Nonlinear Robust PLS Modeling of Wastewater Effluent Quality Indices

Lijie Zhao[1,2], Decheng Yuan[1]
Shenyang University of chemical technology / Provincial key laboratory of control technology of chemical industry process, Shenyang, P.R.China
Email:zlj_lunlun@163.com, yuandecheng@163.com

Jian Tang[2], Wei Wang[2], Tianyou Chai[2]
Northeastern University / Key Laboratory of Integrated Automation of Process Industry, Ministry of Education, Shenyang, P.R.China
Email: freeflytang@gmail.com, ww_wangwei@163.com, tychai@mail.neu.edu.cn

***Abstract*—To solve the strong nonlinearity and data deterioration due to missing, outliers contained in the training data, this paper combines robust EMPCA (Expectation Maximization Principle Component Analysis) and the error-based input weights updating NNPLS (Neural Network Partial Least Square) to build a nonlinear and robust model as a software sensor of effluent quality indices for the anoxic-aeration activated sludge with nitrogen removal process in wastewater treatment pants. As the first step, data preprocessing based on the modified robust EMPCA is used to eliminate gross error, estimate missing data. Then an error-based input weights updating NNPLS (EB-NNPLS) is further used to predict effluent quality indices. This study compares the performance of partial least squares (PLS) regression analysis, polynomial PLS, NNPLS and EB-NNPLS with robust nonlinearity for the prediction of effluent quality. Simulations results for industrial process data show that the proposed method outperforms basic PLS, the polynomial PLS and NNPLS for the prediction of effluent quality indices.**

***Index Terms*—Principal Component Analysis; Partial Least Square (PLS); Expectation Maximization (EM); neural netowrk Partial Least Square (NNPLS)**

I. INTRODUCTION

On-line reliable measurement of effluent quality indices is very important to monitoring, control and operational optimization in the wastewater treatment plants (WWTPs). However, due to big investment, poor reliability of continuous operation, and difficult maintenance, existing on-line hardware sensors in accuracy and reliability is not sufficient to measure the wastewater quality parameters.

Due to the complexity of the biological process, it is difficult to establish a mechanistic model of the activated sludge process. With the popular of DCS in wastewater treatment plants, a volume of measured process variables and laboratory analysis data are recorded. It is possible to forecast the effluent quality indices based on data-driven modeling [1].

Data-driven industrial process modeling methods depend on the reliable and accurate industrial filed data. However, these data not only contains missing data with high incidence, but also contains some outliers (gross errors) far from the typical ranges of the measured values due to the frequency of analysis, sensor failure, or operator mistake, etc. Outliers have a large influence on the regression because the residual magnifies the effects of these extreme data points. To minimize the influence of outliers, robust least squares were used. EM or MI is applied to estimate missing data. Víctor et al. [2] proposed filling in the missing points in the series with arbitrary values and then performing ML estimation of the ARIMA model with additive outliers.

In recent years, multivariate statistical methods are successfully applied to monitor and model the wastewater treatment process [3]. All the PCA and PLS algorithms are based on the assumptions that the data has not been spoiled by the outliers. In practice, real data often contain some outliers and usually they are not easy to be separated from the data set. Expectation maximum EM algorithm combined with PCA is a simple and effective method to treatment missing data [4]. However, the classic EMPCA algorithm is very sensitive to the outliers hidden in the data sets.

In practice, it is not suitable to use linear methods like PCA and PLS for moderate and strong non-linearity. Extra principal component will be retained to fit to the nonlinear relationships when linear PLS method deals with moderate and strong on-linear process, which will add the noise useless to the regression so that model generalization decreases and model performance deteriorates.

Many different nonlinear modeling techniques are used to solve the different types of nonlinear problems. A simple nonlinear PLS method is to use the original PLS algorithm for the augmented data matrix through transforming and containing the non-linear terms to the original data matrix. Its limitation is that non-linear terms

Manuscript received January 1, 2011; revised January 30, 2011; accepted February 20, 2011.

The work is supported by China's National 973 program (2009CB320600), NSF in China (61020106003 and 60874050) and China's Postdoctoral Science Foundation (20100471464).

doi:10.4304/jsw.6.6.1067-1074

among variable combination need a priori knowledge about process nonlinearity. It will lead to the number of nonlinear term too much, which causes input variable dimension excessive so as to compute complexity and explain difficulties. A number of methodologies have been proposed to integrate non-linear features within the linear PLS framework, resulting in the development of nonlinear PLS algorithms. Wold et al. introduced a spline-PLS algorithm where a spline function is used to fit the non-linear mapping between each pair of the latent variables [5]. Qin and McAvoy proposed a neural network algorithm, which uses a sigmoid activation function neural network to fit the inner mapping [6]. The main advantages of using the neural network PLS algorithm are that the neural networks PLS algorithm can handle variable correlations and the data set dimensionality. The drawbacks with building NNPLS and RBF-PLS models are the selection of the network structures and the model can easily be over-fitted.

