International Journal of Database Theory and Application
Vol.9, No.7 (2016), pp.157-168
http://dx.doi.org/10.14257/ijdta.2016.9.7.14

Dynamics System Analysis and Intelligent Identification of Aquaculture Water Quality Data

Xiaofei Huang[1, 2] and Wei Wu[2,3 *]

[1]*Hainan College of Software Technology, Qionghai 571400, China*
[2]*College of Information Science and Technology, Hainan University, China*
[3]*Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China*
*xiaofeihr@126.com, *wuwei@sidsse.ac.cn*

Abstract

Data analysis on environmental factors is crucial to aquaculture. Several significant parameters (temperature, pH and dissolved oxygen) and factors related to it are discussed in this paper. Data preparation including fixing some missing data and inaccurate records by non-linear function approaching in the first sampling process. Then the utilization of deterministic tracking theory is adopted for dynamic system analysis. Furthermore, time series analysis from different water layers based on this theory suits the real environment well, Radial Basis Function Neural Networks is well applied in tracking the parameters trend both globally and locally. The results provide effective references for systemic data analysis and control engineering.

Keywords*: aquaculture water quality; dynamic analysis; nonlinear fitting; artificial neural network identification*

1. Introduction

The rapid development of aquaculture contribute to human food supply. As a big investor of this industry, China accounts for 70% productions in the world approximately. However, more than 150 billion RMB annual economic losses is caused by aquatic deterioration, which encourages us to carry out related theoretical research and engineering application for this issue by monitor and regulation.

Many inspiring results have been achieved by researchers and published to public. The prediction and analysis methods are vary in different real circumstances. It is widely accepted that non-linear modeling receives better results than linear processing [1-2]. Researchers introduce Support Machine Vector (SVM) methods [3-4] to approximate to real characters, improved algorithms such as real-value genetic algorithm support vector regression method [5] was proposed to searches for the optimal parameters and application. From the perspective of Artificial Intelligence (AI), neural networks also indicate worthy prospects in dissolved oxygen prediction and other similar parameters [6-7]. Meanwhile, some fuzzy prediction models have been applied in sea water quality analysis through integration of both the characteristic of learning fuzzy logic and neural networks [8-9].

Many prediction and function approximating processes only targeted at receiving good simulation results while many fundamental aspects were ignored. For example, there is no doubt that aquaculture parameters is changing continuously in a day or a season, which forms a standard time series [10] sequence, it is necessary to reorganize and evaluate original data set. Data are usually acquired by monitoring machines or sensors automatically, phenomenons such as data missing and false recording [11] are inevitable through this sampling experiment, in this case, careful data cleaning and transformation are necessary procedures in data preparation. As a dynamic system in open disturbing

ISSN: 2005-4270 IJDTA

circumstance, the water quality is easily be influenced by complex factors, which requires systemic analysis based on influential causes.

This paper is aiming to give a systemic introduction of aquaculture parameters and interactions between them organized as follows: Section I is a brief background introduction; Section II is about data preparation including monitoring, time series approaching and identification foundation; Dynamics analysis are illustrated in Section III, with non-linear function approaching of specific factors; Section IV is mainly about dynamic system identification based on radial basis function neural networks; the conclusion and some related works are given in Section V.

2. Preliminary Results

2.1. Background of Aquaculture Conditions

PH, Temperature, Dissolved Oxygen (DO), Ammonia Nitrogen (AN) level, Salinity, Nitrite are important influential factors for aquaculture creatures. The coordination level of target aquatic products (fish, crab and shell to these factors highly affect its production and quality, and pH, DO, Temperature are critical of all. For example, grouper lives around rocks under tropical sea, the preferred temperature is between 24℃ and 30℃, for most kinds of grouper, it loses physical balance when the temperature over 32℃, and it dies of starvation actively when the temperature below 15℃. In regarding to the pH, it requires this factor stabilize between 6.8 and 8.0, and it will not survival when the pH value over 9.5 or below 5.0; Unsuitable DO could slow down ingestion and activities, which will undermine the health of aquaculture lives, unless it remains more than 4.0 μg/L.

2.2. Acquisition of Water Quality Data

Accurate Sensors could capture aquaculture water parameters in different positions, and send back observational data according to a preferred time period. Here we use AAQ1183-PRO machine (multifunction meter) to get different data (including Temperature, Salinity, Turbidity, Conductivity, Dissolved Oxygen, pH, Chlorophyll level) in every 0.09 seconds from different depths. After this sampling process, a representable data set has been achieved. An example was given in Table 1. In consideration about the depths, we cut down the data set in each 5cm depths, *i.e.* taking account of the data in 5cm, 10cm... 35cm.

