2016 12th World Congress on Intelligent Control and Automation (WCICA)
June 12-15, 2016, Guilin, China

Estimation of Dissolved Oxygen via PLS and Neural Networks

Wei Wang, Changhui Deng and Jinyan Song

***Abstract*—Dissolved oxygen is one of the most important dependent factors in the process of aquaculture. In this article, a new on-line soft sensing method is proposed for estimation of dissolved oxygen in aquaculture. Based on the characters of process data, two advance techniques are used which is PLS (Partial Least Squares) and neural networks. With this method, real-time control and optimization can be realized in aquaculture. An industrial experimental study is described, and the results show that the proposed soft sensing algorithm is effective.**

I. Introduction

Dissolved oxygen (DO) in aquaculture is very important, and it effects fish growth. At present the analyzer of DO is very expensive and most of the farmers can't afford it or do not want to pay for it. Therefore, one kind of cost-effective measuring methods is needed, and soft sensing method is a good choice.

Aquaculture process is multivariable and nonlinear in nature, and the unknown process mechanism creates difficulties in modeling the concentrations of DO. Recently, data-driven models are used to describe the real conditions in many industry process. In the open literature, many methods have been adopted in aquaculture to measure the concentration of DO. These methods can be divided into two categories: One category is hybrid methods based on neural networks and optimal algorithm[1], [2], such as using MLP or L-M BP neural networks to establish the prediction model of water quality [3], [4]. GA and PSO are also used with neural networks in aquaculture [5], [6]. The second category is hybrid methods based on support vector machines and optimal algorithm, such as PSO and its improved algorithm [7]. However, they are both not adopted into practice due to the complexity and low accuracy.

Multivariate statistical theory is a common method for multivariable data analysis and treatment. One of the most popular techniques is PLS algorithm [8], and it emphasizes the explaining function of input to output when extracts the latent scores and removes the noises which have nothing effect for regression. According to the advantages of PLS method and neural networks, we proposed a new modeling method for the concentration of DO. It embeds the neural networks model into the regression framework of PLS for aquaculture process [9]. Based on the sampling data and compared with the other methods used before, the experimental results indicate that the proposed modeling method can measure the concentration of DO accurately.

This work was supported by National Natural Science Foundation of China (61503054, 51307012); Doctor Startup Foundation of Dalian Ocean University (SYYJ2012005); Marine Fisheries Department Research Programs of Liaoning Province (201512); Research Foundation of Education Department of Liaoning Province (L2015079).

Wei Wang was with the College of Information Engineering, Dalian Ocean University, Liaoning Province, 116023, China. and Key Laboratory of Marine Information Technology of Liaoning Province, 116023, Dalian, China. (e-mail: ww_wangwei@dlou.edu.cn).

Changhui Deng was with the College of Information Engineering, Dalian Ocean University, Liaoning Province, 116023, China. and Key Laboratory of Marine Information Technology of Liaoning Province, 116023, Dalian, China. (e-mail: chdeng9@dlou.edu.cn).

Jinyan Song was with the College of Information Engineering, Dalian Ocean University, Liaoning Province, 116023, China. and Key Laboratory of Marine Information Technology of Liaoning Province, 116023, Dalian, China. (e-mail: songjinyan @dlou.edu.cn).

II. Description of Dissolved Oxygen in Aquaculture

Water is the first condition of survival and growth for aquatic organisms. Water quality affects their growth and development directly. It is crucial to decide the breeding efficiency. There are multiple factors that affect water quality in certain aspects such as physical, biological and chemical, etc. [10-11], including: water temperature, DO, PH value, ammonia, nitrite and so on. Aquaculture animals need to absorb DO in the water, and DO has effect on water quality and sediment quality. It also decides the redox conditions. Thus, DO is the most important dependent factor for breeding biological survival. Real-time detection of DO is important for the aquaculture process and it is the key to control DO. With its real-time detection value, the DO control model and control algorithm could be realized.

With the improvement of industrialization and intensive aquaculture, to grasp the dynamic changes of DO and predict it in advance is an important issue to be solved. After understanding the dynamic rules of DO, we can adjust the control strategy in production by physical, chemical methods, even biological control methods can be used. Then we will create a suitable environment of good water quality for fish growth and culture fish scientifically.

III. Dissolved Oxygen Modeling Based on PLS and Neural Networks

A. *Modeling Strategy of dissolved oxygen*

This method is basically a combination of PLS method and neural networks model [9]. The input variables X and output variables Y are projected into the latent space to remove collinearity, then each latent variable pair is mapped with a neural networks model to capture the nonlinearity of the concentration of DO. The modeling strategy is shown in Fig.1.