Both polynomial PLS and NNPLS method are not updated input weights in the training process. Input weights needn't be updated only when mild nonlinear between input/output data. To overcome the problems associated with the updating of both the inner and outer weights, Wold et al. proposed solution is for the internal model using Newton -Raphson linearization, then worked out with the increment of weights of weights updated, which is applicable to any input/output latent variable relationship is continuous and differentiable [7]. Baffi et al. proposed an error based updating procedure, which resulted in an 'error-based' neural network PLS algorithm [8, 9]. In a similar manner to the original input weights updating procedure, the error based input weights updating procedure can be applied to any non-linear functional relationship between the input and the output latent variables, providing that it is continuous and differentiable with respect to the input weights *w*. Weights updated improved NNPLS internal map neural network modeling ability and model overall performance.

The prediction capacity of data-driven model strongly depends on the quality of the training data. It yields the very unreliable result when training data contains some outlier and missing data. In this paper, a nonlinear robust modeling method is proposed for predicting the effluent quality indices. Data analysis performed on a robust EMPCA model is used to eliminate gross error, estimate missing data. Then an error-based input weights updating NNPLS (EB-NNPLS) model is further used to predict effluent quality indices. Finally, a case study compares the performance of partial least squares (PLS), polyPLS, NNPLS, NNPLS with the robust nonlinear model based on the EMPCA and the EB-NNPLS for the prediction of effluent quality.

II. Materials And Methods

A. Description of Wastewater Treatment Process

The case study is a municipal wastewater treatment plant designed for 1060,000 (p. e.), located in shenyang, China. It aims at removal of pollutants in the wastewater. A schematic diagram of the process is shown in Fig 1.

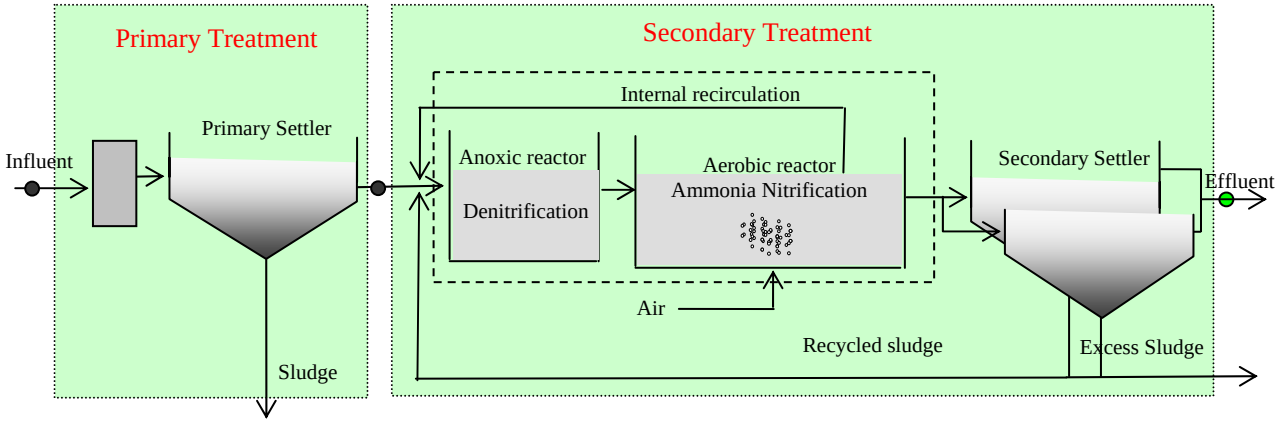


Figure 1 Schematic diagram of the activated sludge process

The plant provides primary and secondary treatment with daily capacity of 400,000 ton. It contains 6 water lines in parallel, where lines NO.1~3# are traditional activated sludge process, lines No.4~6# are anoxic-aeration (A/O) activated sludge process with nitrogen removal. The case study focuses on A/O activated sludge process, which consists of a pre-dentrification system with an anoxic reactor, an aerated reactor, and two secondary settlers. The data used in this study were collected from the A/O activated sludge process.

B. Data Reconciliation based on an Improved EMPCA

Poor-quality data has become a serous problem to the modeling, control and optimization in the wastewater treatment plants. The traditional PCA constructs the rank k subspace approximation to zero-mean training data that is optimal in a least-squares sense. The main disadvantage of least squares is its sensitivity to outliers. Outliers have a large influence on the regression because squaring the residuals magnifies the effects of these extreme data points. To minimize the influence of outliers, robust least squares were used. The mean value for the

corresponding variable is used as the initial value of missing data in EM PCA iterative solution. An improved robust EM PCA algorithm is proposed to solve the problems [10]. The structure of robust estimation with the missing data and outliers is depicted in Fig 2.

