



ELSEVIER

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant



Henri Haimi ^{a,*}, Michela Mulas ^{a,b}, Francesco Corona ^{c,d}, Stefano Marsili-Libelli ^e,
Paula Lindell ^f, Mari Heinonen ^f, Riku Vahala ^a

^a Department of Built Environment, Aalto University, School of Engineering, P.O. Box 15200, FI-00076 Aalto, Finland

^b Department of Chemical Engineering, Federal University of Campina Grande, 58429-140 Campina Grande, Brazil

^c Department of Information and Computer Science, Aalto University, School of Science, P.O. Box 15400, FI-00076 Aalto, Finland

^d Department of Teleinformatics Engineering, Federal University of Ceará, 60455-760 Fortaleza, Brazil

^e Department of Information Technology, University of Florence, Via S. Marta 3, 50139 Florence, Italy

^f HSY Helsinki Region Environmental Services Authority, P.O. Box 100, FI-00066 HSY, Finland

ARTICLE INFO

Article history:

Received 4 September 2015

Received in revised form

30 December 2015

Accepted 7 February 2016

Available online 2 March 2016

Keywords:

Adaptive process monitoring

Anomaly detection

Principal component analysis

Wastewater treatment

ABSTRACT

This work examines real-time anomaly detection and isolation in a full-scale wastewater treatment application. The Viikinmäki plant is the largest municipal wastewater treatment facility in Finland. It is monitored with ample instrumentation, though their potential is not yet fully exploited. One reason that prevents the use of the instrumentation in plant control is the occasional insufficient measurement performance. Therefore, we investigate an intelligent anomaly detection system for the activated sludge process in order to motivate a more efficient use of sensors in the process operation. The anomaly detection methodology is based on principal component analysis. Because the state of the process fluctuates, moving-window extensions are used to adapt the analysis to the time-varying conditions. The results show that both instrument and process anomalies were successfully detected using the proposed algorithm and the variables responsible for the anomalies correctly isolated. We also demonstrate that the proposed algorithm represents a convenient improvement for supporting the efficient operation of wastewater treatment plants.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Wastewater treatment in municipalities has faced considerable developments starting from simple process units and ending up in modern-day plants including numerous highly automated units since the beginning of the 20th century. For instance in Helsinki (Finland), the first wastewater treatment plant (WWTP) built in 1910 consisted of a septic tank and a trickling filter of natural gravel, whereas the first plants using an activated sludge process (ASP) were constructed in different neighbourhoods in the 1930s (Katko, 2000). Today, the treatment of wastewaters from Helsinki and several neighbouring municipalities is centralized to the Viikinmäki WWTP the capacity of which is about 300-fold compared with the first plant in the city. The Viikinmäki central plant is an efficient facility employing several process units supported by an extensive instrumentation and advanced control schemes.

Modern WWTPs are complex facilities where the interactions between several process units and external disturbances take place.

The role of instrumentation, control and automation has become essential for the cost-effective and safe process operation. The advances in information technology and of on-line instrumentation which have occurred in the last few decades have produced sophisticated process control solutions (Olsson et al., 2005; Olsson, 2014). Reliability of the real-time measurements is highly important in the demanding conditions of biological wastewater treatment processes. Even though notable development in on-line instrumentation has taken place during the past decades (Vanrolleghem and Lee, 2003; Campisano et al., 2013), fouling of the instruments, for instance due to solids deposition and slime build-up, impairs their dependability (Olsson, 2014). When the sensors are used for control actions, the reliability of the measurements is even more essential for cost-efficient process operation and for avoiding a break in the feedback loop; this is especially true for aeration control, chemical dosing and pumping. The automatic anomaly detection system aims at providing the operators with timely information on sensor faults and process malfunctioning in general. Therefore, they contribute to the successful WWTP operation by reducing the risks of process malfunctions and by enabling the more dependable use of on-line data in critical control schemes.

* Corresponding author. Tel.: +358 50 407 4214.

E-mail address: henri.haimi@aalto.fi (H. Haimi).

One option for the anomaly detection system development relies on the industries' historical process data where information about both normal and abnormal operations is encoded. Historical data together with mathematical modelling algorithms can be used for designing software that distinguishes with normal and abnormal situations in real-time when incoming data are inputted to the system (Venkatasubramanian et al., 2003). The most popular families of model structures that are used for quantitative data-derived anomaly detection and isolation belong to multivariate statistical and artificial neural network techniques (Venkatasubramanian et al., 2003; Ng and Srinivasan, 2010; Qin, 2011; Ge et al., 2013). For instance in the process industry, intelligent software tools designed based on the historical operation data have been used successfully for monitoring anomalies that manifest themselves as the exceptional variation among the on-line measured variables (Kadlec et al., 2009).

Data-derived approaches, such as multivariate statistics, have also been proposed for anomaly monitoring applications in the biological WWTPs (see Haimi et al., 2013, for references). Considerable efforts at the development of multivariate techniques, for instance principal component analysis (PCA), were made by Rosen (2001) and Lennox (2002), who introduced adaptive and multiscale approaches for monitoring ASPs. Combining PCA and clustering algorithms have also been presented for observing the fluctuation of the process states in both continuous and batchwise wastewater treatment units (Teppola et al., 1999; Aguado et al., 2008, respectively). Later, PCA methods have been proposed for full-scale municipal applications for real-time fault detection and isolation in an ASP (Baggiani and Marsili-Libelli, 2009) and for detecting outliers in the measurement data of a biological post-filtration unit (Corona et al., 2013). PCA techniques have also recently been used for assessing anomalous measurements in the inlet of WWTP (Alferes et al., 2013) and for diagnosing sensor faults in a laboratory-scale wastewater treatment system (Tao et al., 2013).

Even though PCA-based monitoring tools for the municipal wastewater sector have been presented in the literature, the challenges created by the time-evolving process dynamics of the real-life WWTP conditions have not been addressed in the majority of the proposals (Haimi et al., 2013). Most of the investigations where adaptive PCA techniques have been used for dealing with the fluctuating process and influent conditions concern simulated ASPs (Rosen and Yuan, 2001; Lee et al., 2004, 2006; Le Bonté et al., 2005; Aguado and Rosen, 2008). The simulated protocols certainly provide

valuable opportunities for the monitoring methodology development that is demonstrated with the plentiful literature (Jeppsson et al., 2013) and efforts have been made for generating realistic influent wastewater data for the modelling purposes (Martin and Vanrolleghem, 2014). However, the experiments that concern full-scale processes involve additional challenges compared with the simulation platform tests due to the unforeseen and plant-specific features of the influent characteristics. Isolating faults in real-life facilities is also difficult because the occurrences of true anomalies are rarely possible to be extensively verified among a large number of frequently on-line measured process variables, unlike in simulated processes where faults that differ from the normal operations are intentionally encoded. In fact, this also suggests that real operation data are irreplaceable when anomaly monitoring systems are designed and tested for a particular WWTP. For such reasons, the objective of this study is to investigate the applicability of adaptive PCA methodologies for detecting and isolating instrument and process anomalies in a large-scale ASP. One of the general challenges of adaptive PCA techniques is that the length of the historical period considered in the model construction is typically fixed while the process dynamics do change, which often leads to a sub-optimal monitoring performance (Kadlec et al., 2011). Therefore, we examine in this work such adaptive data-derived techniques that are designed to take into account also the varying rapidness of the process changes.

In this paper, we study PCA-based techniques for anomaly detection and diagnosis in the Viikinmäki WWTP. First, the investigated plant with a particular focus on the ASP and the acquired operation data are described. After that, the considered adaptive multivariate methods and the anomaly monitoring algorithm are presented. Finally, we report and discuss the model parameter definition and the results of the research.

2. Material and methods

2.1. Process and instrumentation

The Viikinmäki WWTP (800 000 population equivalent) treats an average influent flow rate of 250 000 m³/d, of which about 85% is domestic and 15% industrial wastewater. The wastewater treatment line consists of bar screening, grit removal, pre-aeration, primary sedimentation, ASP, secondary sedimentation and biological

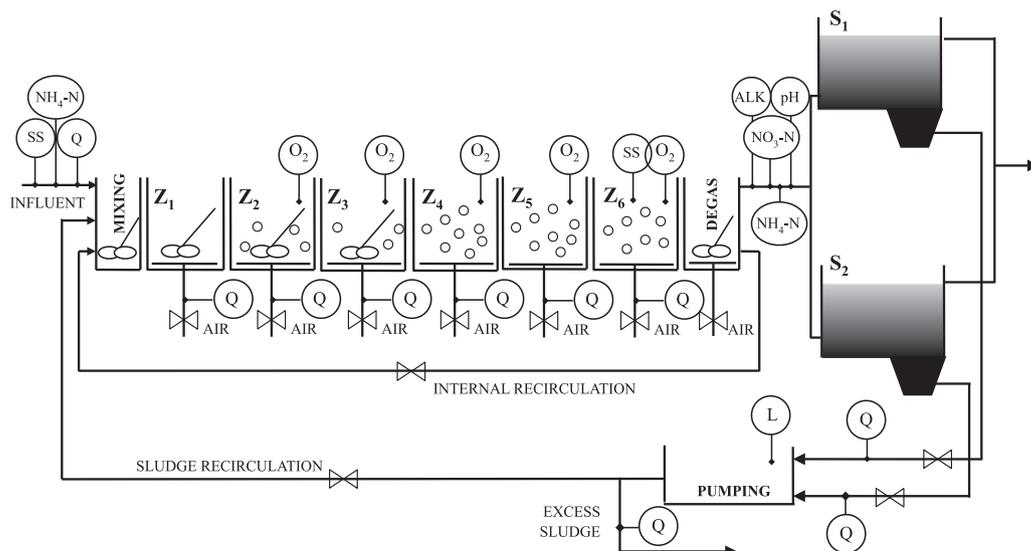


Fig. 1. Simplified layout for a single ASP line and location of on-line measurements.

post-filtration. The sludge treatment is achieved with mesophilic digesters and subsequent dewatering systems. Total nitrogen removal of approximately 90%, total phosphorus removal of 95% and biochemical oxygen demand (BOD₇) removal of 95% of yearly averages are achieved in the plant.

ASP is the core of the treatment process where the biological nitrogen removal is realized, together with the denitrifying post-filtration process at the end of the wastewater treatment line. At the time of the investigation, the ASP consisted of eight treatment lines divided into bioreactor and secondary sedimentation units, one of which being schematically represented in Fig. 1. The ASP employs DN-configuration which means that the zones where denitrification takes place are located before the zones where nitrification is realized. Activated sludge consisting mainly of bacteria and protozoa is recycled in the bioreactor and is needed in the nitrogen removal process executed in different dissolved oxygen conditions. In order to keep the activated sludge in the process, a secondary sedimentation process that sequences the bioreactor is applied for settling the sludge and a desired amount of thickened activated sludge is pumped back to bioreactor. The clarified wastewater from the rectangular sedimentation basins is further led to the denitrifying post-filtration process. Each ASP line begins with a mixing zone where pre-settled wastewater, return sludge from secondary sedimentation and internal recycle sludge from the degassing zone at the end of the bioreactor are fed. After that, the bioreactor is composed of six cascaded zones with the anoxic ones located near the input.

The dissolved oxygen concentrations of the aerated zones are controlled with PID control loops where the air valve positions are the manipulated variables and the dissolved oxygen set-point is 3.5 mg/l. The number of anoxic zones, forming the overall anoxic volume for denitrification, is adjusted according to the nitrification performance. Specifically, the number of anoxic zones depends on the aeration mode, which is controlled in such a way that the effluent ammonium–nitrogen concentration is within the set target range while using the minimum required aerated volume. Time-delays are also included in the aeration mode control scheme in order to increase the stability of the control. In practice, Zone 1 is never aerated and it is mixed mechanically. Zones 2 and 3 are equipped with agitators and are either aerated or non-aerated depending on the aeration mode in use. In contrast, Zones 4–6 are always aerated.

The quality of wastewater entering the bioreactor is monitored continuously in terms of ammonium–nitrogen and suspended solids concentrations in addition to the flow rate measurement. Dissolved oxygen (O₂) levels in Zones 2–6 are measured in real-time as well as mixed liquor suspended solids concentration in the last zone. The gauging station of the bioreactor effluent from the degassing zone covers the measurements of ammonium–nitrogen, nitrate–nitrogen, pH and alkalinity. Additionally, flow rates of sludge recirculation, internal recirculation, excess sludge and air to different zones are monitored.

2.2. Data description and variable selection

The process variables considered in the study concern one ASP line. The collected data covers two years of process operation

Table 1
Process variables considered for the anomaly detection in the ASP.