2.3. Preparation of Water Quality Data

Careful data cleaning is prerequisite conditions before elaborate processing [12]. As to sampling data set from auto recording machines, aging sensors lead to degraded performance, the data transmission line maybe disturbed by mechanical failures, which will cause unpleasant numbers.

2.3.1. Non-Linear Function Approaching

For most data missing problems, adding artificial values of specific data by function approaching is a common way. The mathematical functions are vary according to the real circumstances, it could be expressed as:

$$\begin{aligned} &f(x) = \varsigma(x_i, y_i), i = 1, 2, \cdots n \\ &P(x) \to f(x) \end{aligned} \tag{1}$$

Table 1. Observational Data from Sea Level (0cm) at different Times

Table 1. Observational Data from Sea Level (0cm) at Different Times

Time	Temp ($°C$)	Cond (mS/cm)	Cond25 ($\mu S/cm$)	Salinity (PSU)	Chl-Flu. (ppb)	Chl-a (mg/L)	Turb (NTU)	pH	DO (mg/l)
07:55	29.561	-0.022	-19.9	0	-0.39	-3	1.71	8.01	4.8106
08:13	28.057	0.07	65.3	0.037	1.35	15	1.69	8.05	4.6247
08:28	27.91	0.061	57.4	0.034	-0.77	-7	-1.68	8.1	6.1086
08:42	27.959	0.059	55	0.033	8.51	86	186.38	8.04	9.3707
08:54	28.26	0.04	37.3	0.026	-3.98	-39	2.02	8.06	12.434
09:09	29.08	47.378	43476	28.237	1.73	18	36.87	8.04	11.31
09:22	28.223	0.114	106.7	0.056	19.04	191	-5.9	8.08	13.401
09:34	28.929	0.262	240.7	0.118	20.46	206	16.72	8.06	13.067
09:42	28.211	0.054	50.1	0.031	0.06	2	1.57	8.04	13.835
09:51	28.437	0.031	29.1	0.022	-5.04	-49	6.75	8.04	12.723
09:59	28.39	0.083	77.5	0.043	42.37	425	-6.59	8.08	13.004
10:07	27.982	0.054	50.3	0.031	4.37	45	6.43	8.07	13.564
10:15	28.037	0.056	52.6	0.032	-5.04	-49	-1.96	8.04	14.271
10:23	29.54	55.659	50604	33.5	0.07	2	0.32	8.05	12.934
10:32	28.22	0.137	127.5	0.065	-3.24	-31	17.1	8.06	14.561
10:40	27.643	0.04	37.8	0.026	-5.04	-49	0.28	8.04	15.766
10:53	27.883	0.026	24.8	0.02	-3.7	-36	1.39	8.03	13.833
11:05	29.276	43.877	40105	25.816	-1.27	-12	1.01	8.04	11.523
11:26	29.508	7.519	6840.9	3.769	52.2	523	59.58	8.02	13.263
11:43	28.425	0.026	24.5	0.02	2.15	22	-6.28	7.97	10.984

where $\varsigma(x_i, y_i)$ reflect the given form of real expression of function $f(x)$ (here by discrete data pairs), $P(x)$ is approximation function to $f(x)$. $P(x)$ could be determined by linear format, or non-linear ones in aquaculture water quality data.

2.3.2. Time Series Analysis: Deterministic Tracking

Tracking the data (pH, DO, Temperature) requires time series analysis about stability and accuracy. As to this feasibility confirmation process, some preliminary theories and definitions need to be informed.

Stochastic Process [13] is an ordered sequence of random variables, wherein $x(s, t)$, $s \in S$, $t \in T$, wherein S represents the sample space, T represents the ordinal data sets. $x(\cdot,t)$ is an implementation in stochastic process in the ordinal set of T.

Deterministic Tracking [14]: Let a temporal data sequence S_ξ is produced by system Y_ξ, where Y_ξ is measurable and bounded, Y_ξ \in R, R is a compact set, system status is a regression track.

$$\begin{aligned} &S_\xi \to Y_\xi \\ &S_\xi = [S(1), S(2), \cdots, S(n)]^T \\ &Y_\xi = [Y(1), Y(2), \cdots, Y(L)]^T \in R \end{aligned} \tag{2}$$

The mode from various factors determining target variables is fixed with deterministic tracking method: S_ξ is synthetical vector of influential factors, Y_ξ is the output parameters. The assumption ensures a deterministic mode from complex high-dimensional inputs corresponding to related output to guarantee the system is unique and identifiable.

3. Dynamics Analysis of Aquaculture Water Quality Data

Water quality studies generally focused on factors due to changes in the weather, tide, artificial feeding or adding ingredients in the previous Section, these natural and non-natural factors influence each other and determine the characteristics of the water quality mutually. It is necessary to understand the main factors affecting the key components of water quality.

