Expert Systems With Applications 63 (2016) 8–19

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network

J. Fernandez de Canete[a,*], P. Del Saz-Orozco[b,1], R. Baratti[c,2], M. Mulas[d,e], A. Ruano[f,3], A. Garcia-Cerezo[g,4]

[a] System Engineering Department, University of Malaga, Campus de Teatinos s/n, 29071, Malaga, Spain
[b] System Engineering Department, University of Malaga, Campus de Teatinos s/n, 29071, Malaga, Spain
[c] Dept. of Mechanical, Chemical and Materials Engineering, University of Cagliari, Via Marengo, 2, I-09123 Cagliari, Italy
[d] Dept. of Civil and Environmental Engineering, Aalto University, P.O. Box 12100, FI-00076 Aalto, Finland
[e] Department of Chemical Engineering, Federal University of Campina Grande, 58429-140 Campina Grande, Brazil
[f] Centre for Intelligent Systems, Faculty of Science and Technology, University of Algarve, Campus de Gambelas, 8005-139 Faro, Algarve, Portugal
[g] System Engineering Department, University of Malaga, Campus de Teatinos s/n, 29071, Malaga, Spain

ARTICLE INFO

Article history:
Received 26 February 2015
Revised 13 June 2016
Accepted 14 June 2016
Available online 20 June 2016

Keywords:
Neural networks
Optimal soft sensing
PCA selection
Activated sludge process
Effluent concentrations
GPS-X simulator

ABSTRACT

Recent studies into the estimation and control of an activated sludge process (ASP) at a wastewater treatment plant suggest that artificial-intelligence methods, such as neural networks, fuzzy systems and genetic algorithms, can improve the plant performance in terms of reduced operation costs and improved effluent quality. In this paper, a neural-network-based soft sensor is developed for the on-line prediction of effluent concentrations in an ASP in terms of primary hard-to-measure variables, such as chemical oxygen demand, total nitrogen content and total suspended solids, starting from secondary on-line easy-to-measure variables, such as oxygen and nitrogen compound concentrations in biological tanks, input flow rate and alkalinity, among others. An algorithm based on principal component analysis is applied to select the optimal net input vectors for the soft sensor, using an appropriated number of samples of the secondary variables set. The proposed soft sensor is tested on the ASP of a large-scale municipal wastewater treatment plant running under the GPS-X simulation frame and validated with operational gathered data. Satisfactory low values for mean and maximum absolute prediction errors are obtained, even when high values of sampling time of primary variables are set, as it is frequently done during monitoring operation. In this way, data-driven soft-sensors based on neural networks can become valuable tools for plant operators for the recognition of operational states in terms of low cost and efficient prediction of primary process variables such as chemical oxygen demand, total nitrogen content and total suspended solids, therefore avoiding the acquisition of expensive and sometimes unreliable instruments for measuring nutrient concentrations in plant.

1. Introduction

Biological treatment processes are usually employed for the removal of organic pollutants from wastewaters, with the activated sludge process (ASP) being the most commonly used technology for this purpose. Depending on the specific application, an ASP can achieve the removal of both biological nitrogen and organic carbon substances. From a system dynamics point of view, an ASP can be considered as a complex system, owing to its non-linear dynamics, large uncertainty of uncontrolled inputs, model parameters and structures as well as multiple-input multiple-output structure in general.

The availability of effluent quality indicators or primary variables, such as chemical oxygen demand (COD), total nitrogen content (TN) and total suspended solids (TSS), is essential in order to evaluate the plant performances and design advanced control strategies. These variables are usually hard to measure on-line due to the high cost and the time delay associated to real-time monitoring. In fact, some of these are generally determined off-line

* Corresponding author. Fax: +34 951952540.
E-mail addresses: canete@isa.uma.es (J. Fernandez de Canete), pabsazhua@uma.es (P. Del Saz-Orozco), roberto.baratti@dimcm.unica.it (R. Baratti), michela.mulas@aalto.fi (M. Mulas), aruano@ualg.pt (A. Ruano), alfonso.garcia@isa.uma.es (A. Garcia-Cerezo).
[1] Fax: +34 951952540.
[2] Fax: +39 0706755067.
[3] Fax: +351 289 800066.
[4] Fax: +34 951952540.

http://dx.doi.org/10.1016/j.eswa.2016.06.028
0957-4174/

in laboratory under high uncertainty, making an operational limitation (Chapman, 1992). The use of software sensors enables the estimation of such off-line primary variable throughout the use of easy on-line measured secondary variables, and has been applied to wastewater treatment plants with success (Haimi, Mulas, Corona, & Vahala, 2013). Based the modelling approach, the models used for describing the biological processes in the wastewater treatment can be mainly classified as phenomenological and or data-driven models.

