Received October 3, 2019, accepted October 24, 2019, date of publication October 29, 2019, date of current version November 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950034

Modeling of Adaptive Multi-Output Soft-Sensors With Applications in Wastewater Treatments

JING WU[1,2], HONGCHAO CHENG[1], YIQI LIU[1], (Member, IEEE), BIN LIU[3], (Member, IEEE), AND DAOPING HUANG[1]

[1]School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China
[2]School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China
[3]Department of Management Science, University of Strathclyde, Glasgow G1 1XQ, U.K.

Corresponding authors: Yiqi Liu (aulyq@scut.edu.cn) and Daoping Huang (audhuang@scut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61873096 and Grant 61673181, and in part by the Science and Technology Program of Guangzhou, China, under Grant 201804010256.

ABSTRACT Given the multivariable coupling, strong nonlinearity and time-varying features in the wastewater treatment processes, adaptive strategies, including just-in-time learning (JITL), time difference (TD), and moving window (MW) methods have been chosen in this paper to enhance multi-output soft-sensor models to ensure online prediction for a variety of hard-to-measure variables simultaneously. In the proposed adaptive multi-output soft-sensors, multi-output partial least squares (MPLS), multi-output relevant vector machine (MRVM) and multi-output Gaussian process regression (MGPR) served as the multi-output models. The integration of adaptive strategies and multi-output models not only provides a solution for multi-output prediction, but also offers a potential to alleviate the degradation of multi-output soft-sensors. To further improve the adaptive ability, four adaptive soft-sensors, termed TD-MW, TD-JIT, JIT-MW, and TD-JIT-MW, have been proposed by mixing the three aforementioned adaptive strategies to upgrade multi-output soft-sensors. All the adaptive multi-output soft-sensors are analyzed and compared in terms of simulation data and practical industrial data, which exhibit stationary and nonstationary behaviors, respectively.

INDEX TERMS Adaptive soft sensor, multiple-output, wastewater treatment plants (WWTPs), multiple adaptive mechanisms.

I. INTRODUCTION

Soft-sensors are generally applied in the predictions of difficult-to-measure but quality-related variables, such as the chemical oxygen demand (COD) and biological oxygen demand for five days (BOD_5) in wastewater treatment plants (WWTPs), mainly due to expensive analyzer costs, hostile working surroundings and large time delays of hardware sensors measurement [1]. Compared with the first-principle models, data-driven soft-sensors are easy to implement and do not require much prior knowledge of the process, thus, they have been widely used for decades. Many methods have been proposed to build data-driven soft-sensors for WWTPs and other industrial processes, such as partial least squares (PLS) [2], [3], support vector machines (SVM) [4], relevant vector machines (RVM) [5], [6], and Gaussian process regression (GPR) [7], [8].

Traditionally, data-driven soft-sensors are organized as single-output models, with historical datasets being used to estimate a difficult-to-measure process variable [9]. However, there is always a variety of quality-relevant variables that are difficult-to-measure, such as the BOD_5, COD, total phosphorus (TP) and sludge volume index (SVI) in WWTPs. Applying a single-output model to measure several difficult-to-measure variables simultaneously is tedious and inadequate. Although every single-output soft-sensor can be constructed to estimate each difficult-to-measure variable separately, the solution is premised on independent assumptions within hard-to-measure response variables, thus leading to a deterioration of the prediction performance. Also, reserving different sets of selected samples for each model separately is required, if every single-output soft-sensor can be reasonably constructed to estimate each difficult-to-measure variable separately. Generally, the multi-output regression model not only needs to consider the potential relationships between the input variables and targets, but also take into account the co-relationships among the targets, aiming to

The associate editor coordinating the review of this manuscript and approving it for publication was Xiang Huang.

guarantee a better interpretation and easier generalization. With popularity of statistical learning and advancement of automation, numerous of multiple-output regression methods have been widely applied in recent years [10], [11], such as multivariate partial least squares (MPLS) [2], multi-output relevant vector machine (MRVM) [12], and multi-output Gaussian process regression (MGPR) [8], [13] methods. Liu *et al.* [8] proposed a MGPR model for multi-step prediction with the results demonstrating that the proposed methodology was capable of predicting future SVI with better accuracy than a standard single-ouput GPR. Literature [12] indicated that the performance of MRVM is better than RVM slightly due to the consideration of correlationship among the hard-to-measure variables. Xiao *et al.* [9] developed multi-output soft-sensors using Multivariate Linear Regression model (MLR), MRVM and MGPR models aiming to predict multiple hard-to-measure variables simultaneously and to capture the joint distribution of the response variables. However, these researches are usually limited in off-training and on-line testing.

The performance of multi-output soft-sensors usually degrades after on-line use in process industries over a period of time [9]. The predictive accuracy of a soft-sensor is gradually decreased with the negative effects of uncertainties, including mechanic components failure, catalyst deactivation, process conditions switching, etc. Once the soft-sensors are put into practice without receiving proper maintenance, they will degrade gradually and could further lead to significant losses. However, aforementioned situations will lead to more serious degradation of the multi-output models [2]. Depending on the rate of changes and time of duration, the drifts can be divided into gradual drifts with slow changes and abrupt changes, wherein the concept transits immediately from one state to another. If the model cannot effectively adapt to the gradual drifts or abrupt changes properly in the industrial process, the performance of the soft-sensor will become worse. This, in turn, affects the process monitoring and control performance. In this light, some adaptive learning techniques have been proposed to maintain the predictive accuracy of soft-sensors [14]. Adaptive soft sensors can be constructed based on the moving window (MW) [15], just-in-time (JIT) [5], and time difference (TD) [16] approaches. The MW method is performed by collecting the latest or relatively long-term data, to rebuild the soft-sensor when a new data sample arrives. It handles the drifts of input variables effectively and provides better performance for gradual drifting processes, but it does not handle well the situations of process-state dependent nonlinearities and abrupt changes well [2]. Differently, JIT-based models collect the most similar data from the historical database and are able to adapt to slight shifting changes. However, the performance of models using the JIT method seems unsatisfactory in some situations because JIT does not take into account the potential associations among process variables. Additionally, both the JIT and MW methods require reconstructing the model very frequently. Differently, the TD method not only adapts to gradual drifts of both secondary variables and corresponding targets simultaneously, but is also not affected by online abnormal data. Because models using the TD method do not need to be reconstructed, they can avoid the highly frequent model updating issue. In addition, the low maintenance cost also adds more advantages for a TD model. Kaneko and Funatsu [17] compared and discussed the adaptive mechanisms, including MW, JIT, and TD models, in different scenarios of model degradations. However, multiple kinds of abnormal changes could be coupled and coexistent. Also, these adaptive strategies still have their own pros and cons. Any single adaptive method can fail to respond to the complex process situation effectively. To effectively deal with these issues, multiple adaptive mechanisms, such as MW-JIT [18], [19], TD-MW [7] and TD-MW-JIT [3], have been proposed to solve model degradations in industrial processes. Qi *et al.* [19] proposed MW-JIT-LSSVM for composition quality prediction in chemical distillation processes. Xiong *et al.* [7] proposed TD-MW-GPR soft-sensor to apply to a sulfur recovery unit and an industrial debutanizer column process. Yuan *et al.* [3] proposed a spatio-temporal adaptive soft-sensor modeling framework, which called TD-MW-JIT-LWPLS in sulfur recovery unit and blast furnace ironmaking process. These literatures [3], [7], [18]–[20] had demonstrated that the combination of these adaptive methods can effectively improve the prediction ability of the model in the face of model degradation. All the algorithms are only verified in some specific situations without a unified comparison. Kaneko and Funatsu [20] used a PLS model for constructing regression models to compare the adaptive strategies of TD, MW, JIT, TD-MW and TD-JIT. However, these adaptive mechanisms are rarely implemented to enhance multiple-output models and to make fair comparisons in the field of wastewater treatments plants. Due to the different kinds of characteristics in WWTP, such as the multivariable coupling, strong nonlinearity and time-varying features, an effective adaptive combination method to verify the applicability of the method under the characteristics of WWTP have not been proposed yet. Therefore, the adaptive multi-output soft-sensor was proposed to solve the problem combining the adaptive strategies according to different application scenarios in WWTP.

