Explicit and interpretable nonlinear soft sensor models for influent surveillance at a full-scale wastewater treatment plant

Xiaodong Wang [a, b *], Knut Kvaal [b], Harsha Ratnaweera [b]

[a] School of Environmental and Municipal Engineering, Qingdao University of Technology, Fushun Road 11, Qingdao 266033, China

[a] Faculty of Science and Technology, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Aas, Norway

*Corresponding author: xiaodong.wang01@outlook.com, xiaodong.wang@nmbu.no

Abstract

In wastewater treatment plants, the most adopted sensors are those with the properties of low cost and fast response. Soft sensors are alternative solutions to the hardware sensor for online monitoring of hard-to-measure variables, such as chemical oxygen demand (COD) and total phosphorus (TP). The purpose of this study is to obtain a modelling approach which is able to identify the nonlinearity of influent and explain the correlation of inputs-outputs. Thus, the variation of influent characteristics was investigated at the first stage, which provided the basis to build global and local multiple linear regression models. Secondly, a nonlinear modelling tool multivariate adaptive regression splines (MARS) was applied for influent COD and TP prediction. Satisfactory prediction accuracy was obtained in terms of root mean square error (RMSE) and R^2. Unlike other machine learning techniques which are "black box" models, MARS provided interpretable models which explained the nonlinearity and correlation of inputs-outputs. The MARS models can be used not only for prediction, but also to provide insight of influent variation.

Keyword: Multiple linear regression; Multivariate adaptive regression splines; MARS; Nonlinear model; Soft sensor; Wastewater treatment plant

1. Introduction

The increasing requirement of wastewater treatment efficiency and economics have driven more researchers and practitioners to the field of surveillance and control of wastewater treatment plants (WWTPs) [1]. One of the limitations to achieve advanced control of WWTPs is the lack of robust and affordable online measurement instruments for some water quality variables. Fortunately, lack of hardware sensor measurements does not mean lack of information. The inner correlation and propagation trend of multiple wastewater characteristics opens another window to find low-cost solutions for global public health.

With the development of data science and machine learning techniques, more researchers have attempted to obtain unmeasured variable values by data mining and modelling [2–4]. An indirect data retrieval method, named soft sensor (or virtual sensor) was developed to obtain hard-to-measure variables by manipulating easy-to-measure variables. In WWTPs, the hard-to-measure variables are either associated with a long time-delay or high capital cost [5]. The easy-to-measure process variables are typically pH, conductivity, oxidation/reduction potential (ORP), turbidity, temperature, flowrate and probably ammonium nitrate due to recent technological development [6]. The data-driven soft sensors use easy-to-measure variables as inputs to construct prediction models, and the outputs are usually the hard-to measure variables [7–9]. Chemical oxygen demand (COD) and total phosphorus (TP) are two of the most important water quality variables for wastewater treatment process. However, the online real-time measurement of these two variables were limited due to the cost of the devices and long time-delay for real-time control. Thus, soft sensors become a potential solution for WWTPs to measure these variables online.

In practice, several statistical learning and machine learning algorithms have been applied for soft sensor modelling. The conventional multiple linear regression (MLR) has been used for water quality prediction as interpretable methods [10,11]. MLR models have the advantages of simplicity and easy interpretability, which can be easily programmed for surveillance and control in practice. However, since MLR assumes the linearity of input-output relation, its limits in prediction accuracy have also been addressed in several water related studies [12,13]. Several nonlinear statistical learning and machine learning methods have been tested to build water quality soft sensor models [14–16], and the most studied methods for water quality soft sensor are neural networks [9,17,18]. Nonlinear models can be easily simulated nowadays, but it is still a challenge to apply complicated nonlinear models in practice [1]. Neural networks' capability of prediction and nonlinearity capture were well addressed in various literature, but lack of interpretability has limited their value, because the insight of dataset and natural characteristics of input-output are usually left unexplained as "black boxes" [19]. Another drawback of training the neural networks is the lack of general protocol to determine the structure, i.e. the number of layers and neurons [14,20]. Nevertheless, researchers and practitioners will continue the effort to make data-driven soft sensors more feasible due to their fantastic potential to substitute hardware sensors and overcome measurement delay [21].

Application of sophisticated modelling techniques does not always pay off unless the methods were critically compared with simpler methods [22]. In this study, we attempted to investigate the correlation of predictors and response variables and simultaneously build soft sensors models for chemical oxygen demand (COD) and total phosphorus (TP) of WWTP influent. A global linear model may not work well in different seasonal and climate conditions, even in a selected WWTP. Local models for each of the subsets can be obtained by dividing the general dataset into several subsets based on different conditions. This is the conventional solution to simultaneously deal with nonlinear prediction and to capture the insights of nonlinear relationships. A well established statistical learning method named Multivariate Adaptive Regression Splines (MARS) developed by Friedman is an alternative solution to achieve the goal of this study [23].

Multivariate adaptive regression splines (MARS) can be used in data mining and prediction for complex and high-dimensional nonlinear data. Kuter et al. applied both MARS and multilayer feedforward artificial neural network (ANN) for fractional snow cover estimation [19]. The MARS approach performed the same as ANN, but was more computationally efficient in model building. Moreover, MARS has been proven as an efficient tool in the prediction of phenol and nitrophenol adsorption [13], classification of satellite images [24], rainfall-runoff simulation [25], and identification of dominant interaction of climatic effect on rainfall and water availability [26]. The advantage of MARS is its ability to automatically add knots to the general curve, which break the global model into piecewise linear polynomial splines. By smoothly connecting spline pieces, the MARS model is able to capture both the linearity and nonlinearity. Therefore, MARS retains the interpretability of linear models, and it is also capable to provide insight of the natural phenomenon.

