälä and Lumley, 2006; Cecil and Kozlowska, 2010). Another issue that limits the use of data-derived soft-sensors often lies with the education of the engineers who specialize in water and wastewater treatment. Even though huge amounts of data are measured in the modern WWTPs, the current curricula in many universities ignores statistical and artificial intelligence techniques that would be useful in handling the data. The reports on the successful soft-sensor implementations in WWTPs will most likely be the best promotion for these techniques and increase the interest of plant engineers and consultants to adopt the data-driven technologies. However, it should be borne in mind that for practical solutions, the soft-sensing techniques need to be applied in such a way that their use is straightforward for the operators in order to make the soft-sensors attractive.

Based on these considerations, the future research targeting for practical implementations of soft-sensors in WWTPs would be important for increasing the awareness of these alternatives to the conventional measurement and monitoring solutions. The research related to soft-sensors should also follow the development of the state-of-art hardware sensors, concerning their dependability, feasibility and expense, and the typical measurements used for the process control. Thus, the primary variables whose measurement reliabilities may be bottlenecks in the efficient process operation can be recognized along the technical development that will take place. Based on this knowledge, soft-sensors for the relevant applications can be designed. Moreover, the research should focus on solving the real-life problems in WWTPs rather than developing complicated methodologies motivated by theoretical interests. This is also to say that the soft-sensor solutions should be selected based on the problem at hand and they should be relatively simple.

6. Conclusion

In this review, we focused on data-derived soft-sensor applications in biological wastewater treatment facilities. After introducing briefly the treatment processes used in the reviewed case studies and the typical characteristics of process data in WWTPs, a general guideline for soft-sensor design was provided. The presented soft-sensor applications were divided into prediction, process monitoring and process fault detection, and sensor monitoring categories according to the main objectives of the soft-sensor design.

To summarize, the data-driven soft-sensors have become more popular as a greater number of real-time measurements have been applied in municipal and industrial WWTPs. Typical soft-sensor applications in the treatment plants include the prediction of the primary process variables that are hard to measure reliably or with reasonable costs using hardware instrumentation. Often, these types of soft-sensors are proposed for estimating the concentrations of nutrients or organic matter. Due to the increased amount of real-time process measurements, monitoring individual process variables has become troublesome using, for instance, univariate control charts. Hence, applications where a large amount of data is compressed into visual low-dimensional monitoring tools that are informative and easy to interpret has become valuable for plant operators for the recognition of the operational states. Another purpose of the data-driven soft-sensors is hardware instrumentation monitoring and providing a back-up system during instrument down-time.

The data-derived techniques typically used for soft-sensor design in wastewater treatment include different multivariate statistical and ANN methods. According to the reviewed studies, the conventional approaches of PCA and PLS are usually not able to satisfactorily catch the dynamic and nonlinear behaviour of the biological WWTPs. For this reason, the adaptive and nonlinear extensions of multivariate methods have been proposed for modelling treatment processes in many of the recent studies. As for the batch processes, multiway PCA and PLS have been widely used methods. ANNs, in particular the FFNN methods employing back-propagation learning algorithms, have been the most popular techniques applied for data-driven soft-sensor development in the reviewed publications, especially for estimation tasks. In addition, neuro-fuzzy systems and hybrid methods have become more common in wastewater treatment modelling applications.

Acknowledgement

Henri Haimi and Michela Mulas kindly acknowledge the Finnish “Maa- ja vesitekniikan tuki ry.” for their financial support. Maa- ja vesitekniikan tuki ry. is a non-profit association that supports the research and education in the fields of water engineering, environmental engineering and soil conservation.

References

Aarnio, P., Minkkinen, P., 1986. Application of partial least-squares modelling in the optimization of a waste-water treatment plant. Anal. Chim. Acta 191, 457–460.

Aguado, D., Rosen, C., 2008. Multivariate statistical monitoring of continuous wastewater treatment plants. Eng. Appl. Artif. Intell. 53, 1080–1091.
Aguado, D., Ferrer, A., Seco, A., Ferrer, J., 2006. Comparison of different predictive models for nutrient estimation in a sequencing batch reactor for wastewater treatment. Chemometr. Intell. Lab. 84, 75–81.
Aguado, D., Ferrer, A., Ferrer, J., Seco, A., 2007a. Multivariate SPC of a sequencing batch reactor for wastewater treatment. Chemometr. Intell. Lab. 85, 82–93.
Aguado, D., Zarzo, M., Seco, A., Ferrer, A., 2007b. Process understanding of a wastewater batch reactor with block-wise PLS. Environmetrics 18, 551–560.
Aguado, D., Ribes, J., Montoya, T., Ferrer, J., Seco, A., 2009. A methodology for sequencing batch reactor identification with artificial neural networks: a case study. Comput. Chem. Eng. 33, 465–472.