In this study, several novel adaptive multi-output soft-sensors are proposed and applied in wastewater treatments. The main contributions of this paper are as follows. (*i*) Adaptive strategies including JIT, TD and MW have been applied together with MPLS, MGPR and MRVM to form adaptive multi-output soft-sensors. (*ii*) We compared and discussed the potentials to propose new adaptive strategies, including TD-MW, TD-JIT, JIT-MW, and TD-JIT-MW, by fusing three basic adaptive strategies, with the aim of strengthening the pros and alleviating the cons of basic adaptive strategies. All the proposed adaptive strategies are coordinated with the multi-output regression models, MPLS, MGPR and MRVM methods to enhance the adaptive ability. (*iii*) Three case studies, including two simulated processes and field data datasets from WWTPs, are used to evaluate the effectiveness of the proposed soft-sensors.

Section II reviews the three basic adaptive methods and three multi-output models, including MPLS, MRVM and MGPR. In Section III, the framework of the proposed models and the structures of TD-MW, TD-JIT, JIT-MW, and TD-JIT-MW are elaborated. In Section IV, the proposed adaptive multi-output soft-sensors are applied and compared in the simulation data and practical industrial data, which respectively exhibit the stationary and non-stationary behaviors. The comparison and discussion of the proposed soft-sensors are elaborated in Section V. In Section VI, conclusions are made.

II. PRELIMINARIES

In this section, the methodology of adaptive algorithms is briefly illustrated, along with some descriptions and equations, including the moving windows, time difference, just-in-time learning methods and three multiple-output soft-sensors. More details can be found in [3], [8], [17], [18], [21].

A. ADAPTIVE LEARNING STRATEGIES

1) MOVING WINDOWS

The MW method is commonly suitable for fitting gradual drifts in a process. In the MW method, by adding the newest sample to the window and removing the oldest samples, a window of length L slides along the dataset at each step to select the new dataset. Then, the model can be reconstructed by using the new dataset of the window and updated to a new process state.

In the MW method, the selections of the window length L and step size are important procedures. Therefore, abundant features can be included to promote model construction by setting a suitable length of the window. In contrast, if the parameters are too large or too small, this can lead to degradation [14]. Generally, one sampling interval is preferred for the step size [3].

2) JUST-IN-TIME LEARNING

Just-in-time learning is also called lazy learning [22], or a local learning strategy. This method requires updating at each step to predict outputs. Since only the most similar samples are used to develop the online model in JITL, the local model is able to capture process dynamics. When a new query sample arrives, the desired output can be estimated by building a local model depending on the most relevant data in the database.

In the JIT technology, the Euclidean norm of the distance measurement is used to select the dataset corresponding to $\boldsymbol{x}(t_{new})$. The distance can be used to evaluate the similarity between $\boldsymbol{x}(t_{new})$ and $\boldsymbol{x}(t)$ can be calculated as follows

$$d_j = \left\| \boldsymbol{x}(t_{new}), \boldsymbol{x}_j \right\|^2, \quad j = 1, 2, \ldots, n \tag{1}$$

Then, the JIT local model is constructed using the relevant dataset depending on the similarity factor d_j.

3) TIME DIFFERENCE MODEL

There are two sets of n rows data, $\boldsymbol{x}(t) \in R^{n\times m}$ (with m columns of input variables) and $\boldsymbol{y}(t) \in R^{n\times l}$(with l columns of outputs), representing input and output variables at the time t, respectively. Typically, once new data $\boldsymbol{x}(t_{new}) \in R^{1\times m}$ is incoming, the trained model is used to predict the value of $\boldsymbol{y}(t_{new})$.

In TD modeling, we first calculate $\Delta\boldsymbol{x}(t)$ and $\Delta\boldsymbol{y}(t)$ as

$$\Delta\boldsymbol{x}(t) = \boldsymbol{x}(t) - \boldsymbol{x}(t-i) \tag{2}$$

$$\Delta\boldsymbol{y}(t) = \boldsymbol{y}(t) - \boldsymbol{y}(t-i) \tag{3}$$

Then, the relationship between $\Delta\boldsymbol{x}(t)$ and $\Delta\boldsymbol{y}(t)$ can be modeled using a regression method as follows

$$\Delta\boldsymbol{y}(t) = f(\Delta\boldsymbol{x}(t)) + \boldsymbol{e} \tag{4}$$

where $f(\cdot)$ is a model using the TD method and $\boldsymbol{e} \in R^{n\times 1}$ is the calculation error vector. Then, the model can be trained with sufficient training data. The time difference between inputs can be calculated with the new sample data $\boldsymbol{x}(t_{new})$ as follows

$$\Delta\boldsymbol{x}(t_{new}) = \boldsymbol{x}(t_{new}) - \boldsymbol{x}(t_{new}-1) \tag{5}$$

The output of the TD model can be predicted with the $\Delta\boldsymbol{x}(t_{new})$ as

$$\Delta\boldsymbol{y}(t_{new}) = f(\Delta\boldsymbol{x}(t_{new})) \tag{6}$$

Then, the predicted output is

$$y(t_{new}) = \Delta y(t_{new}) + y(t_{new}-1) \tag{7}$$

B. MULTI-OUTPUT MODELS

Since PLS is essentially a multi-input and multi-output modeling method. Therefore, the derivation of MPLS model can be seen from PLS [2]. The other two multi-output models are illustrated in this section. A set of n observations consisting of input-target vector pairs $\{\boldsymbol{x}_i, \boldsymbol{y}_i|_{i=1}^n\}$ exists, where $\boldsymbol{x} \in R^m$ is an input vector and $\boldsymbol{y} \in R^l$ is an output target vector. The input and output data matrices are denoted by $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]^T \in R^{n\times m}$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n]^T \in R^{n\times l}$, respectively.

1) MULTI-OUTPUT RELEVANT VECTOR MACHINE (MRVM)

Essentially, the RVM is a Bayesian regression framework model developed from the SVM [6]. It was originally only a single-output model. Since then, the RVM has been extended to multivariate outputs by Thayananthan [21] and has been formulated as a general multivariate regression tool. The dependence between input variables $\boldsymbol{x}$ and the target variables $\boldsymbol{y}$ can be formulated as follows:

$$\boldsymbol{y} = \boldsymbol{\Phi}(\boldsymbol{x})\boldsymbol{W} + \boldsymbol{\varepsilon} \tag{8}$$

where $\boldsymbol{\Phi}(\boldsymbol{x})$ is the $N \times (N+1)$ matrix with $\boldsymbol{\Phi}(\boldsymbol{x}) = [\phi(x_1), \phi(x_2), \ldots, \phi(x_N)]^T$, wherein $\phi(x_i) = [1, k(x_i, x_1), k(x_i, x_2), \ldots, k(x_i, x_N)]^{\mathrm{T}}$. $\boldsymbol{W}$ is the weight matrix $\boldsymbol{W} = (\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_l)^T$, which is dominated by a set of hyperparameters. $\boldsymbol{\varepsilon}$ is a vector of the sample noise and is assumed to be mean-zero Gaussian with a diagonal covariance matrix $\boldsymbol{S} = diag(\sigma_1^2, \ldots, \sigma_l^2)$. To guarantee the optimal fit of the parameters, $\boldsymbol{A} = diag(\alpha_0^{-2}, \alpha_1^{-2}, \ldots, \alpha_G^{-2})^T$ is defined as a diagonal matrix including all the hyperparameters $\boldsymbol{\alpha} = \{\alpha_r\}_{r=1}^G$.

