JOURNAL OF ENVIRONMENTAL SCIENCES 32 (2015) 90–101

Available online at www.sciencedirect.com

ScienceDirect

www.journals.elsevier.com/journal-of-environmental-sciences

JOURNAL OF ENVIRONMENTAL SCIENCES www.jesc.ac.cn

Prediction of effluent concentration in a wastewater treatment plant using machine learning models

Hong Guo[1], Kwanho Jeong[1], Jiyeon Lim[2], Jeongwon Jo[2], Young Mo Kim[1], Jong-pyo Park[3], Joon Ha Kim[1], Kyung Hwa Cho[2,*]

1. School of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), 261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Republic of Korea
2. School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Republic of Korea
3. HECOREA. INC, 405, Woori Venture Town II, 70, Seonyu-ro, Yeongdeungpo-gu, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 30 June 2014
Revised 11 December 2014
Accepted 22 January 2015
Available online 20 April 2015

Keywords:
Artificial neural network
Support vector machine
Effluent concentration
Prediction accuracy
Sensitivity analysis

ABSTRACT

Of growing amount of food waste, the integrated food waste and waste water treatment was regarded as one of the efficient modeling method. However, the load of food waste to the conventional waste treatment process might lead to the high concentration of total nitrogen (T-N) impact on the effluent water quality. The objective of this study is to establish two machine learning models—artificial neural networks (ANNs) and support vector machines (SVMs), in order to predict 1-day interval T-N concentration of effluent from a wastewater treatment plant in Ulsan, Korea. Daily water quality data and meteorological data were used and the performance of both models was evaluated in terms of the coefficient of determination (R^2), Nash–Sutcliff efficiency (NSE), relative efficiency criteria (d_{rel}). Additionally, Latin-Hypercube one-factor-at-a-time (LH-OAT) and a pattern search algorithm were applied to sensitivity analysis and model parameter optimization, respectively. Results showed that both models could be effectively applied to the 1-day interval prediction of T-N concentration of effluent. SVM model showed a higher prediction accuracy in the training stage and similar result in the validation stage. However, the sensitivity analysis demonstrated that the ANN model was a superior model for 1-day interval T-N concentration prediction in terms of the cause-and-effect relationship between T-N concentration and modeling input values to integrated food waste and waste water treatment. This study suggested the efficient and robust nonlinear time-series modeling method for an early prediction of the water quality of integrated food waste and waste water treatment process.

Introduction

Following the restrictive landfill legislation passed by the European Union (EU) in 1999, many developed countries have implemented various policies and technical developments for reducing the quantity of biodegradable waste landfill (Burnley et al., 2011; García et al., 2005). The South Korean government also prohibited the landfill of municipal solid sludge (MSS) and food waste (FW) in the early 21st century (S. Cheon et al., 2013). However, this strict regulation causes the dumping of both the sludge and FW water (*i.e.*, leachate) at sea, consequently leading to the prohibition of its disposal in the ocean

* *Corresponding author.*E-mail: khcho@unist.ac.kr (Kyung Hwa Cho).

http://dx.doi.org/10.1016/j.jes.2015.01.007
1001-0742/

by the Marine Environment Management Act (Behera et al., 2010; S. Cheon et al., 2013). At this time, basic environmental treatment facilities such as wastewater treatment plants (WWTPs) appear to be an alternative inland treatment to resolve the problem. In the inland treatment with wastewater treatment plant, over 80% of FW, which is recyclable organic waste of municipal solid wastes (MSW), is dehydrated, and the remaining waste goes through recycling processes such as composting, feed, and anaerobic digestion to generate biomass energy (Chelliapan et al., 2012; Li et al., 2012). In particular, methane gas, one of the biogases, can be utilized as a biomass energy source (Lee et al., 2009). However, a large amount of FW leachate inevitably occurs in all recycling processes because of a high moisture content from the FW leachate (Han et al., 2012), resulting in a significant burden on the wastewater treatment systems.