Aguilar-Martin, J., López de Mántaras, R., 1982. The process of classification and learning the meaning of linguistic descriptors of concepts. In: Gupta, M., Sanchez, E. (Eds.), Approximate Reasoning in Decision Analysis. North-Holland Publishing Company, New York, pp. 165–175.
Äijälä, G., Lumley, D., 2006. Integrated soft sensor model for flow control. Wat. Sci. Tech. 53, 473–482.
Anderson, T., 2003. An Introduction to Multivariate Statistical Analysis, third ed. John Wiley & Sons.
Ansari, A.A., Gill, S.S., Lanza, G.R., Rast, W., 2010. Eutrophication: Causes, Consequences and Control. Springer.
Atkeson, C., Moore, A., Schaal, S., 1997. Locally weighted learning. Artif. Intell. Rev. 11, 11–73.
Baffi, G., Martin, E., Morris, A., 1999. Non-linear projection to latent structures revisited: the quadratic PLS algorithm. Comput. Chem. Eng. 23, 395–411.
Baggiani, F., Marsili-Libelli, S., 2009. Real-time fault detection and isolation in biological wastewater treatment plants. Water Sci. Technol. 60, 2949–2961.
Banadda, N., Nhapi, I., Kimwaga, R., 2011. A review of modeling approaches in activated sludge systems. Afr. J. Environ. Sci. Technol. 5, 397–408.
Bang, Y., Yoo, C., Lee, I.B., 2003. Nonlinear PLS modeling with fuzzy inference system. Chem. Int. Lab. Sys. 64, 137–155.
Baruch, I.S., Mariaca-Gaspar, C.R., 2009. A Levenberg–Marquardt learning applied for recurrent neural identification and control of a wastewater treatment bioprocess. Int. J. Intell. Syst. 24, 1094–1114.
Baruch, I., Flores, J., Thomas, F., Garrido, R., 2001. Adaptive neural control of nonlinear systems. In: Dorffner, G., Bischof, H., Hornik, K. (Eds.), Artificial Neural Networks – ICANN 2001. Springer, Berlin, pp. 930–936.
Baruch, I.S., Georgieva, P., Barrera-Cortes, J., 2005. Adaptive recurrent neural network control of biological wastewater treatment. Int. J. Intell. Syst. 20, 173–193.
Belanche, L.A., Valdés, J.J., Comas, J., Roda, I.R., Poch, M., 1999. Towards a model of input–output behaviour of wastewater treatment plants using soft computing techniques. Environ. Model. Softw. 14, 409–419.
Belanche, L., Valdés, J.J., Comas, J., Roda, I.R., Poch, M., 2000. Prediction of the bulking phenomenon in wastewater treatment plants. Artif. Intell. Eng. 14, 307–317.
Bezdek, J., Ehrlich, R., Full, W., 1984. FCM: the fuzzy c-means clustering algorithm. Comput. Geosci. 10, 191–203.
Bisschops, I., Spanjers, H., Keesman, K., 2006. Automatic detection of exogenous respiration end-point using artificial neural network. Water Sci. Technol. 53, 273–281.
Blom, H.A., 1996. Indirect measurement of key water quality parameters in sewage treatment plants. J. Chemometr. 10, 697–706.
Boger, Z., 1992. Application of neural networks to water and wastewater treatment plant operation. ISA T. 31, 25–33.
Bongards, M., 2001. Improving the efficiency of a wastewater treatment plant by fuzzy control and neural networks. Water Sci. Technol. 43, 189–196.
Borowa, A., Brdys, M.A., Mazur, K., 2007. Modelling of wastewater treatment plant for monitoring and control purposes by state – space wavelet networks. Int. J. Comput. Commun. Control 2, 121–131.
Brault, J.M., Labib, R., Perrier, M., Stuart, P., 2010. Prediction of activated sludge filamentous bulking using ATP data and neural networks. Can. J. Chem. Eng. 9999, 1–13.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.
Bro, R., 1997. PARAFAC: tutorial and applications. Chemometr. Intell. Lab. 38, 149–171.
Caccavale, F., Digiulio, P., Iamarino, M., Masi, S., Pierri, F., 2010. A neural network approach for on-line fault detection of nitrogen sensors in alternated active sludge treatment plants. Water Sci. Technol. 62, 2760–2768.
Capodaglio, A.G., Jones, H.V., Novotny, V., Feng, X., 1991. Sludge bulking analysis and forecasting: application of system identification and artificial neural computing technologies. Water Res. 25, 1217–1224.
Caraman, S., Sbarciog, M., Barbu, M., 2007. Predictive control of a wastewater treatment process. Int. J. Comput. Commun. Control 2, 132–142.
Cecil, D., Kozlowska, M., 2010. Software sensors are a real alternative to true sensors. Environ. Model. Softw. 25, 622–625.
Choi, D.J., Park, H., 2001. A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process. Water Res. 35, 3959–3967.
Çinar, Ö., 2005. New tool for evaluation of performance of wastewater treatment plant: artificial neural network. Process. Biochem. 40, 2980–2984.