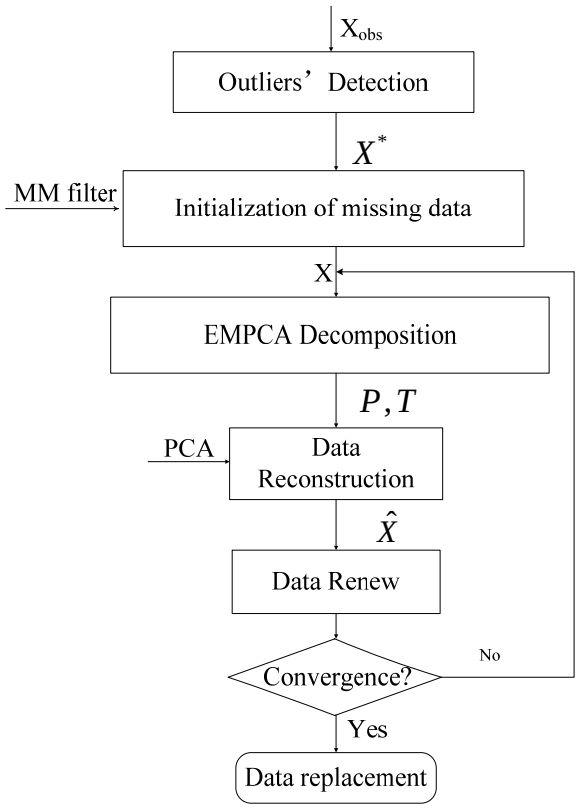


Figure 2. Structure of robust estimation algorithm

It consists of the following parts: the outliers' detection, initialization of missing data, PCA decomposition using EM algorithm and data reconstruction. In the outliers' detection part, the outliers' positions in the original data are firstly determined according to centre limit thermo. Then they are regarded as the missing points. Due to the dynamic nature of wastewater treatment, the moving median (MM) filter is used as the initial values of missing points. PCA decomposition is used to solve the scores and loading of the incomplete data using EM optimization algorithm. In reconstruction part, estimation of the missing data is given according to EMPCA model parameters scores vector T and loading vector P.

This can be summarized as follows:

Step 1: Determining outliers' positions according to center limit theorem and let the outliers missing points. Given data sample X_{obs}(m×n), m is the number of sample data, n is the number of variable.

$$\overline{X} - 3\delta \le X_{obs} \le \overline{X} + 3\delta \tag{1}$$

where $\overline{X}_j$ is median estimation of the observed sample and δ its median absolute deviation.

$$\overline{X}_j = median(X_{obs}(:, j)) = \begin{cases} \sum_I X_{obs}((i+1)/2, j), & i = 1,3,5,... \\ (\sum_I X_{obs}((i+1)/2, j) + \sum_I X_{obs}(i/2, j))/2, & i = 2,4,6,... \end{cases} \tag{2}$$

$$I = \{(i, j) : X_{obs}, 1 \le i \le m, 1 \le j \le n\}$$

$$\delta_j = 1.483 median(X_j - median(X_i)) \mid i = 1,\Lambda\ m,\ j = 1,\Lambda\ n \tag{3}$$

The observed variable $X_{obs}(i, j)$ at the ith sample is outlier when it excesses its confidence limit. The outliers are deleted, which can be seen as missing points. Therefore, the sample data only contains missing points, not contaminated by the outliers.

$$X*(i, j) = \begin{cases} X_{obs}(i, j), & \overline{X}_j - 3\delta_j \le X_{obs}(i, j) \le \overline{X}_j + 3\delta_j \\ \text{missing}, & \begin{cases} X_{obs}(i, j) \le \overline{X}_j - 3\delta_j, & or \quad X_{obs}(i, j) \ge \overline{X}_j + 3\delta_j \\ not \quad observed \end{cases} \end{cases} \tag{4}$$

Step 2: Replace the missing elements in data matrix X with their initial estimations using the Moving Median (MM) filter for the corresponding variables. The filter is described as follows:

$$X(i, j) = median(X^*(i, j - \omega/2),\Lambda\ , X^*(i, j, n + \omega/2)), \quad j = \omega/2 + 1, \omega/2 + 2,\Lambda\ , I_j - \omega/2 \tag{5}$$

where I_j, is the number of observations of the real signal $X^*(:, j)$, $w+1$ is the window width, and $X(i, j)$ is the output signal of the filter. MM filter is pronounced if the signal is contaminated with outliers.

Step 3: Perform PCA of the completed data set X using EM algorithm. Model parameters including score vector T and loading vector P are solved iteratively by E step and M step.

$$\text{E-step: } P = (T^T T)^{-1} T^T X \tag{6}$$

$$\text{M-step: } T = XP^T (P^T P)^{-1} \tag{7}$$

Step 4: Reconstruct the data matrix $\hat{X} = TP' + E$ with the predefined number of significant principal components.

Step 5: Replace the missing elements in the matrix X^* with their predicted values from $\hat{X}$.

Step 6: Repeat steps 2 to 5 till convergence.