TAG	Description	Unit
<i>I-NH4</i>	Ammonium–nitrogen in the influent to the bioreactor	mg/l
<i>I-SS</i>	Suspended solids in the influent to the bioreactor	mg/l
<i>I-Q</i>	Influent flow rate to the bioreactor	m ³ /s
<i>OX-SS</i>	Suspended solids in the bioreactor	g/l
<i>E-NH4</i>	Ammonium–nitrogen in the effluent of the bioreactor	mg/l
<i>E-NO3</i>	Nitrate–nitrogen in the effluent of the bioreactor	mg/l
<i>E-pH</i>	pH in the effluent of the bioreactor	–

(January 1, 2009–December 31, 2010), recorded as hourly average values. Analysers that are used to measure concentrations and other chemical properties were considered in the study for facilitating their supervision in the demanding environment. Instruments that produce information, for instance, about flow rates have been proved to work reliably in WWTPs and they do not require repeated maintenance like the analysers do (Thomsen and Önnérth, 2009). Therefore, the monitoring of the analysers was a priority in this case. The primary criterion in variable selection was their potential use in future advanced control schemes, for instance, such as the one proposed for the considered ASP by Mulas et al. (2015).

From all the acquired data, the variables selected for anomaly monitoring are grouped in Table 1, where the TAGs in column 1 are later used for identification. The only investigated sensor that is currently used in the aeration control is *E-NH4*. However, it was also included in the study because the initial inspection of the data showed frequent unexpected peaks in the *E-NH4* signal. The occurrence of measurement reliability problems for the selected analysers, such as unjustified drifts and peaks, were detected in the data inspection step. Therefore, an adequate anomaly detection system would increase the feasibility of the investigated analysers for process control purposes. Dissolved oxygen sensors were not considered in the study because they are already successfully used in the aeration control and the data inspection did not reveal relevant signs of unreliability. The effluent alkalinity measurement was excluded from the investigation because its information was very well described by *E-pH*. *I-Q* was the only flow rate measurement included in the study. The purpose of its inclusion was to provide information about the flow dynamics of the process and, on the other hand, to guarantee the high-grade performance of the instrument, which would be even more important in such a potential feedforward control scheme where influent ammonium load (kg/h) would be used. The internal recirculation and the sludge recirculation being controlled proportionally to *I-Q*, it was considered informative about the overall flow conditions, that was also confirmed by the statistical analysis.

In the pre-processing of the acquired data, only the obvious outliers that violated the technological limitations of hardware instruments were discarded. Such observations were considered the measurements that exceeded the instrument measuring range or that were associated with the unfeasible zero-values. In addition, data were synchronized in such a way that *I-NH4*, *I-SS* and *I-Q* were shifted 3 h back in time which approximately corresponds the hydraulic retention time of the bioreactor.

2.3. Methods for anomaly detection

2.3.1. General procedure

Principal component analysis (Jolliffe, 2002) is a multivariate statistical technique for extracting the information from the data by eliminating the information redundancy due to variables cross-correlation. PCA identifies the principal directions of the transformed data and ranks the contribution of each original variable in explaining the observed variability. Let \mathbf{X} indicate a data matrix with the K observations each comprising D process variables. Each of the K observations $\mathbf{x}(k) = [x_1(k), \dots, x_d(k), \dots, x_D(k)]^T$ at time k represents a point in the D -dimensional data space. PCA factorizes the $K \times D$ data matrix \mathbf{X} using eigenvalue decomposition, to obtain

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (1)$$

where \mathbf{T} is a $K \times S$ score matrix, \mathbf{P} a $D \times S$ loading matrix and \mathbf{E} a $K \times D$ residual matrix. S is for the number retained principal components (PCs) and each of the K measurements at time k is modelled as a S -dimensional point $\mathbf{t}(k) = \mathbf{x}(k)\mathbf{P}$. The scores are the new coordinates of the observations in a (sub)space whose directions are defined by the set of loadings $\{\mathbf{p}_1, \dots, \mathbf{p}_S, \dots, \mathbf{p}_S\}$, or PCs. Typically, most of the variation in the data can be explained

by retaining a small number of PCs compared with the original dimension of \mathbf{X} (i.e. $S \ll D$).

The Hotelling's T^2 statistic and the Q statistic (Jackson and Mudholkar, 1979) and their confidence limits, T_{lim}^2 (Atkinson et al., 2004) and Q_{lim} (Nomikos and MacGregor, 1995) calculated for a certain confidence level, are often employed in monitoring tasks. T^2 measures the distance of the projected observation $\mathbf{t}(k)$ from the origin of the principal component subspace:

$$T^2(k) = \mathbf{t}(k)\mathbf{\Lambda}^{-1}\mathbf{t}(k) \quad (2)$$

where $\mathbf{\Lambda}^{-1}$ denotes a diagonal matrix with the reciprocal of the eigenvalues associated with the retained PCs. Q measures the distance of an observation $\mathbf{x}(k)$ from its reconstruction $\hat{\mathbf{x}}(k) = \mathbf{t}(k)\mathbf{P}^T = [\hat{x}_1(k), \dots, \hat{x}_d(k), \dots, \hat{x}_D(k)]^T$ on the subspace:

$$Q(k) = \sum_{d=1}^D (x_d(k) - \hat{x}_d(k))^2 \quad (3)$$

For the anomaly detection, a PCA model is first constructed and the thresholds T_{lim}^2 and Q_{lim} are calculated using training data that are supposed to be anomaly-free. Then, the model is used for monitoring faults in testing data for which the T^2 and Q statistics are calculated. The testing samples whose T^2 and Q are less than T_{lim}^2 and Q_{lim} are considered representing normal process behaviour, whereas the observations whose T^2 and Q exceed T_{lim}^2 and/or Q_{lim} are assumed to denote a possible anomaly. The Q statistic has ability to indicate changes in the correlation structure of the measured variables and, thus, to detect sensor faults whereas the T^2 statistic is more sensitive to significant process variations (Liefucht et al., 2006).

Once a violation of T_{lim}^2 or Q_{lim} is detected, the variables' contributions to the statistics are studied. The contributions along the d th PC to the T^2 and Q statistics are calculated as $\mathbf{c}(k) = \mathbf{x}(k)\text{diag}(\mathbf{p}_d)$ and $\mathbf{e}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k)$, respectively (MacGregor et al., 1994). $\text{diag}(\mathbf{p}_d)$ denotes the diagonal matrix of the column vector \mathbf{p}_d , $\mathbf{x}(k)$ denotes the vector of original data at time k and $\hat{\mathbf{x}}(k)$ denotes its reconstruction using a model with d PCs.

2.3.2. Moving-window procedure

A major limitation of PCA-based fault analysis is that the model once built, it is time-invariant while the processes are time-varying. When such models are used, false alarms might result. This is because a PCA model describes the process conditions represented by the training period and is applicable to testing in corresponding conditions. However, if the conditions change considerably during the testing period, the trained model is no longer valid. PCA methods based on moving-windows have been proposed for monitoring tasks when processes with considerable dynamic behaviour are considered in order to overcome some of the deficiencies of the static PCA approach (Ku et al., 1995; Baggiani and Marsili-Libelli, 2009).

In the moving-window approach, historical data from a time period defined by the *window-length* L are used for building PCA

models. New PCA models are built at the time intervals of a *shift-size* Z . In such a manner, a window shifts along time and a new model is trained at each step by including the newest data and excluding the oldest ones. The unseen testing data sets associated with each PCA model are of the length Z . Testing data are monitored using the continuously calculated T^2 and Q statistics and the cut-offs T_{lim}^2 and Q_{lim} determined for the latest model available. The contributions for the moving-window PCA approach can be calculated in a corresponding manner as for the conventional PCA. The fault monitoring procedure using the moving-window PCA technique with fixed L is shown for $1 \dots n$ models in Fig. 2(a). Conventionally, in the moving-window applications each model covers the same window-length and the shift-size is fixed (Kadlec et al., 2011).

Based on the value of Z , the moving-window techniques can be categorized as sample-wise and block-wise approaches (Kadlec et al., 2011). In the sample-wise techniques, $Z=1$ which means that the PCA model is recalculated after every new sample coming in. When the process operating conditions change abruptly, sample-wise moving-window models are efficient in monitoring (Choi et al., 2006). In the block-wise techniques, Z corresponds for a certain number of samples, or samples of a certain time period, after which the PCA model is recalculated. The advantages of the block-wise moving-window techniques include a low computational cost in comparison with the sample-wise techniques. The blockwise techniques also reduce the risk of recalculating the model based on an anomalous observation because the detected faulty samples can be discarded from the next training matrix prior recalculating the model (Choi et al., 2006).

2.3.3. Adaptive window-length procedure

Even though the moving-window PCA extension provides considerable advantages over the static PCA approach in monitoring of time evolving processes, one of its limitations is the fixed window-length. This is due to the fact that rapidness of the process changes varies. In general, if the process changes rapidly, the window-length should be shortened and when the changes are slow, the large window-length should be preferred. Adaptive window-lengths have been considered (Kadlec et al., 2011) and here we apply two window-length adaptation methods originally presented by He and Yang (2008) and Aych et al. (2012), denoted AMW_1 and AMW_2 hereafter (when used with the PCA technique, AMWPCA_1 and AMWPCA_2). Contrary to the adaptive window-lengths, the shift-size is fixed.

Using the AMW_1 method (He and Yang, 2008), window-length L is defined for each model $\{1, \dots, n, \dots, N\}$ as follows:

$$L(n) = L_{min} + (L_{max} - L_{min}) \exp \left\{ - \left(\alpha \frac{\|\Delta \mathbf{b}(n-1)\|}{\|\Delta \mathbf{b}_0\|} + \beta \frac{\|\Delta \mathbf{R}(n-1)\|}{\|\Delta \mathbf{R}_0\|} \right)^\gamma \right\} \quad (4)$$

where L_{min} and L_{max} are minimum and maximum window-lengths, respectively. $\|\Delta \mathbf{b}(n-1)\|$ is the Euclidean vector norm of difference between the previous two consecutive $1 \times D$ mean vectors, $\mathbf{b}(n-1)$ and $\mathbf{b}(n-2)$, calculated from training data. For a $M(n) \times D$

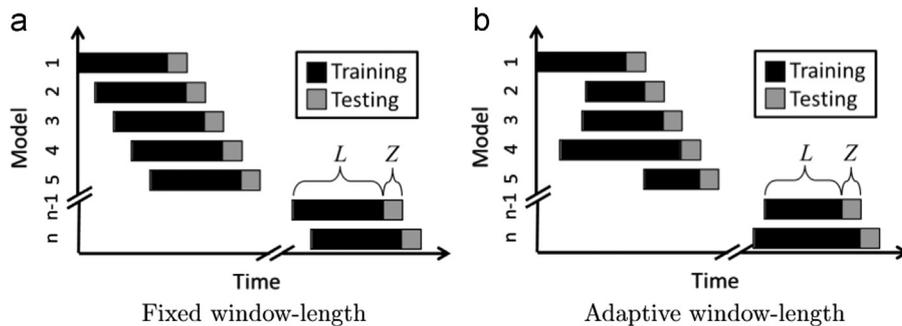


Fig. 2. Moving-window monitoring procedures using fixed (a) and adaptive (b) window-lengths L with $1, \dots, n$ PCA models. Shift-size Z is fixed in both (a) and (b).

training data matrix $\mathbf{X}_{tm}(n)$, where $M(n)$ represents the number of observations that is specific to each training matrix, the mean vector $\mathbf{b}(n)$ is computed according to the following equation:

$$\mathbf{b}(n) = \frac{1}{M(n)} \sum_{i=1}^{M(n)} \mathbf{x}_{tm_i} \quad (5)$$

where \mathbf{x}_{tm_i} is the i th row vector of $\mathbf{X}_{tm}(n)$. Because of the data pre-processing, observations may have been discarded from training data matrices and therefore $M(n) \leq L(n)$. Correspondingly, $\|\Delta \mathbf{R}(n-1)\|$ is the Euclidean matrix norm of the difference between the two consecutive $D \times D$ correlation matrices, $\mathbf{R}(n-1)$ and $\mathbf{R}(n-2)$. For a data matrix $\mathbf{X}_{tm}(n)$, the correlation matrix $\mathbf{R}(n)$ is calculated as follows:

$$\mathbf{R}(n) = \frac{1}{M(n)} \sum_{i=1}^{M(n)} (\mathbf{x}_{tm_i} - \mathbf{b}(n))(\mathbf{x}_{tm_i} - \mathbf{b}(n))^T \quad (6)$$

$\|\Delta \mathbf{b}_0\|$ and $\|\Delta \mathbf{R}_0\|$ represent the Euclidean vector norm of difference between two consecutive mean vectors and the Euclidean matrix norm of the difference between two consecutive correlation matrices in reference conditions, respectively. They are calculated correspondingly as $\|\Delta \mathbf{b}(n-1)\|$ and $\|\Delta \mathbf{R}(n-1)\|$, using two sets of reference data that associate with normal process conditions without anomalous observations. Three parameters are used for tuning the function; α and β are weights given for $\|\Delta \mathbf{b}(n-1)\|/\|\Delta \mathbf{b}_0\|$ and $\|\Delta \mathbf{R}(n-1)\|/\|\Delta \mathbf{R}_0\|$, respectively, and γ is an exponential parameter that affects the sensitivity of L to the process change.