Çinar, Ö., Hasar, H., Kinari, C., 2006. Modeling of submerged membrane bioreactor treating cheese whey wastewater by artificial neural network. J. Biotech. 123, 204–209.
Civelekoglu, G., Perendeci, A., Yigit, N.O., Kitis, M., 2007. Modeling carbon and nitrogen removal in an industrial wastewater treatment plant using an adaptive network-based fuzzy inference system. Clean – Soil, Air, Water 35, 617–625.
Clara, N., 2008. Neural networks complemented with genetic algorithms and fuzzy systems for predicting nitrogenous effluent variables in wastewater treatment plants. WSEAS Trans. Syst. 7, 695–705.
Cohen, A., Janssen, G., Brewster, S.D., Seeley, R., Boogert, A.A., Graham, A.A., Mardani, M.R., Clarke, N., Kasabov, N.K., 1997. Application of computational intelligence for on-line control of a sequencing batch reactor (SBR) at Morrinsville sewage treatment plant. Water Sci. Technol. 35, 63–71.
Copp, J., 2002. The COST Simulation Benchmark: Description and Simulator Manual. Technical Report. Office for official publications of the European Communities, Luxembourg.
Corominas, L., Villez, K., Aguado, D., Rosén, C., Rieger, L., Vanrolleghem, P.A., 2009. Evaluation of fault-detection strategies performance in wastewater treatment processes. In: Proc. 10th IWA Conference on Instrumentation, Control and Automation, Cairns, Australia.
Corona, F., Mulas, M., Haimi, H., Sundell, L., Heinonen, M., Vahala, R., 2013. Monitoring nitrate concentrations in the denitrifying post-filtration unit of a municipal wastewater treatment plant. J. Process. Contr. 23, 158–170.
Côté, M., Grandjean, B.P.A., Lessard, P., Thibault, J., 1995. Dynamic modelling of the activated sludge process: improving prediction using neural networks. Water Res. 29, 995–1004.
Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.
de Vries, B., Principe, J., 1991. A theory for neural networks with time delays. In: Lippmann, R., Moody, J., Touretzky, D. (Eds.), Advances in Neural Information Processing Systems 3. Morgan Kaufmann Publishers Inc., San Mateo, pp. 162–168.
Dellana, S.A., West, D., 2009. Predictive modeling for wastewater applications: linear and nonlinear approaches. Environ. Model. Softw. 24, 96–106.
Delnavaz, M., Ayati, B., Ganjidoust, H., 2010. Prediction of moving bed biofilm reactor (MBBR) performance for the treatment of aniline using artificial neural networks (ANN). J. Hazard. Mater. 179, 769–775.
Dochain, D., Vanrolleghem, P.A., 2001. Dynamical Modelling and Estimation in Wastewater Treatment Processes. IWA Publishing.
Douglas, J., Usländer, T., Schimak, G., Esteban, J., Denzer, R., 2008. An open distributed architecture for sensor networks for risk management. Sensors 8, 1755–1773.
Duda, R., Hart, P., Stork, D., 2000. Pattern Classification, second ed. John Wiley & Sons.
Dürrenmatt, D., Gujer, W., 2012. Data-driven modeling approaches to support wastewater treatment plant operation. Environ. Model. Softw. 30, 47–56.
Elman, J., 1990. Finding structure in time. Cogn. Sci. 14, 179–211.
Everitt, B.S., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis, fifth ed.. In: Wiley's Series in Probability and Statistics John Wiley & Sons.
Fernandez, F.J., Seco, A., Ferrer, J., Rodrigo, M.A., 2009. Use of neurofuzzy networks to improve wastewater flow-rate forecasting. Environ. Model. Softw. 24, 686–693.
Fisher, R., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7, 179–188.
Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G., 2007. Soft Sensors for Monitoring and Control of Industrial Processes. In: Advances in Industrial Control. Springer.
Freund, Y., Schapire, R., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55, 119–139.
Fu, C., Poch, M., 1995. System identification and real-time pattern recognition by neural networks for an activated sludge process. Environ. Int. 21, 57–69.
Fuente, M.J., Vega, P., 1999. Neural networks applied to fault detection of a biotechnological process. Eng. Appl. Artif. Intell. 12, 569–584.
Fujiwara, K., Kano, M., Hasebe, S., Takinami, A., 2009. Soft-sensor development using correlation-based just-in-time modeling. AIChE J. 55, 1754–1765.
Fullér, R., 2000. Introduction to Neuro-fuzzy Systems. In: Advances in Intelligent and Soft Computing, vol. 2. Springer.
Gernaey, K.V., van Loosdrecht, M.C.M., Henze, M., Lind, M., JØrgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. Environ. Model. Softw. 19, 763–783.
González, I.M., López García, H., 2006. End-point detection of the aerobic phase in a biological reactor using SOM and clustering algorithms. Eng. Appl. Artif. Intell. 19, 19–28.
Grau, P., Beltrán, S., de Gracia, M., Ayesa, E., 2007. New mathematical procedure for the automatic estimation of influent characteristics in WWTPs. Water Sci. Technol. 56, 95–106.