The predictive distribution of every new input $\boldsymbol{x}_{new}$ can be calculated as follows:

$$p(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\alpha}^{opt}, (\sigma^{opt})^2) = \int p(\boldsymbol{y}^*|\boldsymbol{W}, (\sigma^{opt})^2) p(\boldsymbol{W}|\boldsymbol{y}, \boldsymbol{\alpha}^{opt}, (\sigma^{opt})^2) d\boldsymbol{W} \quad (9)$$

where the optimal values of the hyperparameters $\boldsymbol{\alpha}^{opt} = \{\boldsymbol{\alpha}_r^{opt}\}_{r=1}^G$ and noise parameters $\boldsymbol{\sigma}^{opt} = \{\boldsymbol{\sigma}_r^{opt}\}_{r=1}^l$ are estimated by maximizing the marginal likelihood. Since both terms of the integrand are Gaussian, the Eq. (9) can be calculated as

$$p(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\alpha}^{opt}, (\sigma^{opt})^2) = N(\boldsymbol{y}^*|\boldsymbol{z}^*, (\sigma^*)^2) \quad (10)$$

where $\boldsymbol{z}^* = [z_1^*, \ldots z_r^*, \ldots, z_l^*]^T$ is the predictive mean with $z_r^* = (\mu_r^{opt})^T \boldsymbol{\Phi}(\boldsymbol{x}^*)$ and $(\sigma^*)^2 = [(\sigma_1^*)^2, \ldots, (\sigma_r^*)^2, \ldots, (\sigma_l^*)^2]^T$ is the predictive variance with $(\sigma_r^*)^2 = (\sigma_r^{opt})^2 + \boldsymbol{\Phi}(\boldsymbol{x}^*)^T \boldsymbol{\Sigma}_r^{opt} \boldsymbol{\Phi}(\boldsymbol{x}^*)$. More details of the MRVM model can be found in [21] and [23].

2) MULTI-OUTPUT GAUSSIAN PROCESS REGRESSION (MGPR)

The multi-output Gaussian processes model has been used to develop linear and nonlinear regression models in dynamic processes. It induces the dependencies of latent variables in a highly correlated model to efficiently approach accurate GPR models [8].

The MGPR model predicts the outputs through a weighted combination of an individual latent function $\{f_h\}_{h=1}^H$ and a number O of shared latent functions $\{g\}_{j=1}^O$, which have the independent Gaussian process priors $g_j(x) \sim GP(0, k_j(\cdot,\cdot))$. As mentioned above, with the input $\boldsymbol{x}$ and noisy outputs $\boldsymbol{y}$, each single latent function of an output also has a GP prior like $f_h(x) \sim GP(0, k_h^f(\cdot,\cdot))$. To appoint these processes, a shared group of inducing variables b_j is used to induce each $g_j(x)$.

The collective variables are as follows: $g = \{g_j\}, f = \{f_h\}, b = \{b_j\}, o = \{o_j\}, B = \{B_j\}$, and $B^f = \{B_h^f\}$, where $g_j = \{g_j(x_i)\}$ and $f_h = \{f_h(x_i)\}$. The subscript $h = (1, \ldots, H)$ is the index of the outputs for each corresponding processes, $j = (1, \ldots, O)$ denotes the index of the shared latent processes, and $i = (1, \ldots, N)$ is the index of the inputs.

Given the predictive distribution the h-th output with the new sample data x^* as follows:

$$p(y_*|y, x^*) = N(y_*; \sum_{j=1}^{O} w_{hj}\mu_{j*} + \mu_{h*}^f, \sum_{j=1}^{O} w_{hj}^2 s_{j*} + s_{h*}^f) \quad (11)$$

where μ_{j*} and s_{j*} are the mean and variance of the prediction for $g_{j*} = g_j(x_*)$, i.e., $p(g_j|y, x^*) = N(g_{j*}; \mu_{j*}, s_{j*})$. Similarly, the mean and variance of the prediction for $f_{h*} = f_h(x_*)$ are μ_{h*}^f and s_{h*}^f, i.e., $p(f_{h*}|y, x_*) = N(f_{h*}; \mu_{j*}, s_{j*})$. Finally, the predictions of the mean and variances can be calculated as

$$\mu_*^h = \sum_{j=1}^{Q} w_{hj}\mu_{j*} + \mu_{h*}^f \quad (12)$$

$$s_*^h = \sum_{j=1}^{Q} (w_{hj}^2 s_{j*} + s_{h*}^f) \quad (13)$$

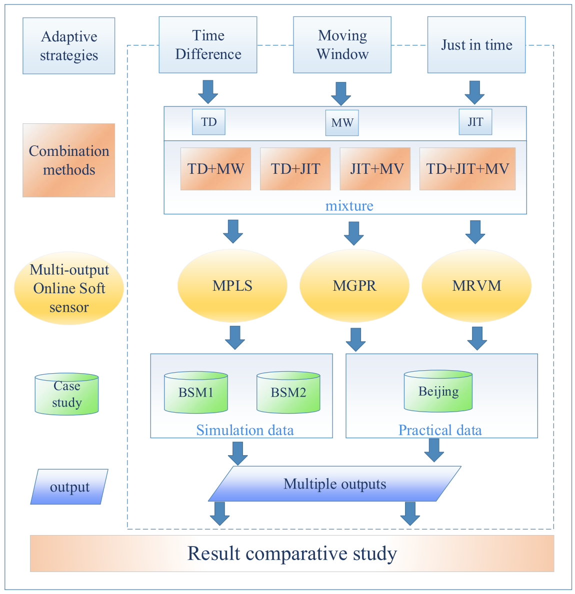

FIGURE 1. The framework of general technology.

III. FRAMEWORK OF ADAPTIVE MULTI-OUTPUT SOFT-SENSOR MODELING

To reduce the data gradual drifts or abrupt changes of the industrial processes, three adaptive strategies (TD, MW and JIT) are used to formulate the adaptive capacity of the MPLS model, MRVM and MGPR models. As shown in Fig. 1, the general technical framework is implemented as follows. (1) Three basic adaptive technologies including TD, MW and JIT, are mixed and formulated into four hybrid methods: TD + MW, TD + JIT, JIT + MW, and TD + JIT + MW, which are detailed in the subsequent sections. (2) These methods are combined with the three multi-output models to build online multi-output models. (3) The performances of all adaptive multi-output models are compared for both simulation data and practical industrial data.

A. TD-MW-MODELS

In this section, TD-MW models are mixed using two adaptive technologies, aiming to reinforce the reliability of soft-sensors to deal with the dynamics of processes. Fig. 2 displays the training and prediction process of the soft-sensor model based on the TD-MW method. The main steps of the modeling process are illustrated as follows:

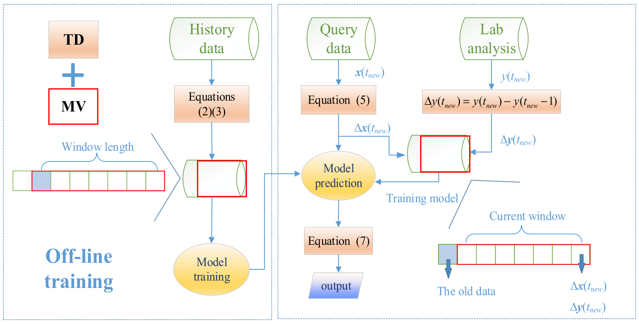

FIGURE 2. The schematic of the TD-MW model.

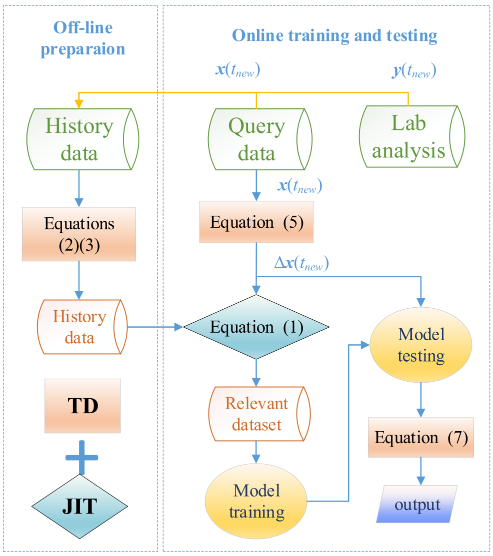

FIGURE 3. The schematic of the TD-JIT model.

a) Both input and output variables are converted into $\Delta\boldsymbol{x}(t)$ and $\Delta\boldsymbol{y}(t)$ using Eq. (2)-(3).
b) The model is then trained by an initial moving window, which contains L samples according to the TDs dataset.
c) Once a new sample arrives (query data), t can be calculated by formula (5). Then, the model can predict the time-differenced outputs $\Delta\hat{\boldsymbol{y}}(t_{new})$ by using Eq. (4). The predictive output $\hat{\boldsymbol{y}}(t_{new})$ can obtain by (7).
d) After the output $\boldsymbol{y}(t_{new})$ of lab analysis has obtained, $\Delta\boldsymbol{y}(t_{new})$ can be calculated by the TD process.
e) The model is rebuilt according to the updated window by including the newly sample $\Delta\boldsymbol{x}(t_{new})$ and $\Delta\boldsymbol{y}(t_{new})$, while dropping the oldest sample, and the process is then repeated (c) until the query data ends [20].