The primary goal of this study is to derive explicit soft sensor models, which can be used to predict hard-to-measure WWTP influent variables. However, training and validating nonlinear models requires a large quantity of measured data to perform sufficient model runs [27]. Although running these models may not be computationally expensive, it is too expensive and time consuming to obtain a large dataset of influent wastewater quality. Therefore, the secondary goal of this study is to obtain interpretable nonlinear models with limited full-scale WWTP influent data.

2. Materials and methods

2.1 Dataset and problem description

A multivariate dataset of WWTP influent characteristics was obtained by sampling and analyzing the hourly influent quality and quantity. The dataset was supposed to cover influent characteristics of both warm and cold season in Norway. Due to the high cost of laboratory analysis and time limitation, the sampling lasted six days in warm season, and five days in cold season. Samples were collected in every hour and 24 samples were collected in each sampling day. In warm season, all of the samples were collected in dry climate condition (no storm event). Melting snow resulted in occasional wet climate condition in cold season. The cold season data were collected in both dry climate and wet climate conditions. The general dataset can be divided into three subsets (Warm-Dry, Cold-Dry and Cold-Wet) by applying a classifier developed in a previous study [28].

There are both easy-to-measure variables and hard-to-measure variables in the dataset. In the context of soft sensor development, the easy-to-measure variables are pH, flow rate (Flow), total suspended solid (TSS), water temperature (WaterTemp), and ammonium nitrate (NH_4-N). The hard-to-measure variables are chemical oxygen demand (COD) and total phosphorus (TP), because online measurement devices for these two variables are expensive and have long time-delays which prevent them from being used for real-time surveillance and control. Therefore, we are interested to study the correlation of these easy-to-measure variables and hard-to-measure variables, and to obtain soft sensor models as alternatives of hardware sensors.

2.2 Global and piecewise multiple linear regression (MLR)

Multiple linear regression (MLR) is the linear model trained based on least square estimation, which is a simple way to interpret the correlations of inputs and outputs. To increase prediction accuracy, high order terms and interaction terms can be involved as inputs. In this study, square terms of original variables and two-effect-interaction terms were applied in MLR models. All the original variables, square terms and interaction terms were included to train an over-fitted model at the first step. Secondly, backward stepwise selection method was applied to eliminate non obligatory prediction terms based on Akaike's information criterion (AIC) to obtain the "shrunk model" with only significant variables [29]. The global MLR models were trained based on the general dataset, and the piecewise local models were trained based on corresponding subsets. At last, every 10-12 observations were randomly selected to form several folds, and these folds were employed to perform cross validation to verify the final model.

2.3 Multivariate adaptive regression splines

Multivariate adaptive regression splines (MARS) was initially presented by Frieman as a nonlinear regression method [23]. It can be viewed as an integration of piecewise linear regression, which captures nonlinearity by adding knots to input variables to break the curve into piecewise basis functions. The general form of the MARS model is expressed as:

$$Y = \beta_0 + \sum_{i=1}^{M} \beta_i \cdot h_i(\boldsymbol{X^n}) + \varepsilon \qquad (1)$$

where $h_i(\boldsymbol{X})$ is the basis function representing each piece of local linear regression, and β_i is the associated coefficient. The coefficients β were estimated based on the least squares method. The basis functions have the following form:

$$h(x)_+ = \begin{cases} x-k, & if\ x>k \\ 0, & otherwise \end{cases};$$

$$h(x)_- = \begin{cases} k-x, & if\ x<k \\ 0, & otherwise \end{cases} \quad (2)$$

where k is a univariate knot. Thus, the MARS method produces continuous models. The determination of basis functions was a data-driven process. MARS can apply both forward stepwise and backward stepwise to select inputs. In this study, second order terms and interaction terms are involved and being selected automatically. The cross validation method was applied to verify MARS models in the same way as it is used for MLRs. The "earth" package was applied in R to build MARS models [30].

Unlike neural network based methods, statistical learning methods such as MARS should not be viewed as black boxes [31]. MARS retained the interpretability to explain the nonlinear correlation of inputs and outputs.

3. Results and discussion

3.1 Variation of wastewater characteristics

The variation of influent quantity and quality of wastewater treatment plants contributes major uncertainties to process operation, designing and modelling [32,33]. Our previous study had shown the daily, weekly and seasonal variation of influent, caused by both human activity and climate [28]. To further investigate the correlation of wastewater quality and quantity, the variation of flowrate and contaminants under different conditions was monitored, as shown in Figure 1. Figure 1(a) showed 5-day continuous measurement of influent flowrate (Flow), COD, TSS and ammonium nitrogen (NH_4-N) during warm season. Only dry climate condition was observed during the sampling days (Warm-Dry condition). Generally, the wastewater contaminants concentration increased or decreased at the same time as the flowrate, which indicated the linearity of influent quality and quantity. The linearity of influent variation in dry climate may be traced back to regular human activity, and it provides the possibility to establish explicit soft sensor models.