In the aquaculture process, the input variable X is a combination of t_T, PH, c_N, c_{NO} and c_{NH}, where t_T is the water temperature, PH is the water PH value, c_N is the concentration of nitrite, c_{NO} is the concentration of ammonia, c_{NH} is the concentration of total nitrogen. The output variable Y is the

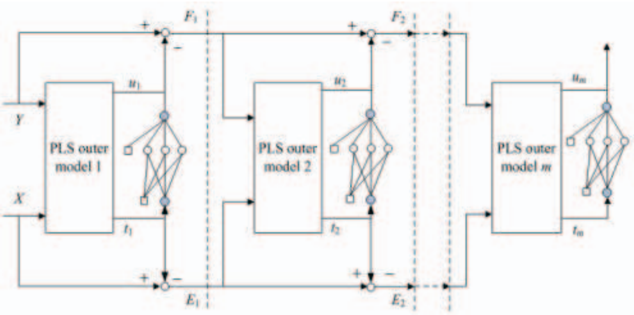

Fig.1 The schematic of neural networks and PLS modeling structure

concentration of DO.

B. Outer PLS model of dissolved oxygen

PLS is a method that projects the input matrix X (X=[t_T, PH, c_N, c_{NO}, c_{NH}])and output matrix Y (Y= [DO]) down into a latent space, extracting a number of principal factors (input score vector t and output score vector u) with orthogonal structure and capturing most of the variance in the original data. The detail of it is introduced as follows.

$$X = \sum_{h=1}^{m} t_h p_h^T + E \tag{1}$$

$$Y = \sum_{h=1}^{m} u_h q_h^T + F \tag{2}$$

where h = 1,2,...,m, and m is the number of latent variables, p and q are loading vectors, E and F are residual error matrices, Score vectors (t_h and u_h) of the same factor h are used to train the inner model. The original PLS algorithm was developed as a linear regression method that uses a linear inner relation on the latent space. Then various nonlinear PLS algorithms have been proposed to cope with the problems introduced by nonlinearity. Here we use neural networks model.

C. Inner neural networks model of dissolved oxygen

Neural networks is a type of commonly used nonlinear model, and has been successfully used to model a class of nonlinear systems. For each factor h of inner neural networks model, t_h is the input signal, u_h is the output variable. It obeys the following relation:

$$u_h = f_h(t_h) \tag{3}$$

where $f_h(.)$ stands for the inner neural networks model.

D. Modeling Algorithm of dissolved oxygen

The modeling algorithm of DO can be formulated as follows:

Step 1: Scale input variable X=[t_T, PH, c_N, c_{NO}, c_{NH}] and output variable Y= [DO] to have zero mean and unit variance. Let $E_0 = X$, $F_0 = Y$ and h =1.

Step 2: Call PLS algorithm, extract the score vectors t_h , u_h for the input and output variables of the inner neural networks model.

Step 3: Build the inner neural networks model f_h.

Step 4: Calculate the loadings of X and Y.

$$\begin{aligned} p_h^T &= t_h^T E_{h-1} / t_h^T t_h \\ q_h^T &= \hat{u}_h^T F_{h-1} / \hat{u}_h^T \hat{u}_h \end{aligned} \tag{4}$$

Step 5: Calculate the residuals for factor h.

$$\begin{aligned} E_h &= E_{h-1} - t_h p_h^T \\ F_h &= F_{h-1} - \hat{u}_h q_h^T \end{aligned} \tag{5}$$

Step 6: Let h = h +1 , then return to step 2 until all m principal factors are calculated and there is no useful information in matrix E_m and F_m. To determine the optimal number of latent variables, cross-validation is used here.

IV. INDUSTRY EXPERIMENTS

A. Data sampling

In order to collect data, a measurement system is developed. The system is mainly to control the circulation of water, and monitoring the water quality parameters. The system uses liquid level sensor, pH sensor, conductivity sensor and ammonia nitrogen and temperature integrated sensor to collect data. 20 groups data are sampled, and 12 groups are used to model and 8 groups are used to predict.

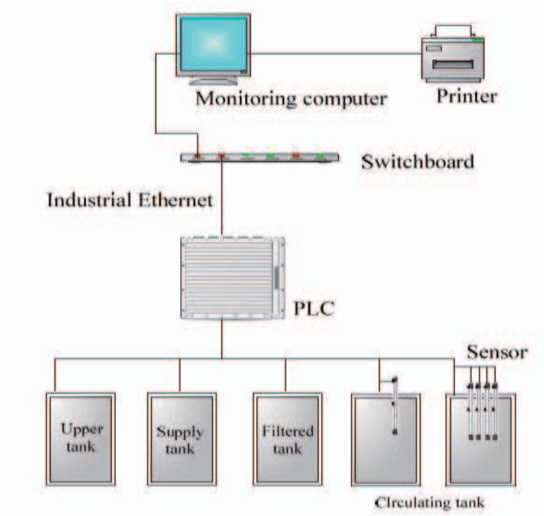

Fig. 2 Intensive aquaculture system.