B. TD-JIT-MODELS

The TD-JIT model is to combine the TD and JIT techniques. Even though literature [20] has established this mixed model, it has not given a specific framework and process. Fig. 3 is represented the schematic of the TD-JIT model. As shown in Fig. 3, the historical dataset has been formed by TD processing. When a new sample comes, the historical

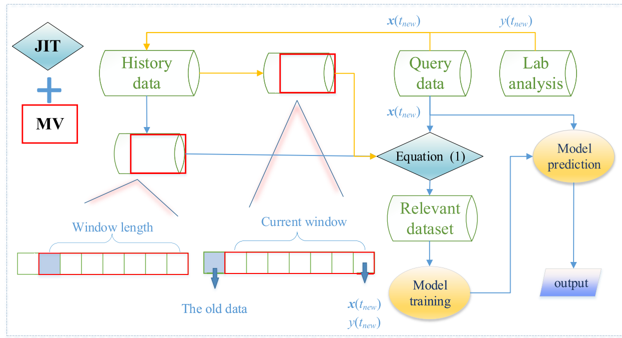

FIGURE 4. The schematic of the JIT-MW model.

data that are mostly similar to the query data put into a new database, which is used to train the model sequentially. We use the Euclidian distance as the similarity index (Eq. (1)). Then, the trained model uses the new data for prediction and derives the time-differenced outputs $\Delta\hat{\boldsymbol{y}}(t_{new})$. The prediction of the output $\hat{\boldsymbol{y}}(t_{new})$ can be estimated by Eq. (7). After the lab analysis output is obtained, the historical dataset can be updated with the new measurement data $\boldsymbol{x}(t_{new})$ and the new lab analysis data $\boldsymbol{y}(t_{new})$.

C. MW-JIT-MODELS

As increasing amounts of historical data contained in the dataset, it will become very time-consuming because the JIT method has to select a relevant dataset from the database based on predefined similarity criteria. In addition, it is well recognized that the JIT strategy is only effective for the operating conditions described by the current local query data. In this light, it is very inefficient if we construct a similar set by searching all the data from the database [19]. Therefore, a MW-JIT model structure can be built as seen in Fig. 4.

As shown in Fig. 4, first, the window length and the data size inside the window are set up. When a new sample arrives (query data), Eq. (1) is used to compare it with the current window data to obtain the relevant dataset. Then, the model uses the most similar data in the current window for training. The new data will be delivered to the trained model for prediction. After the lab analysis output is obtained, the window can be updated with $\Delta\boldsymbol{x}(t_{new})$ and $\Delta\boldsymbol{y}(t_{new})$, while dropping the oldest data. Then, the model can be rebuilt by using the most relevant data according to the updated window.

D. TD-JIT-MW-MODELS

As in the JIT-MW part of the TD-JIT-MW model in the previous section, the JIT-MW model performs the same function in the model, while TD processing is added to the JIT-MW model as the preprocessing procedure (Fig. 5).

Since JIT and MW are both online training models, there are few studies that combine the JIT and MW methods [18], [19]. At present, no one has combined these three adaptive strategies into a single model. This paper attempts to compare the predictive performances of the three combined models.

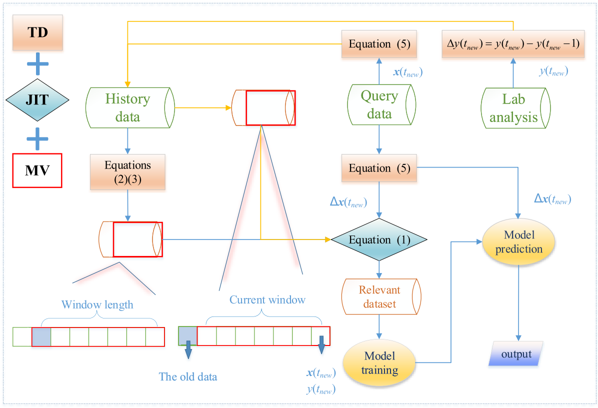

FIGURE 5. The schematic of the TD-JIT-MW model.

TABLE 1. The list of the compared models.

Multi-output models	Adaptive multi-output models
MPLS	TD-MPLS MV-MPLS JIT-MPLS TD-MW-MPLS TD-JIT-MPLS JIT-MW-MPLS TD-JIT-MW-MPLS
MRVM	TD- MRVM MV- MRVM JIT- MRVM TD-MW- MRVM TD-JIT- MRVM JIT-MW- MRVM TD-JIT-MW- MRVM
MGPR	TD- MGPR MV- MGPR JIT- MGPR TD-MW- MGPR TD-JIT- MGPR JIT-MW- MGPR TD-JIT-MW- MGPR

IV. CASE STUDY

The prediction performance is assessed in three cases studies consisting of two simulation studies and a practical industrial dataset. The comparative models are tabulated as shown in Table 1.

Two commonly used indexes, namely the root mean square error (RMSE) and the correlation coefficient (r), are used to measure the fit of each output. In addition, the root-mean sum of squares of the diagonal (RMSSD) and multiple correlation coefficient (MR) served as the criteria in multi-output models. The formulas of those evaluation criteria are shown in the Supplementary Material A.

The computer configuration is: OS - Windows 7 (64 bit), CPU - *i*5, RAM - 8G, and MATLAB 2012a.

A. THE SIMULATION DATA

Two simulation studies that named Benchmark Simulation Model No. 1 (BSM1) and Benchmark Simulation Model No. 2 (BSM2), respectively, are presented in this section to assess the performance of the proposed adaptive multi-output soft-sensors. The simulation platforms are developed by the first International Association Water Quality Task Group (http://www.benchmarkwwtp.org/) to offer an unbiased benchmarking wastewater treatment model for comparing the performances of different control strategies.

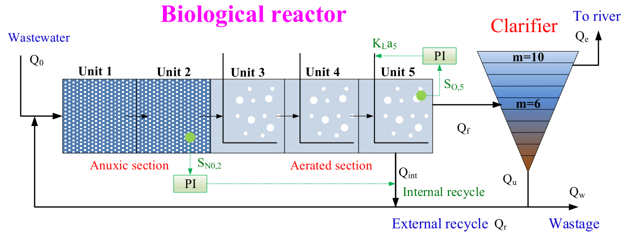

FIGURE 6. General overview of the BSM1.

1) BENCHMARK SIMULATION MODEL NO.1

The platform of the first case study is BSM1, which is a sophisticated dynamic mathematical model that can simulate the biological, chemical and physical processes of a WWTP. As shown in Fig.6, the platform consists of five compartment biological tanks (including two anoxic sections and three aerated sections) and a 10-layered secondary settler. The aim of this process is to remove organic matter and to perform nitrification and denitrification. It is designed for an average flow of 20,000 m^3 per day and an average COD concentration of 300 mg/L.

The purpose of this study is to establish adaptive multi-output soft-sensors to simultaneously estimate the COD, BOD_5, total nitrogen (TN) and total amount of solids (TSS), which are typically difficult-to-measure online and can indicate the performance of a WWTP. In BSM1, the simulated data covering 14 days and sampled every 15 minutes in stormy weather were chose. The selection of input variables is shown in Table S1 (Supplementary Material B). Half of the selected samples are used for training, and the remaining samples are used for online predictive testing. The influent flow rate, BOD_5, etc. under stormy weather conditions change dramatically [24]. Because the dramatic changes of the BOD_5, COD, etc. imposes an enormous challenge on the performance of an adaptive soft-sensor. We try to compare the performance of the multi-output models with different adaptive strategies under the same conditions. Therefore, the same Gaussian kernel function and the same kernel parameters are selected, even though the choice of the kernel function and the setting of the width parameter greatly affect the prediction performance of the MRVM model. The setting of these parameters is not necessarily optimal for every model, but these parameters are necessary to set up identically and optimally relatively if using the same models. In this light, it is fair to show the pros and cons of each method. The main parameters of the compared models are given in Table S2 (Supplementary Material B).

