Environmental Modelling & Software 47 (2013) 88–107

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

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

Review

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

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

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

A R T I C L E I N F O

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

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

A B S T R A C T

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

1. Introduction

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

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

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

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

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

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

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

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

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

2. Wastewater treatment plants

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