In this Section, we explain observational data according to real circumstances and give traditional prediction based on non-linear function approaching without consideration of interactions among factors aforementioned.

3.1. Dynamics Analysis of Temperature

Water temperature affects aquatic species fundamentally by changing organism metabolic rate and determining their food demand, thus influences the quality of aquatic products. Meanwhile, the temperature also affect phytoplankton and other variables in the system. Figure 1 indicates the temperature characters between 07:55 and 11:43.

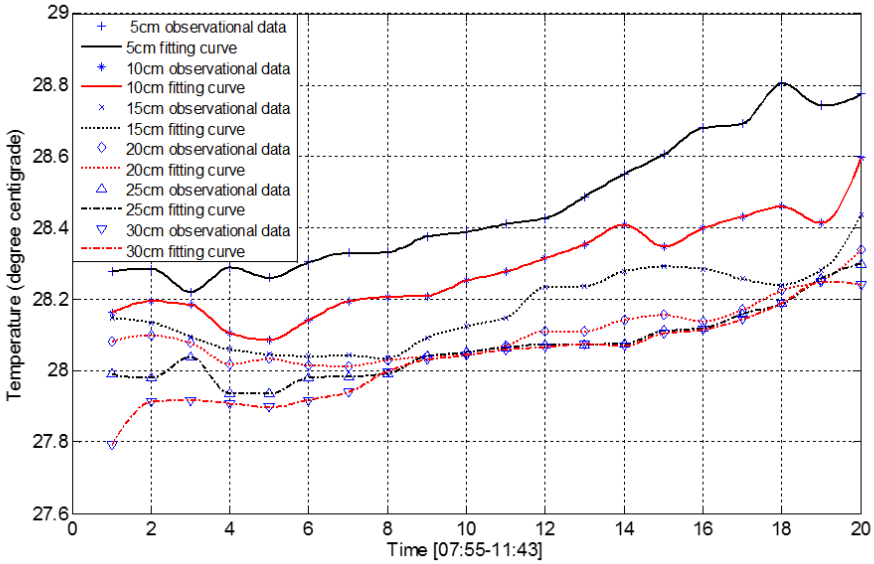

Figure 1. Observational Temperature Data from Different Depths

The temperature increases slowly and stable form the morning to the noon. As is indicated in Figure 1, the control/monitor parameter is influenced by other parameters and weather conditions, in which temperature is a significant one changing all over the time. The average temperature of South China Sea fluctuate between 18℃ and 28℃, particularly between 28℃ and 29℃ in summer.

We compared the temperature from different depths at 09:42 and 10:07, the results in given in Figure 2.In Figure 2a, it is clearly unstable at 0cm, the monitor sensors may suffer from ripple, makes the recorded data step apparently. After removing this points, the function approximating results would be reasonable. The temperature dropped steadily from 10cm to 20cm, and kept this trend more slowly at 20cm to 30cm, then it dropped rapidly afterwards. We use 6-order polynomial function to simulate this process, the red line is the results, and its parameters are shown in Table 2.

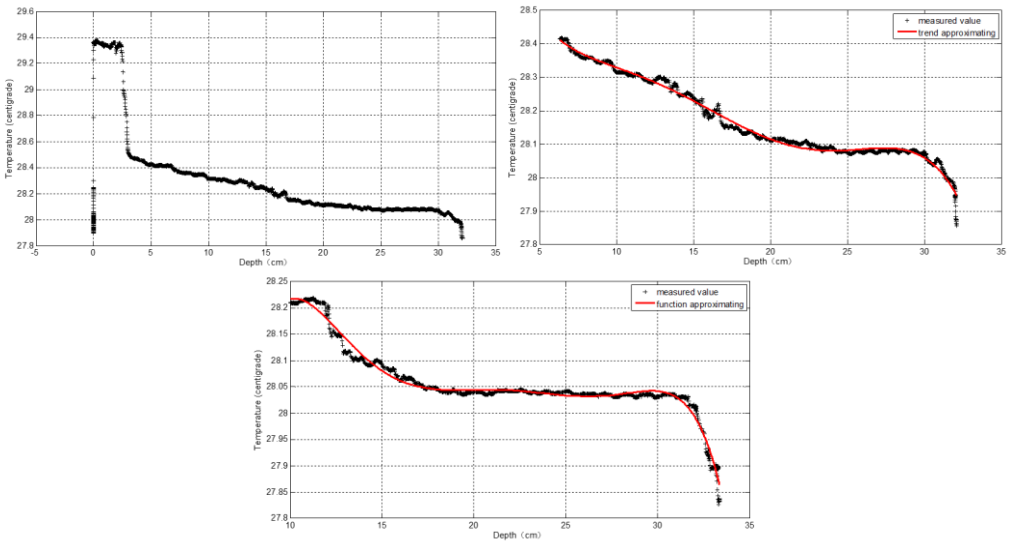

Figure 2. Observational Temperature Data and Fitting Result at Different Depths: (a) Observational Data at 09:42; (b) Fitting after Fault Data Eliminated at 09:42; (c) Fitting and Observational Data at 10:07