TABLE 2. The comparation of the RMSSD between the adaptive multi-output models in BSM1.

MPLS	RMSSD	MR	MRVM	RMSSD	MR	MGPR	RMSSD	MR
TD	0.4363	0.9979	TD-JIT	0.3336	0.9990	TD-JIT	0.4260	0.9981
TD-MW	0.5991	0.9994	TD	0.3339	0.9990	TD	0.4280	0.9981
TD-MW-JIT	0.7208	0.9952	TD-MW-JIT	0.4140	0.9985	TD-MW-JIT	0.9866	0.9909
TD-JIT	0.8276	0.9938	TD-MW	0.4152	0.9985	TD-MW	0.9869	0.9908
MW-JIT	1.5404	0.9835	MW-JIT	5.0354	0.7071	MW-JIT	2.0109	0.9642
MW	1.6393	0.9813	MW	5.8084	0.5438	MW	2.0193	0.9649
basic	6.6426	0.4756	JIT	7.3502	0.3286	basic	4.0511	0.8313
JIT	6.7317	0.4591	basic	7.4413	0.1769	JIT	7.0748	0.8313

The prediction performances of 24 models are displayed in Table 2. The results of four outputs are showed in the Table S3 (Supplementary Material B). It is obvious that all basic models (MPLS, MRVM and MGPR) achieved poor predictions in terms of the RMSSD. This was mainly because the basic models cannot adapt to the dramatic changes that occur in stormy weather, which results in performance degradation. Additionally, multivariate coupling adds more potentials for the prediction degradation. The basic MGPR produced better performance than the other two basic models, MPLS and MRVM, in terms of the RMSSD and MR. As tabulated in Table 2, the TD-JIT-MRVM model achieved the best performance with the RMSSD being 0.33. The RMSSD values of models with TD were significantly lower than those models without TD. Both the TD models and the TD-JIT models yielded a better prediction performance compared to other models. The TD-JIT models provided the best performance in the MRVM and MGPR. In contrast, the TD-MPLS achieved the best performance in the MPLS models. To further explain the performances of the proposed models, the predicted results of the BOD_5 and COD are profiled in Fig. 7, by comparing the MPLS, MRVM, and MGPR under seven adaptive strategies: TD, JIT, MW, TD-MW, TD-JIT, MW-JIT and TD-MW-JIT. As seen in Fig. 7, the models based on TD method fitted the real value effectively, illustrating that the TD method can enhance the prediction performance of multi-output models effectively. It is important to note that the JIT models achieved approximately the worst performance. At the same time, the MW-JIT models produced better performance than that of the JIT models, but the effect reflected in the peak is not as good as that produced by mixed TD models. The main factor that compromises the performance of the models using JIT and MW is the fact that the JIT models cannot capture the process dynamics effectively with uniform parameter settings under the dramatic changes of process variable in stormy weather.

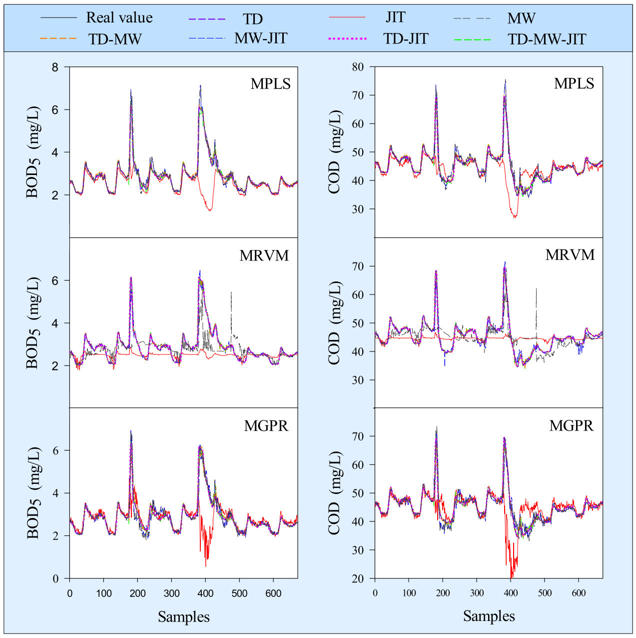

FIGURE 7. Predicted results of the MPLS, MRVM and MGPR with the four adaptive strategies for the BOD5 and COD in the BSM1.

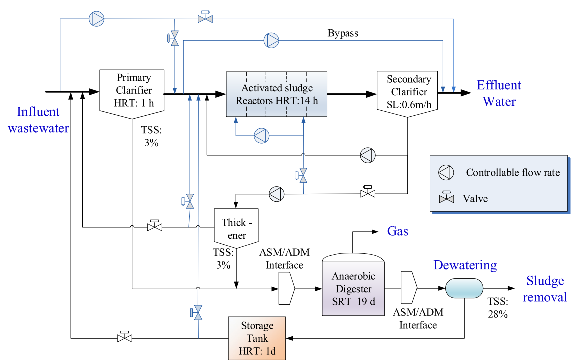

FIGURE 8. The General overview of the BSM2.

2) BENCHMARK SIMULATION MODEL NO.2

The BSM2 is the platform of the second case study, including the biological treatment processing of wastewater of BSM1 and the sludge treatment process. As shown in Fig. 8 [25], it not only includes a biological reactor and a secondary clarifier like BSM1 but also includes a primary clarifier, a secondary thickener for the sludge wasted from the BSM1 clarifier, a dewatering unit, and different possible control handles [26]. Unlike BSM1, BSM2 has influent dynamics covering 609 days, which include the effects of rainfall and temperature over the entire year. The first part of the influent data (254 days) has been used to attain a pseudo-steady state.

TABLE 3. The comparation of the RMSSD between the adaptive multi-output models in the BSM2.

MPLS	RMSSD	MR	MRVM	RMSSD	MR	MGPR	RMSSD	MR
TD-MW-JIT	0.0401	0.9999	TD	0.2070	0.9968	TD-JIT	0.1989	0.9971
TD-JIT	0.0550	0.9998	TD-JIT	0.2182	0.9967	TD	0.2040	0.9970
TD-MW	0.0750	0.9997	TD-MW	0.2384	0.9959	TD-MW-JIT	0.2078	0.9969
MW-JIT	0.1415	0.9991	TD-MW-JIT	0.2385	0.9959	TD-MW	0.2085	0.9968
MW	0.1415	0.9991	MW	2.5268	0.6402	MW-JIT	0.6665	0.9574
TD	0.2080	0.9968	MW-JIT	2.5354	0.6551	MW	0.7349	0.9512
JIT	1.9797	0.8786	basic	3.1340	0.3661	basic	2.2483	0.6853
basic	1.9840	0.8763	JIT	3.1374	0.3542	JIT	2.2589	0.7259

The last 355 days are used to evaluate the performance of the plant.

In this case study, the purpose is to compare the performances of adaptive multi-output soft-sensors when the sludge bulking has occurred. The phenomenon of sludge bulking is caused by the excessive proliferation of filamentous bacteria in sludge. It will lead to a decline of the effluent quality and endanger the operation of the whole biochemical system. To simulate sludge bulking in BSM2, the settling velocity in the layer should be reduced [27]. Therefore, we simulated the sludge bulking process by modifying the five settler parameters at the time of initializing the second settler, as shown in Table S4 (Supplementary Material B). The initialization file of the second settler is *settler1dinit_ bsm2.m*, which can be found in the benchmark files of BSM2. When the sludge bulking occurs, the inputs and outputs data drift slowly. Through this case study, we can compare the prediction performances of adaptive multi-output soft-sensors for the COD, BOD_5, TN and TSS. All input variables used for the model construction are sampled every 15 minutes and tabulated in Table S5 (Supplementary Material B). The parameters of the compared models are defined in Table S6 in Supplementary Material B.

