Prediction of Dissolved Oxygen Concentration for Shrimp Farming Using Quadratic Regression and Artificial Neural Network

Kasorn Galajit
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
kasorn.galajit@nectec.or.th

Pitisit Dillon
Faculty of Applied Science
King Mongkut's University of Technology North Bangkok
Bangkok, Thailand
s5804021620048@email.kmutnb.ac.th

Suradej Duangpummet
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
suradej.duangpummet@nectec.or.th

Jakkaphob Intha
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
jakkaphob.intha@nectec.or.th

Prachumpong Dangsakul
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
prachumpong.dangsakul@nectec.or.th

Khongpan Rungprateepthaworn
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
khongpan.rungprateepthaworn@nectec.or.th

Rachaporn Keinprasit
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
rachaporn.keinprasit@nectec.or.th

Jessada Karnjana
NECTEC
National Science and Technology Development Agency
Pathum Thani, Thailand
jessada.karnjana@nectec.or.th

Abstract—In aquaculture, one of the most critical factors for sustaining life under the water is the dissolved oxygen (DO) since it affects not only the animal survival rate but also the growth rate. Therefore, in smart aquafarming, the DO content should be monitored thoroughly. As a consequence, in practice, many DO sensors are installed in systems, and they contribute markedly to the system cost. This work aims to reduce the cost by replacing some DO sensors with a model that can describe the dynamics of DO content in a specific controlled environment. Thus, we propose two predictive models: one based on the quadratic regression and another based on an artificial neural network. Experimental results show that, under the limitation of the number of data used in the model construction, both models perform equally. Also, both prediction fitted more to observed data when the DO level is low. This finding supports the practical model usage since in practice we more concern with the efficiency of the model in the case of low DO concentration.

Index Terms—dissolved oxygen model, shrimp farming, quadratic regression, neural network

I. Introduction

During 2002 and 2006, industrial shrimp farming in Thailand shifted from domesticating the native black tiger shrimp (Penaeus monodon) to domesticating the Pacific whiteleg shrimp (Litopenaeus vannamei). From then on, the whiteleg shrimp (both frozen and processed) has become one of the primary export products of Thailand [1], [2]. When the demand increases due to the world population explosion [3], the aquaculture sector adopts the information, communication, and embedding technologies for intensive farming. In order to operate the intensive farming productively, automatic control systems are unavoidable [4]–[7] because environmental parameters need to be carefully controlled to meet the requirement for sustaining life under the water, as well as to prevent losses due to infectious diseases [8], [9].

For shrimp farming, one of the most critical water-quality parameters is the dissolved oxygen (DO) since it affects the survival rate, growth rate, and, consequently, the gross yield [10], [11]. Therefore, the DO concentration is thoroughly monitored in smart aquaculture. Hence, in practice, many DO sensors are deployed in the automatic control system [12]. These sensors, especially ones with the high accuracy, are expensive. As it goes without saying, they contribute markedly to the implementation cost of the system.

978-1-7281-0164-4/18/$31.00 ©2018 IEEE

This work aims to reduce the cost by replacing some DO sensors with a model that can describe the dynamics of DO content in a specific controlled environment, i.e., a suitable environment for shrimp domestication. Such the model can be used as a source of information (or knowledge) in data fusion techniques, such as Kalman filtering and theory-of-evidence-based methods. However, in literature, many previously proposed models are complicated and take a lot of factors into their dynamic equations [13]–[15]. Alternatively, this work experiments with data-driven approaches.

The rest of this paper is organized as follows. Section II describes the overview of our previously proposed automatic aerator-control system for shrimp farming and states the problem statement including its assumption explicitly. Section III gives details of two proposed predictive models: the one that is based on the quadratic regression and another that is based on an artificial neural network. Section IV details our experiments and results. Discussion and conclusion are made in Section V and Section VI, respectively.

II. Overview of Our Previously Proposed, Automatic Aerator-Control System for Shrimp Farming

Recently, we proposed an automatic aerator-control system for shrimp farming [12]. The architecture of our proposed system is illustrated in Fig. 1. The system consists of two node types: pond control node and aerator node. Many sensors are attached to the pond control node, and they are used to measure water-quality parameters, such as DO content, pH, and temperature. Data read from those sensors are sent to a database server through the Universal Mobile Telecommunication System (UMTS) and the Internet. All data are kept for analysis, and they can be viewed or accessed by logging in to a monitoring web page. Also, based on the values of the data, the pond control node decides which aerators are to be turned on or off. For example, if the pond control node finds that the DO content is lower than the requirement for sustaining life, it demands some aerators to be in action. In order to turn on or off aerators, the pond control node sends a command to aerator nodes attached to the aerators. These nodes communicate wirelessly conforming to our proprietary protocol via the DigiMesh network. Then, the aerator nodes respond to the received command accordingly.