Table 2. Fitting result at Different Time (MC: Multi-Nominal Coefficient; TC: Test Coefficient; CI: Confidence Interval

Time	09:42	10:07
MC	6-Order, $p \times 10^3$: 0.0000; -0.0000; 0.0001; -0.0025; 0.0374; -0.2794; 29.2031	6-Order, $p \times 10^3$: -0.0000; 0.0000; -0.0014; 0.0383; -0.5556; 4.1091; 16.1670
2-3 TC	R: [7x7 double]; df: 1727; normr: 0.5360	R: [7x7 double]; df: 1582; normr: 0.3944
[0.4pt]2-3 Prediction	27.5074	27.4738
[0.4pt]2-3 95% CI	[27.4674, 27.5474]	[27.4510, 27.4966]

The function of P(x) and prediction results are reliable, it fluctuate within a very close Confidence Intervals. In real circumstances, the temperature changes steadily and regularly.

3.2. Dynamics Analysis of Dissolved Oxygen

Quantity of dissolved oxygen is a multiple result of production and consumption. Phytoplankton photosynthesis is a main source, but it is closely related to weather factors. Dissolved air diffusion is also an important source of natural, some man-made factors including aerators and water updates. Accordingly, the oxygen consumption is mainly reflected in breathing, metabolism, decomposition and escaping gas. The DO varied with time is shown in Figure 3.

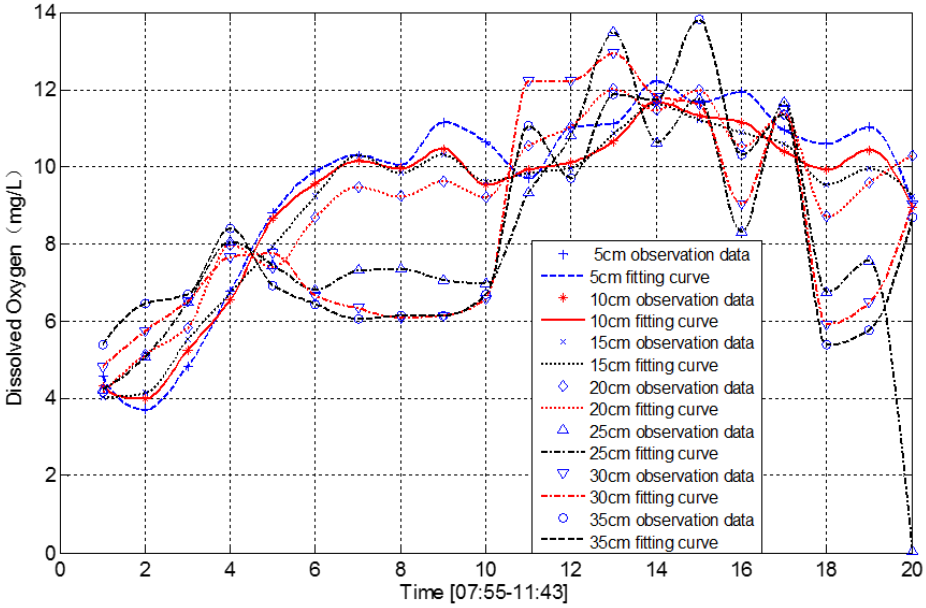

Figure 3. Observational DO Data from Different Depths

Illumination intensity strengthens from the morning, as a result, phytoplankton photosynthesis exceeds respiration, the increasing mechanism does not altered even take different depths into account. From Figure 3, this trend represents the *x* axis from 0 to 5, *i.e.* 07:00 to 08:00 approximately. However, After 08:00, the fish wake up with stronger respiration, accompanied by ingestion, this phenomenon is more obvious in 15cm to 35cm depths, which leads the oxygen consumption quantity increased dramatically, DO indicates a decreasing trend. This current weakens when the ingestion finished after 09:00, followed by increasing trend again between 09:00-11:00, but the ingestion phase will re-occur at noon, and reflected by DO decreasing in this figure.

Aquaculture water is an open system, it could be influenced by weather, waste discharge and many other artificial or natural factors. Consider it as a feedback control system, taking Temperature as a target variable, one possible control flow chart is shown in Figure 4, and the system description details is shown in Figure 5.