When using Eq. (4), the values of the window-lengths L within the range defined by L_{min} and L_{max} depend on the differences between the consecutive training data sets and the settings of the function parameters. If $\|\Delta \mathbf{b}(n-1)\|$ and/or $\|\Delta \mathbf{R}(n-1)\|$ reduce (s) implying that the variation between two previous consecutive training data sets decreases, L of the next training data set increases. Also the reduction of the weight(s) α and/or β causes a raise in L . On the contrary, a decrease of the exponential parameter γ reduces the length L and leads to less aggressive responses of L to the variations within the measurement data.

With the AMW_2 approach (Ayeche et al., 2012), the window-lengths are determined accordingly:

$$L(n) = L_{max} - (L_{max} - L_{min}) [1 - \exp(-\delta(\|\Delta \mathbf{R}_{ref}(n-1)\|))] \quad (7)$$

where $\|\Delta \mathbf{R}_{ref}(n-1)\|$ is the Euclidean matrix norm of the difference between $\mathbf{R}(n-1)$ and \mathbf{R}_{ref} . Otherwise $\|\Delta \mathbf{R}_{ref}(n-1)\|$ is calculated like $\|\Delta \mathbf{R}(n-1)\|$, but instead of using the second previous correlation matrix in its calculation, \mathbf{R}_{ref} representing the correlation matrix of a reference data set is used. The parameter δ controls the sensitivity of the change in L .

When using Eq. (7), the window-lengths L vary between L_{min} and L_{max} depending on the differences between the training data and the reference data and on the value of the preset the function parameter δ . In the case of the decreasing value of $\|\Delta \mathbf{R}_{ref}(n-1)\|$, the variability between the previous training data set and the reference data set diminishes and, consequently, the length L of the next training data set increases. A reduction of the δ value results in a raise of L .

The anomaly monitoring procedure with an adaptive moving-window PCA follows the same principles, for instance recomputing T_{lim}^2 and Q_{lim} for each model n , as in the fixed window-length case. The monitoring procedure for 1, ..., n models is visualized in Fig. 2(b).

2.3.4. Anomaly monitoring algorithm

A wide scale of methods has been presented for selecting a sufficient subset of PCs, including heuristic and statistical approaches (Valle et al., 1999; Jolliffe, 2002). It has been suggested that also for the adaptive PCA approaches the number of retained PCs should be individually determined for each model (Venkatasubramanian et al., 2003). In this work, we apply the eigengap technique (Davis and Kahan, 1970) for selecting an appropriate number of PCs for the

models. When the eigenvalues sorted in descending order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_D$, the eigengap is defined as $\mu_d = \lambda_d - \lambda_{d+1}$, with $d = \{1, \dots, D-1\}$. The index of the eigenvalue associated with the largest eigengap defines the dimensionality S of the projection subspace. The use of the eigengap method, that has also recently been applied in other studies for choosing the retained PCs (Shen et al., 2012), is based on the dominant standpoint that suggests the first PCs associating with large eigenvalues to contain the most significant information about the original variables (Bro and Smilde, 2014). Alternative approaches that rely both on the fault detection sensitivity and on the fault direction, such as the signal-to-noise ratio, have been proposed for selecting the significant PCs in order to monitor the specific fault types (Tamura and Tsujita, 2007). Such approaches could be further investigated for potentially improving the anomaly detection algorithm presented in this study as well.

The window-lengths were defined in time instead of the number of samples. This implies that the number of samples M in a training data matrix \mathbf{X}_{tm} of each model n may have been less than the value of L because of the discarding procedure in the data pre-processing step. For this reason, a certain proportion P of the samples in a training data set was required. If the requirement was not fulfilled, the PCA model was not considered representative and the previous valid model was maintained. The required P value varied depending on L , the criterion being stricter for shorter windows in order to have sufficient samples for building descriptive PCA models. For a data set of length L_{min} , a requirement of 90% of the samples available was applied whereas for a data set of length L_{max} , the limit was set at 50%. In particular, the required proportion of samples P_{lim} in any data set n was determined as follows:

$$P_{lim}(n) = (0.5 - 0.9) \frac{L(n) - L_{min}}{L_{max} - L_{min}} + 0.9 \quad (8)$$

A 30-day period at the beginning of the acquired data set was used for defining the reference values for the AMW approaches. Data from the reference period were divided in N_{ref} sets, which in this case equaled 30. PCA was performed for each $n = \{1, \dots, N_{ref}\}$ 1-day reference data set $\mathbf{X}_{ref}(n)$. Before performing PCA, data were standardized, i.e. made zero mean and unit variance. $\mathbf{X}_{ref}(n)$ were further cleaned from the samples that violated T_{lim}^2 or Q_{lim} . The confidence level of 97.5% was employed for defining T_{lim}^2 and Q_{lim} for each model separately. For the AMW_1 method, $\|\Delta \mathbf{b}_0(n)\|$ and $\|\Delta \mathbf{R}_0(n)\|$ were calculated for each model except the first one ($n = 1$) that cannot be determined. Finally, the references $\|\Delta \mathbf{b}_0\|$ and $\|\Delta \mathbf{R}_0\|$ were defined as the averages of the corresponding values associated with the individual models $n = \{2, \dots, N_{ref}\}$. As for the AMW_2 approach, the same data sets $\mathbf{X}_{ref}(n)$ were used for calculating the reference matrix \mathbf{R}_{ref} . Specifically, \mathbf{R}_{ref} was defined as an element-by-element average matrix of the correlation matrices $\mathbf{R}(n)$ from the reference period.

The algorithms for calculating the reference(s) for determining the window-lengths adaptively and for detecting and isolating anomalies in testing data sets \mathbf{X}_{test} are sketched below. Algorithm 1 is performed only once for defining the references $\|\Delta \mathbf{b}_0\|$ and $\|\Delta \mathbf{R}_0\|$ (AMW_1) and \mathbf{R}_{ref} (AMW_2). Algorithm 2 is executed at the time intervals of Z for defining L , T_{lim}^2 and Q_{lim} and, then, for monitoring the T^2 and Q statistics of each incoming sample. The anomaly monitoring procedure of the moving-window PCA with fixed window-lengths (MWPCA) follows Algorithm 2 except for step 2 that concerns the calculation of window-length L individually for each model n according to Eq. (4) (AMWPCA_1) or to Eq. (7) (AMWPCA_2).

Algorithm 1. Calculation of the reference output(s) for the window-length determination.

Input: Reference data matrices \mathbf{X}_{ref} , confidence level, number of reference data matrices N_{ref}

Output: Vector norm $\|\Delta \mathbf{b}_0\|$ and matrix norm $\|\Delta \mathbf{R}_0\|$, or correlation matrix \mathbf{R}_{ref}

```

1: for  $n = 1 : N_{ref}$  do
2: Standardize  $\mathbf{X}_{ref}$ 
3: Calculate the samples retained in the window after the
   data pre-processing,  $P$ 
4: Calculate the limit for the required proportion of samples
   in the window,  $P_{lim}$ 
5: Determine confidence limits for the statistics
6: if Enough samples are retained in the window, i.e.  $P \geq P_{lim}$ 
7: Perform PCA
8: Determine the dimensionality  $S$ , i.e. the number of the
   retained PCs
9: for  $d = 1 : D - 1$  do
10: Calculate the eigengap  $\mu_d$ 
11: end for
12:  $S = \operatorname{argmax} \mu_d$ 
13: Calculate  $T_{lim}^2$  and  $Q_{lim}$ 
14: else Use  $T_{lim}^2$  and  $Q_{lim}$  of the previous valid model
15: end if
16: Clean  $\mathbf{X}_{ref}$  from the samples exceeding cut-off limits
17: for each observation  $\mathbf{x}_{ref}$  at time  $k$  in matrix  $\mathbf{X}_{ref}$  do
18: Calculate  $T^2$  and  $Q$ 
19: if  $T^2 > T_{lim}^2$  and/or  $Q > Q_{lim}$  then
20: Discard sample
21: end if
22: end for
23: if  $n > 1$  then
24: Calculate the reference output(s) using non-standardized
    $\mathbf{X}_{ref}$ 
25: if AMWPCA_1 is employed then
26: Calculate  $\|\Delta \mathbf{b}_0\|$  and  $\|\Delta \mathbf{R}_0\|$ 
27: else AMWPCA_2 is employed
28: Calculate  $\mathbf{R}_{ref}$ 
29: end if
30: end if
31: end for
32: Calculate the mean reference output(s) over models
    $\{2, \dots, N_{ref}\}$ 
33: if AMWPCA_1 is employed then
34: Calculate mean  $\|\Delta \mathbf{b}_0\|$  and mean  $\|\Delta \mathbf{R}_0\|$ 
35: else AMWPCA_2 is employed
36: Calculate mean  $\mathbf{R}_{ref}$ 
37: end if

```

Algorithm 2. Anomaly detection and isolation.

Input: Training data matrices \mathbf{X}_{trn} , testing data matrices \mathbf{X}_{tst} , confidence level, function parameters α , β and γ , or function parameter δ

Output: Confidence limits T_{lim}^2 and Q_{lim} , statistics T^2 and Q , contributions \mathbf{c} and \mathbf{e}

```

1: for  $n = N_{ref} + 1 : \text{end}$  do
2: Calculate window-length  $L$  using  $\|\Delta \mathbf{b}_0\|$  and  $\|\Delta \mathbf{R}_0\|$ , or  $\mathbf{R}_{ref}$ 
3: Standardize  $\mathbf{X}_{trn}$ 
4: Calculate the samples retained in the window after the
   data pre-processing,  $P$ 
5: Calculate the limit for the required proportion of samples
   in the window,  $P_{lim}$ 
6: Determine confidence limits for the statistics
7: if Enough samples are retained in the window, i.e.  $P \geq P_{lim}$ 
then
8: Perform PCA
9: Determine the dimensionality of  $S$ , i.e. the number of
   retained PCs
10: for  $d = 1 : D - 1$  do

```

```

11: Calculate the eigengap  $\mu_d$ 
12: end for
13:  $S = \operatorname{argmax} \mu_d$ 
14: Calculate  $T_{lim}^2$  and  $Q_{lim}$ 
15: else Use  $T_{lim}^2$  and  $Q_{lim}$  of the previous valid model
16: end if
17: for each observation  $\mathbf{x}_{tst}$  at time  $k$  in matrix  $\mathbf{X}_{tst}$  do
18: Standardize  $\mathbf{x}_{tst}$ 
19: Calculate  $T^2$  and  $Q$ 
20: Check the respect of confidence limits
21: if  $T^2 > T_{lim}^2$  and/or  $Q > Q_{lim}$  then
22: Anomalous sample: calculate  $\mathbf{c}$  and  $\mathbf{e}$ , i.e., the variables'
   contributions to  $T^2$  and  $Q$ 
23: end if
24: end for
25: end for

```

3. Results and discussion

In this section, we first describe the selection of the parameters for the anomaly detection systems, where window-lengths adjust (AMWPCA_1 and AMWPCA_2) and where the window-length is fixed (MWPCA). Then, examples of their implementation in an ASP are presented with a particular consideration assigned to their performances considering different types of anomalies. Finally, we summarize the results for process operation during the entire testing period.

3.1. Selection of the parameters

Several parameters were needed to be adjusted in the anomaly monitoring procedure. The first 4000 samples of the acquired data set were used for selecting the appropriate shift-sizes and the parameters required for the window-length calculations according to Eqs. (4) and (7). The parameter selection for the different monitoring approaches is described in the following order: AMWPCA_1, AMWPCA_2 and MWPCA. We discuss first the parameters that were selected based on the process knowledge or the evident properties of the acquired operational data. After that, the selection of those parameters that were more intensively investigated is reported.

3.1.1. Adaptive window-length approach AMWPCA_1

The range that limits the L values was selected based on *a priori* knowledge about the influent behaviour in municipal WWTPs. L_{min} was set at 24 h (1 day) and L_{max} at 168 h (7 days). In particular, L_{min} represents the diurnal trends and L_{max} the weekly trends which both are typical for the influent flow rate and concentrations in municipal WWTPs (Henze et al., 2008). These trends were also evidently present in the operational data of the Viikinmäki WWTP.