Grieu, S., Thiery, F., Traoré, A., Nguyen, T.P., Barreau, M., Polit, M., 2006. KSOM and MLP neural networks for on-line estimating the efficiency of an activated sludge process. Chem. Eng. J. 116, 1–11.
Gustafson, E., Kessel, W., 1979. Fuzzy clustering with a fuzzy covariance matrix. In: Proc. of IEEE Conference on Decision and Control, San Diego, USA.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.
Häck, M., Köhne, M., 1996. Estimation of wastewater process parameters using neural networks. Water Sci. Technol. 33, 101–115.
Hamed, M.M., Khalafallah, M.G., Hassanien, E.A., 2004. Prediction of wastewater treatment plant performance using artificial neural networks. Environ. Model. Softw. 19, 919–928.
Hamoda, M.F., Al-Ghusain, I.A., Hassan, A.H., 1999. Integrated wastewater treatment plant performance evaluation using artificial neural networks. Water Sci. Technol. 40, 55–65.
Hanbay, D., Turkoglu, I., Demir, Y., 2007. Prediction of chemical oxygen demand (COD) based on wavelet decomposition and neural networks. Clean – Soil, Air, Water 35, 250–254.

Hartigan, J., Wong, M., 1979. A k-means clustering algorithm. Appl. Stat. 28, 100–108.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed.. In: Springer Series in Statistics Springer.
Hauduc, H., Gillot, S., Rieger, L., Ohtsuki, T., Shaw, A., Takács, I., Winkler, S., 2009. Activated sludge modelling in practice: an international survey. Water Sci. Technol. 60, 1943–1951.
Hauduc, H., Rieger, L., Ohtsuki, T., Shaw, A., Takács, I., Winkler, S., Héduit, A., Vanrolleghem, P.A., Gillot, S., 2011. Activated sludge modelling: development and potential use of a practical applications database. Water Sci. Technol. 63, 2164–2182.
Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, second ed. Prentice Hall.
Heikkinen, M., Poutiainen, H., Liukkonen, M., Heikkinen, T., Hiltunen, Y., 2011. Subtraction analysis based on self-organizing maps for an industrial wastewater treatment process. Math. Comput. Simulat. 82, 450–459.
Henze, M., Gujer, W., Mino, T., van Loosedrecht, M., 2000. Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. Scientific and Technical Reports, No. 9. IWA Publishing.
Hill, D., Minsker, B., 2010. Anomaly detection in streaming environmental data: a data-driven modeling approach. Environ. Model. Softw. 25, 1014–1022.
Hill, D., Minsker, B., Amir, E., 2009. Real-time Bayesian anomaly detection in streaming environmental data. Water Resour. Res. 45. http://dx.doi.org/10.1029/2008WR006956.
Hong, Y.S.T., Rosen, M.R., Bhamidimarri, R., 2003. Analysis of a municipal wastewater treatment plant using a neural network-based pattern analysis. Water Res. 37, 1608–1618.
Hong, S.H., Lee, M.W., Lee, D.S., Park, J.M., 2007. Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks. Biochem. Eng. J. 35, 365–370.
Hotelling, H., 1936. Relations between two sets of variants. Biometrika 28, 321–377.
Huang, M.Z., Wan, J.Q., Ma, Y.W., Li, W.J., Sun, X.F., Wan, Y., 2010. A fast predicting neural fuzzy model for on-line estimation of nutrient dynamics in an anoxic/oxic process. Bioresour. Technol. 101, 1642–1651.
Hubert, M., Van Kerckhoven, J., Verdonck, T., 2012. Robust PARAFAC for incomplete data. J. Chemometr. 26, 290–298.
Jang, J.S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE T. Syst. Man. Cy 23, 665–685.
Jeppsson, U., Rosen, C., Alex, J., Copp, J., Gerneay, K.V., Pons, M.N., Vanrolleghem, P.A., 2006. Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs. Water Sci. Technol. 53, 287–295.
Jolliffe, I.T., 2002. Principal Component Analysis. In: Springer Series in Statistics. Springer.
Kadlec, P., 2009. On Robust and Adaptive Soft Sensors. Ph.D. thesis. Bournemouth University, School of Design, Engineering & Computing, Bournemouth, UK.
Kadlec, P., Gabrys, B., 2008. Soft sensor based on adaptive local learning. In: Proc. of the 15th International Conference On Neural Information Processing, Auckland, New Zealand, pp. 1172–1179.
Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. Comput. Chem. Eng. 33, 795–814.
Kadlec, P., Gabrys, B., Grbić, R., 2011. Review of adaptation mechanisms for data-driven soft sensors. Comput. Chem. Eng. 35, 1–24.
Kariya, T., Kurata, H., 2004. Generalized Least Squares. In: Wiley Series in Probability and Statistics. John Wiley & Sons.
Khataee, A.R., Kasiri, M.B., 2011. Modeling of biological water and wastewater treatment processes using artificial neural networks. Clean – Soil, Air, Water 39, 742–749.