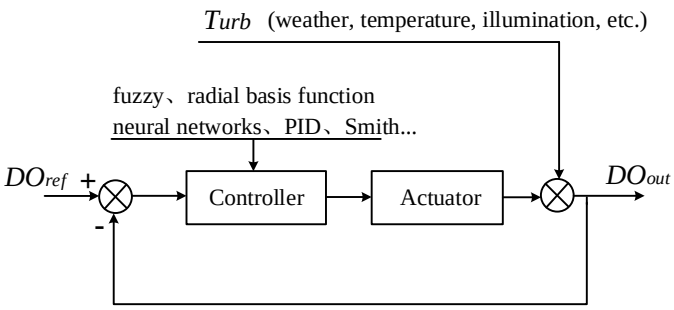

Figure 4. Dynamic Control Model of DO (Control Flow Chart)

DO_{ref} is our desired reset limited value, if there is no T_{urb} from acid rain, waste discharge or any other factors, the system would be in a self-regulating condition. As some turbulence affect the DO_{out}, the feedback signal and DO_{ref} will calculate a proper error value, which will be sent to controller for further computing based on strategies including fuzzy, PID and Smith compensation control. A control parameter will be

produced and added to actuator to regulate DO finally, and repeat the process periodically.

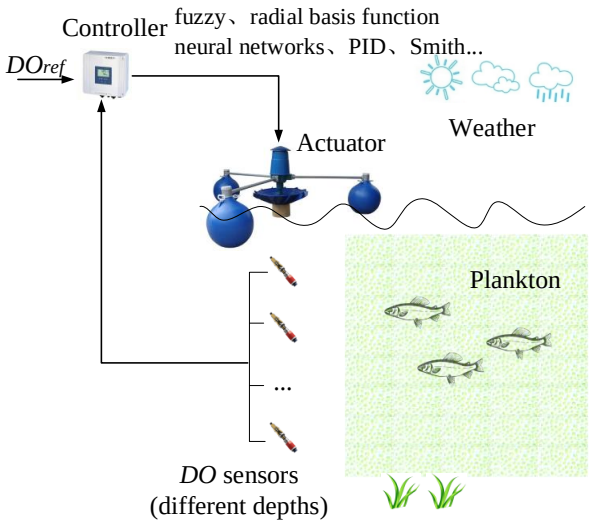

Figure 5. Dynamic Control Model of DO (System Ecosystem Fundamentals)

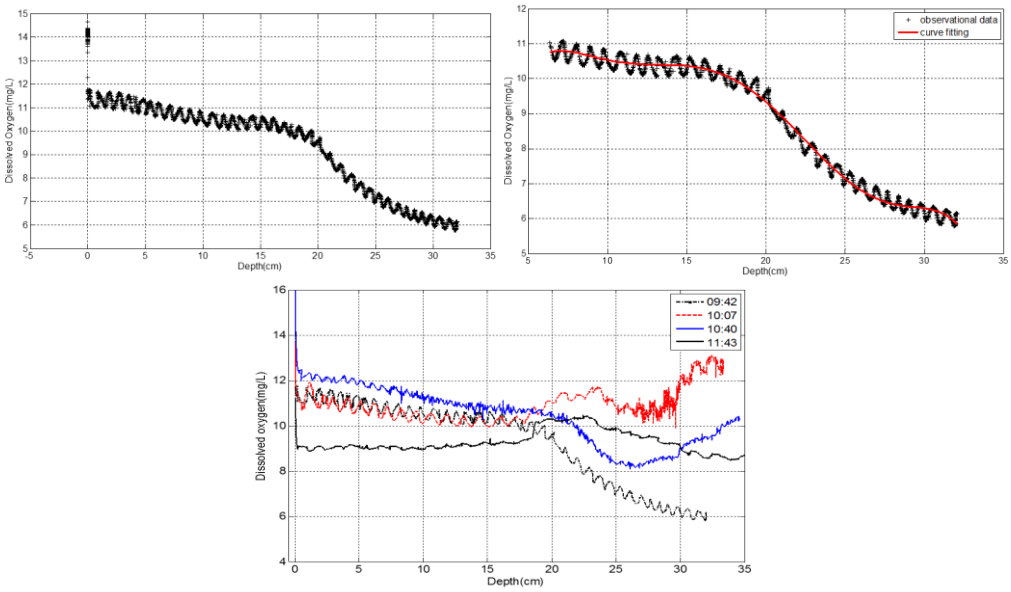

Figure 6 Observational DO Data and Fitting Result from Different Depths: (a) Observational Data at 09:42; (b) Fitting Result after Data Cleaning at 09:42; (c) Fitting Result and Observational Data at Different Times

DO dynamic balance details at 09:42 and 10:07 are indicated in Figure 6, As is described above, the illumination intensity affects it significantly, results in the surface and shallow depths contain rather high DO quantity. From Fig 6a, the DO is not stable because of turbulence from the air, in order to avoid interference, the function approximating result in Fig 6b based on data cleaning is reasonable. High-order nonlinear equation can fit the data by overall trends, but not behaves well in tracking periodical caused by periodic changes in sea conditions rhythm locally. In Figure 6c, the DO varied at 20cm because of controller interfering, with the judging criteria of the DO below 10mg/L.