The parameters α and β are the weights of $\|\Delta \mathbf{b}(n-1)\|/\|\Delta \mathbf{b}_0\|$ and $\|\Delta \mathbf{R}(n-1)\|/\|\Delta \mathbf{R}_0\|$, respectively. In order to restrict the variation of α and β , we defined their relationship as follows: α is in the range 0–1 and $\beta = 1 - \alpha$. The magnitudes of $\|\Delta \mathbf{b}(n-1)\|/\|\Delta \mathbf{b}_0\|$ and $\|\Delta \mathbf{R}(n-1)\|/\|\Delta \mathbf{R}_0\|$ were found to be of the same level and no reason for weighting the changes in mean vectors over the changes in correlation matrices, or *vice versa*, was recognized. For these reasons, we selected the weights in the window-length calculation procedure to be $\alpha = 0.5$ and $\beta = 0.5$.

The selection of the shift-size Z and the function parameter γ are investigated in the rest of this subsection. The potential shift-sizes Z were studied within the range 1–24 h, at the intervals of 1 h. The range $\leq L_{min}$ is motivated by the findings that indicate that unforeseen changes in the process or the instrument signals often take place within the time frames of a few hours. Therefore,

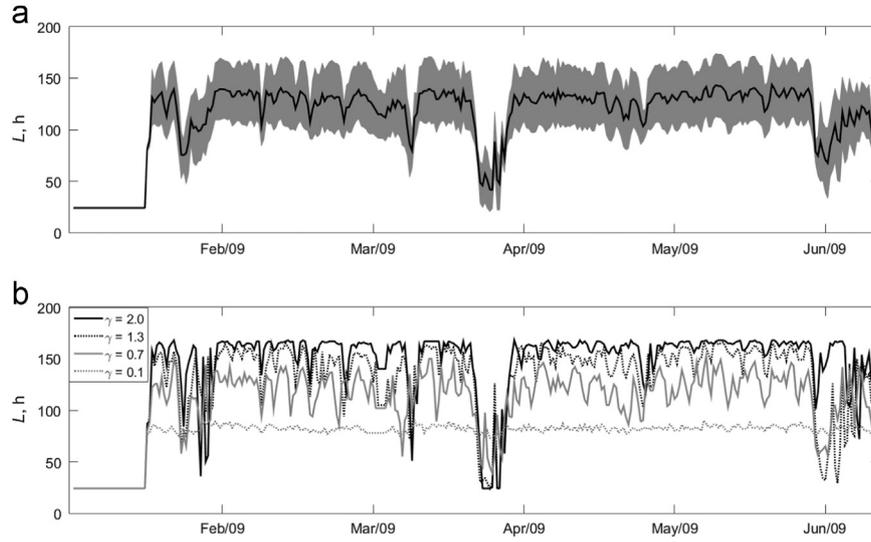


Fig. 3. Time-series of the average L over γ between 0.0 and 2.0 for AMWPCA_1. The grey area is defined by $\mu_L(n) + \sigma_L(n)$ and $\mu_L(n) - \sigma_L(n)$ (a). Time-series of L resulting with four different γ values shown in the legend (b).

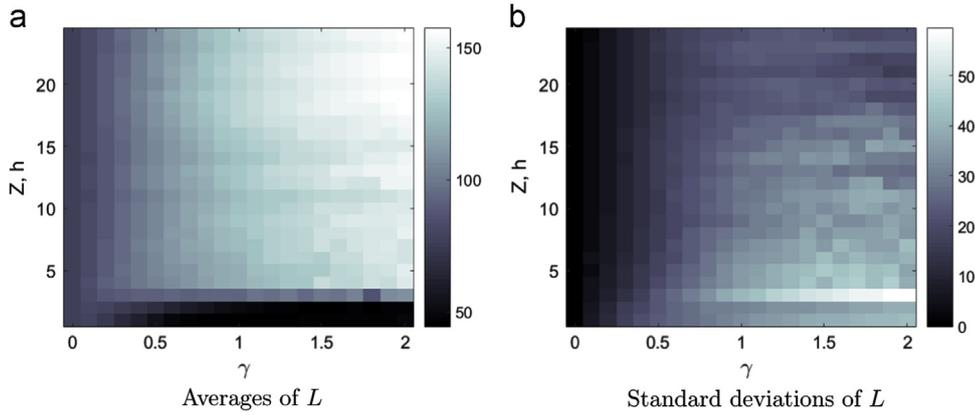


Fig. 4. Averages (a) and standard deviations (b) of L with the investigated combinations of Z and γ for AMWPCA_1. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

recomputing the PCA model more frequently than once a day is expected to be beneficial. The function parameter γ affects the intensity of the changes in L as the consequences of the variations in the operational conditions. It was examined among the values $\gamma = [0.0, 0.1, 0.2, \dots, 2.0]$. Particularly, if $\gamma = 0.0$, there is no variation in L which in the given situation always equals 77.0 h.

The time-series of the average L over all γ values and using $Z = 12$ is shown in Fig. 3(a). $Z = 12$ is applied in the example because it is the middle of the investigated γ range. The grey shading in the figure is delimited by the average L over the γ values, denoted $\mu_L(n)$, and the standard deviation of L over the γ values, denoted $\sigma_L(n)$, for each model n as follows: maximum is $\mu_L(n) + \sigma_L(n)$ and minimum is $\mu_L(n) - \sigma_L(n)$. Variation in the average L can be easily observed. The minimum and maximum average L are 41.3 h and 143.7 h, respectively. The time-series of L when using $\gamma = [0.1, 0.7, 1.3, 2.0]$ and $Z = 12$ are presented in Fig. 3(b), in order to demonstrate the effect of γ on the L values. The large γ values give rise to typically large L values. They also cause drastic drops and variation of L when the properties of the consecutive training matrices differ from each other. On the other hand, the small γ values cause only minor differences in L and, therefore, their use leads practically close to fixed window-lengths.

Next, the effect of the different Z and γ combinations on the window-lengths L were examined. The average L values from the considered Z and γ combinations, denoted $\mu_L(Z, \gamma)$, are used to dye

the rectangles in Fig. 4(a) as shown in the colourbar. The longest average windows are located in the upper right corner of the plot, while short windows are found in the bottom. The standard deviations of L , denoted $\sigma_L(Z, \gamma)$, were found to be the largest with small Z and large γ , as shown on the bottom right of Fig. 4(b). The lowest variation of L using any Z was associated with the small γ values. The maximum and minimum $\mu_L(Z, \gamma)$ and $\sigma_L(Z, \gamma)$ among the investigated combinations of Z and γ are presented in Table 2, where the associated Z and γ values are also reported.

In the selection of Z and γ , we aimed at such a combination that results in window-lengths that clearly vary when changes between the consecutive training matrices arise. Therefore, too small γ values were not appealing (see the time-series of L with $\gamma = 0.1$ in Fig. 3(b)). On the other hand, too abrupt a variation among the window-lengths, that is associated with the large γ values, was not desired in order to maintain stability in the monitoring system. For these reasons, the γ values that result in the variation behaviour that is close to the average L time-series over the considered γ values were searched for, concerning each Z individually. The root mean squared differences D_γ between average L and L with a certain γ value were quantified as follows:

$$D_\gamma = \left(\frac{1}{N} \sum_{n=1}^N (L_\gamma(n) - \mu_L(n))^2 \right)^{1/2} \quad (9)$$

where $L_\gamma(n)$ denotes the L defined using a specific γ concerning model n and $\mu_L(n)$ denotes the average of the L values over the investigated γ values for the models with serial number n . The γ values that yield minimum D_γ for each considered Z are indicated with red dots in Fig. 5(a), where the D_γ values of the different combinations are shown in the colourbar. The majority of the γ values resulting in a minimum D_γ are in the range 0.7–0.9.

From the practical point of view, a too large share of observations is not desired to be labelled as anomalous and, thus, the monitoring system to cause too frequent alarms. For this reason, the shares of anomalous samples were investigated. The anomaly shares associated with the different Z and γ combinations are visualized in Fig. 5(b). A share of anomalies corresponding with the maximum of 0.25 was considered acceptable and the blocks that satisfy this condition are marked with red dots. The largest anomaly shares are found in the top left corner of the figure, along with the combinations of $Z = 2$ and $\gamma \geq 1.5$. The maximum anomaly share among the studied combinations is 0.37 ($Z = 24$, $\gamma = 0.0$). The smaller anomaly shares are located in the bottom left corner, the minimum value being 0.18 ($Z = 1$, $\gamma = 0.1$).

The combination of Z and γ to be used in the testing of AMWPCA_1 was selected among those that result in minimum D_γ for any Z and that associate with a share of the detected anomalies of maximum 0.25. In other words, the possible combinations are marked with a red dot both in Fig. 5(a) and in Fig. 5(b). Among seven combinations that fulfilled the criteria, the one that resulted in the smallest D_γ was selected ($Z = 6$, $\gamma = 0.7$). All the parameters that were chosen for testing the AMWPCA_1 approach are summarized in Table 3.

3.1.2. Adaptive window-length approach AMWPCA_2

A number of parameters were also selected for the other adaptive window-length approach, AMWPCA_2. Based on the criteria explained in Section 3.1.1, L_{min} was set at 24 h and L_{max} at 168 h. The adjustment of the shift-size Z and the function parameter δ was approached correspondingly as the selection of Z and

Table 2
Maximum and minimum of the average L values ($\mu_L(Z, \gamma)$) and of the standard deviations in L ($\sigma_L(Z, \gamma)$) among the investigated combinations of Z and γ for AMWPCA_1.

	$\mu_L(Z, \gamma)$ (h)	Z (h)	γ
Max $\mu_L(Z, \gamma)$	157.4	21	2.0
Min $\mu_L(Z, \gamma)$	44.6	1	0.9
	$\sigma_L(Z, \gamma)$ (h)	Z (h)	γ
Max $\sigma_L(Z, \gamma)$	59.4	3	1.9
Min $\sigma_L(Z, \gamma)$	0.0	{1,2,...,24}	0.0

γ for the AMWPCA_1 approach. The influence of Z on the variation of L was studied within the range 1–24 h at the intervals of 1 h and of δ among the values $\delta = [0.0, 0.1, 0.2, \dots, 2.0]$. When $\delta = 0.0$, L is fixed at the size of 168 h, i.e. L_{max} .

The time-series of the average L computed over the investigated δ values and $Z = 12$ is shown in Fig. 6(a). The grey shading is defined follows: maximum is $\mu_L(n) + \sigma_L(n)$ and minimum is $\mu_L(n) - \sigma_L(n)$. The minimum and maximum of the average L values are 52.3 h and 100.4 h, respectively. Hence, the variation of L when using AMWPCA_2 is more moderate than when using AMWPCA_1. The time-series of L with $\delta = [0.1, 0.7, 1.3, 2.0]$ and $Z = 12$ are presented in Fig. 6(b). The δ value evidently sets the general level of L . For instance, the large δ values force the L values close to L_{min} and the small δ values close to L_{max} . Moreover, it can be observed that the small δ values give rise only to small deviation in L . The largest fluctuation in L takes place when δ is moderate.

The impact of the different Z and δ combinations on the average window-lengths is shown in Fig. 7(a). The colours of the rectangles associate with the L values as indicated in the colourbar. The largest average window-lengths, $\mu_L(Z, \delta)$, equaling L_{max} are located in the left of the plot whereas the $\mu_L(Z, \delta)$ values get gradually smaller when moving to the right side, along with the increasing δ values. The largest standard deviations of L , $\sigma_L(Z, \delta)$, were produced with moderate δ for any Z , as it is illustrated in Fig. 7(b), while the lowest $\sigma_L(Z, \delta)$ were associated with the small δ values. The maximum and minimum $\mu_L(Z, \delta)$ and $\sigma_L(Z, \delta)$ are collected in Table 4, together with the associated Z and δ values.

The differences between the time-series of average of the L values over the considered δ values and the time-series of L calculated with a certain δ value were studied. The δ values associated with the minimum D_δ values (calculated correspondingly as D_γ in Eq. (9)) are marked with red dots in Fig. 8(a). In all except one of the situations, the δ value resulting in a minimum D_δ was 0.8. These cases typically correspond also to the largest variation in L , as it is depicted in Fig. 7(b).

The shares of anomalous samples are shown in Fig. 8(b), where the ones representing values ≤ 0.25 are marked with red dots. The largest anomaly shares are among the combinations in the top right corner, the maximum being as high as 0.56 ($Z = 19$, $\delta = 2.0$). The combinations with the smallest anomaly shares are located in the bottom left corner, with the minimum value of 0.17 ($Z = 1$, $\delta = 0.2$).