However, the variation of influent in dry climate condition and wet climate condition follows different statistical distribution [34]. To capture the influent characteristics in wet climate condition, continuous sampling and measurement of influent quality and quantity was conducted in snow season for six days. This was done as unpredictable snow melting may happen at any time during the day [35,36]. In this study, as shown in Figure 1(b), the wastewater quality propagated differently than influent flowrate when the climate factor was shifted to wet condition due to snowmelt (Figure 1b). The drop and rise feature of influent characteristics during freezing cold time (Cold-Dry condition) was similar to that in Warm-Dry condition. While there was snow melting (Cold-Wet condition), contaminants concentration decreased due to dilution because flowrate increased dramatically. Figure 1(b) indicates that the climate effect increased the uncertainty of influent variation.

Overall, the nonlinearity of influent characteristics was the combined effect of both human activity and climate. To develop soft sensor models for influent quality prediction, the nonlinearity in different scenarios should be considered.

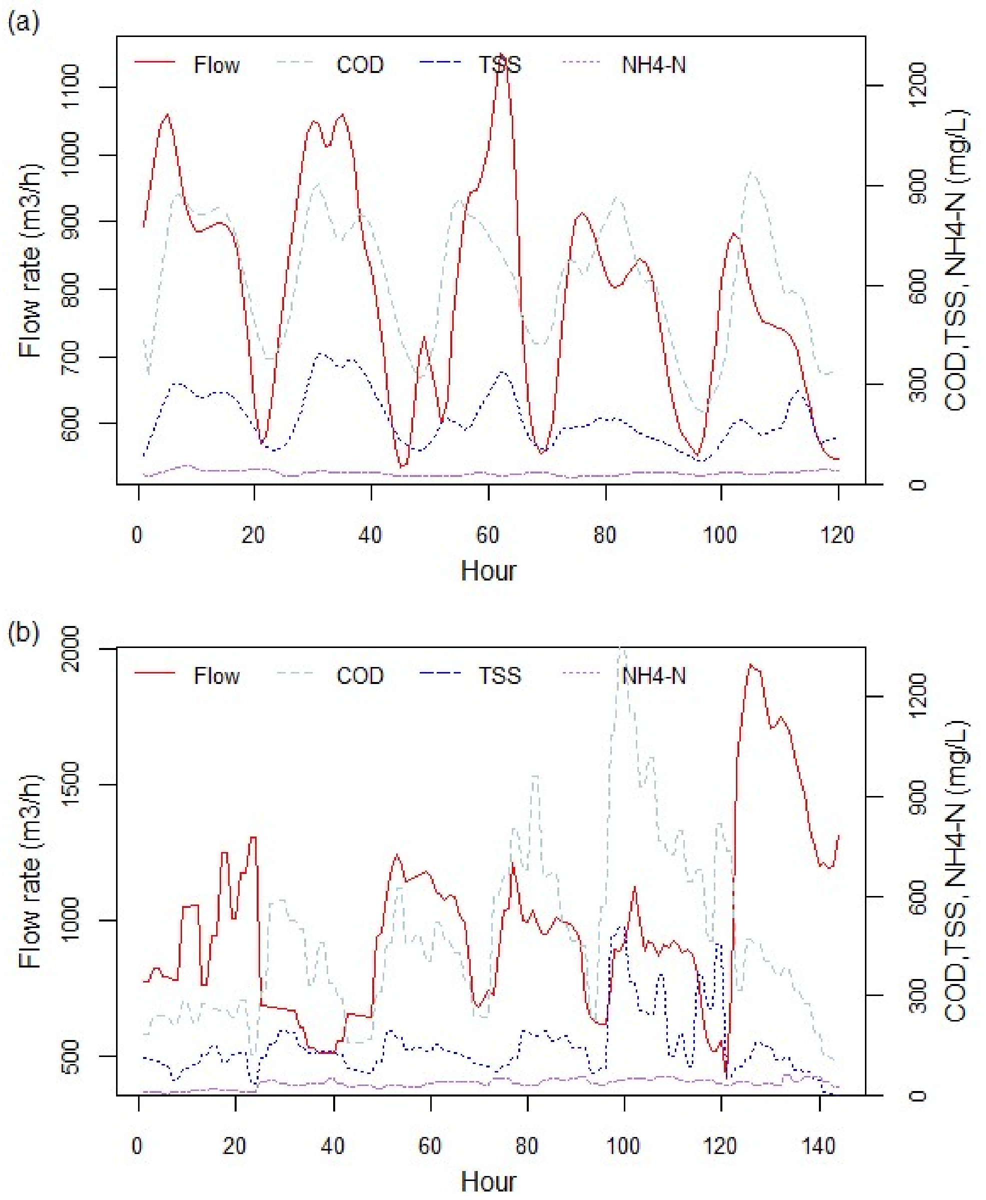

Figure 1 The hourly data of influent flowrate, COD, TSS and ammonium in (a) warm season, no storm events happened during five days observation (Warm-Dry condition);(b) cold season, with both dry and climate conditions.

3.2 Global and piecewise multiple linear regression

Though it is difficult to detect when the wet climate starts due to snow melting, the general dataset can be split into Warm-Dry, Cold-Dry and Cold-Wet subsets by applying statistical learning algorithm according to the previous study [28]. The next step is to investigate the correlation between influent characteristic factors. Since COD and total phosphorus (TP) are not easy to measure directly by online sensors, it is more essential to study the prediction accuracy of COD and TP.