3.3. Dynamics Analysis of PH

PH is a passive water quality factor subjected to combined effects from others, fundamentally determined by CO_3^{2-} in all dissociation forms. A large pH value will cause alkaline water environment, excessive ammonia nitrogen elements will poison the aquatic lives; a small pH value means it is too acidic, which will weaken aquatic metabolism mechanism function. Acid rain, discharge of sewage are natural sources could affect the quality of acid-base balance. Biological activities (respiration and metabolism are internal reasons could alter this pH value. Figure 7 describes the rough conditions of the morning.

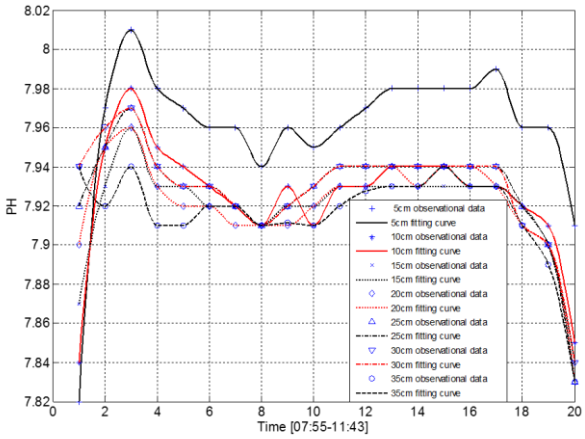

Figure 7. Observational PH Data from Different Depths

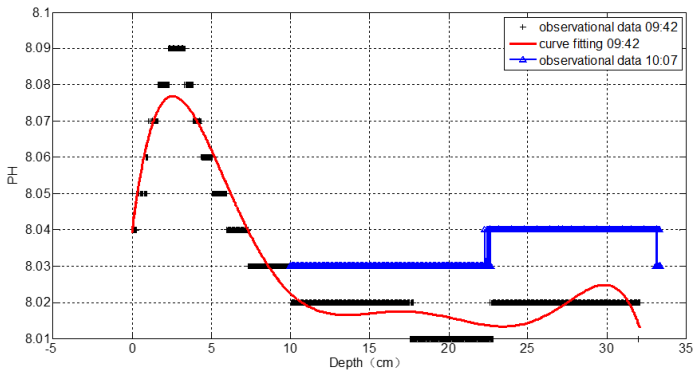

Figure 8. Observational PH Data and Fitting Result from Different Depths

Phytoplankton photosynthesis strengthens from the morning because of enhancing illumination intensity, absorb large quantity of carbon dioxide, leads to the pH value increased; However, aquatic lives wake up at about 08:00 and ingestion reinforced, carbon emission is increased, and pH dropped by a tiny level, but by the help of stronger illumination intensity from 09:00 to 11:00, a balance is achieved dynamically between the decomposition and carbon dioxide emissions, the pH stabilized in this period. But the ingestion phase at noon triggered the water acidity again.

pH measured at different depth values as shown in Figure 8, the surface area (0cm to 2.5cm) is rich in algae, photosynthesis surpasses dioxide reduction process, the dissociation forms of CO_3^{2-} are limited, thus lead a high pH value relatively. Deeper water (less than 150m) means more lives, more consumption of oxygen and more carbon dioxide emissions and lower pH.

4. Identification Based on Dynamics system and Artificial Intelligence

4.1. Dynamics Model of System Parameters

According to dynamics analysis from the previous Section, interconnected relationship (constraint or promote) does exist among parameters including DO, temperature and pH. Meanwhile, several natural disturbances such as weather, waste discharge also affect the aquatic bio-balance status. Though curve fitting based on non-linear function approximating receives worthy results globally, it cannot reflect detailed information very well.

Artificial Intelligence provides an effective computing mode derived from biology and neural science, and even pushed forward by computer science. It receives amazing achievements in all engineering areas like nonlinear control, pattern recognition, optimization, signal analysis and processing, aerospace and intelligent monitoring. These inspiring results are playing more important roles in our daily lives, companies and national projects.

In regarding to aquaculture water quality data analysis, the merits of AI are accurate tracking with error tolerances, and fast simulation.

4.2. Radial Basis Function Neural Networks Identification

Nonlinear structure is accepted to describe the relationship between influential factors due to the interconnection among neurons, identification modeling based on dynamic neural networks could be adopted to simulate it hereby, the basic form of RBFNN is shown in Figure 9.