It is important to emphasize that the values for the missing data are optimized in order to be further analyzed by PCA. The convergence *f* was calculated as:

$$f = \frac{\mathrm{SS}_{\mathrm{miss}}(\mathrm{r}) - \mathrm{SS}_{\mathrm{miss}}(\mathrm{r-1})}{\mathrm{SS}_{\mathrm{miss}}(\mathrm{r-1})} \tag{8}$$

where $\mathrm{SS}_{\mathrm{miss}}(\mathrm{r}) = \sum_{i=1}^{n} (X_{pi}^*)^2$ and T is the estimated value for the missing element X_{pi}^*. The EMPCA parameters fulfill the least square criterion since the

EMPCA approach minimizes the sum of the squared residuals:

$$\min SPE = \sum (X^*_{i,j} - X^*_{p_{i,j}})^2 \quad (9)$$

C. Error-based Input Weights Updating Neural Network PLS Algorithms

The neural network PLS algorithm keeps the robust generalization property by using linear PLS regression as of outer models and universal approximation capabilities by using neural networks as inner models. Outer PLS model decomposes the ($n\times M$) matrix of zero-mean variables $X(n\times M)$ and the ($n\times p$) matrix of zero-mean variables $Y(n\times p)$ into the form

$$\begin{cases} X = TP' + E \\ Y = UQ' + F \end{cases} \quad (10)$$

where the $\boldsymbol{T}$, $\boldsymbol{U}$ are ($n \times h$) matrices of the h extracted score vectors, the ($M \times h$) matrix $\boldsymbol{P}$ and the ($p \times h$) matrix $\boldsymbol{Q}$ represent matrices of loadings and the ($n \times M$) matrix $\boldsymbol{E}$ and the ($n \times p$) matrix $\boldsymbol{F}$ are the matrices of residuals. Neural network is used as inner regression model. Though there are some advantages of NNPLS, the agorithm doesn't update the outer input weights w, Since a inner nonlinear function impacts upon mappings of the inner and the outer model, NNPLS without updating the input weights may no longer be acceptable for highly nonlinearity.

The error-based input weights updating procedure assumes that the non-linear function between the input and the output latent variables is continuous and differentiable with respect to the input weights w and that the non-linear mapping can be approximated by means of a Newton–Raphson linearization of the non-linear function [9]. The input weights updating procedure is performed within the NIPALS algorithm and replaces the step for the calculation of the input weights w. A non-linear functional relationship $f(\cdot)$ between the input and the output latent variables *t* and *u*:

$$u = f(t) + e = f(X, w) + e \quad (11)$$

where $f(\cdot)$ stands for the nonlinear relation represented by a neural network, input score vectors t, output score vectors u and residual matrices e. $f(\cdot)$ is a continuous function differentiable with respect to w for each pair of latant variables. A two-layer feed-forward network is used to fitted the nonlinear relation with one centred sigmoidal activation function σ hidden layer with *NC* neurones and one linear activation function output layer. The non-linear inner relationship provided by the neural network can be written in explicit form as:

$$u = \omega_2 \cdot \sigma(\omega_1 \cdot t + \beta_1) + \beta_2 + e \quad (12)$$

Replacing t with $X \cdot w$, (12) becomes:

$$u = \omega_2 \cdot \sigma(\omega_1 \cdot (X \cdot w) + \beta_1) + \beta_2 + e \quad (13)$$

where X denotes the input data matrix when defining the first latent variable, or the deflated input data matrix when referring to subsequent latent variables. The above relationship can be approximated by means of Newton–Raphson linearization:

$$u = f_{00} + \frac{\partial f}{\partial w}\Big|_{00} \cdot \Delta w = f_{00} + \sum_{m=1}^{M} \frac{\partial f}{\partial w_m}\Big|_{00} \cdot \Delta w_m \quad (14)$$

Where:

$$f_{00} = \hat{u} = f(X, w, \omega_1, \beta_1, \omega_2, \beta_2) \quad (15)$$

$$\frac{\partial f(X \cdot w)}{\partial w_k} = \frac{\partial f(t)}{\partial w_k} = \frac{1}{2} \cdot \omega_1 \cdot \omega_2 \cdot (1 - \sigma^2(\omega_1 \cdot t + \beta_1)) \cdot x_k \quad (16)$$

$f_{00} = \hat{u} = f(t)$ is the prediction of the output latent variable. Δw is the column vector compring the finite increments Δw_m. The overall Newton–Raphson approximation can be written as:

$$u = f_{00} + \frac{1}{2} \cdot \omega_1 \cdot \omega_2 \cdot (1 - \sigma^2(\omega_1 \cdot t + \beta_1)) \cdot x_k \cdot \Delta w_k \quad (17)$$

A matrix Z in the error-based updating procedure is defined

$$Z = [Z_k] = \left[\frac{\partial f(t)}{\partial \omega_k}\right] = \left[\frac{1}{2}\omega_1 \cdot \omega_2 \cdot (1 - \sigma^2(\omega_1 \cdot t + \beta_1)) \cdot x_k\right] \quad (18)$$

Placing the updating parameters Δw_k in a column vector Δw, the Taylor series expansion can be written as:

$$u = f_{00} + Z \cdot \Delta w \quad (19)$$

The mismatch e between the output score vectors *u* and the nonlinear estimation by neural network model was given by:

$$e = u - \hat{u} = Z \cdot \Delta w \quad (20)$$

where:

$$u = Y \cdot q \quad (21)$$

$$\hat{u} = \omega_2 \cdot \sigma(\omega_1 \cdot t + \beta_1) + \beta_2 \quad (22)$$

New input weights *w* in the PLS outer mapping can be calculated by updating parameters Δw according to the matrix Z and mismatch e:

$$w = w + \Delta w \quad (23)$$

where :

$$\Delta w = (Z^T \cdot Z)^{-} \cdot Z^T \cdot e \quad (24)$$

III. Results And Discussion

A. Analysis of Historical Data

Wastewater treatment plants (WWTPs) are known to be highly non-linear systems subject to large perturbations in influent flow rate and pollutant load, together with uncertainties concerning the composition of the incoming wastewater. Reliable estimates of effluent quality are of great value for different operational tasks such as process monitoring, online simulation, and advanced control. The case study focuses on A/O activated sludge process.

