Water Research 128 (2018) 304–313

Contents lists available at ScienceDirect

Water Research

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

Predicting influent biochemical oxygen demand: Balancing energy demand and risk management

Jun-Jie Zhu [a,*], Lulu Kang [b], Paul R. Anderson [a]

[a] Department of Civil, Architectural and Environmental Engineering, Illinois Institute of Technology, Chicago, IL 60616-3793, USA
[b] Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616-3793, USA

ARTICLE INFO

Article history:
Received 6 August 2017
Received in revised form
18 October 2017
Accepted 24 October 2017

Keywords:
Biochemical oxygen demand (BOD)
Decision making
Energy saving
Risk management
Soft sensor
Wastewater

ABSTRACT

Ready access to comprehensive influent information can help water reclamation plant (WRP) operators implement better real-time process controls, provide operational reliability and reduce energy consumption. The five-day biochemical oxygen demand (BOD_5), a critical parameter for WRP process control, is expensive and difficult to measure using hard-sensors. An alternative approach based on a soft-sensor methodology shows promise, but can be problematic when used to predict high BOD_5 values. Underestimating high BOD_5 concentrations for process control could result in an insufficient amount of aeration, increasing the risk of an effluent violation. To address this issue, we tested a hierarchical hybrid soft-sensor approach involving multiple linear regression, artificial neural networks (ANN), and compromise programming. While this hybrid approach results in a slight decrease in overall prediction accuracy relative to the approach based on ANN only, the underestimation percentage is substantially lower (37% vs. 61%) for predictions of carbonaceous BOD_5 ($CBOD_5$) concentrations higher than the long-term average value. The hybrid approach is also flexible and can be adjusted depending on the relative importance between energy savings and managing the risk of an effluent violation.

1. Introduction

Soft sensors, mathematical/statistical approaches based on easily-acquirable information and historical data, offer an effective way to predict real-time and future information, which can result in more cost-effective controls at a water reclamation plant (WRP). Although our previous assessment (Zhu and Anderson, 2016) demonstrated multiple benefits of soft-sensors including reliable predictions and flexibility, in that work we focused on predicting average values of the target variables. With that kind of focus, predictions of values that deviate from the average can be poor, which is a concern because failure to predict high influent BOD_5 concentration can increase the risk of violating an effluent permit. One way to eliminate this risk is to provide excess aeration, but that approach leads to excess energy consumption. To address these issues, we developed a hybrid approach that integrates a soft-sensor methodology with a multi-criteria decision analysis (MCDA) technique. The overall objective was to balance energy demand and risk management when predicting influent carbonaceous BOD_5 ($CBOD_5$) at the Metropolitan Water Reclamation District of Greater Chicago (MWRDGC) Calumet WRP. We combined an artificial neural network (ANN) for accurate predictions of low $CBOD_5$ concentrations to minimize excess energy demand, with the upper confidence limit (UCL) at 95% confidence level of multiple linear regression (MLR) to overestimate high $CBOD_5$ concentrations, to minimize the risk of insufficient treatment. This paper begins with a description of the advantages of ANN and MLR (95% UCL) in dealing with these kinds of predictions. Subsequently, we present how compromise programming (CP), one of many MCDA techniques, can help to determine a critical value that splits the $CBOD_5$ data into two groups: low and high concentrations. The paper concludes with an evaluation of this hybrid model, and discussion of how this hybrid approach can be applied to different local conditions, such as permit requirements and energy costs.

1.1. Prediction of influent conditions at the Calumet WRP

The MWRDGC Calumet plant, one of the oldest plants in the Chicago area, began operations in 1922. Based on ten years of recent (2002-2011) historical data (MWRDGC, 2013), the plant treats

* Corresponding author.
E-mail address: jzhu21@iit.edu (J.-J. Zhu).

https://doi.org/10.1016/j.watres.2017.10.053
0043-1354/

about 11.4 m^3/s of influent wastewater with average concentrations of 20 mg TKN (total Kjeldahl nitrogen)/L (NH_3-N/TKN ≈ 0.55), 74 mg $CBOD_5$/L, and 138 mg SS (suspended solids)/L (VSS (volatile SS)/SS ≈ 0.67). A more detailed inventory of these historical data was presented by Zhu (2015).

Many existing WRPs currently use excess energy for aeration to satisfy effluent permit requirements; therefore, effluent quality is not sensitive to influent conditions. That sensitivity, however, increases when the amount of aeration is decreased (Zhu and Anderson, 2017). As a result, a better understanding of the dynamic variations of influent characteristics can help WRP operators plan more cost-effective process management. Soft sensor applications at WRPs are attractive because of their relatively low cost, short response time, and compatibility with conventional hard sensor networks (Fortuna et al., 2007). Our initial evaluation of a soft-sensor approach at the Calumet WRP demonstrated reliable predictions of the current day's influent ammonia concentration (R^2 ≈ 0.83 with mean relative error (MRE) ≈ 13.3%) and the next day's influent flow (0.83 and 11.9%) using an iterated stepwise MLR (*ISMLR*) methodology (Zhu and Anderson, 2016). However, the ISMLR approach yielded much lower accuracy for influent $CBOD_5$ predictions (0.40 and 26.5%), especially for higher $CBOD_5$ concentrations. Prediction of higher $CBOD_5$ values is important because adequate treatment demands more energy for aeration, and inadequate treatment increases the risk of an effluent permit violation.