Table 3
Parameters selected for testing AMWPCA_1.

L_{min} (h)	L_{max} (h)	α	β	γ	Z (h)
24	168	0.5	0.5	0.7	6

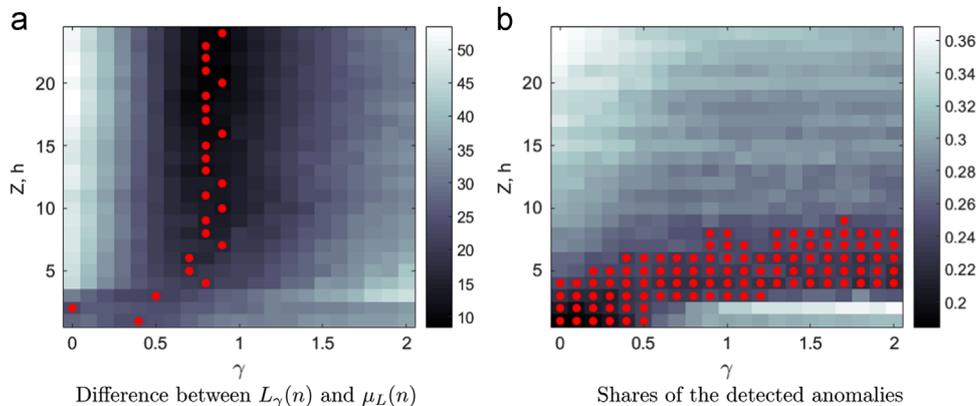


Fig. 5. Differences D_γ between $L_\gamma(n)$ and $\mu_L(n)$ (a) and shares of detected anomalies (b) with the investigated combinations of Z and γ for AMWPCA_1. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

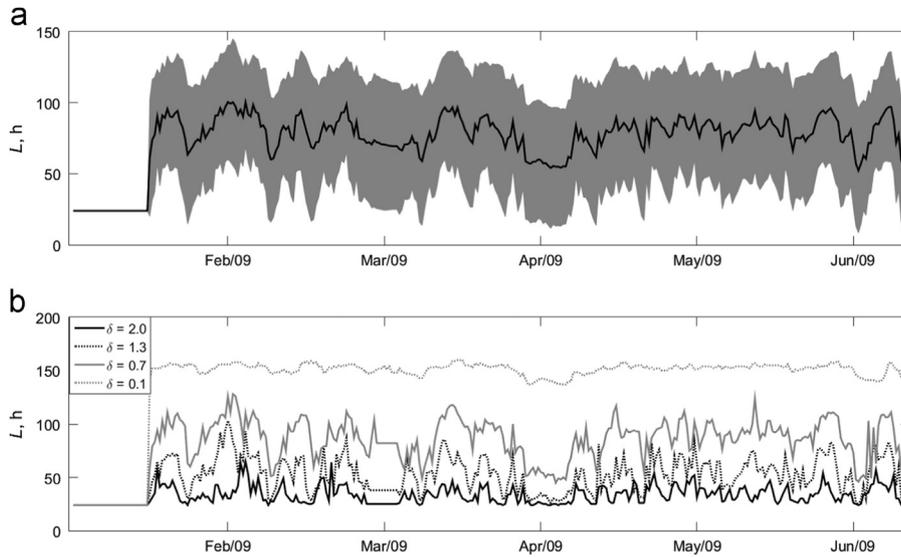


Fig. 6. Time-series of the average L over δ between 0.0 and 2.0 for AMWPCA_2. The grey area is defined by $\mu_L(n) + \sigma_L(n)$ and $\mu_L(n) - \sigma_L(n)$ (a). Time-series of L with four different δ values shown in the legend (b).

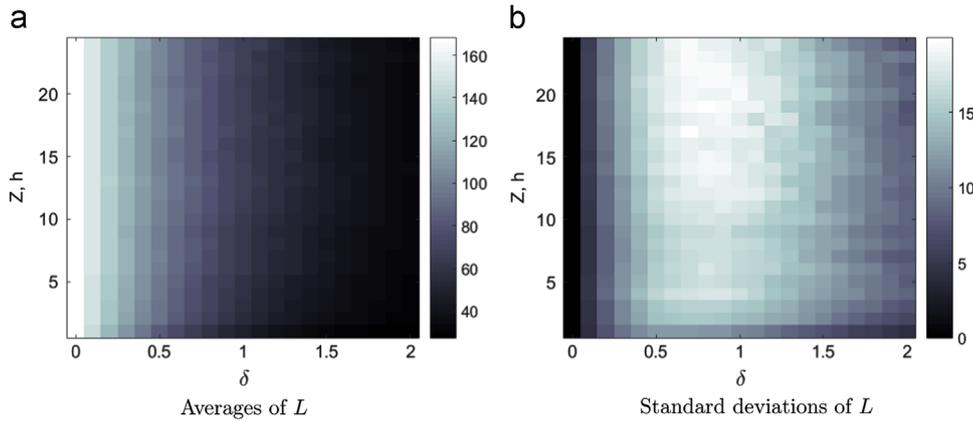


Fig. 7. Averages (a) and standard deviations (b) of L with the investigated combinations of Z and δ for AMWPCA_2. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Table 4

Maximum and minimum of the average L values ($\mu_L(Z, \delta)$) and of the standard deviations in L ($\sigma_L(Z, \delta)$) among the investigated combinations of Z and δ for AMWPCA_2.

	$\mu_L(Z, \delta)$, (h)	Z , (h)	δ
Max $\mu_L(Z, \delta)$	168.0	{1,2,...,24}	0.0
Min $\mu_L(Z, \delta)$	27.7	1	2.0
	$\sigma_L(Z, \delta)$ (h)	Z (h)	δ
Max $\sigma_L(Z, \delta)$	19.9	19	0.8
Min $\sigma_L(Z, \delta)$	0.0	{1,2,...,24}	0.0

The combination of Z and δ to be used in the testing of the anomaly detection performance of AMWPCA_2 was selected among the ones that associate with minimum D_δ for any Z and with the share of the detected anomalies of maximum 0.25. Three combinations satisfied the criteria and among those the one that yielded in the smallest D_δ was selected ($Z = 1, \delta = 0.7$). The parameters chosen for testing the AMWPCA_2 approach are reported in Table 5.

3.1.3. Fixed window-length approach MWPCA

Anomaly monitoring using fixed window-length MWPCA was also considered for the comparison purposes. Two parameters had to be

selected for MWPCA: window-length L and shift-size Z . The considered window-lengths values were $L = [24 \text{ h}, 48 \text{ h}, \dots, 168 \text{ h}]$. In other words, the smallest investigated L corresponds with L_{min} used in the adaptive window-length approaches and the largest one with L_{max} . The set of the studied shift-size values was $Z = [1 \text{ h}, 2 \text{ h}, \dots, 24 \text{ h}]$ that is the same as with the techniques with adapting L values.

The shares of the observations detected as anomalous with the different Z and L combinations are shown in Fig. 9. The rectangles connected with the specific combinations are dyed according to the anomaly shares as indicated in the colourbar. The large anomaly shares result when the small window-lengths are employed, the maximum value being 0.63 ($Z = 23 \text{ h}, L = 24 \text{ h}$). The long windows provide the least sensitive models that are linked with the small shares of the anomalous samples, with the minimum of 0.17 ($Z = 1 \text{ h}, L = 168 \text{ h}$). The Z and L combinations that associate with the anomaly share of maximum 0.25 are marked with red dots.

Among all the 168 combinations of Z and L , 76 satisfied the criterion of the anomaly detection share ≤ 0.25 . The models with the various combinations fulfilling the criterion result in considerably different overall anomaly monitoring performances (Fig. 9). Because the motivation for testing MWPCA was to compare its performance with the AMWPCA approaches, abundantly different Z and L values for the MWPCA approach from the levels of those when using the adapting techniques were not desired.

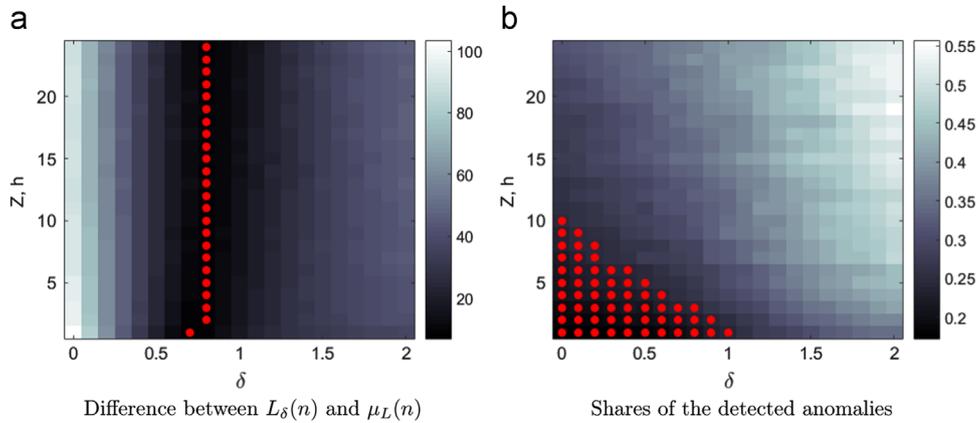


Fig. 8. Differences D_δ between $L_\delta(n)$ and $\mu_L(n)$ (a) and shares of detected anomalies (b) with the investigated combinations of Z and δ for AMWPCA_2. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Table 5
Parameters selected for testing AMWPCA_2.

L_{min} (h)	L_{max} (h)	δ	Z (h)
24	168	0.7	1

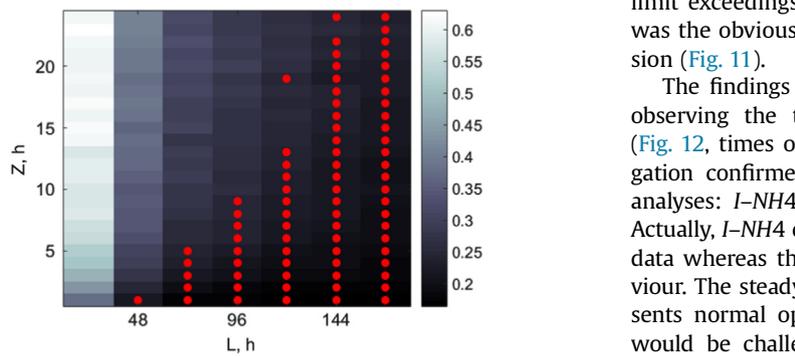


Fig. 9. Shares of detected anomalies with the investigated combinations of Z and L for MWPCA. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

When considering both AMWPCA_1 and AMWPCA_2 together, the following average values with the chosen parameters during the reference period were calculated: $Z = 3.5$ h and $L = 89.9$ h. Based on these averages and the criterion of the maximum accepted level of the anomaly detection share, the selected parameters for testing the MWPCA approach were $Z = 4$ h and $L = 96$ h.

3.2. Example of anomaly detection

After the parameter selection, the rest of the collected data, consisting of 13 520 hourly samples, were employed for testing the anomaly monitoring approaches. Particularly, two instrument anomaly cases and one process anomaly case were investigated using the AMWPCA_1, AMWPCA_2 and MWPCA techniques with the parameter set-ups reported in Section 3.1.

3.2.1. Instrument anomalies

Case 1 – drifting measurement: The performances of the AMWPCA_1, AMWPCA_2 and MWPCA anomaly detection systems were explored during a period of three days. The T^2 and Q statistics were studied with respect to T_{lim}^2 and Q_{lim} (Fig. 10). The first violations of T_{lim}^2 or Q_{lim} during the examined period for the monitoring techniques are shaded in grey. Q_{lim} was the threshold that was first violated

considering each anomaly detection approach. That happened using AMWPCA_1 (L between 30 and 120 h during the period) at 8 pm on October 16, using AMWPCA_2 (L between 52 and 69 h) at 5 pm on the following day, October 17, and using MWPCA at 11 pm on October 16.

The explanations for the detected Q_{lim} violations were investigated by analysing variables' contributions to the Q statistic considering a model with one PC at the moments of the confidence limit exceedings. The contribution bar plots indicate that $I-NH4$ was the obvious reason for the threshold violations in each occasion (Fig. 11).

The findings of contribution analyses were further studied by observing the time-series of the standardized input variables (Fig. 12, times of the first Q_{lim} violation highlighted). The investigation confirmed the information provided by the contribution analyses: $I-NH4$ deviates clearly from the rest of the variables. Actually, $I-NH4$ drifts from about 50 mg/l to 80 mg/l in the original data whereas the other variables represent normal diurnal behaviour. The steady drifting of $I-NH4$ that, on the other hand, represents normal operational area in the beginning of the anomaly would be challenging to detect timely using univariate control charts. With multivariate techniques such as PCA, also the slow changes in one variable with respect to the simultaneous changes in the others can be recognized as demonstrated in this example.

