Engineering Applications of Artificial Intelligence 52 (2016) 65–80

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.

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).

http://dx.doi.org/10.1016/j.engappai.2016.02.003
0952-1976/

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 m^3/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

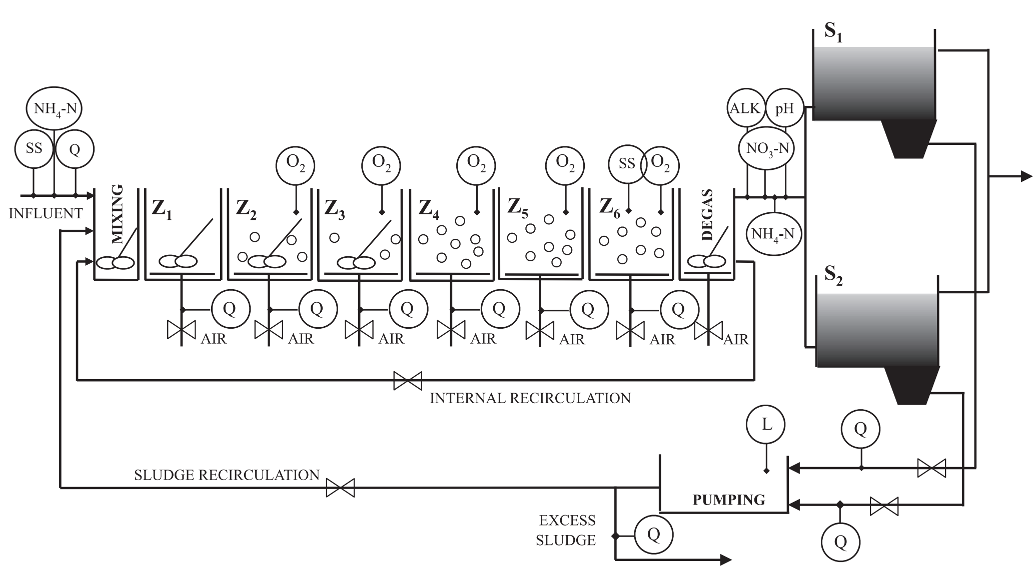

Fig. 1. Simplified layout for a single ASP line and location of on-line measurements.