The results of four outputs of all models are showed in the Table S7 (Supplementary Material B). The comparation of the RMSSD between the adaptive multi-output models are profiled in Table 3. As the table shown, all basic models produce the lowest predictive results in terms of the RMSSD. This is mainly because the general models cannot adapt to drifting changes between secondary and quality variables under the condition of sludge bulking which results in performance degradation. It is recognizable that the linear model MPLS has better performance than the other two nonlinear models in terms of the RMSSD and MR. This is mainly because that the MPLS can approach a moderate nonlinear process with local linearity. As shown in the table, the TD-JIT-MPLS model produces the best performance compared with all other models with the RMSSD being 0.0401. The RMSSD value of the model with TD is significantly lower than that of the models without TD in this case. The TD-JIT models yielded the best performance in the MRVM and MGPR. Differently, the TD-MW-JIT-MPLS achieved the best performance in the MPLS models. This demonstrates

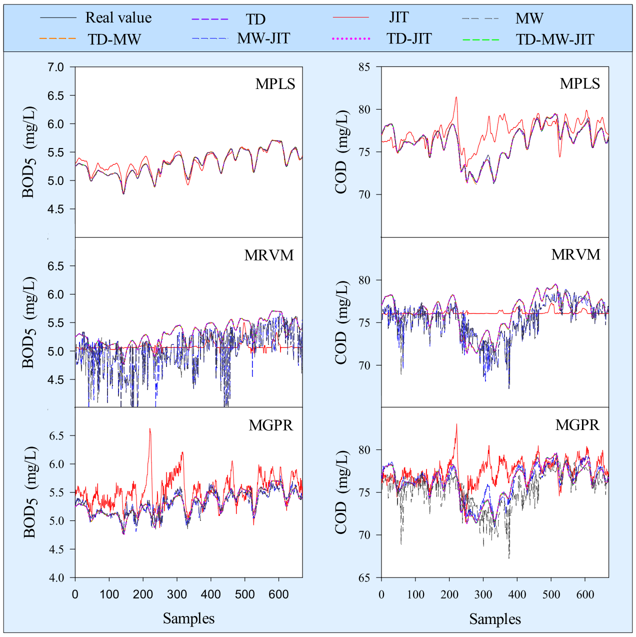

FIGURE 9. Predicted results of the MPLS, MRVM and MGPR with the four adaptive strategies for the BOD5 and COD in BSM2.

that the TD method not only can adapt to the drifts among input variables and corresponding variables with high stability but can also enhance the performance of multi-output models effectively. To better confirm the performance of the MPLS, MRVM, and MGPR under the four adaptive strategies, the predicted results of the BOD5 and COD are shown in Fig. 9. As displayed in Fig. 9, the MW-RVM and MW-JIT-RVM produced jagged curves due to the improper kernel parameters of the MRVM model. At the same time, we can also see that the JIT models achieved the worst predictive performance. The performance of the models with JIT method is not always satisfactory because the JIT model is more suitable for stable processes.

B. PRACTICAL INDUSTRIAL DATA

Different from the previous two simulation cases, a practical WWTP is considered in this section with a bulking sludge phenomenon. The final scenario is a full-scale WWTP in Beijing, China, which uses an oxidation ditch (OD) process

TABLE 4. The comparation of RMSSD between the adaptive multi-output models in Beijing.

MPLS	RMSSD	MR	MRVM	RMSSD	MR	MGPR	RMSSD	MR
TD-MW	7.6779	0.9854	TD	7.8369	0.9239	TD-MW	10.8320	0.8319
MW-JIT	10.8441	0.8693	TD-JIT	7.9815	0.9120	TD-JIT	10.8763	0.8148
MW	10.8512	0.8700	TD-MW	8.4871	0.9136	TD-MW-JIT	11.3895	0.7985
TD-JIT	12.8478	0.7964	TD-MW-JIT	8.7728	0.8925	MW	12.5043	0.8565
TD	12.8706	0.7927	MW	15.6913	0.7296	TD	12.6416	0.7912
TD-MW-JIT	13.4201	0.7820	MW-JIT	16.3283	0.7075	MW-JIT	12.7657	0.7985
basic	26.2710	0.8586	JIT	43.1288	0.4900	JIT	13.6737	0.7988
JIT	26.2925	0.8561	basic	44.5531	0.4285	basic	16.4879	0.8318

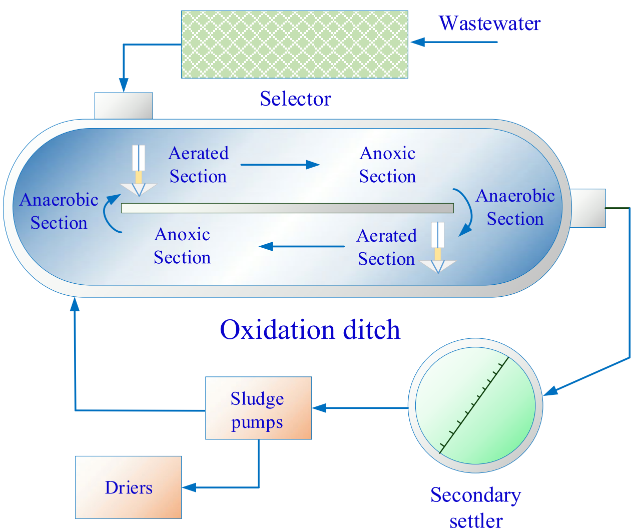

FIGURE 10. The wastewater plant for validation.

to treat municipal wastewater. Fig. 10 shows the structure of the OD process for the plants, and the more details can be seen in [8]. Filamentous bulking sludge was observed due to the low COD loading rate of the influent. In total, 212 data points are sampled from the process at one-day intervals. The first 112 of those data for samples are used for training, while the remaining data are used for testing to develop and validate the model in this study.

The purpose of this case study is to develop a multi-output soft-sensor to estimate the prediction for the COD, BOD_5, total phosphorus (TP) and SVI. The SVI is a difficult-to-measure variable and an empirical index used for characterizing the sludge bulking problem. Since the phenomenon of bulking sludge existed for around half a year in this plant, the input and output data collected from the field exhibits slow drifting, as seen in the second case, which affects the prediction performance among the adaptive multi-output soft sensors. For the MRVM models, "Gaussian" is selected as the kernel function, the setting of the width parameter is "5", and the maxIts is set to 100. The similarity factor of the JIT models d_j is set to 10. Other parameters are defined as the same as in the previous case.

The prediction performances of 24 models are shown in Table 4. The results of four outputs of all models are showed in the Table S8 (Supplementary Material B). Due to process change, it is obvious that all basic models produce relatively poor predictions in terms of the RMSSD. As in the case of BSM1, the basic MGPR yielded better performance than other two basic models in terms of the RMSSD and MR. As shown in Table 4, among all the models, the best prediction result is obtained from the TD-MW-MPLS model, with the RMSSD being 7.6779. In the MPLS and MGPR, the TD-MW models achieved the best performance, whereas the TD-JIT yielded the best performance when using the MRVM models. At the same time, we can also see that the JIT models achieved the lowest prediction performance among all models. The models using the TD method provide a relatively better prediction performance than either of the nonlinear models, MRVM and MGPR, without TD. In addition, the TD-MRVM can provide better performance improvement than the TD-MGPR. However, in the MPLS, the TD-MW yielded the best performance, whereas the MW-JIT and MW provided better performance than the other adaptive algorithms using TD models. To further explain this situation, the predicted results of the BOD5 and SVI are displayed in Fig. 11, with the comparison of the MPLS, MRVM, and MGPR under the seven adaptive strategies. For the TD-JIT-MPLS model, the prediction deviation occurs in the result of the BOD5. This is mainly due to the time delay caused by the correlation among the hard-to-measure variables. Therefore, the RMSSD of the TD-JIT-MPLS has been reduced. We can also see that the drifting pattern of the SVI is not tracked well by the JIT models.