Multiple linear regression (MLR) method was used to build models for COD and TP prediction and study the influence of interaction term on prediction accuracy. Independent variables flowrate (Flow), water temperature (WaterTemp), total suspended solid (TSS), ammonium nitrogen (NH_4-N) and pH were selected as predictors. This is because these variables are relatively easy-to-measure variables in WWTPs. Moreover, quadratic terms and interaction terms were also included to build an over-fitted model with all the possible features as prediction terms. Secondly, backward stepwise selection method was applied to eliminate non obligatory prediction terms based on Akaike's information criterion (AIC) to avoid over-fitting [29]. The selected prediction terms and coefficients for the final models of COD and TP were listed in Table S1 and Table S2 in supplementary material. In this study, we intended to obtain soft sensor models that are applicable in different seasonal and climate conditions. Therefore, regression models for COD and TP prediction was established based on the general dataset (Table S1) and subsets of different conditions (Table S2).

The results of linear multiple linear regression (MLR) modelling was sown in Table 1. For COD predictions, the model built on the general dataset without interaction term was a second order polynomial equation (MLR 1). While the COD model with interaction terms (MLR 2) contains both second order and interaction terms. In total 13 prediction terms were selected by AIC backward stepwise selection. The model performance in terms of RMSE were quite close for MLR1 and MLR2. The cross-validated R^2 of MLR 1 (0.835) was approximately equal to that of MLR 2 (0.847), which indicates limited contribution from interaction term for COD prediction.

Considering the uncertainty of full-scale WWTPs and the criteria of prediction accuracy for wastewater soft sensors in literature [11,15,37], it would be sufficient to build soft sensor for influent monitoring if R^2 is higher than 0.80. Therefore, Model 1 is simple and robust enough to serve as a global model for COD prediction. However, the second order polynomial without interaction term for TP prediction (MLR 3) was not as satisfactory as MLR 1 for COD prediction, but Model 4 with three square terms and seven interaction terms performed much better in terms of RMSE and R^2. The cross-validated R^2 was slightly lower than R^2, which indicated that the model was not over fitted. Therefore, the interaction term was necessary to train a global MLR model for TP prediction.

Unlike global COD models, there was a clear difference in the prediction accuracy for global TP which was caused by interaction terms. Local MLR models were built to study the linearity of TP and its correlation with easy-to-measure variables. RMSE would no longer serve as a fair comparison index, because the TP values in wet climate were significantly lower than that in dry climate. Therefore, the comparison of piecewise MLR models was based on R^2 only in this case.

As shown in Table 1, the MLR model built on Warm-Dry subset (MLR 5) appeared with satisfactory prediction accuracy. While the R^2 of the other piecewise local MLRs did not reach the criteria of 0.80, regardless of interaction terms. Although R^2 of MLRs in cold season can be improved by including interaction terms, the cross-validated R^2 values of MLR 7 and MLR 9 were too low to serve as soft sensor models. The results in Table 1 revealed the higher nonlinearity of TP variation in the influent.

A satisfactory global COD prediction model can be obtained by applying multiple linear regression. For TP prediction, interaction terms were necessary to obtain similar prediction accuracy as the COD MLR model. The piecewise MLRs for TP performed poorly. The nonlinearity of TP requires further study.

Table 1 Mutiple linear regression (MLR) performance for global (using general dataset) and piecewise (subsets of general dataset) COD and TP prediction.

Model	Target variable	Interaction term	Dataset	RMSE	R^2	Cross validation R^2
MLR 1	COD	No	General	98.128	0.840	0.835
MLR 2		Yes	General	92.360	0.859	0.847
MLR 3	TP	No	General	0.919	0.714	0.707
MLR 4		Yes	General	0.658	0.853	0.845
MLR 5		No	Warm-Dry	0.607	0.821	0.813
MLR 6		No	Cold-Dry	0.809	0.532	0.490
MLR 7		Yes	Cold-Dry	0.572	0.766	0.715
MLR 8		No	Cold-Wet	0.365	0.705	0.643
MLR 9		Yes	Cold-Wet	0.317	0.774	0.696

3.3 Multivariate adaptive regression splines

The piecewise MLR models in the last section were built based on the classification of season-climate conditions. Multivariate adaptive regression splines (MARS) can provide smooth piecewise function without specific assumptions of the relationship between predictors and target variables [23]. The knots and splines are automatically selected based on the property of the data rather than being pre-defined. One of the objectives of this work is to develop interpretable models for COD and TP prediction, which can deal with both linear and nonlinear problems with sufficient accuracy. Although MARS has advantage in interpretability [19], MARS has been used less in the field of water quality and wastewater treatment compared with neural network. Soft sensor models that can provide more explicit information were preferred for the surveillance and control of wastewater treatment process.