Kim, M.H., Kim, Y.S., Prabu, A.A., Yoo, C.K., 2009. A systematic approach to data-driven modeling and soft sensing in a full-scale plant. Water Sci. Technol. 60, 363–370.
Kohavi, R., John, G., 1997. Wrappers for feature subset selection. Artif. Intell. 97, 273–324.
Kohonen, T., 2001. Self-organizing Maps. In: Springer Series in Information Sciences, vol. 30. Springer.
Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. Phys. Rev. E 69, 066138–1–066138–16.
Krishnapuram, R., Keller, J., 1993. A possibilistic approach to clustering. IEEE Trans. Fuzzy Sys. 1, 98–110.
Lee, J., Verleysen, M., 2007. Nonlinear Dimensionality Reduction. In: Information Science and Statistics. Springer.
Lee, D.S., Jeon, C.O., Park, J.M., Chang, K.S., 2002. Hybrid neural network modeling of a full-scale industrial wastewater treatment process. Biotechnol. Bioeng. 78, 670–682.
Lee, C., Choi, S.W., Lee, I.B., 2004. Sensor fault identification based on time-lagged PCA in dynamic processes. Chemometr. Intell. Lab. 70, 165–178.
Lee, D.S., Vanrolleghem, P.A., Park, J.M., 2005. Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. J. Biotechnol. 115, 317–328.
Lee, C., Choi, S.W., Lee, I.B., 2006. Sensor fault diagnosis in a wastewater treatment process. Water Sci. Technol. 53, 251–257.
Lee, H.W., Lee, M.W., Park, J.M., 2007. Robust adaptive partial least squares modeling of a full-scale industrial wastewater treatment process. Ind. Eng. Chem. Res. 46, 955–964.
Lee, M.W., Hong, S.H., Choi, H., Kim, J.H., Lee, D.S., Park, J.M., 2008. Real-time remote monitoring of small-scaled biological wastewater treatment plants by a multivariate statistical process control and neural network-based software sensors. Process. Biochem. 43, 1107–1113.
Lee, H.W., Lee, M.W., Park, J.M., 2009. Multi-scale extension of PLS algorithm for advanced on-line process monitoring. Chemometr. Intell. Lab. 98, 201–212.
Lee, J.W., Hong, Y.S.T., Suh, C.W., Shin, H.S., 2011. Performance prediction of the full-scale biological nutrient removal process using an artificial neural network. In: Proc. of Watermatex 2011, 8th IWA Symposium on Systems Analysis and Integrated Assessment, San Sebastián, Spain.
Lennox, J., Rosen, C., 2002. Adaptive multiscale principal components analysis for online monitoring of wastewater treatment. Water Sci. Technol. 45, 227–235.
Liitiäinen, E., Verleysen, M., Corona, F., Lendasse, A., 2009. Residual variance estimation in machine learning. Neurocomputing 72, 3692–3703.
Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data, second ed.. In: Wiley Series in Probability and Statistics John Wiley & Sons.
Luccarini, L., Porra, E., Spagni, A., Ratini, P., Grilli, S., Longhi, S., Bortone, G., 2002. Soft sensors for control of nitrogen and phosphorus removal from wastewaters by neural networks. Water Sci. Technol. 45, 101–107.
Lumley, D., 2002. On-line instrument confirmation: how can we check that our instruments are working? Water Sci. Technol. 53, 469–476.
Machón, I., López, H., Rodríguez-Iglesias, J., Maranón, E., Vázquez, I., 2007. Simulation of a coke wastewater nitrification process using a feed-forward neuronal net. Environ. Model. Softw. 22, 1382–1387.
Marsili-Libelli, S., 1990. Adaptive estimation of bioactivities in the activated sludge process. IEE Proc-D 137, 349–356.
Marsili-Libelli, S., 2006. Control of SBR switching by fuzzy pattern recognitions. Water Res. 40, 1095–1107.
Marsili-Libelli, S., Müller, A., 1996. Adaptive fuzzy pattern recognition in the anaerobic digestion process. Pattern Recogn. Lett. 17, 651–659.
Masson, M., Canu, S., Grandvalet, Y., Lynggaard-Jensen, A., 1999. Software sensor design based on empirical data. Ecol. Model. 120, 131–139.
Miettinen, T., Hurse, T.J., Connor, M.A., Reinikainen, S.P., Minkkinen, 2004. Multivariate monitoring of a biological wastewater treatment process: a case study at Melbourne water's western treatment plant. Chemometr. Intell. Lab. 73, 131–138.
Mjalli, F.S., Al-Asheh, S., Alfadala, H.E., 2007. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. J. Environ. Manage. 83, 329–338.
Molvar, A.E., Roesler, J.F., Babcock, R.H., 1976. Instrumentation and Automation Experiences in Wastewater-treatment Facilities. volume EPA-600/2-76-198 of Environmental Protection Technology. U.S. Environmental Protection Agency, Office of Research and Development, Municipal Environmental Research Laboratory.