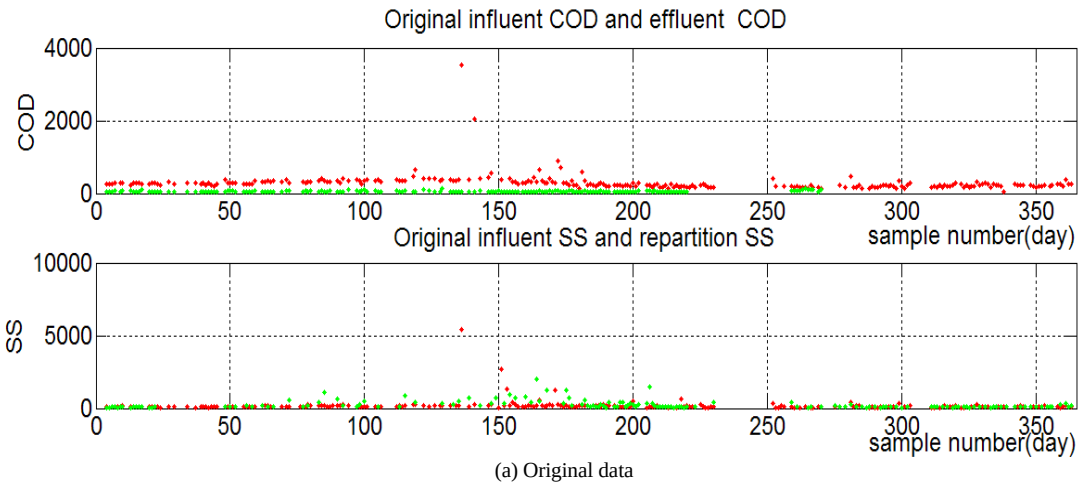
The following variables were sampled at the points from influent from sewage, primary clarifiers, bioreactor to effluent from secondary clarifiers. Basic statistical descriptors of the variables comprised in the database and its statistical analysis are shown in Table1. The database covers 365 consecutive days, each day as a sample.

TABLE I.
Basic Statistical Analysis And Measured Variables

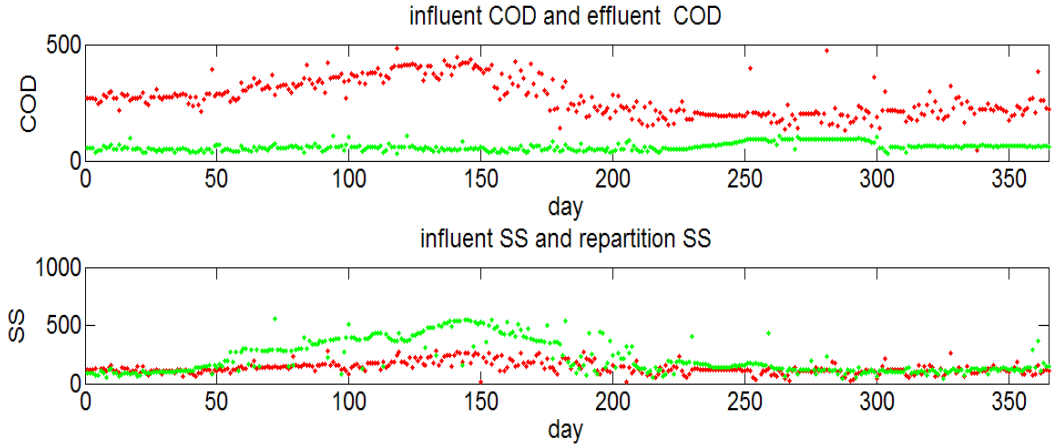
No.	Variable	Missing (%)	Outlier (%)	Mean	StDev	Unit
Influent from sewage						
1	Influent COD	16.9	0.5	269.84	78.61	mg/l
2	Influent SS	15.5	1.1	134.85	51.18	mg/l
3	Influent pH	23.4	0	7.25	0.09	-
4	ammonia nitrogen	34.4	2.5	33.01	7.23	mg/l
Primary clarifiers						
5	COD at repartition 2#	31.1	3.6	348.86	173.65	mg/l
6	SS at repartition 2#	29.8	3.6	219.46	141.51	mg/l
7	Flowrate	7.1	0.3	1641.50	880.59	m^3/d
Bioreactor						
8	Return sludge flow	7.1	0	3031.10	1117.60	m^3/d
9	ORP in Aeration tank	7.7	0.8	-122.92	145.37	mV
10	ORP in anoxic tank	8.1	1.6	-47.43	173.67	mV
11	Aeration flow	13.2	2.7	3233.00	2595.60	m^3/d
12	DO in aerobic tank	7.3	0.3	4.73	2.58	mg/l
13	DO in anoxic tank	8.2	0	3.64	3.14	mg/l
14	MLSS	17.7	0	6.18	2.32	g/l
15	Sludge volume	17.7	0	592.81	152.87	ml/l
16	SVI	17.7	0.3	96.34	23.77	-
17	pH	24.7	0	6.80	0.21	-
Effluent from secondary clarifiers						
18	BOD_5	39.3	4.4	14.47	5.29	mg/l
19	COD	27.1	3	60.08	16.50	mg/l
20	SS	27	1.4	15.03	8.68	mg/l
21	ammonia-nitrogen	40	4.1	24.48	5.98	mg/l