As shown in Table S3 in supplementary material, MARS models were built for global influent COD and TP prediction. The MARS model for COD prediction employed square terms but interaction terms were not applied. On the contrary, the application of interaction terms in the TP model reflected that the joint effects of influent variables were significant for TP prediction. The scatter plot of predicted COD and TP by MARS model were shown in Figure 2(a) and Figure 2(b), respectively. The model predicted COD and TP fit the measured data well in a wide range. As shown in Figure 2, the data points were located close to the trend line even for extreme high values, i.e. COD$\geq 800\ mg/L$ and TP$\geq 6\ mg/L$. The prediction performance in terms of RMSE and R2 were listed in Table 2. The RMSE of the MARS model for COD prediction (MARS 1) was 80.4 (Table 2), which is much smaller than that of MLR 1 (98.128) and MLR 2 (92.360). A similar result was also found in TP modelling. The R^2 of MARS 2 and MLR 4 were almost equal for global TP prediction, but the decrease of RMSE was significant due to the better capture of nonlinearity. Table 2 also showed the necessary numbers of knots and prediction terms for MARS model construction. The MARS model for TP prediction required more knots and interaction terms to achieve the same R^2 level as that of COD, which indicated the higher nonlinearity of TP propagation. Overall, the accuracy in terms of RMSE for COD and TP were improved by applying MARS method.

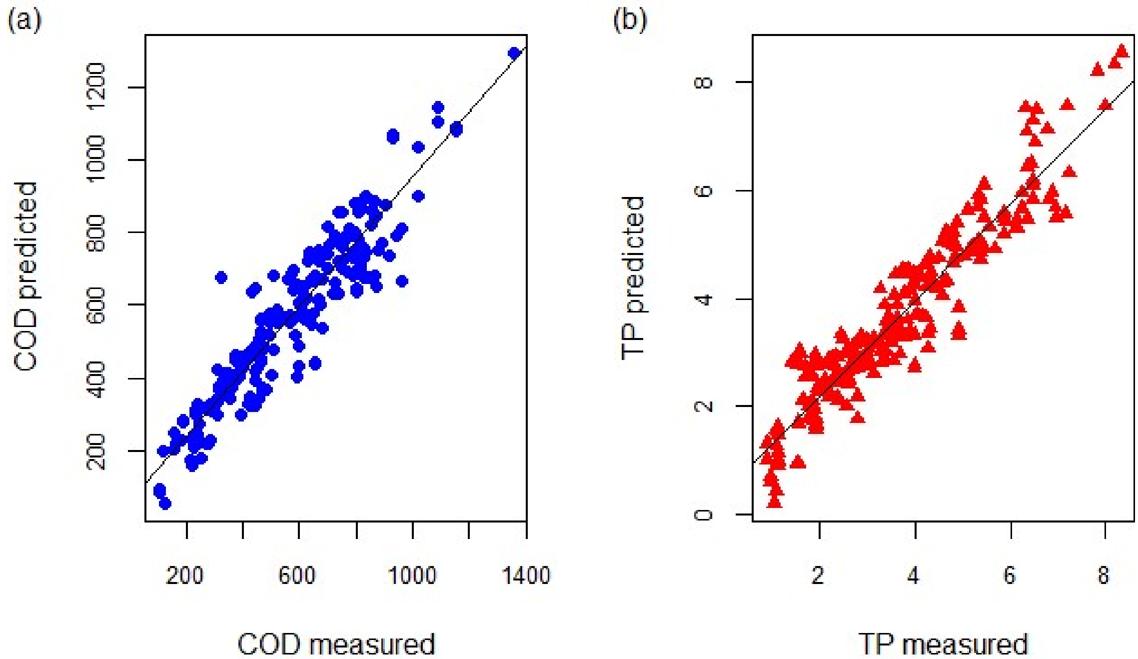

Figure 2 (a) Predicted COD by MARS method versus measured COD; (b) Predicted TP by MARS versus measured TP.

Table 2 Performance of MARS models for global influent COD and TP prediction

Model	Target	Interaction term	Knots	Basis function	RMSE	R^2	Cross Validation R^2
MARS 1	COD	No	6	13	80.439	0.892	0.851
MARS 2	TP	Yes	9	16	0.573	0.889	0.841

3.4 Discussion and outlook of soft sensor models

The fitted COD and TP by global MLR and MARS were compared with the measurement, as is shown in Figure 3. Generally, the difference of prediction performance can be hardly detected from Figure 3. However, the COD prediction by MARS was closer to the measurement for peak hours (Figure 3(a)), which may be the reason of significant lower RMSE of MARS than MLR. For TP prediction, MARS also showed better approaching to the measurement in fluctuating situations in Figure 3(b). Therefore, we can conclude that MARS performed better than MLR for the prediction of extremal values. MARS breaks the general data into several pieces of splines and allows the slope of each piece to be different. The MARS models can capture the extremal situations due to the allowance of several splines. For the global MLR models with higher order and interaction terms, they always need to be continuous and their first derivatives also need to be continuous. In general, global MLR fit smoother curves (Figure 3), but it did not help improve the fitting of real world data due to nonlinear fluctuation [12]. The improvement of nonlinearity capture is important to reduce measurement error and increase control stability.

It is also instructive to evaluate the numbers of knots and basis functions of the two MARS models. MARS 1 was constructed with 6 knots and 13 basis functions, while MARS 2 has 9 knots and 16 basis functions. To achieve a satisfactory prediction, the TP dataset was broken into smaller pieces of splines by adding more numbers of knots, which reflected higher nonlinearity of TP than COD. The basis functions are second order polynomials. As listed in Table S3, the COD model (MARS 1) has first order and square terms, while the TP model (MARS 2) included interaction terms. The interaction terms suggested that the correlation of TP and an easy-to-measure variable were dependent on the third easy-to-measure variable [31]. Thus, the interaction terms in basis functions were significant for TP prediction, but the interaction effect was not statistical significant for COD prediction.