Partial of the sampling data are listed in Table I.

TABLE I
PARTIAL OF ORIGINAL DATA

Variables	1	2	…	19	20
t_T(°C)	19.4	19.5	…	19.4	19.0
PH	8.01	8.01	…	7.99	8.13
c_N (mg/l)	0.0021	0.0028	…	0.0182	0.0122
c_{NO}(mg/l)	0.1002	0.1003	…	0.941	0.401
c_{NH}(mg/l)	0.0612	0.0711	…	0.2390	0.2336
DO(mg/l)	8.0198	8.0192	…	7.6883	8.1008

B. Modeling and predicting

There are five input variables t_T, PH, c_N, c_{NO}, c_{NH} and one output variable DO in the sampling data set. According to the proposed modeling steps of DO, the percent variance calculated by cross validation are shown in Table II. From the results, we extract two score vectors from the inputs.

TABLE II
PERCENT VARIANCE CAPTURED BY NNPLS MODEL

	X-Block		Y-Block	
LV #	This LV	Total	This LV	Total
1	59.96	59.96	98.59	98.59
2	24.60	84.56	0.77	99.36
3	11.41	95.97	0.02	99.38
4	2.72	98.69	0.06	99.44
5	1.31	100	0.04	99.48

Based on the results of cross validation , the relationship of input and output with DO soft sensing model are shown in Fig3.

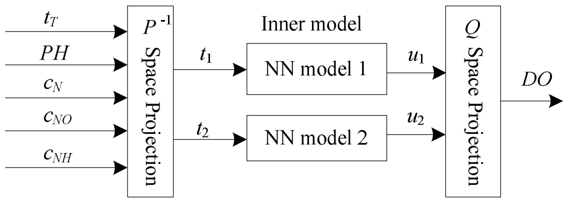

Fig.3 The structure of NNPLS modeling for dissolved oxygen

The specific formula of this model is as follows:

$$\begin{cases} Y = U\boldsymbol{Q}^T \\ u_h = f_h(t_h) \\ t_h = X(P_h^T)^{-1} \end{cases} \tag{6}$$

where $U = [u_1, u_2]$, $Q = [q_1^T, q_2^T]$, t_h and u_h are the input and output variables of *ith* (i=1,2) inner model. At last, the well trained model is used to predict the concentration of DO, and the overall output is calculated by:

$$\hat{Y} = \hat{\boldsymbol{U}}\boldsymbol{Q}^T \tag{7}$$

where $\hat{U} = [\hat{u}_1, \hat{u}_2]$, $Q = [q_1^T, q_2^T]$.

The training and predicting results are shown in Fig.4 and Fig.5. From the comparison of the curves, we see the accuracy of the proposed model is relatively high, and good effect is achieved by using outer PLS model.

The performance of each model is evaluated in terms of the root-mean-square-error criterion, and the performance indexes is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

where $\hat{y}$ and y are values of prediction and measurement respectively. When using this formula for prediction, n is 8. The auto-correlation function of errors is shown in Fig.6.

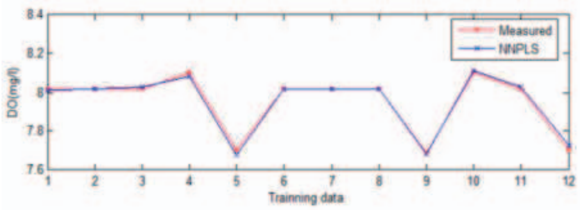

Fig.4 The trainning results of DO by NNPLS method

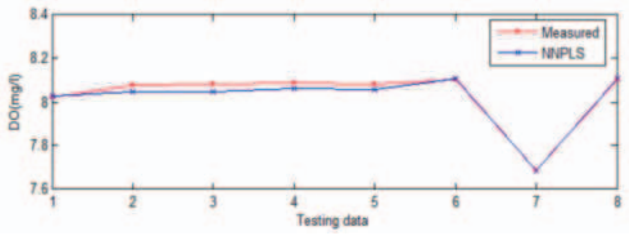

Fig.5 The testing results of DO by NNPLS method.