V. COMPARISON AND DISCUSSION

This section aims to compare and discuss the specific characteristics and differences of the three cases (Table S9 in Supplementary Material) for different multi-output soft-sensors. In BSM1, due to the significant dynamics (stormy weather), the dramatic changes lead to a heavily nonlinear evolution of the output variables. The basic MGPR achieves the best performance in BSM1 compared to the remaining two basic models, MPLS and MRVM. This is mainly because the prediction distribution of the MGPR at a set of test points is simply assumed to be a multivariate Gaussian distribution, so the MGPR has better nonlinear predictive and adaptive abilities for abrupt changes in processes than the other two models. BSM2 is a simulated sludge bulking process

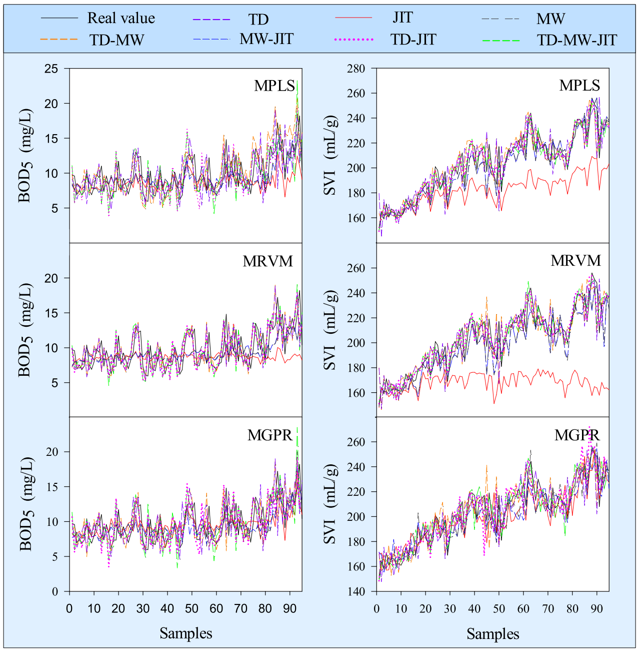

FIGURE 11. Predicted results of the MPLS, MRVM and MGPR with the four adaptive strategies for the BOD5 and SVI in Beijing.

achieved through the artificial changes of the sedimentation coefficient, which can cause the data to drift slowly. In the BSM2 case study, it can be found that although the corresponding variables gradually drift, the data exhibits a globally linear growth. As a linear model, the MPLS can provide better performance than the other two basic models in terms of the RMSSD by utilizing the linear correlation among the response variables. The linear correlation adds further sensitiveness to the predictive model. Therefore, the MPLS is suitable for slowly drifting scenarios. However, the basic MPLS may not be applicable when the data drift severely. In contrast, the MRVM, as a nonlinear model, yielded the worst performance in the simulated cases. This is most likely due to an improper kernel parameter. Different from the simulated sludge bulking, the scenario in the last case study is sludge bulking from a Beijing WWTP with a fault lasting for over half a year. Nonlinearity, heavily dynamics, significant uncertainty (including changes of the process input (raw) materials, wear of mechanical components, changes in the external environment, etc.) and multiple time scales add further complexity for modeling. At this time, the MGPR still achieved better performance. It is recognizable that the MGPR is used for modeling dynamic processes of both linear and nonlinear systems, and can handle cases of abrupt changes and gradual drift better.

In addition, the prediction results of the adaptive strategies for each multi-output model under the three cases are compared. As seen in Table S10, Table S11, and Table S12 in the Supplementary Material, the comparisons of the three cases of the MPLS, MRVM, and MGPR, are shown respectively. The performances of the seven adaptive strategies (including the TD, MW, JIT, TD-MW, TD-JIT, MW-JIT and TD-MW-JIT strategies) for the MPLS model and basic MPLS model under the three scenarios are displayed in Table S10.

JIT-MPLS yielded the lowest prediction performance in the three cases. At the same time, the JIT-MRVM and JIT-MGPR also produce the lowest predicted performances in all of the three cases. It is recognizable that the basic JIT strategy cannot adapt to the dynamic characteristics of abrupt changes and drifting processes effectively under the framework of multi-output models. In contrast, the models using the TD method have better performances than the models using the MW and JIT strategies for the simulation occasion. This is mainly because variables after TD processing not only can present a stable state but also do not affect the inherent distribution among the input and output. Meanwhile, the multi-output structure considers the correlations among the targets, which improves the prediction performance of models using the TD method. Therefore, the TD method is more suitable for multi-output models than the JIT method. On the other hand, the models using the TD method can be considered as a special type of MW-based model with the window length being 1. In addition, the default time lag is set as 1 without requiring any further parameters. However, some models using the TD method underperform for real cases. As shown in Table S10 and Table S12, the MW-MPLS and MW-MGPR achieve better performances than the TD-MPLS and TD-MGPR in the third case. This shows that the MW method can adapt to the drifting situation when coupled with the linear multi-output model better. Additionally, both MW and JIT are limited in the setting of the window size and similarity parameter; thus, parameters setting research is usually needed for special data to achieve the best performance.

The combination of TD with MW or TD with JIT can improve the adaptive ability of the basic MW and JIT strategies significantly, through the more stable data produced after TD processing. Although JIT-based models always produce the worst performance, after TD processing, the TD-JIT models provide very high-quality performance in dealing with data drifting and abrupt changes. As the results shown, both the TD-MW and TD-JIT mixed methods can adapt to data drifting and abrupt changes effectively. The difference between these two hybrid methods is that they are applicable to different multi-output models. For nonlinear models (MRVM and MGPR), the TD-JIT models are the best choice for both simulated data and actual data. Meanwhile, the TD-MW combined models are more practical for linear models such as the MPLS. However, since both MW and JIT are more suitable for gradual data drifting, the performances of models based on MW-JIT under the situations with abrupt changes are not as good as those of mixed TD models. The TD-MW-JIT models did not show an absolute excellent performance in all experiments, and only the TD-MW-JIT-MPLS obtained the optimal performance compared with other models in the BSM2 case. It can be found that "the more combinations of adaptive strategies, the better the performance" is a false cognizance.

In a summary, we can choose appropriate adaptive strategies and multi-output models according to different application scenarios through the following guidelines:

1) In the process with data drifting slowly, it is better to choose the linear model MPLS based on TD-MW method. TD-MW-MPLS can achieved the better performance than other two nonlinear multi-output models.
2) In the process with data abrupt changes, nonlinear multi-output models such as MRVM base on TD-JIT will be the better choice.
3) In the process with a mixture of data drifting and abrupt changes, it is necessary to justify that the drifting or abrupt changes govern the major patterns. Then, the proper combination of adaptive strategies and multi-output models can be chose accordingly.

Finally, it is important to notice that the multi-output model has more important requirements than the single-output model, considering not only the relationships between the features of input variables and targets but also the relationships among the outputs. However, it only applies the same sets of input features to predict all the responses simultaneously in this study. Therefore, all of the results are not satisfactory compared with those of single-output models or offline models. In further research, the interaction between variables will be taken into consideration, and the variables will be selected carefully before using multiple outputs prediction.

VI. CONCLUSION

In this paper, two bunches of adaptive multi-output soft sensors are proposed and applied in the wastewater treatments. Furthermore, the mixtures of different adaptive strategies, including the TD-MW, TD-JIT, JIT-MW, TD-JIT-MW with the three basic adaptive strategies, are coordinated with the multi-output regression models, the MPLS, MRVM, and MGPR methods to form 24 adaptive multi-output soft-sensors. By means of three case studies (two simulated processes and a field data sets from WWTPs), all the proposed soft-sensors are compared in terms of the RMSSD and MR. In conclusion, this paper provides a reference for applying different adaptive combinatorial multi-output models in different scenarios.

REFERENCES

[1] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, Apr. 2009.

[2] W. Shao, S. Chen, and C. J. Harris, "Adaptive soft sensor development for multi-output industrial processes based on selective ensemble learning," *IEEE Access*, vol. 6, pp. 55628–55642, 2018.

[3] X. Yuan, Z. Ge, and Z. Song, "Spatio-temporal adaptive soft sensor for nonlinear time-varying and variable drifting processes based on moving window LWPLS and time difference model," *Asia–Pacific J. Chem. Eng.*, vol. 11, no. 2, pp. 209–219, 2016.