Concluding, all the methods were found to detect the drifting measurement, AMWPCA_1 providing the most effective anomaly monitoring performance throughout the episode. However, the AMWPCA_2 approach recognized the drifting failure substantially later (18–21 h) than the other studied monitoring systems. If $I-NH4$ was included in a feedforward control scheme of the ASP, this long time delay in detecting a measurement drift would be substantially inconvenient for the solid process operation. The window-lengths, *i.e.* historical operational periods on which the monitoring procedure compared the incoming observations, do not explain the late detection of the drift with AMWPCA_2 in this case (Fig. 13). In fact, AMWPCA_1 operated with a shorter L and MWPCA with a longer L than AMWPCA_2 at the moment of the anomaly isolation with those approaches. However, it is likely that AMWPCA_2 adapted to the constant deviation of the $I-NH4$ signal because of the shorter shift-size than in the other examined systems. The other approaches actually result in larger differences between the consecutive training matrixes in the cases of drifting measurements because of the larger Z values.

Case 2 – peaks in measurements: Another example concerns a 3-week period in winter time. The time-series of T^2 and Q of the AMWPCA_1, AMWPCA_2 and MWPCA approaches are depicted in Fig. 14. The window-lengths of AMWPCA_1 ranged between 47 and 153 h during the episode and of AMWPCA_2 between 46 and 82 h. Several high peaks particularly in the Q statistic appeared on

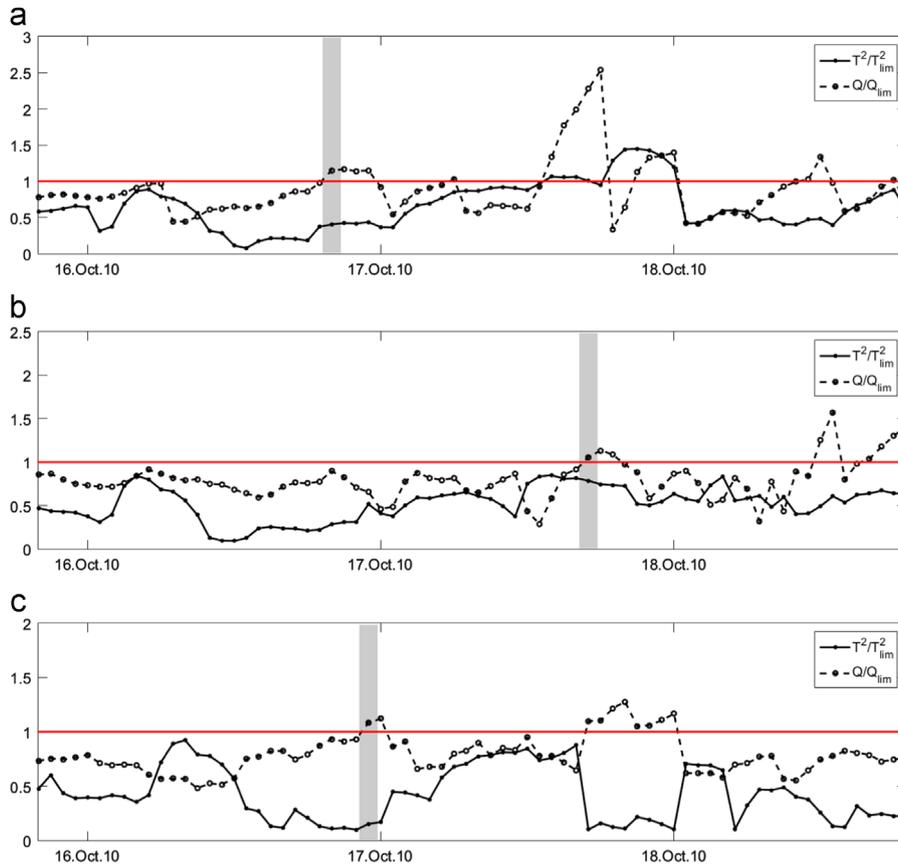


Fig. 10. Time-series of T^2 and Q using AMWPCA_1 (a), AMWPCA_2 (b) and MWPCA (c) during a 3-day period.

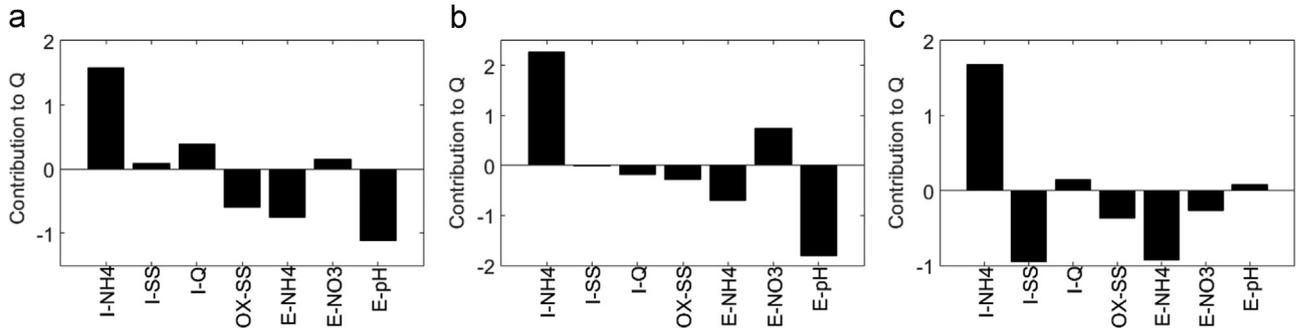


Fig. 11. Variables' contributions to Q using AMWPCA_1 (a) at 8 pm on October 16, using AMWPCA_2 (b) at 5 pm on October 17 and using MWPCA (c) at 11 pm on October 16.

December 20–24, partially depending on the employed anomaly monitoring technique.

The variables' contributions to the violated statistics during the highest peak (at 2 am on December 22 for AMWPCA_1 and AMWPCA_2 and at 5 am on December 22 for MWPCA; the first highlights in Fig. 14) were examined. In the cases of the AMWPCA_2 and MWPCA approaches only Q_{lim} was exceeded, whereas T_{lim}^2 too was violated for AMWPCA_1. The contributions indicate that the peaks were due to *I-SS* (Fig. 15). The peaks were clearly detected and isolated using each of the approaches.

The standardized time-series of the variables are represented in Fig. 16(a) with a focusing on three days in Fig. 16(b). The first high peak in *I-SS* (from about 80 mg/l up to 300 mg/l in the original data) correspond to the investigated peaks in statistics (Fig. 14). In addition, most of the other exceedings of the statistics' confidence limits on December 20–24 were connected with the largest contributions among the process variables associated with *I-SS*. During these days, the monitoring behaviour of AMWPCA_2 proved to be unstable

resulting in T^2 and Q values that frequently violated their cut-off limits and then quickly dropped below the limits. The observed instability is presumably connected with the small shift-size Z , which necessitates updating Q_{lim} and T_{lim}^2 at each time step, that was selected for the particular anomaly detection approach.

The time-series of the window-lengths during the studied 3-day episode are shown in Fig. 17. Considerable variation in L cannot be observed with the AMWPCA_2 technique. By contrast, the window-length using AMWPCA_1 gets evidently smaller (from 133 h to 96 h) as the consequence of the *I-SS* peak on December 22. Actually, AMWPCA_1 not only detected the considered peak most timely, but the sufficient adaptation of L to the process dynamics also resulted with the most adequate performance among the techniques concerning the entire period.

3.2.2. Process anomalies

Case 3 – process disturbance: A longer-term violation first on Q_{lim} and then on T_{lim}^2 took place on December 25–28 (Fig. 14). The

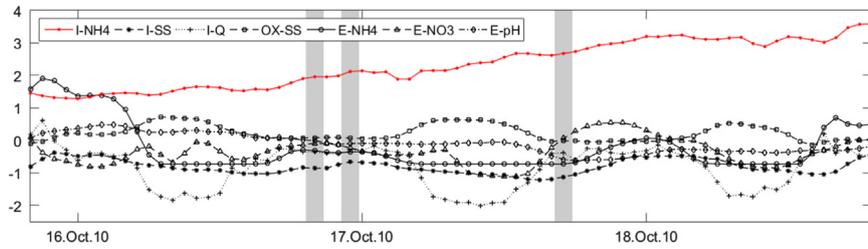


Fig. 12. Time-series of the standardized variables during a 3-day period.

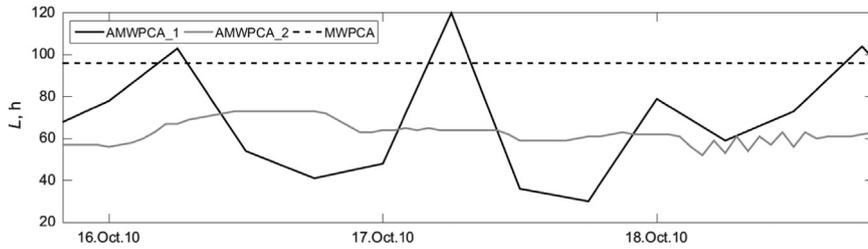


Fig. 13. Time-series of the window-lengths of the investigated anomaly monitoring approaches during a 3-day period.

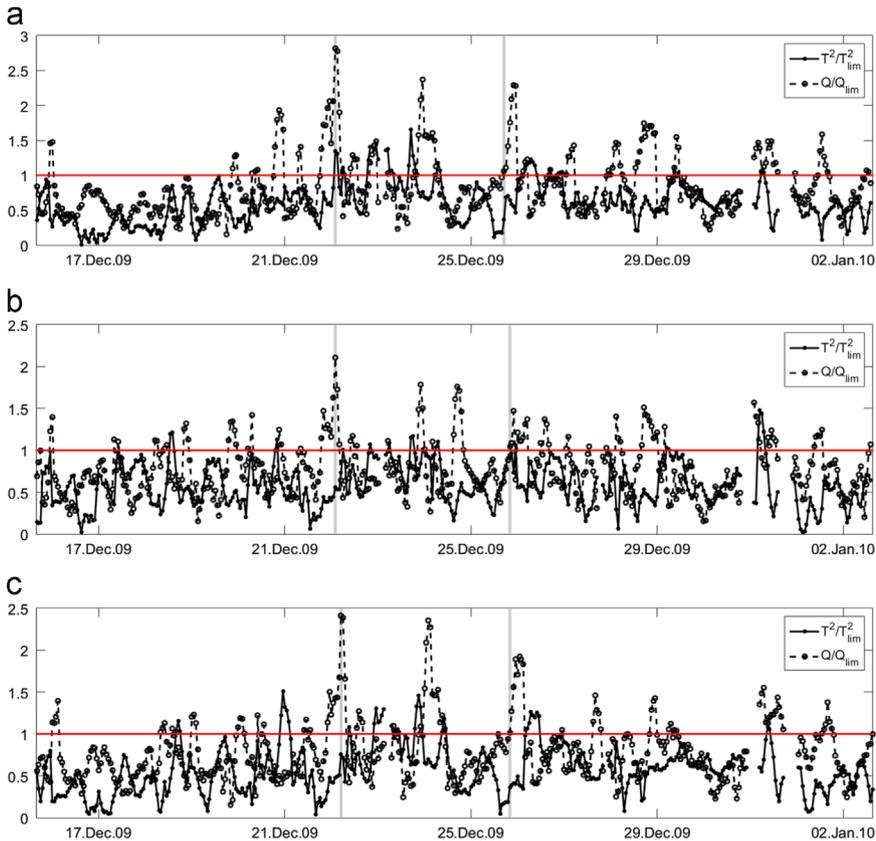


Fig. 14. Time-series of T^2 and Q using AMWPCA_1 (a), AMWPCA_2 (b) and MWPCA (c) during a 3-week period.

violations were detected at slightly differing times using the different approaches (at 5 pm on December 25 for AMWPCA_1; at 8 pm on December 25 for AMWPCA_2 and for MWPCA). Q_{lim} was violated first in each case and therefore the contributions to Q at the onset of the break were studied (the second grey bands in Fig. 14). $E-pH$ corresponded to the largest contribution in the case of each explored anomaly monitoring method (Fig. 18) and, in fact, its contribution to Q was even more evident during the next few hours when more drastic exceedings of Q_{lim} occurred.