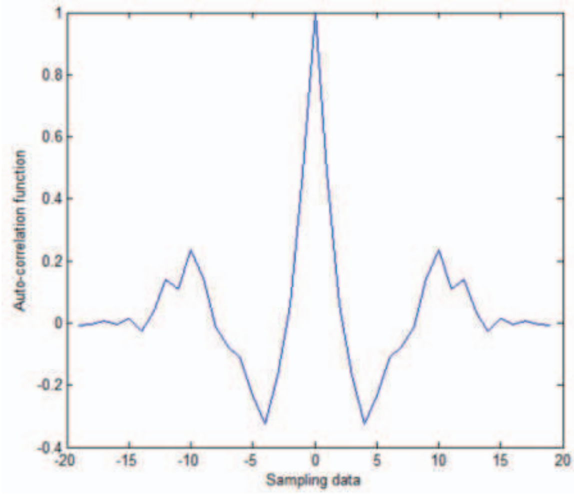

Fig. 6 Auto-correlation function of errors

Compare the proposed method with other methods used for DO [12], the predict results are shown in TABLE III.

TABLE III
ACCURACY COMPARISON OF DIFFERENT METHODS

Method	PLS	BP	ELM	NNPLS
RMSE	0.0795	0.2884	0.2879	0.0211

From the comparison of the results, we can see our method can correctly predict the concentrations of DO in aquaculture. It is also a guideline for process engineers to arrive at the control and optimum operational strategy.

V. CONCLUSION

According to the nonlinear and dynamic characters in the aquaculture process, we formulate a new modeling approach for estimation the concentration of DO. It is based upon the integration of neural networks model within the PLS framework. The proposed method uses a PLS framework to solve the problem of high dimensionality and collinearity, and the inner neural networks model is used to approximate the nonlinearity. The industrial experiment results indicate that the proposed method gives a better prediction performance and future research would address using the developed model for aquaculture process control and optimization.

REFERENCES

[1] Bernhard H. Schmid, M.Asce, Jari Koskiaho, "Artificial Neural Network Modeling of Dissolved Oxygen in a Wetland Pond: The Case of Hovi, Finland," *Journal of Hydrologic Engineering*, 2006, 11(2): 188-192.

[2] Zhao Ying, Nan Jun, Cui Fuyi, etal. "Water quality forecast through application of BP neural network at Yuqiao reservoir," *Journal of Zhejiang University Science A*, 2007, 8(9): 1482-1487.

[3] X.Y. Miao, C.H. Deng, X.J. Li, Y.P. Gao, D.G. He, "A Hybrid Neural Network and Genetic Algorithm Model for Predicting Dissolved Oxygen in an Aquaculture Pond," *2010 International Conference on Web Information Systems and Mining*, 2010, pp. 520-531.

[4] Wang Ruimei, Fu Zetian, He Yyouyuan, "Prediction model of dissolved oxygen fuzzy system in aquaculture pond based on neural network," *Agricultural Science and Technology*, 2010, 11(8): 14-18.

[5] C.H. Deng, X.J. Wei, L.X. Guo, "Application of Neural Network Based on PSO Algorithm in Prediction Model for Dissolved Oxygen in Fishpond," *Proceedings of the 6th World Congress on Intelligent Control and Automation*, pp. 9401-9405, June 2006.

[6] Hu Xuemei, Hu Yingzhan, Yu Xingzhi, "The Soft Measure Model of Dissolved Oxygen Based on RBF Network in Ponds," *The Fourth International Conference on Information and Computing*, 2011: 38-41.

[7] S.Y. Liu, H.J. Tai, Q.S.Ding, D.L. Li, L.Q. Xu, Y.G. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction," *Mathematical and Computer Modelling*, Volume 58, Issues 3-4, August 2013, pp. 458-465.

[8] Svante Wold, Michael Sjostrom and Lennart Eriksson. "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, pp. 109-130, 2001.

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

[10] Michael L. Cuenco, Robert R. Sticheney and William E. Grant, "Fish bioenergetics and growth in aquaculture ponds: II. Effects of interactions among, size, temperature, dissolved oxygen, unionized ammonia and food on growth of individual fish," *Ecological Modelling*, 1985, 27: 191-206.

[11] Edwin J. Niklitschek, David H. Secor, "Dissolved oxygen, temperature and salinity effects on the ecophysiology and survival of juvenile Atlantic sturgeon in estuarine waters: II. Model development and testing," *Journal of Experimental Marine Biology and Ecology*, 2009, 381: S161–S172.

[12] Wei Wang, Changhui Deng, Xiangjun Li, "Soft Sensing of Dissolved Oxygen in Fishpond via Extreme Learning Machine," *The 11th World Congress on Intelligent Control and Automation*, Shenyang, China, June 27-30, 3391-3393.