Table I lists on-line and off-line variables and their means and standard deviations, percent of missing and outliers. Partial original data with the high missing percent and a small amount of outliers, such as influent and effluent quality BOD_5, COD, SS et al. Data Analysis of missing elements and outliers on the real data set shows the original history data contains measurements of pollutants' concentration performed in different sampling sites over a certain period of time.

B. Data Processing



(a) Original data



(b) Rectificated data

Figure 3. Data rectification results comparisons

To preprocess the original data prior to this application, outliers are firstly detected according to center limit theorem in the incomplete database. Then missing data is initialized using MM filters. Robust expectation-maximization principal component analysis (EMPCA) decomposes the initialized data into scores (T) and loading (P) to reconstruct the missing data and outliers through $\hat{X} = TP'$. Comparisons between the original data and rectified data in influent SS and effluent SS, influent COD and repartition COD were shown in Fig. 3. Results of data rectification using robust EMPCA show the robust EMPCA can better detect outliers and estimate the missing data.

C. Soft Sensor of Effluent Quality

There are lots of factors, including inlet water quality and quantity, operating conditions, and external environment, that affect effluent quality in the wastewater treatment process. The secondary variables with the strongest relationships are chosen to establish a soft sensing model of effluent quality. The first 17 variables from No.1 to 17 were used as predictors X to explain response variable Y-block from No. 18 to 21. Original data set exists not only a few missing point, meanwhile some data points were measured or not recorded. History data is indived into two groups: 200 samples for model training and 165 samples for model test.

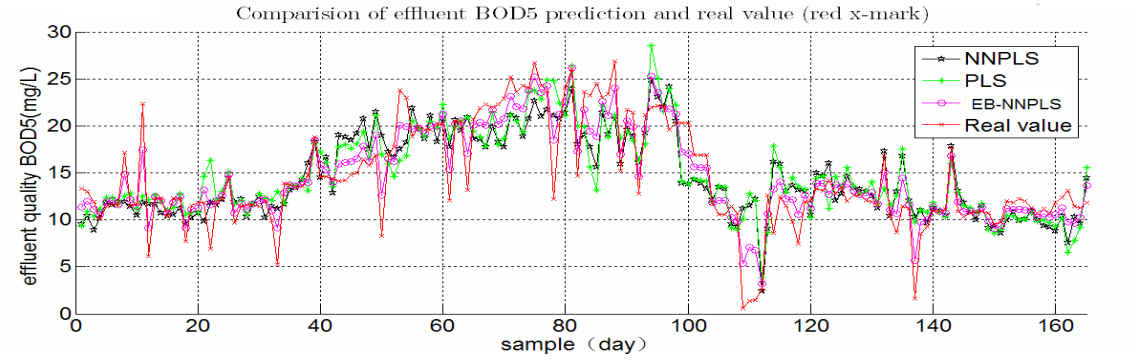
TABLE II.