Compared with machine learning algorithms (e.g. neural networks), MARS is less “black box” for nonlinear predictive modelling. This is because MARS is more informative and interpretable to retrieve the real world knowledge [12,26]. In this study, the knots of MARS were capable of informing where significant changes of TP or COD may happen. Moreover, the three local piecewise MLRs for TP prediction were built on the basis of pre-known knowledge of season-climate conditions, which are highly dependent on the accuracy of data classification. In other words, the TP dataset was firstly classified into three subsets and each of the local MLR models was built on the corresponding subsets. The local MLRs turned out to perform unsatisfactory as global models. Moreover, the MARS model for TP prediction included 9 knots rather than 3 knots, which indicates that the causes of nonlinearity of TP was beyond the effects of seasonal variation or climate impact.

Overall, the MARS models are not only flexible in prediction of hard-to-measure variables, but also provided explicit knowledge for the downstream operation of wastewater treatment process.

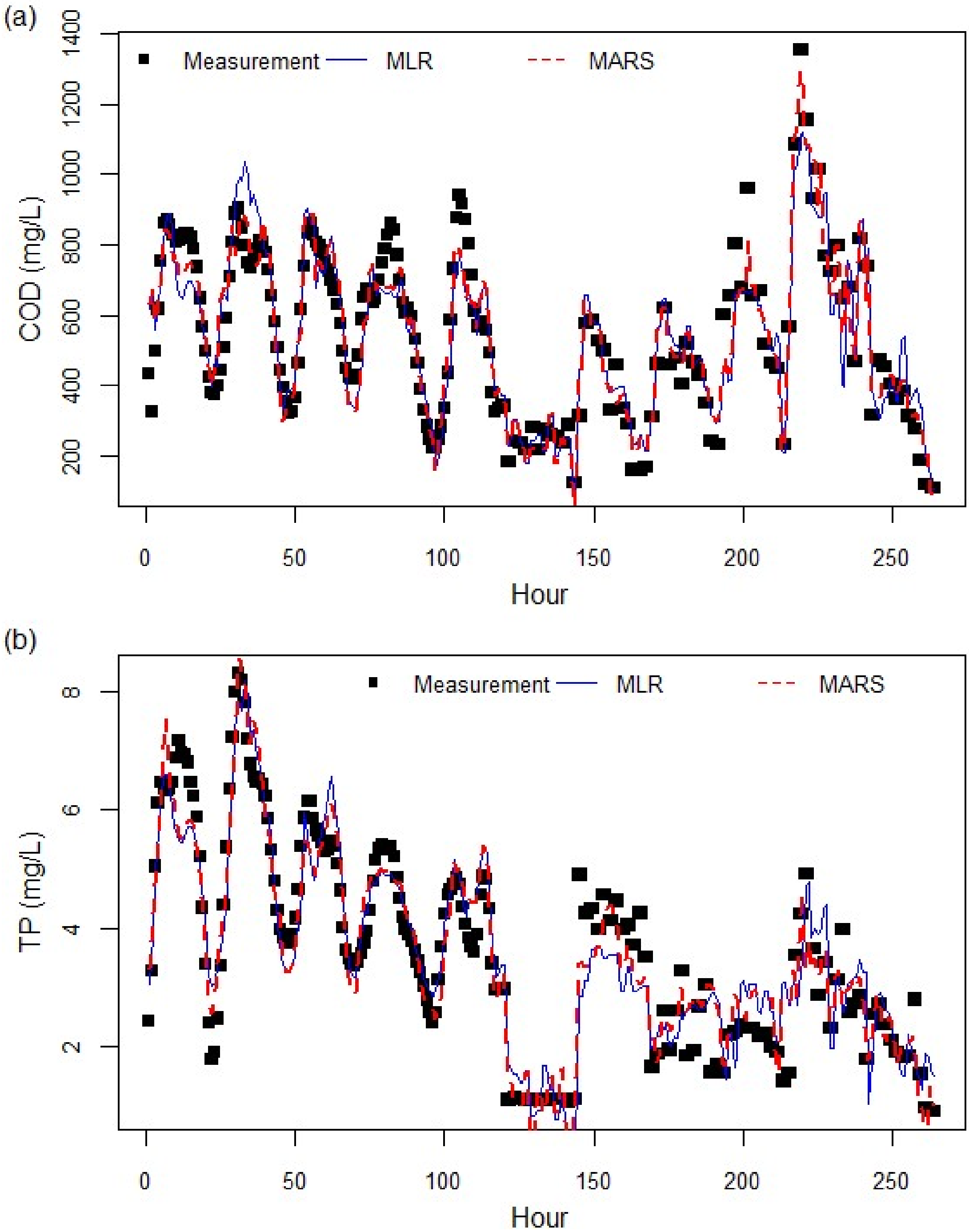

Figure 3 (a) The measured COD and predicted COD by global MLR and MARS; (b) The measured TP and predicted TP by global MLR and MARS.