According to several research reports (Kim et al., 2008; Sosnowski et al., 2003), the water treatment process can be more effective by using FW leachate in WWTP. This is because the FW leachate contains a large amount of acid fermentation liquid (AFL) which can be utilized as an organic carbon source for removing nitrogen and phosphorus in advanced waste-water treatment (AWT) processes (Han et al., 2012; Lee et al., 2003). The digestion process with only sewage sludge could be less effective due to the low carbon/nitrogen (C/N) ratio and low level of biodegradable organic compounds. FW leachate contains a high amount of solid contents as well as a high C/N ratio, while containing a low amount of the nutrient-type elements (Mata-Alvarez, 2003). Therefore, the combined treatment of sewage sludge and FW improves the removal efficiency of nitrogen and phosphorus in AWT, enhancing the stability of the digestion process. Furthermore, higher production of methane gas is an additional benefit from the co-digestion with FW leachate (Cecchi et al., 1988; Hamzawi et al., 1998; Mata-Alvarez et al., 1990; Poggi-Varaldo and Oleszkiewicz, 1992; Schmit and Ellis, 2001). Owing to these advantages, the anaerobic digestion process of sewage sludge with FW has been increased in WWTPs in Korea. However, this process also faces critical issues which are associated with the side effects of co-digestion. One of the issues is that the influent water quality is degraded by mixing with the returned FW leachate from the anaerobic co-digestion process, so it tends to increase liquor suspended solids (MLSS) and causes a large amount of scum in the activated sludge reactor (Kim and Shin, 2009; Mahmoud et al., 2003). As well, a sudden increase of the FW leachate could cause an unstable digestion process and lower the level of effluent water quality from WWTPs (S. Cheon et al., 2013).

Generally, water quality of a WWTP is sensitive to parameters such as pH, temperature, concentrations of specific substrates, and contaminants. This is because wastewater is treated by the metabolism processes of microorganisms. However, biological treatment still exhibits time-varying and highly nonlinear characteristics affected by various known and unknown parameters (Hamed et al., 2004; Hong et al., 2003; Mjalli et al., 2007). Due to these complicated features, many previous studies evaluated and diagnosed the performance of WWTP by using a mathematical model for the process simulation and control (Gernaey et al., 2004; Hamed et al., 2004; Hong et al., 2003; Iacopozzi et al., 2007; Mjalli et al., 2007; Rivas et al., 2008; Wintgens et al., 2003). Thereinto, a machine learning model has proved to be a useful tool because it has a relatively high accuracy for dealing with complicated systems. Furthermore, a key advantage of these models to the evaluation of WWTP performance is that these can directly predict output values from input values only after training and validation step. Artificial neural networks (ANNs) and support vector machines (SVMs) are representative machine-learning techniques (Dreyfus, 2005; Shon and Moon, 2007). Two machine learning models' performance studies have been widely discussed before (Hamed et al., 2004; Palani et al., 2008; Singh et al., 2009; Yoon et al., 2011). However, only black box modeling has the limitation on the process control and there has yet to elucidate the cause-and-effect relationship for input and output value for process control.

In this study, two machine learning models would be developed for predicting effluent T-N concentration for the integrated food waste and waste water treatment plant in Ulsan Metropolitan city, Korea. Moreover, by sensitivity analysis between input values and output values, the cause-and-effect relationship would be elucidated for the future process control and selection of the prior machine learning model for integrated food waste and waste water treatment. The objective of this study is: a) development of reliable 1-day interval early T-N concentration prediction model by parameter optimization method; b) evaluation of the building model by sensitivity analysis to find the cause–effect based reasonable model as future decision-making tool; c) to propose an early warning prediction tool to avoid the impact of FW leachate loading to the integrated food waste and waste water treatment.

1. Method and materials

1.1. Field sampling

We collected water samples in an attempt to investigate the effect of FW leachate on Yong-yeon (YY) WWTP in Ulsan. The samples were collected from 6 different spots, including influent, flow-distribution tank, aeration tank, effluent, FW leachate, and pre-treated FW leachate (Fig. 1). The collected samples were delivered to a laboratory at the Ulsan National Institute of Science and Technology (UNIST) and were analyzed in terms of total suspended solids (TSS), chemical oxygen demand (COD), total nitrogen (T-N), and total phosphorus (T-P); water temperature and pH were measured *in-situ* at the sampling stations.

1.2. Sample analysis

TSS of a water sample was measured by filtering a 20 mL sample through pre-weighed 47 mm Glass-Fiber paper (with 1.2 μm pore size), then weighing the filter again after drying to remove all water in the sample. COD, T-P, and T-N were measured through absorptiometric analysis. COD and T-P were measured for 4 sampling locations: influent, flow-distribution tank, aeration tank, and effluent. T-N was measured for 6 sampling locations including 2 additional stations (*i.e.*, pre- and post-aerobic transamination of FW leachate). The absorbance of samples, which were mixed with the proper reagents was quantified under the 200-900 nm wavelength and the target