CUMULATIVE VARIANCE FOR DIFFERENT PLS ALGORITHM

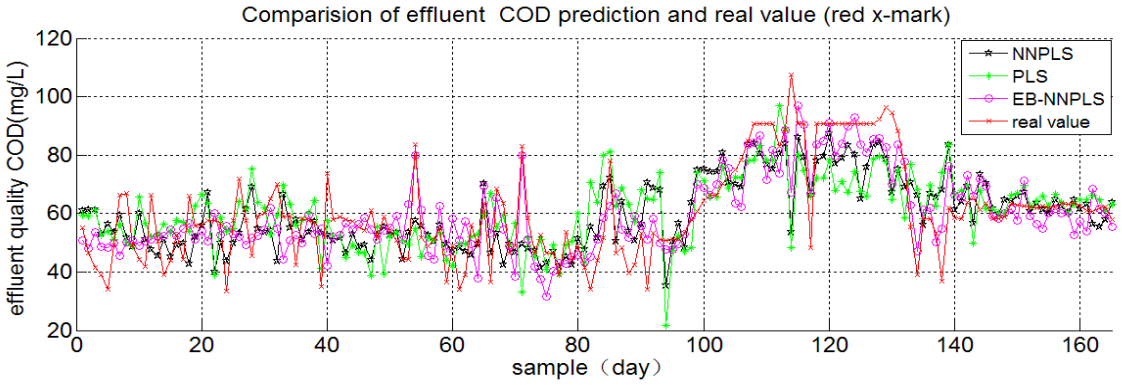
LV	*PLS*		*POLYPLS*		*NNPLS*		*EB-NNPLS*	
	X-Block	*Y-Block*	*X-Block*	*Y-Block*	*X-Block*	*Y-Block*	*X-Block*	*Y-Block*
1	12.67	17.53	36.57	14.76	36.57	14.95	27.39	45.4
2	34.92	25.46	58.69	26.95	58.69	27.12	43.64	55.79
3	59.71	27.74	66.11	31.7	66.13	31.84	52.61	61.1
4	69.68	29.8	70.77	36.36	70.2	38.16	65.02	63.24
5	75.32	31.09	77.82	39.09	77.78	40.75	71.01	64.51
6	80.01	31.57	81.57	40.98	81.53	42.84	74.3	65.94
7	83.24	31.93	85.84	42.00	85.82	43.84	77.72	66.33
8	85.39	32.05	89.21	43.13	89.2	45.39	81.18	67.26
9	87.05	32.13	91.45	43.93	91.46	46.07	85.21	68.27
10	88.74	32.22	93.14	44.37	92.86	46.64	87.73	68.34

In Table II the cumulative variances captured by the input and output variables for each model for COD prediction are given. Comparing the error-based NNPLS algorithm with the PLS, polynominal PLS and NNPLS, it can be observed that variance captured in Y-block for error-based NNPLS is more than the variance in Y-block of the other methods due to the input weight updating procedure when 10 latent variables are retained. NNPLS and EB-NNPLS is performed using conjugate gradient optimization procedure or optimization tool box procedure “leastsq”. The maximum number of sigmoids for each latent variable for inner neural nets models is set to 6.

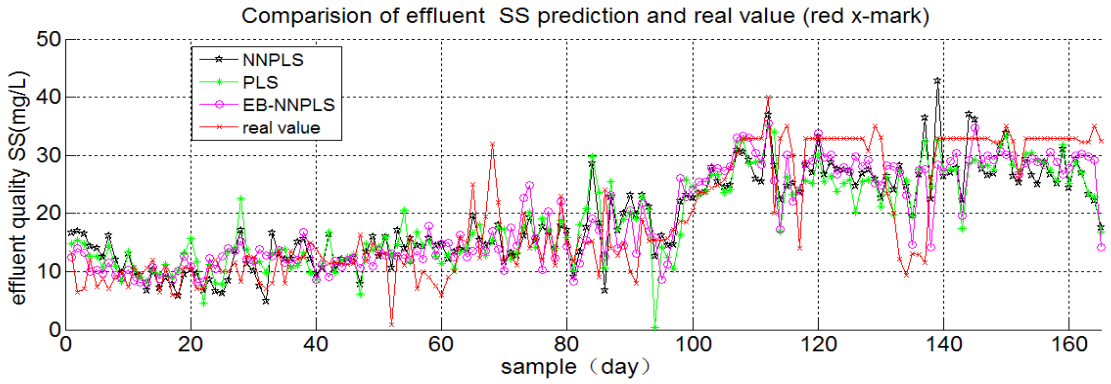
Fig. 4 shows comparison results of the predicted values and real values for EB-NNPLS, NNPLS and linear PLS model. It concludes that the prediction performance of nonlinear PLS model is better than prediction performance of linear model, nonlinear EB-NNPLS model better than NNPLS and polynomial PLS model.



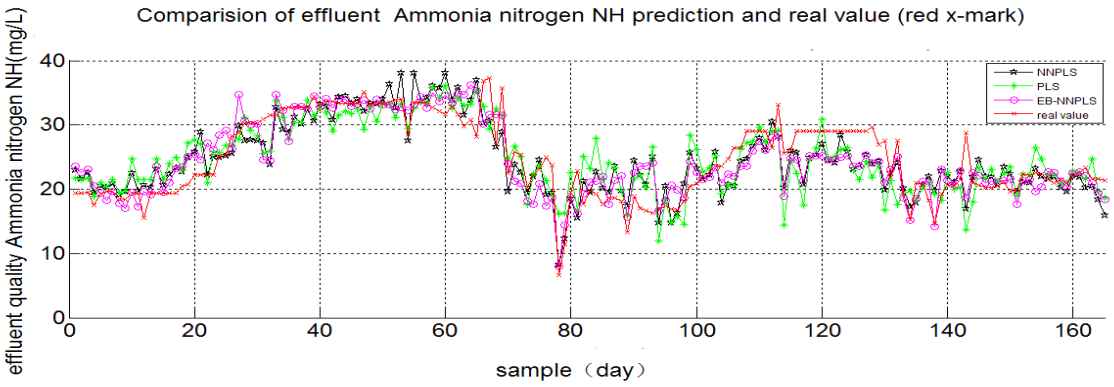
(a) BOD_5



(b) COD



(c) SS



(d) Ammonia nitrogen NH

Figure 4. Comparison results of the predicted values and real values

Root mean square error (RMSE) of EB-NNPLS is compare with RMSE of the linear, polynomial PLS, NNPLS for test data (Table III). It can be seen from Table III that the performances of nonlinear PLS outperform linear PLS and error-based input weight updating method is better than the methods with updating input weights.