Moon, T.S., Kim, Y.J., Kim, J.R., Cha, J.H., Kim, D.H., Kim, C.W., 2009. Identification of process operating state with operational map in municipal wastewater treatment plant. J. Environ. Manage. 90, 772–778.
Moral, H., Aksoy, A., Gokcay, C.F., 2008. Modeling of the activated sludge process by using artificial neural networks with automated architecture screening. Comput. Chem. Eng. 32, 2471–2478.
Mujunen, S.P., 1999. Multivariate Monitoring of Wastewater Treatment Processes in Pulp and Paper Industry. Ph.D. thesis. Lappeenranta University of Technology, Department of Chemical Technology, Lappeenranta, Finland.
Mujunen, S.P., Minkkinen, P., Teppola, P., Wirkkala, R.S., 1998. Modeling of activated sludge plants treatment efficiency with PLSR: a process analytical case study. Chemometr. Intell. Lab. 41, 83–94.
Olsson, G., 1977. State of the Art in Sewage Treatment Control. In: American Inst. of Chemical Engineers, Symp Series 159, pp. 52–76.
Olsson, G., 2012. ICA and me – a subjective review. Water Res. 46, 1585–1626.
Olsson, G., Newell, B., Rosen, C., Ingildsen, P., 2003. Application of information technology to decision support in treatment plant operation. Water Sci. Technol. 47, 35–42.
Olsson, G., Nielsen, M., Yuan, Z., Lynggaard-Jensen, A., Steyer, J.P., 2005. Instrumentation, Control and Automation in Wastewater Systems. Scientific and Technical Reports, No. 15. IWA Publishing.
Pai, T.Y., Tsai, Y.P., Lo, H.M., Tsai, C.H., Lin, C.Y., 2007. Grey and neural network prediction of suspended solids and chemical oxygen demand in hospital wastewater treatment plant effluent. Comput. Chem. Eng. 31, 1272–1281.
Pai, T.Y., Chuang, S.H., Wan, T.J., Lo, H.M., Tsai, Y.P., Su, H.C., Yu, L.F., Hu, H.C., Sung, P.J., 2008. Comparisons of grey and neural network prediction of industrial park wastewater effluent using influent quality and online monitoring parameters. Environ. Monit. Assess. 146, 51–66.
Pai, T.Y., Yang, P.Y., Wang, S.C., Lo, M.H., Chiang, C.F., Kuo, J.L., Chu, H.H., Su, H.C., Yu, L.F., Hu, H.C., Chang, Y.H., 2011. Predicting effluent from the wastewater treatment plant of industrial park based on fuzzy network and influent quality. Appl. Math. Model. 35, 3674–3684.
Petersen, B., Gernaey, K., Henze, M., Vanrolleghem, P.A., 2003. Calibration of activated sludge models: a critical review of experimental designs. In: Agathos, S.N., Reineke, W. (Eds.), Biotechnology for the Environment: Wastewater Treatment and Modeling, Waste Gas Handling. Kluwer Academic Publishers.
Phillips, H.M., Sahlstedt, K.E., Frank, K., Bratby, J., Brennan, W., Rogowski, S., Pier, D., Anderson, W., Mulas, M., Copp, J.B., Shirodkar, N., 2009. Wastewater treatment modelling in practice: a collaborative discussion of the state of the art. Water Sci. Technol. 59, 695–704.

Pokhrel, D., Viraraghavan, T., 2004. Treatment of pulp and paper mill wastewater – a review. Sci. Total Environ. 333, 37–58.
Poutiainen, H., Niska, H., Heinonen-Tanski, H., Kolehmainen, M., 2010. Use of sewer on-line total solids data in wastewater treatment plant modelling. Water Sci. Technol. 62, 743–750.
Pu, H.C., Hung, Y.T., 1995. Use of artificial neural networks: predicting trickling filter performance in a municipal wastewater treatment plant. Environ. Manage. Health 6, 16–27.
Puig, S., Corominas, L., Vives, M., Balaguer, M., Colprim, J., Colomer, J., 2005. Development and implementation of a real-time control system for nitrogen removal using OUR and ORP as end points. Ind. Eng. Chem. Res. 44, 3367–3373.
Qiao, J., Chen, Q., Han, H., 2011. The chemical oxygen demand modelling based on a dynamic structure neural network. In: Einschlag, F.S.G. (Ed.), Waste Water – Evaluation and Management. InTech.
Qin, S., McAvoy, T., 1992. Nonlinear PLS modeling using neural networks. Comput. Chem. Eng. 16, 379–391.
Ráduly, B., Gernaey, K.V., Capodaglio, A.G., Mikkelsen, P.S., Henze, M., 2007. Artificial neural networks for rapid WWTP performance evaluation: methodology and case study. Environ. Model. Softw. 22, 1208–1216.
Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P.A., Comeau, Y., 2010. Data reconciliation for wastewater treatment plant simulation studies – planning for high-quality data and typical sources of errors. Water Environ. Res. 82, 426–433.