4. Conclusion

The online monitoring of influent wastewater characteristics is essential for wastewater treatment process surveillance and control. Soft sensor is an alternative solution for online measurement of COD and total phosphorus (TP) in an economic manner. In this study, we investigated the possibility of using easy-to-measure variables as predictors to construct both global and local soft sensor models for COD and TP prediction. The goal is to build interpretable nonlinear models to serve as soft sensors for the surveillance of wastewater treatment process.

The global MLR models performed similar to the MARS models in terms of R^2 for COD and TP prediction. However, the RMSEs of MARS models were smaller than that of the corresponding MLRs. MARS has the advantage of capturing nonlinearity in fluctuating situations.

Compared with other “black box” modeling techniques, such as neural network, useful information and knowledge can be retrieved from MARS models. The MARS models indicated the points where significant changes happened. Moreover, splines may also suggest the number of groups for pre-classification of the dataset.

Acknowledgements

This work has been supported by the RECOVER project, granted by the Research Council of Norway (grant number 247612).

Appendix A. Supplementary material

Supplementary material related to this article can be found in the online version.

References

[1] G. Olsson, B. Carlsson, J. Comas, J. Copp, K. V Gernaey, P. Ingildsen, U. Jeppsson, C. Kim, L. Rieger, I. Rodríguez-Roda, J.-P. Steyer, I. Takács, P. a Vanrolleghem, A. Vargas, Z. Yuan, L. Åmand, Instrumentation, control and automation in wastewater - from London 1973 to Narbonne 2013, Water Sci. Technol. 69 (2014) 1373–1385. doi:10.2166/wst.2014.057.

[2] W. Yan, P. Guo, Y. Tian, J. Gao, A Framework and Modeling Method of Data-Driven Soft Sensors Based on Semisupervised Gaussian Regression, Ind. Eng. Chem. Res. 55 (2016) 7394–7401. doi:10.1021/acs.iecr.5b04118.

[3] J. Shiri, A. Keshavarzi, O. Kisi, S. Karimi, Using soil easily measured parameters for estimating soil water capacity: Soft computing approaches, Comput. Electron. Agric. 141 (2017) 327–339. doi:10.1016/j.compag.2017.08.012.

[4] H. Seshan, M.K. Goyal, M.W. Falk, S. Wuertz, Support vector regression model of wastewater bioreactor performance using microbial community diversity indices: Effect of stress and bioaugmentation, Water Res. 53 (2014) 282–296. doi:10.1016/j.watres.2014.01.015.

[5] F.A.A. Souza, R. Araújo, J. Mendes, Review of soft sensor methods for regression applications, Chemom. Intell. Lab. Syst. 152 (2016) 69–79. doi:10.1016/j.chemolab.2015.12.011.

[6] H. Haimi, M. Mulas, F. Corona, R. Vahala, Data-derived soft-sensors for biological wastewater treatment plants: An overview, Environ. Model. Softw. 47 (2013) 88–107. doi:10.1016/j.envsoft.2013.05.009.

[7] Y. Liu, J. Chen, Z. Sun, Y. Li, D. Huang, A probabilistic self-validating soft-sensor with application to wastewater treatment, Comput. Chem. Eng. 71 (2014) 263–280. doi:10.1016/j.compchemeng.2014.08.008.