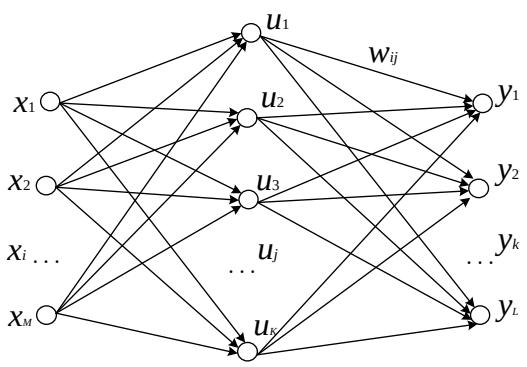

Figure 9. Basic Structure of Radial Basis Function Neural Networks

Radial Basis Function Neural Networks (RBFNN) is well applied due to non-linear parameter structure, massively parallel distributed architecture and self-learning adaptive ability [15-17].

The input layer is perception(s) here similar to sensors of these variables, the hidden layer mapped information to high-dimensional space relatively, complete the operation of weight adjusting and computing; the output layer is similar to the convergence unit, let pH$, DO, temperature as three target outputs variables.

Identification results of DO and pH are shown in Figure 10 and Figure 11, with parameters of RBFNN initialized as: learning rate η=0.05, spreading speed=0.05, M=3 and L=1.

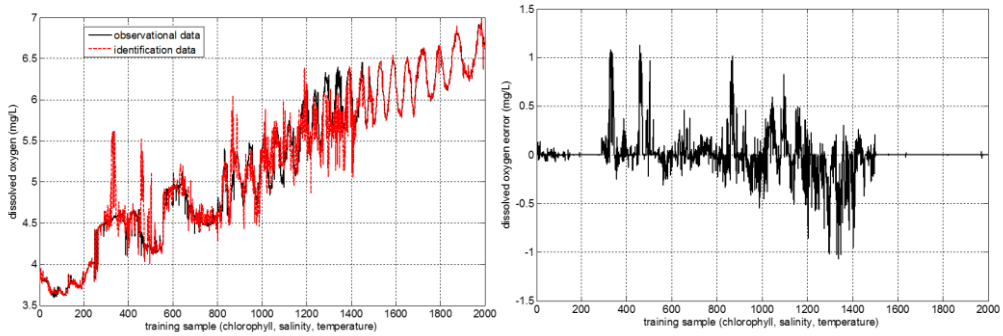

Figure 10. Observational DO Data and RBFNN Identification Result (2,000 Samples 100% Trained): (a) Identification Result Based on RBFNN; (b) Identification Errors

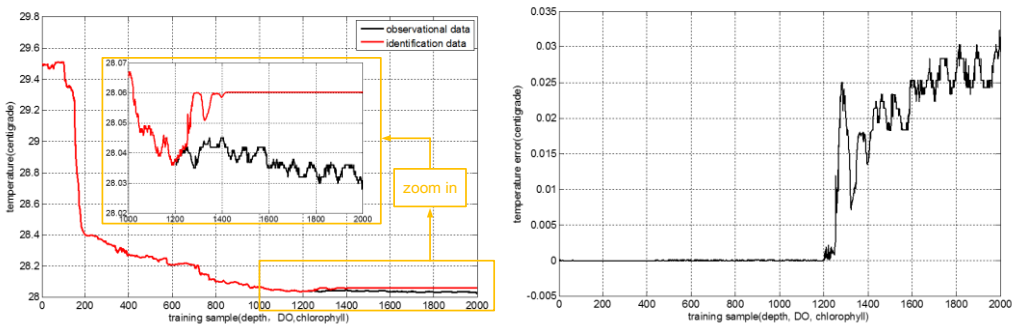

Figure 11. Observational Temperature Data and RBFNN Identification Result (2,000 Samples, with 60% Trained and 40% Simulated): (a) Identification Result Based on RBFNN; (b) Identification Errors

From Figure 10a, the DO prediction indicates good performance, especially after about 1,500 samples training, its prediction error is limited within a small range in the last 500 samples in Figure 10b, even though DO suffers from small fluctuations, the RBFNN model can also tracking it with tiny errors.

In order to test the robust character and ability to adapt to unknown dynamics, we use only 60% samples (1,200 items) to train RNBFNN model, and use the remaining 40% in temperature prediction, the results is given in Figure 11a, it still reflects better tracking characteristic with the error restricted within 0.3 degree centigrade. This result has benefited from a relatively stable temperature range and excellent RBFNN tracking ability.

5. Conclusions & Future Works

In this paper, we emphasized dynamic analysis of major aquaculture water quality factors, relations between these indicators have been discussed. Then artificial intelligence identification verification systemically, with traditional non-linear prediction methods have been given for comparisons. For most time series data, it satisfied with bounded and regression conditions, which gives a solid foundation for radial basis function neural networks simulation modeling, the testing experiment verified dynamic system aforementioned finally.