Rieger, L., Gillot, S., Langergraber, G., Ohtsuki, T., Shaw, A., Takacs, I., Winkler, S., 2012. Guidelines for Using Activated Sludge Models. Scientific and Technical Reports, No. 22. IWA Publishing.
Robinson, R.B., Cox, C.D., Odom, K., 2005. Identifying outliers in correlated water quality data. J. Environ. Eng. 131, 651–657.
Rosen, C., 2001. A Chemometric Approach to Process Monitoring and Control – with Applications to Wastewater Treatment Operation. Ph.D. thesis. Lund University of Technology, Department of Industrial Electrical Engineering and Automation, Lund, Sweden.
Rosen, C., Lennox, J.A., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. Water Res. 35, 3402–3410.
Rosen, C., Olsson, G., 1998. Disturbance detection in wastewater treatment plants. Water Sci. Technol. 37, 197–205.
Rosen, C., Yuan, Z., 2001. Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. Water Sci. Technol. 43, 147–156.
Rosen, C., Larsson, M., Jeppsson, U., Yuan, Z., 2002. A framework for extreme-event control in wastewater treatment. Water Sci. Technol. 45, 299–308.
Rosen, C., Röttorp, J., Jeppsson, U., 2003. Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. Water Sci. Technol. 47, 171–179.
Rosen, C., Jeppsson, U., Vanrolleghem, P.A., 2004. Towards a common benchmark for long-term process control and monitoring performance evaluation. Water Sci. Technol. 50, 41–49.
Rosipal, R., Trejo, L.J., 2001. Kernel partial least squares regression in reproducing kernel hilbert space. J. Mach. Learn. Res. 2, 97–123.
Rousseeuw, P.J., Hubert, M., 2011. Robust statistics for outlier detection. WIREs Data Mining Knowl. Discov. 1, 73–79.
Ruiz, M., Villez, K., Sin, G., Colomer, J., Vanrolleghem, P., 2006. Influence of scaling and unfolding in PCA based monitoring of nutrient removing batch process. In: Proc. of the 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Beijing, China, pp. 114–119.
Rustum, R., 2009. Modelling Activated Sludge Wastewater Treatment Plants Using Artificial Intelligence Techniques (Fuzzy Logic and Neural Networks). Ph.D. thesis. Heriot-Watt University, School of the Built Environment, Edinburgh, Scotland.
Rustum, R., Adeloye, A.J., 2007. Replacing outliers and missing values from activated sludge data using Kohonen self-organizing map. J. Environ. Eng. 133, 909–916.
Rustum, R., Adeloye, A.J., Scholz, M., 2008. Applying Kohonen self-organizing map as a software sensor to predict biochemical oxygen demand. Water Environ. Res. 80, 32–40.
Sin, G., 2004. Systematic Calibration of Activated Sludge Models. Ph.D. thesis. Ghent University, Faculty of Agricultural and Applied Biological Sciences, Ghent, Belgium.
Smilde, A., Bro, R., Geladi, P., 2004. Multi-way Analysis: Applications in the Chemical Sciences. John Wiley & Sons.
Spérandio, M., Queinnec, I., 2004. Online estimation of wastewater nitrifiable nitrogen, nitrification and denitrification rates, using ORP and DO dynamics. Water Sci. Technol. 49, 31–38.
Stone, C., 1977. Consistent nonparametric regression. Ann. Stat. 5, 595–620.
Strang, G., 1994. Wavelets. Am. Sci. 82, 250–255.
Tay, J.H., Zhang, X., 1999. Neural fuzzy modeling of anaerobic biological wastewater treatment systems. J. Environ. Eng. ASCE 125, 1149–1159.
Tay, J.H., Zhang, X., 2000. A fast predicting neural fuzzy model for high-rate anaerobic wastewater treatment systems. Water Res. 34, 2849–2860.
Teppola, P., 1999. Multivariate Process Monitoring of Sequential Process Data – a Chemometric Approach. Ph.D. thesis. Lappeenranta University of Technology, Department of Chemical Technology, Lappeenranta, Finland.
Teppola, P., Minkkinen, P., 1999. Possibilistic and fuzzy c-means clustering for process monitoring in an activated sludge waste-water treatment plant. J. Chemometr. 13, 445–459.
Teppola, P., Mujunen, S.P., Minkkinen, P., 1997. Partial least squares modeling of an activated sludge plant: a case study. Chemometr. Intell. Lab. 38, 197–208.
Teppola, P., Mujunen, S.P., Minkkinen, P., 1998. A combined approach of partial least squares and fuzzy c-means clustering for the monitoring of an activated-sludge waste-water treatment plant. Chemometr. Intell. Lab. 41, 95–103.
Teppola, P., Mujunen, S.P., Minkkinen, P., 1999a. Adaptive fuzzy c-means clustering in process monitoring. Chemometr. Intell. Lab. 45, 23–38.