[8] H. Haimi, M. Mulas, F. Corona, S. Marsili-Libelli, P. Lindell, M. Heinonen, R. Vahala, Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant, Eng. Appl. Artif. Intell. 52 (2016) 65–80. doi:10.1016/j.engappai.2016.02.003.
[9] J. Fernandez de Canete, P. Del Saz-Orozco, R. Baratti, M. Mulas, A. Ruano, A. Garcia-Cerezo, Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network, Expert Syst. Appl. 63 (2016) 8–19. doi:10.1016/j.eswa.2016.06.028.
[10] S. Akilandeswari, B. Kavitha, Comparison of ANFIS and statistical modeling for estimation of chemical oxygen demand parameter in textile effluent, Der Chem. Sin. 4 (2013) 96–99.
[11] M. Ay, O. Kisi, Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques, J. Hydrol. 511 (2014) 279–289. doi:10.1016/j.jhydrol.2014.01.054.
[12] A.T.C. Goh, W.G. Zhang, An improvement to MLR model for predicting liquefaction-induced lateral spread using multivariate adaptive regression splines, Eng. Geol. 170 (2014) 1–10. doi:10.1016/j.enggeo.2013.12.003.
[13] E. Z-Flores, M. Abatal, A. Bassam, L. Trujillo, P. Juárez-Smith, Y. El Hamzaoui, Modeling the adsorption of phenols and nitrophenols by activated carbon using genetic programming, J. Clean. Prod. 161 (2017) 860–870. doi:10.1016/j.jclepro.2017.05.192.
[14] Y. Qiu, Y. Liu, D. Huang, Date-Driven Soft-Sensor Design for Biological Wastewater Treatment Using Deep Neural Networks and Genetic Algorithms, J. Chem. Eng. Japan. 49 (2016) 925–936. doi:10.1252/jcej.16we016.
[15] Y. Liu, Adaptive just-in-time and relevant vector machine based soft-sensors with adaptive differential evolution algorithms for parameter optimization, Chem. Eng. Sci. 172 (2017) 571–584. doi:10.1016/j.ces.2017.07.006.
[16] B. Szelag, J. Gawdzik, Assessment of the Effect of Wastewater Quantity and Quality, and Sludge Parameters on Predictive Abilities of Non-Linear Models for Activated Sludge Settleability Predictions, Polish J. Environ. Stud. 26 (2017) 315–322. doi:10.15244/pjoes/64810.
[17] M. Zhang, X. Liu, A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction, Chemom. Intell. Lab. Syst. 126 (2013) 83–90. doi:10.1016/j.chemolab.2013.04.018.
[18] D.-J. Choi, H. Park, A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process, Water Res. 35 (2001) 3959–3967. doi:10.1016/S0043-1354(01)00134-8.
[19] S. Kuter, Z. Akyurek, G.W. Weber, Retrieval of fractional snow covered area from MODIS data by multivariate adaptive regression splines, Remote Sens. Environ. 205 (2018) 236–252. doi:10.1016/j.rse.2017.11.021.
[20] Y. Liu, J. Guo, Q. Wang, D. Huang, Prediction of Filamentous Sludge Bulking using a State-based Gaussian Processes Regression Model, Sci. Rep. 6 (2016) 31303. doi:10.1038/srep31303.
[21] H. Haimi, F. Corona, M. Mulas, L. Sundell, M. Heinonen, R. Vahala, Shall we use hardware sensor measurements or soft-sensor estimates? Case study in a full-scale WWTP, Environ. Model. Softw. 72 (2015) 215–229. doi:10.1016/j.envsoft.2015.07.013.
[22] O. Samuelsson, A. Björk, J. Zambrano, B. Carlsson, Gaussian process regression for monitoring and fault detection of wastewater treatment processes, Water Sci. Technol. 75 (2017) 2952–2963. doi:10.2166/wst.2017.162.
[23] J.H. Friedman, Multivariate Adaptive Regression Splines, Ann. Stat. 19 (1991) 1–67. http://www.jstor.org/stable/224183.
[24] E. Quirós, Á.M. Felicísimo, A. Cuartero, Testing multivariate adaptive regression splines (MARS) as a method of land cover classification of TERRA-ASTER satellite images, Sensors. 9 (2009) 9011–9028. doi:10.3390/s91109011.
[25] M. Rezaie-Balf, Z. Zahmatkesh, S. Kim, Soft Computing Techniques for Rainfall-Runoff Simulation: Local Non–Parametric Paradigm vs. Model Classification Methods, Water Resour. Manag. 31 (2017) 3843–3865. doi:10.1007/s11269-017-1711-9.

[26] W. Xing, W. Wang, Q. Shao, B. Yong, Identification of dominant interactions between climatic seasonality, catchment characteristics and agricultural activities on Budyko-type equation parameter estimation, J. Hydrol. 556 (2018) 585–599. doi:10.1016/j.jhydrol.2017.11.048.

[27] D. Shahsavani, S. Tarantola, M. Ratto, Evaluation of MARS modeling technique for sensitivity analysis of model output, Procedia - Soc. Behav. Sci. 2 (2010) 7737–7738. doi:10.1016/j.sbspro.2010.05.204.

[28] X. Wang, K. Kvaal, H. Ratnaweera, Characterization of influent wastewater with periodic variation and snow melting effect in cold climate area, Comput. Chem. Eng. 106 (2017) 202–211. doi:10.1016/j.compchemeng.2017.06.009.

[29] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control. 19 (1974) 716–723. doi:10.1109/TAC.1974.1100705.

[30] Milborrow Stephen, Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's Fortran utilities with Thomas Lumley's leaps wrapper. earth: Multivariate Adaptive Regression Splines, Version 4.6.1, R Packag. (2018). https://cran.r-project.org/package=earth.

[31] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning: with Applications in R, Springer New York, 2014. https://books.google.no/books?id=at1bmAEACAAJ.

[32] E. Belia, Y. Amerlinck, L. Benedetti, B. Johnson, G. Sin, P.A. Vanrolleghem, K. V. Gernaey, S. Gillot, M.B. Neumann, L. Rieger, A. Shaw, K. Villez, Wastewater treatment modelling: Dealing with uncertainties, Water Sci. Technol. 60 (2009) 1929–1941. doi:10.2166/wst.2009.225.

[33] G. Olsson, ICA and me--a subjective review., Water Res. 46 (2012) 1585–624. doi:10.1016/j.watres.2011.12.054.

[34] M. Talebizadeh, E. Belia, P.A. Vanrolleghem, Influent generator for probabilistic modeling of nutrient removal wastewater treatment plants, Environ. Model. Softw. 77 (2016) 32–49. doi:10.1016/j.envsoft.2015.11.005.

[35] B.G. Plósz, H. Liltved, H. Ratnaweera, Climate change impacts on activated sludge wastewater treatment: a case study from Norway., Water Sci. Technol. 60 (2009) 533–541. doi:10.2166/wst.2009.386.

[36] S. Moghadas, a-M. Gustafsson, T.M. Muthanna, J. Marsalek, M. Viklander, Review of models and procedures for modelling urban snowmelt, Urban Water J. 9006 (2015) 1–16. doi:10.1080/1573062X.2014.993996.

[37] S. Puig, M.C.M. van Loosdrecht, J. Colprim, S.C.F. Meijer, Data evaluation of full-scale wastewater treatment plants by mass balance, Water Res. 42 (2008) 4645–4655. doi:10.1016/j.watres.2008.08.009.