Relations among the aquaculture water quality factors has been confirmed, our next target is to understand it quantitatively, this will be a further confirmation about our current results. Several mathematical theories and tools are requires for this target, including principle component analysis and correlation matrix analysis. Meanwhile, it is imperative to dig more about RBFNN model from prediction model to a control model, to

realize theoretical knowledge from data mining to control area, which will benefit both control theory and engineering cases.

References

[1] B. Gourevitch, R. B. Jeannes and G. Faucon, "Linear and nonlinear causality between signals: methods, examples and neurophysiological applications", Biological cybernetics, vol. 95, no. 4, **(2006)**, pp. 349-369.
[2] J. N. Reddy, "An introduction to nonlinear finite element analysis: with applications to heat transfer, fluid mechanics, and solid mechanics", in: Wyatt, Oscar S. (Ed.), an Introduction to Nonlinear Finite Element Analysis. Oxford University press, London, **(2014)**, pp. 235-271.
[3] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt and B. Scholkopf, "Support vector machines", Intelligent Systems and their Applications, vol. 13, no. 4, **(1998)**, pp. 18-28.
[4] G. Buyukyildiz and M. Tezeland, "Monthly evaporation forecasting using artificial neural networks and support vector machines", Theoretical and Applied Climatology, **(2015)**, pp. 1-12.
[5] S. Y. Liu, H. J. Tai, Q. S. Ding, D. L. Li, L. Q. Xu and Y. Q. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction", Mathematical and Computer Modelling, vol. 58, no. 3, **(2013)**, pp. 458-465.
[6] F. F. Li, D. L. Li, Y. G. Wei, D. D. Ma and Q. Ding, "Dissolved oxygen prediction in apostichopus japonicus aquaculture ponds by bp neural network and ar mode", Sensor Letters, vol. 8, no. 1, **(2010)**, pp. 95-101.
[7] M. Sun, J. Chen and D. Li, "Water temperature prediction in sea cucumber aquaculture ponds by rbf neural network model", International Conference on Systems and Informatics (ICSAI), **(2012)**, pp. 1154-1159.
[8] Y. L. Liu, Y. G. Wei and Y. G. Chen, "Dissolved oxygen prediction model which based on fuzzy neural network", In: Daoliang Li, Yingyi Chen (Ed.), Computer and Computing Technologies in Agriculture VII. Springer, Berlin Heidelberg, **(2014)**, pp. 544-551.
[9] Z. B. Li, J. Yu, J. Yue, L. F. Zhang and D. L. Li, "An improved gray model for aquaculture water quality prediction", Intelligent Automation and Soft Computing, vol. 18, no. 5, **(2012)**, pp. 557-567.
[10] M. S. Shahriar, A. Rahman and J. McCulloch, "Predicting shellfish farm closures using time series classification for aquaculture decision support", Computers and Electronics in Agriculture, vol. 102, **(2014)**, pp. 85-97.
[11] C. Sowder and E. A. Steel, "A note on the collection and cleaning of water temperature data", Water, vol. 4, no. 3, **(2012)**, pp. 597-606.
[12] J. Williamson, Q. Liu, F. L. Lu, W. Mohrman, K. Li and S. Li, "Data sensing and analysis: Challenges for wearables", 20th Asia and South Pacific Design Automation Conference, **(2015)**, pp. 136-141.
[13] Y. Shen, X. Steven, A. H. Haghani and H. Y. A. Hao, "Data-driven monitoring for stochastic systems and its application on batch process", International Journal of Systems Science, vol. 44, no. 7, **(2013)**, pp. 1366-1376.
[14] C. Wang and D. J. Hill, "Deterministic learning and rapid dynamical pattern recognition", IEEE Transactions on Neural Networks, vol. 18, no. 3, **(2007)**, pp. 617-630.
[15] L. Yingwei, N. Sundararajan and P. Saratchandran, "A sequential learning scheme for function approximation using minimal radial basis function neural networks", Neural computation, vol. 9, no. 2, **(1997)**, pp. 461-478.
[16] M. V. Ayouri1, M. Shahbazian, B. Moslemi and M. Taheri, "Robust identification of hydrocarbon debutanizer unit using radial basis function neural networks", Journal of Automation and Control, vol. 3, no. 1, **(2015)**, pp. 10-17.
[17] K. B. Cho and B. H. Wang, "Radial basis function based adaptive fuzzy systems and their applications to system identification and prediction", Fuzzy sets and systems, vol. 83, no. 3, **(1996)**, pp. 325-339.

Authors

Xiaofei Huang, joined Hainan College of Software Technology as a staff college, her research interests include mathematical computing Technology.

Wei Wu, joined Institute of Deep-sea Science and Engineering, CAS as a staff college. His research interest covers data and knowledge engineering.