[4] H. Kaneko and K. Funatsu, "Adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants," *Chemometr. Intell. Lab. Syst.*, vol. 137, pp. 57–66, Oct. 2014.

[5] Y. Liu, "Adaptive just-in-time and relevant vector machine based soft-sensors with adaptive differential evolution algorithms for parameter optimization," *Chem. Eng. Sci.*, vol. 172, pp. 571–584, Nov. 2017.

[6] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[7] W. Xiong, Y. Li, Y. Zhao, and B. Huang, "Adaptive soft sensor based on time difference Gaussian process regression with local time-delay reconstruction," *Chem. Eng. Res. Des.*, vol. 117, pp. 670–680, Jan. 2017.

[8] Y. Liu, Y. Pan, D. Huang, and Q. Wang, "Fault prognosis of filamentous sludge bulking using an enhanced multi-output Gaussian processes regression," *Control Eng. Pract.*, vol. 62, pp. 46–54, May 2017.

[9] H. Xiao, B. Bai, X. Li, J. Liu, Y. Liu, and D. Huang, "Interval multiple-output soft sensors development with capacity control for wastewater treatment applications: A comparative study," *Chemometrics Intell. Lab. Syst.*, vol. 184, pp. 82–93, Jan. 2019.

[10] H. Borchani, G. Varando, C. Bielza, and P. Larrañaga, "A survey on multi-output regression," *Data Mining Knowl. Discovery*, vol. 5, no. 5, pp. 216–233, 2015.

[11] Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017," *Can. J. Chem. Eng.*, vol. 96, pp. 2073–2085, Oct. 2018.

[12] Y. Liu, B. Liu, X. Zhao, and M. Xie, "Development of RVM-based multiple-output soft sensors with serial and parallel stacking strategies," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2727–2734, Nov. 2019.

[13] G. Leen, J. Peltonen, and S. Kaski, "Focused multi-task learning in a Gaussian process framework," *Mach. Learn.*, vol. 89, pp. 157–182, Oct. 2012.

[14] P. Kadlec, R. Grbić, and B. Gabrys, "Review of adaptation mechanisms for data-driven soft sensors," *Comput. Chem. Eng.*, vol. 35, no. 1, pp. 1–24, 2011.

[15] L. Yao and Z. Ge, "Moving window adaptive soft sensor for state shifting process based on weighted supervised latent factor analysis," *Control Eng. Pract.*, vol. 61, pp. 72–80, Apr. 2017.

[16] H. Kaneko and K. Funatsu, "Maintenance-free soft sensor models with time difference of process variables," *Chemometr. Intel. Lab. Syst.*, vol. 107, no. 2, pp. 312–317, 2011.

[17] H. Kaneko and K. Funatsu, "Classification of the degradation of soft sensor models and discussion on adaptive models," *AIChE J.*, vol. 59, no. 7, pp. 2339–2347, 2013.

[18] W. Xiong, W. Zhang, B. Xu, and B. Huang, "JITL based MWGPR soft sensor for multi-mode process with dual-updating strategy," *Comput. Chem. Eng.*, vol. 90, pp. 260–267, Jul. 2016.

[19] L. Qi, L. Xing, W. Liu, and B. Wei, "Adaptive soft sensor based on a moving window just-in-time learning LS-SVM for distillation processes," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 51–56, 2015.

[20] H. Kaneko and K. Funatsu, "Moving window and just-in-time soft sensor model based on time differences considering a small number of measurements," *Ind. Eng. Chem. Res.*, vol. 54, no. 2, pp. 700–704, 2015.

[21] A. Thayananthan, "Template-based pose estimation and tracking of 3D hand motion," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., 2006.

[22] G. Cybenko, "Just-in-time learning and estimation," *Nato ASI F, Comput. Syst. Sci.*, vol. 153, pp. 423–434, Jan. 1996.

[23] A. Thayananthan, R. Navaratnam, B. Stenger, P. H. Torr, and R. Cipolla, "Pose estimation and tracking using multivariate regression," *Pattern Recognit. Lett.*, vol. 29, pp. 1302–1310, Jul. 2008.

[24] Y. Qiu, Y. Liu, and D. Huang, "Date-driven soft-sensor design for biological wastewater treatment using deep neural networks and genetic algorithms," *J. Chem. Eng. Jpn.*, vol. 49, no. 10, pp. 925–936, 2016.

[25] U. Jeppsson, C. Rosen, J. Alex, J. Copp, K. V. Gernaey, P. A. Vanrolleghem, and M.-N. Pons, "Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs," *Water Sci. Technol.*, vol. 53, no. 1, pp. 287–295, 2006.

[26] M. Meneses, H. R. Concepcion, D. Vrecko, and R. Vilanova, "Life cycle assessment as an environmental evaluation tool for control strategies in wastewater treatment plants," *J. Cleaner Prod.*, vol. 107, pp. 653–661, Nov. 2015.

[27] D. Garcia-Alvarez, "Fault detection using principal component analysis (PCA) in a wastewater treatment plant (WWTP)," in *Proc. Int. Student's Sci. Conf.*, 2009, pp. 1–6.

JING WU was born in Songtao, Guizhou, China, in 1988. She received the B.E. degree in textile engineering from Qingdao University, Qingdao, in 2010, and the M.E. degree in computer application technology from Guizhou University, Guiyang, in 2013. She is currently pursuing the Ph.D. degree with the South China University of Technology. From 2013 to 2015, she was a Teaching Assistant with the College of Information Engineering, Guizhou Minzu University, Guiyang, where she has been a Lecturer with the College of Data Science and Information Engineering, since 2016. Her research interests include machine learning with industrial applications and data-driven soft sensors in wastewater treatment.

HONGCHAO CHENG was born in Xinyang, Henan, China. He received the B.E. degree in statistics from Kashi University, in 2016. He is currently pursuing the Ph.D. degree in control science and engineering with the Department of Automation, South China University of Technology, Guangzhou, China. His research interests include statistical theory research, fault diagnosis, soft sensors, process monitoring, and wastewater treatment.

YIQI LIU was born in Haikou, China, in 1983. He received the B.S. and M.S. degrees in control engineering from the Chemical University of Technology, Beijing, in 2009, and the Ph.D. degree in control engineering from the South China University of Technology, Guangzhou, China, in 2013. From 2013 to 2016, he was a Lecturer with the South China University of Technology, where he has been an Associate Professor with the Department of Automation, since 2016. He has authored more than 50 articles. He holds three patents. His research interests include soft sensors, fault diagnosis, and wastewater treatment. He is an IEEE Member and an Associate Editor of IEEE Access. He was a recipient of the Hong Kong Scholar Award, in 2016, the Chinese Scholarship Council Award, in 2011, and the Deutscher Akademischer Austausch Dienst Award, in 2015.

BIN LIU received the B.S. degree in automation from Zhejiang University, China, and the Ph.D. degree in industrial engineering from the City University of Hong Kong, Hong Kong. He was a Postdoctoral Fellow with the University of Waterloo, Canada. He is currently a Lecturer with the Department of Management Science, University of Strathclyde, Glasgow, U.K. His research interests include risk analysis, reliability and maintenance modeling, decision making under uncertainty, and data analysis.

DAOPING HUANG received the B.Eng. degree in chemical automation and instruments and the M.Eng. and Ph.D. degrees in automatic control theory and applications from the South China University of Technology (SCUT), Guangzhou, China, in 1982, 1986, and 1998, respectively.

In 1982, he started his teaching and research career with the Department of Automation, SCUT, where he is currently a Full Professor and the Vice Dean of the School of Automation Science and Engineering. From 1995 to 1996, he was with the University of Gent, Belgium, as a Government-Sponsored Visiting Scholar. He has authored three academic books and more than 150 conference and journal articles. His research interests include intelligent detection and control and soft-sensing technology as well as fault diagnosis, and accident prediction for industrial processes. He serves as the Vice Director of the Education Committee and a member of the Process Control and Application Committees at the Chinese Association of Automation. He has directed over ten research projects. He was granted the Third Prize from the National Education Committee, in 1992, and the Second Prize from the Guangdong Provincial Government, in 2005, for his contributions to science and technology.