1.2. Prediction of high BOD concentration

In contrast with other common influent parameters such as ammonia and phosphorus, the conventional BOD measurements involves a five-day test. The resulting lack of recent BOD data could be the major cause of the relatively lower accuracy in real-time (daily resolution) BOD predictions using the conventional ISMLR method. Although several previous studies have addressed this concern, their approaches often relied on real-time data that a conventional WRP may not have. For example, Oliveira-Esquerre et al. (2004) applied MLR or partial least squares (PLS) methods to predict inlet BOD concentration in an aerated lagoon at a pulp and paper mill. Their models, based on real-time COD information, gave reasonably accurate predictions ($0.57 \le R^2 \le 0.70$). Dogan et al. (2008) compared ANN and MLR methods to predict inlet BOD concentrations at a WRP in Adapazari, Turkey. Although both ANN (R^2 up to 0.92) and MLR (up to 0.91) models achieved very good accuracy, their approach relied on real-time COD, TN, and TP concentrations. Similarly, Rustum et al. (2008) used a Kohonen self-organizing map, a type of ANN algorithm, to predict influent BOD_5 concentration at three WRPs in Scotland. Their model provided good agreement (R^2 ≈ 0.81) with measured data, but it also relied on both real-time COD and ammonia data. Kusiak et al. (2013) developed an integrated model to predict $CBOD_5$ concentrations at a WTP in Des Moines, Iowa. They first used genetic programming to estimate the most recent four BOD daily data without relying on real-time COD or nitrogen compounds, and then tested several conventional soft sensors for predicting $CBOD_5$ based on the estimated BOD data and other available historical data. Their approach obtained fairly good overall results, but did not provide good predictions on higher $CBOD_5$ values. In general, information provided by these studies cannot be directly used to address concerns at the Calumet WRP because:

- Other studies used real-time information (COD, nitrogen compounds, and phosphorus) or recent (within four days) BOD data in their models; this information may not be readily available for a conventional WRP.
- Other studies looked for overall accuracy in their BOD predictions; they did not emphasize the high BOD concentrations.

To understand how to better manage higher influent BOD concentrations, we reviewed and compared merits and drawbacks of typical soft sensor approaches, as described below.

1.3. Soft-sensor methodologies

Soft-sensor methodologies fall into four major categories: Multivariate statistics (Oliveira-Esquerre et al., 2004; Appels et al., 2011), machine learning (Wilkes et al., 2011; Seshan et al., 2014), fuzzy logic, and hybrid (Haimi et al., 2013). Based on the assumption of a normal distribution (Haimi et al., 2013), multivariate statistics can provide clear linear relationships between input regressors and the predicted variable. Machine learning methods are attractive because of their ability to address nonlinear problems, even though their internal structures can be difficult to interpret, and they can be subject to the risk of overfitting (Haimi et al., 2013). Fuzzy logic methods are increasingly used, often combined with ANN as network-based fuzzy inference system (ANFIS) (Perendeci et al., 2008; Huang et al., 2010). The ANFIS method combines advantages from ANN and fuzzy logic, so it can be readily adapted and easier to interpret relative to a typical ANN (Haimi et al., 2013). ANFIS is a hybrid method, combining different modeling algorithms, which can lead to more accurate predictions. Some studies have demonstrated that hybrid methods can address a specific problem. For example, Lee et al. (2005) tested five different non-parametric models that worked with the ASM1 model to predict common wastewater parameters (MLSS, SS, COD, and CN) at a full-scale coke-manufacturing wastewater treatment plant. They found that the NNPLS (PLS + ANN) achieved the best performance, and could help to interpret the operation behavior for fault detection and isolation. More recently, Khan et al. (2013) proposed a hybrid method (MLR + fuzzy logic) to predict D.O. concentrations in Bow River, Calgary. Rather than giving deterministic results, the hybrid method provided a range of probabilistic outputs. They determined that the risk of low D.O. concentrations (below the suggested guideline of 5 mg O_2/L) was between 3.9 and 4.9%. The advantage of a hybrid method is that it can provide a way to manage high $CBOD_5$ values. In other words, even if a soft sensor approach cannot predict high $CBOD_5$ values at the desired level of accuracy, it is possible to slightly overestimate these $CBOD_5$ values, which can result in better risk management.

We developed a hierarchical hybrid method (ISMLR + MLR + ANN) that combined advantages from each individual method. ISMLR can identify a subset of important regressors. MLR provided a prediction interval, which can help to manage high $CBOD_5$ concentrations. Compared to other common methods, ANN provided the highest global prediction accuracy based on our preliminary investigation (detailed information is provided in the Methods Section). Therefore, it can be used to provide better predictions for low $CBOD_5$ concentrations.

To implement this approach, it is necessary to determine a critical threshold to separate high and low $CBOD_5$ concentrations. The choice of this threshold is important, because if it is too high, only a small fraction of the high concentrations will be accounted for, increasing the risk of violating effluent permit requirements. Conversely, a critical threshold that is set too low could generate too many overestimates, leading to excess energy consumption. MCDA techniques can help address this concern.

1.4. Multi-criteria decision analysis techniques

Dozens of MCDA techniques have been developed; they can be