The suggested anomalous $E-pH$ values were confirmed by observing the standardized time-series (the latter shaded 4-day period of Fig. 16(a) focused in Fig. 19; times of the first Q_{lim} violations highlighted) where a sudden drop in $E-pH$ took place (from 6.1 to 5.3 in the original data). Actually, the $E-pH$ reduction caused a malfunction in the biological nitrogen removal process resulting in high $E-NH_4$ peaks and in the increased concentration of $E-NO_3$ (see Fig. 16(a)). $E-NH_4$ increased from the normal concentration level of 0–7 mg/l all the way to 20 mg/l, which is the maximum of

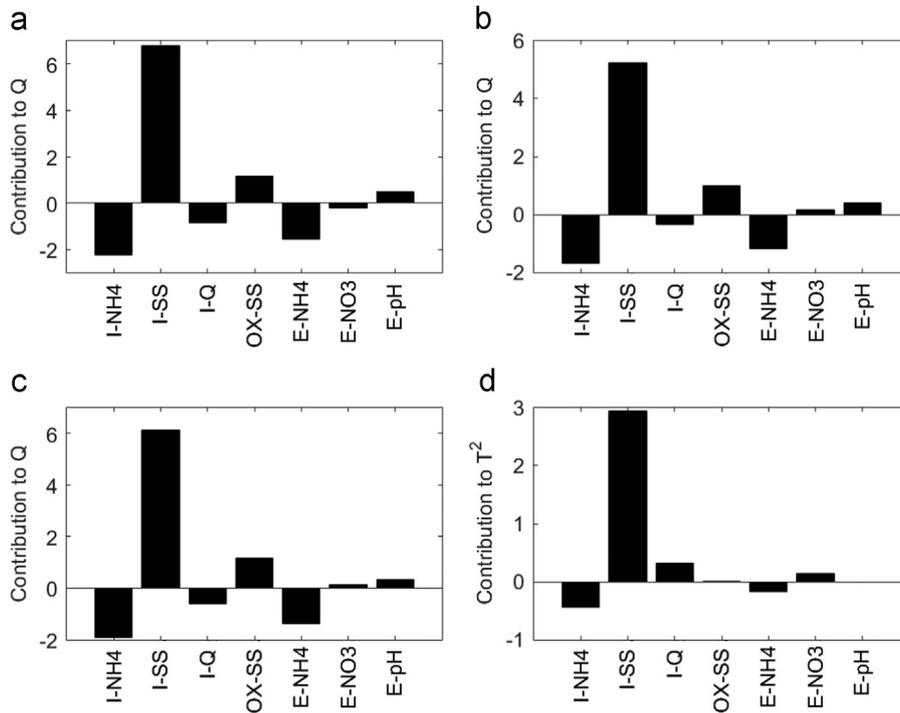


Fig. 15. Variables' contributions to Q using AMWPCA_1 (a) and AMWPCA_2 (b) at 2 am on December 22, using MWPCA (c) at 5 am on December 22 and contributions to T^2 using AMWPCA_1 (d) at 2 am on December 22.

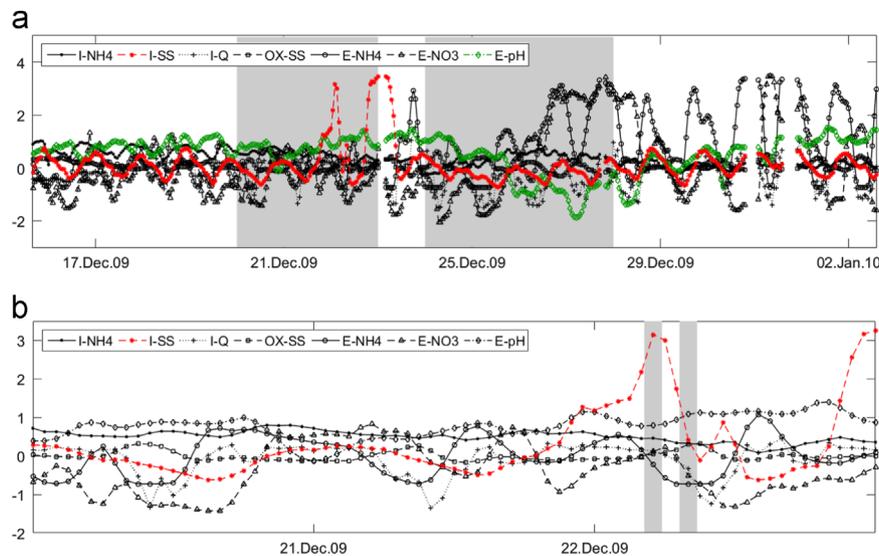


Fig. 16. Time-series of the standardized variables during a 3-week period (a) with a focusing on the first highlighted 3-day episode associating with Case 2 (b).

the instrument range, whereas $E\text{-NO}_3$ rose from the normal level of 6–14 mg/l to over 20 mg/l. The abnormal $E\text{-NH}_4$ and $E\text{-NO}_3$ concentrations also manifest themselves as evident changes in their contributions to the T^2 and Q statistics. Therefore, a monitoring system giving an early warning about abrupt changes among the measured variables would significantly improve the prevention of the corresponding process anomalies.

When the window-length time-series of anomaly monitoring techniques during the 4-day episode are examined (Fig. 20), it is noticed that with AMWPCA_2 the L values are rather close to each other. Again, the AMWPCA_1 reacts more explicitly to the changes among the process variables, manifested as the more intensively varying L values. In fact, L with AMWPCA_1 lowers from 108 h to 56 h when the problems on the nitrogen removal take place and

the abnormal behaviour of $E\text{-pH}$, $E\text{-NH}_4$ and $E\text{-NO}_3$ occurs. The lowered L gives rise to an improved monitoring performance compared with the fixed window-length method in the situation where rapid process changes take place due the pH drop and the resulting nitrifier inhibition.

3.3. Summary of anomaly detection

Investigation of normal and anomalous observations among the testing data shows that AMWPCA_1 and AMWPCA_2 provided slightly differing results, the AMWPCA_1 approach detecting more anomalies (Table 6). This did not only concern the total share of the anomalous samples, but also the shares of both the T^2_{lim} and Q_{lim} violations among them. The AMWPCA_1 and AMWPCA_2 approaches had significantly

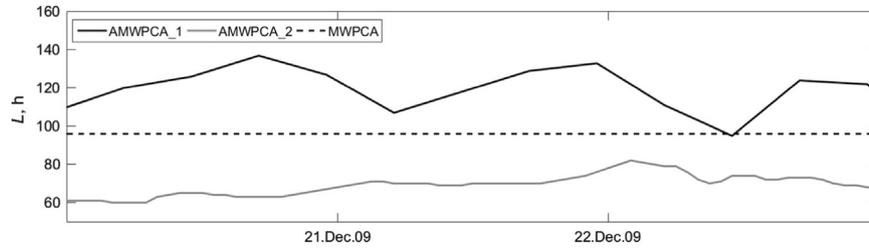


Fig. 17. Time-series of the window-lengths of the investigated anomaly monitoring approaches during a 3-day period.

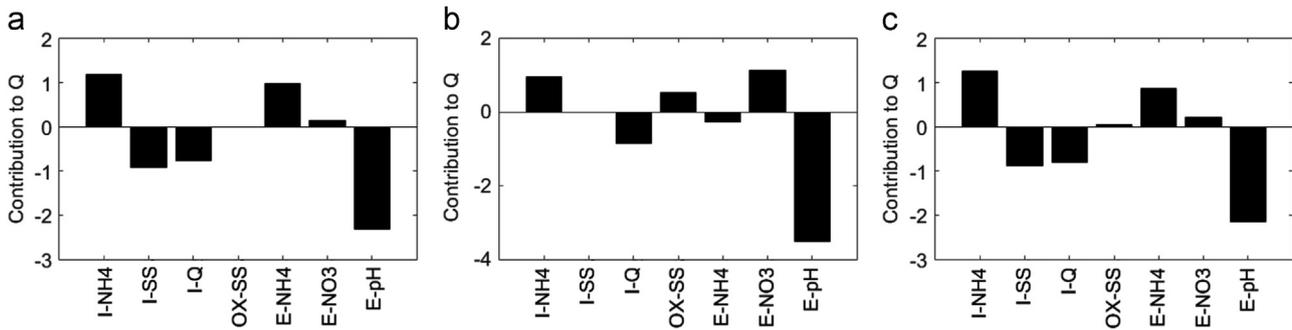


Fig. 18. Variables' contributions to Q using AMWPCA_1 (a) at 5 pm on December 25, and using AMWPCA_2 (b) and MWPCA (c) at 8 pm on December 25.

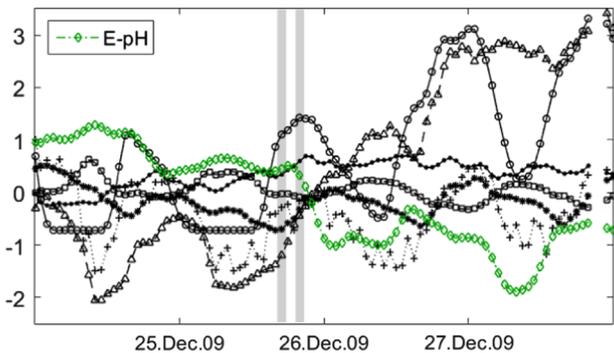


Fig. 19. Time-series of the standardized variables during 4-day episode associating with Case 3. The period corresponds for the latter highlighted area in Fig. 16(a).

dissimilar average window-lengths, 101.6 h and 58.3 h, respectively. A corresponding difference occurred also between their standard deviations, those being 31.3 h for AMWPCA_1 and 13.5 h for AMWPCA_2. This is easily noticed in Fig. 21 where their window-lengths are depicted. The anomaly monitoring performance of MWPCA, defined in terms of the total number of the detected anomalies, corresponded with the AMWPCA_2 approach. It must be emphasized that the shares of the normal and anomalous samples do not describe the correctness of anomaly detection. Because PCA is an unsupervised method, no labels representing the normality of the samples has been used in the model training. Moreover, such labels are not available for evaluating the correct and wrong detections using the investigated techniques. The final decision of the desired anomaly monitoring policy is for the plant management to be made and the detection sensitivities can be further fine-tuned by adjusting the confidence level.

The variables most frequently responsible for anomalies did not differ between the approaches when the largest contributions during T_{lim}^2 and Q_{lim} violations were examined. With each considered anomaly detection technique, *I-NH4* was isolated most often as the fault source, followed by *I-SS*. *I-NH4* is a potential measurement to be used in feedforward control of the ASP because the incoming ammonium load into the bioreactor significantly influences the required aerated

volume. That emphasizes the benefits of installing an anomaly monitoring system to support the efficient process operation under the circumstances where relatively frequent abnormal *I-NH4* signals were found to be present. Considering all the approaches, *I-Q* and *OX-SS* caused the smallest number of anomalies.

As for the model dimension S , two-PC models were the most popular being favoured in 76.2–80.0% of the situations, depending on the monitoring method. The largest model dimension considering each investigated approach was five, compared with the original dimension D of seven. The average subsets of PCs for the studied approaches ranged between 2.25 and 2.30 (Table 6). The correlation between the model dimensions and the window-lengths of the AWMPCA approaches was found to be weak. On average, the models reconstructed 72.3–75.5% of the total variation with different approaches.

The cases when the previous valid model was maintained due to a limited number of samples in the training data ranged between 16.2% and 22.5% for the considered approaches. Apparently, such situations were the most common for the methods with small L values, and therefore for the AMWPCA_2 approach in this study, because the required share of samples P_{lim} is negatively correlated with the window-length (Eq. (8)).

The computational burden of the different approaches was highly dependent on the applied shift-sizes. This was expected since a small Z necessitates frequent recalculation, for instance, of the window-length, of the PCA model and of the statistics' confidence limits. Hence, the shortest computing times were associated with AMWPCA_1. The calculation of the MWPCA ($Z = 4$) and AMWPCA_2 ($Z = 1$) procedures required approximately 83% and 805% more time than of AMWPCA_1 ($Z = 6$), respectively.

The techniques that enable adaptivity in the window-lengths were shown to provide an increased flexibility for the anomaly monitoring. Specifically, they were demonstrated to possess properties to tune the models adequate for applications that concern time-varying operational conditions. Using different criteria for selecting the parameters or altering their potential ranges would have provided a different anomaly detection capability, for instance, if a less strict detection policy was desired. In particular, the AMWPCA_1 technique was shown to be widely adjustable whereas

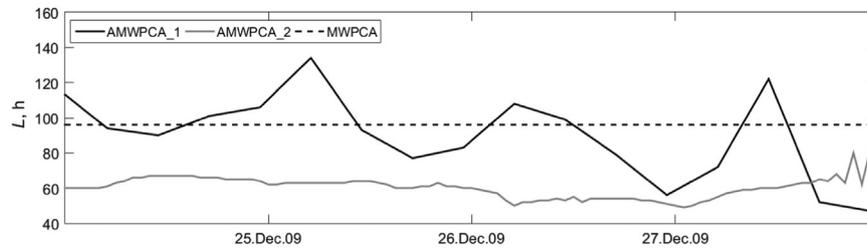


Fig. 20. Time-series of the window-lengths of the investigated anomaly monitoring approaches during a 4-day period.