The most commonly used phenomenological models belong to the activated sludge model (ASM) family (Henze, Gujer, Mino, & Van Loosdrecht, 2000) proposed by the IWA Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment. Different simulation platforms, either commercial (e.g., WEST, SIMBA, BIOWIN and GPS-X) or general purposes software (e.g., Matlab and Simulink), employ these models as main skeleton (Gernaey, Van Loosdrecht, Henze, Lind, & Jørgensen, 2004). The simulation protocol developed by the IWA Task Group on Good Modelling Practice (Rieger et al., 2012) gives guidelines for achieving accurate modelling results. The phenomenological approach has definitely proved to be valuable, for example, in wastewater treatment process design, revamp, employee training, optimisation of the plant operation and understanding the system's behaviour (Phillips et al., 2009, Olsson, 2012). It has also been used successfully for the on-line estimation of the nitrogen removal process by Spérandio and Queinnec (2004) and the suspended solids in a membrane bioreactor through an extended Kalman filter (Beltrán, Irizar, Monclús, Rodríguez-Roda, and Ayesa (2009), for the automatic influent characterisation of a wastewater treatment plant by Grau, Beltrán, De Gracia, and Ayesa (2007) and recently for the state estimation problem for a large scale wastewater treatment plant by Busch et al., (2013). However, real-time applications of the phenomenological approach still shows major challenges related the successful model calibration (Petersen, Gernaey, Henze, & Vanrolleghem, 2003; Sin, 2004; Hauduc et al., 2009, Nasr, 2012), the computational requirements and the interaction between fast and slow dynamics that result in and ill-conditioned problems (Dochain & Vanrolleghem, 2001).

The data-driven modelling approach involves the use of artificial neural networks (ANN), fuzzy logic and genetic algorithms (GA) and their associated hybrids schemes (Jagielska, Matthews, & Whitfort, 1999). Artificial neural networks can be considered from an engineering point of view, as a nonlinear heuristic model useful to make predictions and data classifications starting from I/O data performance (Haykin 2008), and have been used both for modelling and control of nonlinear system in process engineering (Xiong & Jutan 2002) and also as soft sensors for process control (Fortuna, Graziani, Rizzo, & Xibilia, 2007).

Artificial neural networks have been employed successfully for online prediction and process monitoring in a number of wastewater treatment plant applications. An example is given by Mjalli, Al-Asheh, and Alfadala (2007), where the performance of a wastewater treatment plant based on the available historical data was predicted. Pai et al., 2008 used grey model and artificial neural network to predict effluent suspended solids and chemical oxygen demand. Dogan, Ates, Yilmaz, and Eren (2008) demonstrate the capabilities of the ANN model for BOD modelling. Hybrid schemes were proposed by Choi and Park (2001) and Civelekoglu, Perendec, Yigit, and Kitis (2007) for estimating primary variables starting from secondary measurements of a biological WWTP. Some others soft sensors applications for primary variables have been referred to using neural networks (Hamed, Khalafallah, & Hassanien, 2004) multivariate statistical processes (Woo et al., 2009) or both (Lee et al., 2008) all of them based on expensive on-line information comprising, for instance, influent measurements of COD, TN and TSS.

Recently, Haimi et al., (2013) presented a comprehensive review on the use of soft sensors in biological wastewater treatment plant and pointed out that supervised and unsupervised ANN are typically very popular in this field of applications. The trend is confirmed also by the more recently contributions. Such as for instance the k-means clustering and multi-layer perceptron methods used for modelling chemical oxygen demand concentration by Ay and Kisi (2014). Han, Qian, and Qiao, (2014) used a self-organizing radial basis function neural network prediction in a nonlinear predictive control scheme. Neural nonlinear adaptive control design techniques were presented by Lin and Luo (2105) to solve the dissolved concentration control problem for an uncertain wastewater treatment process.

In this paper, neural networks are used for the on-line prediction of hard-to-measure primary chemical oxygen demand, total nitrogen content and total suspended solids at the effluent stream starting from exclusively on-line easy-to-measure secondary variables from an ASP simulation model running under GPS-X framework, in which a PCA-based selection procedure is used to determine the significant secondary variables and its delayed values. A validation process has been performed in order to adjust the neural network parameters of the estimator, starting from the off-line values of the primary effluent variables. The proposed soft sensor was tested on the ASP of a large-scale municipal wastewater treatment plant running under the GPS-X simulation frame. Satisfactory estimation results are obtained, even when high values of off-line sampling time are used during the ASP monitoring process, whereas no nutrient concentrations were used as input information for the soft-sensing task.

This paper is organised as follows. In Section 2, the ASP together with the ASP simulation model under GPS-X are described. The neural-network estimation process is detailed in Section 3, together with the variable selection procedure used to build the structure of the sensor. The performance of the neural soft estimation of primary variables that was applied to the referred ASP is shown in Section 4 while Section 5 presents some discussion regarding the methodology used and the results obtained. A summary is also included in Section 6 in order to explain the motivation that was followed during the proposed approach, along with future work that is yet to be carried out.

2. Wastewater process description

In this work, we considered an ASP operated at the TecnoCasic wastewater treatment plant (WWTP) that was reported previously (Mulas, 2006). The wastewater treatment includes a mechanical treatment to remove floating and settleable solids, a biological treatment with activated sludge for the removal of nitrogen and organic pollutants as well as other operations such as sludge treatment and water chemical treatment.

The bioreactor consisted of a pre-denitrification anoxic zone followed by a nitrification aerobic zone, with both oxygen and nitrogen sensors located in the anoxic and aerobic zones to provide information about the aeration process. To maintain the microbiological population, the sludge from the settler was recirculated into the anoxic basin while the sludge concentration was kept constant by means of sludge withdrawn from the settler (Mulas, 2006).

The ASP layout was composed of a biological treatment reactor of 2000 m^3, in which the aeration was obtained with fine-pore air diffusers located at the bottom of the bioreactor. Then, it follows a secondary settler with a surface area of 707 m^2 and 4 m depth, from which the output flow is partly recycled from the secondary settler to the front end of the plant and the remaining waste flow is pumped out of the secondary settler. The average nominal influent conditions for dry, rainy and stormy weather for the considered plant are reported in Table 1.