TABLE III.

BASIC STATISTICAL ANALYSIS AND MEASURED VARIABLES

Methods / *RMSE*	*PLS*	*POLYPLS*	*NNPLS*	*EB-NNPLS*
BOD	3.80	3.54	3.16	2.94
COD	14.45	12.42	12.10	9.86
SS	6.74	6.31	5.72	5.56
Ammonia nitrogen NH	4.06	3.49	3.19	2.87

IV. CONCLUTIONS

The soft sensor technique is to solve prediction of some key variables. This paper deals with the development of software sensor techniques that estimate the effluent quality parameters from easy measurable secondary variables using robust EMPCA and error-based weight updating NNPLS. Based on the error input weights updating NNPLS method is to establish the effluent quality model for estimation of effluent quality. The nonlinear robust soft sensor technique proposed here may also be used for fault detection of processes, the estimation of toxicity, and automation of other wastewater treatment processes. Simulations for industrial process data show that the proposed method outperforms basic PLS, the polynomial PLS and NNPLS for the prediction of effluent quality.

REFERENCES

[1] Lee D S, Vanrolleghem P A. “Monitoring of a Sequencing Batch Reactor Using Adaptive Multiblock Principal Component Analysis,” *Biotechnology Bioeng*, 2003, 82(4):489–497.

[2] Víctor Gómez, Agustin Maravall, Daniel Pena, “Missing observations in ARIMA models:Skipping approach versus additive outlier approach,” Journal of Econometrics, vol.88, pp. 341-363, 1999.

[3] Wise B M, Gallagher N B. “The Process Chemometrics Approach to Process Monitoring and Fault Detection,” *J Proc Control*, 1996, 6:329-348.

[4] Dempster A. P., N. M. Laird, D. B. Rubin. (1977). “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society*, Series B, 34 1-38 .

[5] S. Wold, “Non-linear partial least squares modelling II. Spline inner relation,” *Chemical and Intelligent Laboratory Systems*, 1992(14):71-84.

[6] S.J. Qin, T.J. McAvoy. “Nonlinear PLS modeling using neural networks,” *Computers & Chemical Engineering*. 1992, 16:44, 379-391.

[7] S. Wold, Kettaneh-Wold, N., Skagerberg, B. “Non-linear PLS modelling,” *Chemometrics and International Laboratory Systems*, 1989,7: 53–65.

[8] G. Baffi, E. B. Martin , and A. J. Morris. “Non-linear projection to latent structures revisited: the quadratic PLS algorithm,” *Computers & Chemical Engineering*,1999,23:395-411.

[9] G. Baffi, E. B. Martin and A. J. Morris. “Non-linear projection to latent structures revisited: the neural network PLS algorithm,” *Computers & Chemical Engineering*, 1999, 23(9):1293-1307.

[10] L.J. Zhao, Tianyou Chai. “Wastewater BOD forecasting model for optimal operation using robust time-delay neural network,” *Letter Notes in Computer Science*, 2005, 3498:1028-1033.

Lijie Zhao was born in xingcheng, Liaoning Province, P.R. China in 1972. She received bachelor degree and MS degrees from Shenyang Institue of chemical technology in 1996 and 1999 respectively, and Ph.D. degree in control theory and control engineering speciality from Northeastern University in 2006.

Now she is a postdoctor in Northeastern University. Her current research interests include modeling, fault diagnosis and optimization of complex industrial processes.

Decheng Yuan received bachelor degree and MS degrees from Beijing Institute of Technology in 1982 and 1988 respectively, and Ph.D. degree from Shenyang Institue of Automation, Chinese Academy of Science in 2004. Now he is the vice-presedent of Shenyang University of Chemical Technology, and the professor and Ph.D. supervisor of Northeastern University. His research interests are modeling, monitoring and optimization of wastewater treatment.

Tang Jian received bachelor degree from Naval College of engineering in 1998 and MS degrees from Northeastern University in 2006 respectively. Now he is a Ph.D. candidate in Northeastern University. His current research interests are data driven soft sensor modeling.

Wei Wang was born in 1982. She is a Ph.D. candidate with Northeastern University, China. Her research interests include soft sensing and complex industrial process modeling.

Tianyou CHAI is Member of Chinese Academy of Engineering, IEEE Fellow，IFAC Fellow as well as Academician of International Eurasian Academy of Sciences. He received his Ph. D. degree in Control Theory and Engineering from Northeastern University in 1985 and became a Professor in 1988 in the same university.

He is the founder and Director of the Automation Research Center. Now He serves as head of Department of Information Sciences of National Natural Science Foundation of China (NSFC) since 2010. His research interests are adaptive control, intelligent control, and integrated automation of industrial processes.