Table 6

Shares of normal and anomalous samples, average number of retained PCs, average window-lengths and standard deviation (std) of the window-lengths for the examined methods.

	AMWPCA_1	AMWPCA_2	MWPCA
Normal	0.767	0.789	0.789
Anomalous	0.233	0.211	0.221
PCs	2.30	2.25	2.25
Average L (h)	101.6	58.3	96.0
Std L (h)	31.3	13.5	0.0

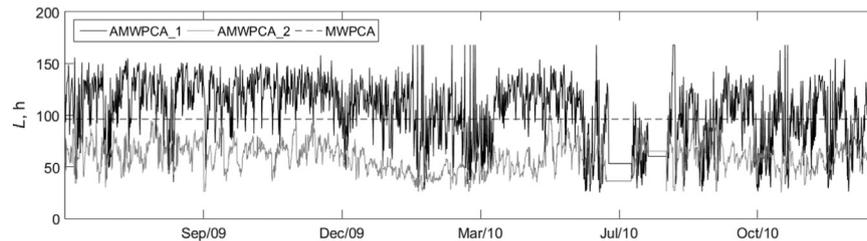


Fig. 21. Time-series of the window-lengths of the investigated anomaly monitoring approaches during the testing period.

the tuning capacity of AMWPCA_2 was indicated to be more limited. In Section 3.2, it was also indicated that the window-length L using AMWPCA_1 was more capable of responding to the process changes, which was the primary motivation of applying adaptive window-lengths instead of the conventional fixed ones. Therefore, the use of AMWPCA_1 for overcoming the challenges created by the varying process dynamics in WWTPs is more practicable than the use of AMWPCA_2. In the provided examples, it was also proved that AMWPCA_1 detected the anomalies more timely than the other methods, including MWPCA.

The advantage of the fixed window-length procedure is that it is simpler to put into operation than the AMWPCA techniques. MWPCA does not require effort and competence for tuning the function parameters of the window-length definition equations to suit for the considered application. However, in the MWPCA method the selection of window-length and shift-size affect considerably the anomaly detection sensitivity, as it was indicated in Section 3.1. Therefore, their selection needs to be paid significantly attention to. The window-length in the MWPCA technique had an impact especially on the detection of the changes in the relations between the variables, or in the covariance structure, which connects the Q statistic. Particularly, the models with short windows exceeded Q_{lim} more often. Moreover, it is obvious that MWPCA suffers from the fixed historical window-lengths in comparison with an adequately tuned AMWPCA approach in such applications where the rapidness of process changes fluctuates.

The results of this work showed that different types of anomalies taking place in WWTPs can be isolated with the tested methods, the AMWPCA_1 providing the most timely detection capability. Typically, the measurement drifts were more demanding for isolation than, for instance, the individual outlying measurement peaks. The studied anomaly monitoring techniques were shown to provide the operators

with the early warnings of process disturbances that are challenging to detect by observing simultaneously several univariate control charts. The parameter selection is a crucial step for all the investigated monitoring approaches and it requires moderate efforts when a high dimension of parameters to be set is involved. In addition, the definition of the shift-size Z was shown to be of significant importance, the overly small values being linked with the adaptation of the models to measurement drifts and with the large computational costs.

4. Conclusions

In this work, we investigated an anomaly detection system in a large-scale municipal WWTP. The methodologies employed were based on moving-window PCA extensions with adaptive and fixed window-lengths. The experimental results showed when monitoring systems with the adequate sets of parameters were defined, drifts and peaks in measurements as well as process anomalies can be detected. Also, the correct isolation of the variables causing the anomalies was demonstrated. The results indicated that other of the examined adaptive window-length approaches successfully modified the window-lengths according to the changes taking place among the relationship of the considered process variables. For the techniques with adapting window-lengths, the tuning of the parameters of the window-length definition equations and of the shift-sizes specifying the model recalculation intervals proved to be the critical factors for the anomaly monitoring performances. In practice, the proposed techniques could be installed as an inexpensive software tool for monitoring sensor and process abnormalities. This would also increase the potential of sensors to be used in advanced control systems, because the risk of using, for instance, faulty influent measurements in

schemes that include feedforward control would be diminished due to the automatic alarms and to the isolation of deviating instruments. The presented algorithm could easily be extended to include more sensors and process units as well as be adapted to other industries, where sufficient on-line instrumentation is available and where the dynamics of the process changes varies.

Acknowledgement

Henri Haimi gratefully acknowledges Maa- ja vesitekniiikan tuki ry. for their financial support. Maa- ja vesitekniiikan tuki ry. is a Finnish non-profit association that supports the research and education of water engineering and of the related environmental engineering and soil conservation.

References

- Aguado, D., Montoya, T., Borrás, L., Seco, A., Ferrer, J., 2008. Using SOM and PCA for analysing and interpreting data from a P-removal SBR. *Eng. Appl. Artif. Intell.* 21, 919–930.
- Aguado, D., Rosen, C., 2008. Multivariate statistical monitoring of continuous wastewater treatment plants. *Eng. Appl. Artif. Intell.* 21, 1080–1091.
- Alferes, J., Tik, S., Copp, J., Vanrolleghem, P.A., 2013. Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection. *Water Sci. Technol.* 68, 1022–1030.
- Atkinson, A.C., Riani, M., Cerioli, A., 2004. Exploring multivariate data with the forward search. *Springer Series in Statistics*. Springer, New York.
- Ayech, N., Chakour, C., Harkat, M.F., 2012. New adaptive moving window PCA for process monitoring fault detection. In: *Proceedings of the 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*. Mexico City, Mexico. pp. 606–611.
- Baggiani, F., Marsili-Libelli, S., 2009. Real-time fault detection and isolation in biological wastewater treatment plants. *Water Sci. Technol.* 60, 2949–2961.
- Bro, R., Smilde, A.K., 2014. Principal component analysis. *Anal. Methods* 6, 2812–2831.
- Campisano, A., Ple, J.C., Muschalla, D., Pleau, M., Vanrolleghem, P.A., 2013. Potential and limitations of modern equipment for real time control of urban wastewater systems. *Urban Water J.* 10, 300–311.
- Choi, S.W., Martin, E.B., Morris, A.J., Lee, I.B., 2006. Adaptive multivariate statistical process control for monitoring time-varying processes. *Ind. Eng. Chem. Res.* 45, 3108–3118.
- 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 Control* 23, 158–170.
- Davis, C., Kahan, W.M., 1970. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* 7, 1–46.
- Ge, Z., Song, Z., Gao, F., 2013. Review of recent research on data-based process monitoring. *Ind. Eng. Chem. Res.* 52, 3543–3562.
- Haimi, H., Mulas, M., Corona, F., Vahala, R., 2013. Data-derived soft-sensors for biological wastewater treatment plants: an overview. *Environ. Modell. Softw.* 47, 88–107.
- He, X.B., Yang, Y.P., 2008. Variable MWPCA for adaptive process monitoring. *Ind. Eng. Chem. Res.* 47, 419–427.
- Henze, M., van Loosdrecht, M.C.M., Ekama, G.A., Brdjanovic, D., 2008. *Biological Wastewater Treatment—Principles, Modelling and Design*. IWA Publishing, London.
- Jackson, J.E., Mudholkar, G.S., 1979. Control procedures for residual associated with principal component analysis. *Technometrics* 21, 341–349.
- Jeppsson, U., Alex, J., Batstone, D.J., Benedetti, L., Comas, J., Copp, J.B., Corominas, L., Flores-Alsina, X., Gernaey, K.V., Nopens, I., Pons, M.N., Rodríguez-Roda, I., Rosen, C., Steyer, J.P., Vanrolleghem, P.A., Volcke, E.I.P., Vrecko, D., 2013. Benchmark simulation models, quo vadis? *Water Sci. Technol.* 68, 1–15.
- Jolliffe, I.T., 2002. *Principal Component Analysis*, 2nd ed Springer, New York.
- Kadlec, P., Gabrys, B., Grbic, R., 2011. Review of adaptation mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* 35, 1–24.
- Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.* 33, 795–814.
- Katko, T.S., 2000. Long-term development of water and sewage services in Finland. *Publ. Works Manag. Policy* 4, 305–318.
- Ku, W., Storer, R.H., Georgakis, C., 1995. Disturbance detection and isolation by dynamic principal component analysis. *Chemometr. Intell. Lab. Syst.* 30, 179–196.
- Le Bonté, S., Potier, O., Pons, M.N., 2005. Toxic event detection by respirometry and adaptive principal components analysis. *Environmetrics* 16, 589–601.
- Lee, C., Choi, S.W., Lee, I.B., 2004. Sensor fault identification based on time-lagged PCA in dynamic processes. *Chemometr. Intell. Lab. Syst.* 70, 165–178.
- Lee, C., Choi, S.W., Lee, I.B., 2006. Sensor fault diagnosis in a wastewater treatment process. *Water Sci. Technol.* 53, 251–257.
- Lennox, J., 2002. *Multivariate Subspaces for Fault Detection and Isolation: With Application to the Wastewater Treatment Process* (Ph.D. thesis). University of Queensland, Brisbane, Australia.
- Lieftucht, D., Kruger, U., Irwin, G.W., 2006. Improved reliability in diagnosing faults using multivariate statistics. *Comput. Chem. Eng.* 30, 901–912.
- MacGregor, J.F., Jaeckle, C., Kiparissides, C., Koutoudi, M., 1994. Process monitoring and diagnosis by multiblock PLS methods. *AIChE J.* 40, 826–838.
- Martin, C., Vanrolleghem, P.A., 2014. Analysing, completing, and generating influent data for WWTP modelling: a critical review. *Environ. Modell. Softw.* 60, 188–201.
- Mulas, M., Tronci, S., Corona, F., Haimi, H., Lindell, P., Heinonen, M., Vahala, R., Baratti, R., 2015. Predictive control of an activated sludge process: an application to the Viikinmäki wastewater treatment plant. *J. Process Control* 35, 89–100.
- Ng, Y.S., Srinivasan, R., 2010. Multi-agent based collaborative fault detection and identification in chemical processes. *Eng. Appl. Artif. Intell.* 23, 934–949.
- Nomikos, P., MacGregor, J.F., 1995. Multivariate SPC charts for monitoring batch processes. *Technometrics* 37, 41–59.
- Olsson, G., 2014. ICA and me—a subjective review. *Water Res.* 46, 1585–1624.
- Olsson, G., Nielsen, M., Yuan, Z., Lynggaard-Jensen, A., Steyer, J.P., 2005. *Instrumentation, Control and Automation in Wastewater Systems*. Scientific and Technical Report No. 15. International Water Association, London.
- Qin, S.J., 2011. Survey on data-driven industrial process monitoring and diagnosis. *Annu. Rev. Control* 36, 220–234.
- Rosen, C., 2001. *A Chemometric Approach to Process Monitoring and Control: With Applications to Wastewater Treatment Operation* (Ph.D. thesis). Lund University, Lund, Sweden.
- 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.
- Shen, H.W., Cheng, X.Q., Wang, Y.Z., Chen, Y., 2012. A dimensionality reduction framework for detection of multiscale structure in heterogeneous networks. *J. Comput. Sci. Technol.* 27, 341–357.
- Tamura, M., Tsujita, S., 2007. A study on the number of principal components and sensitivity of fault detection using PCA. *Comput. Chem. Eng.* 31, 1035–1046.
- Tao, E.P., Shen, W.H., Liu, T.L., Chen, X.Q., 2013. Fault diagnosis based on PCA for sensors of laboratorial wastewater treatment process. *Chemometr. Intell. Lab. Syst.* 128, 49–55.
- Teppola, P., Mujunen, S.P., Minkkinen, P., 1999. Adaptive fuzzy C-means clustering in process monitoring. *Chemometr. Intell. Lab. Syst.* 45, 23–38.
- Thomsen, H.R., Önnérth, T.B., 2009. Results and benefits from practical application of ICA on more than 50 wastewater systems over a period of 15 years. In: *Proceedings of the 10th IWA Conference on Instrumentation, Control and Automation*, Cairns, Australia (Keynote paper).
- Valle, S., Li, W., Qin, S.J., 1999. Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods. *Ind. Eng. Chem. Res.* 38, 4389–4401.
- 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.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., Yin, K., 2003. A review of process fault detection and diagnosis. Part III: process history based methods. *Comput. Chem. Eng.* 27, 327–346.