Teppola, P., Mujunen, S.P., Minkkinen, P., 1999b. Kalman filter for updating the coefficients of regression models. A case study from an activated sludge waste-water treatment plant. Chemometr. Intell. Lab. 45, 371–384.
Tharrault, Y., Mourot, G., Ragot, J., 2009. WWTP diagnosis based on robust principal component analysis. In: Proc. of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Barcelona, Spain, pp. 1342–1347.
Thompson, G., Swain, J., Kay, M., Forster, C.F., 2001. The treatment of pulp and paper mill effluent: a review. Bioresour. Technol. 77, 275–286.
Timmerman, M., 2006. Multilevel component analysis. Brit. J. Math. Stat. Psychol. 59, 301–320.
Tomita, R.K., Park, S.W., Sotomayor, O.A.Z., 2002. Analysis of activated sludge process using multivariate statistical tools – a PCA approach. Chem. Eng. J. 90, 283–290.
Tyagi, R.D., Du, Y.G., 1992. Kinetic model for the effects of heavy metals on activated sludge process using neural networks. Environ. Technol. 13, 883–890.
Vanrolleghem, P.A., Lee, D.S., 2003. On-line monitoring equipment for wastewater treatment processes: state of the art. Water Sci. Technol. 47, 1–34.
Villez, K., Ruiz, M., Sin, G., Colomer, J., Rosen, C., Vanrolleghem, P.A., 2008. Combining multiway principal component analysis (MPCA) and clustering for efficient data mining of historical data sets of SBR processes. Water Sci. Technol. 57, 1659–1666.
Villez, K., Rosen, C., D'hooge, E., Vanrolleghem, P., 2010. Online phase length optimization for a sequencing batch reactor by means of the Hotelling T^2 statistic. Ind. Eng. Chem. Res. 49, 180–188.
Walczak, B., Massart, D., 2001a. Dealing with missing data: part I. Chemometr. Intell. Lab. 58, 15–27.
Walczak, B., Massart, D., 2001b. Dealing with missing data: part II. Chemometr. Intell. Lab. 58, 29–42.
Wan, J., Huang, M., Ma, Y., Guo, W., Wang, Y., Zhang, H., Li, W., Sun, X., 2011. Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system. Appl. Softw. Comput. 11, 3238–3246.
Water Environment Federation, 2008. Industrial Wastewater Management, Treatment and Disposal. WEF Manual of Practice No. FD-3, third ed. WEF Press.
West, D., Mangiameli, P., 2000. Identifying process conditions in an urban wastewater treatment plant. Int. J. Oper. Prod. Man. 20, 573–590.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. 58, 109–130.
Woo, S.H., Jeon, C.O., Yun, Y.S., Choi, H., Lee, C.S., Lee, D.S., 2009. On-line estimation of key process variables based on kernel partial least squares in an industrial cokes wastewater treatment plant. J. Hazard. Mater. 161, 538–544.
Woodard & Curran, Inc., 2006. Industrial Waste Treatment Handbook, second ed. Butterworth-Heinemann.
Yetilmezsoy, K., Ozkaya, B., Cakmakci, M., 2011. Artificial intelligence-based prediction models for environmental engineering. Neural Netw. World 21, 193–218.
Yoo, C.K., Kim, D.S., Cho, J.H., Choi, S.W., Lee, I.B., 2001. Process system engineering in wastewater treatment process. Korean J. Chem. Eng. 18, 408–421.
Yoo, C.K., Vanrolleghem, P.A., Lee, I.B., 2003. Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical methods in biological wastewater treatment plants. J. Biotechnol. 105, 135–163.
Yoo, C.K., Bang, Y.H., Lee, I.B., Vanrolleghem, P.A., Rosen, C., 2004. Application of fuzzy partial least squares (FPLS) modeling nonlinear biological processes. Korean J. Chem. Eng. 21, 1087–1097.
Yoo, C.K., Villez, K., Van Hulle, S.W.H., Vanrolleghem, P.A., 2008. Enhanced process monitoring for wastewater treatment systems. Environmetrics 19, 602–617.
Zamarrenoa, J., Vega, P., 1998. State space neural network. Properties and applications. Neural Netw. 11, 1099–1112.
Zhao, L., Chai, T., 2005. Wastewater BOD forecasting model for optimal operation using robust time-delay neural network. In: Proc. ISNN 2005, Second International Symposium on Neural Networks, Part III, Chongqing, China, pp. 1028–1033.
Zhao, L., Chai, T., Cong, Q., 2006. Operating condition recognition of pre-denitrification bioprocess using robust EMPCA and FCM. In: Proc. WCICA 2006, The 6th World Congress on Intelligent Control and Automation, Dalian, China, pp. 9386–9390.
Zhu, Z., Corona, F., Lendasse, A., Baratti, R., Romagnoli, J., 2011. Local linear regression for soft-sensor design with application to an industrial deethanizer. In: Proc. of The18th IFAC World Congress, Milano, Italy, pp. 6212–6217.