As mentioned in Section I, this work aims to construct a model of the DO concentration that changes with time in a specific controlled environment, i.e., a suitable environment for shrimp domestication. Thus, in this work, we deploy only the parts inside the red, dashed box, as shown in Fig. 1, to monitor the DO concentration and some other water-quality parameters. Then, we use the monitoring data to build models. Besides, we assume that the DO content, as well as other relevant factors, is well and appropriately controlled.

Note that this assumption is not that far from the truth since the environment of our experiment are under controlled by a team of experts and researchers from the Aquaculture Product Development and Services (AAPS) laboratory of the National Center of Genetic Engineering and Biotechnology (BIOTEC), Thailand.

III. Proposed Predictive Models

In this work, we adopt two data-driven approaches to predict the DO content and compare their performance. The first approach is based on the quadratic regression, and the second predictive model is based on an artificial neural network. The following subsections summarize our construction of these two models.

A. Model Based on the Quadratic Regression

The quadratic regression is a process for estimating a relationship between the independent variables x and the dependent variable y, and this relationship is modeled by a second-order polynomial in x, i.e., $y = \beta_0 + \beta_1 x + \beta_2 x^2$, where β_0, β_1, and β_2 are model parameters. This equation can be found by using the method of least squares. That is, we look for the values of β_0, β_1, and β_2 such that the squared vertical distance between each point (x_i, y_i) and the parabola $y = \beta_0 + \beta_1 x + \beta_2 x^2$ is minimal. In other words, we find the values of β_0, β_1, and β_2 by minimizing S, where

$$S = \sum_{i=1}^{n} \left(y_i - \left(\beta_0 + \beta_1 x_i + \beta_2 x_i^2\right)\right)^2, \qquad (1)$$

given that there are n data points.

In this work, y is the DO concentration that varies with the (discrete) time x.

B. Model Based on an Artificial Neural Network

An artificial neural network (ANN) used in this work is of the feedforward class, called the multilayer perceptron. Since its function is to estimate the DO concentration varying with time, the model has one input and one output. We implement a shallow ANN that consists of two hidden layers, in which the numbers of nodes are 32 and 16, respectively. The structure of the ANN is shown in Fig. 2. The sigmoid function is deployed as the activation function in this model, and the rmsprop, which is an adaptive learning rate method, is used to train the network.

IV. Experiment and Results

Details of our experimental setup, experiments, and results are provided in this section.

A. Experimental Setup

Two cylindrical plastic containers, as shown in Fig. 3(b), were used as our experiment ponds. Their height and diameter, as well as a sketch of the pond, are illustrated in Fig. 3(a). These two experiment ponds were used to domesticated 120 whiteleg shrimps (Litoenaeus vannamei)

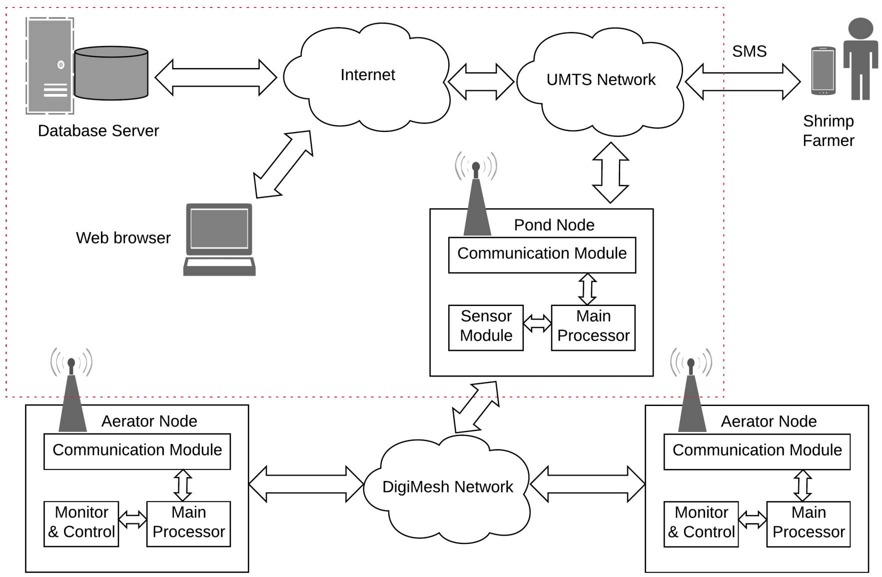

Fig. 1. Architecture of the automatic aerator-control system.

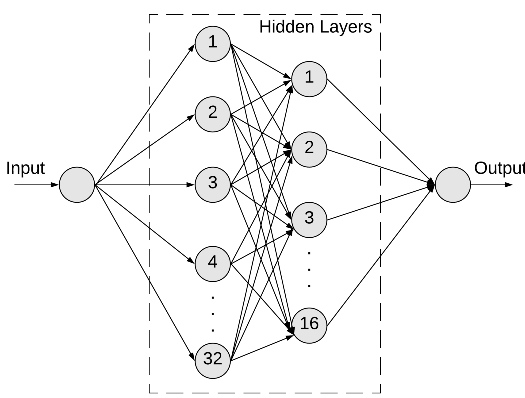

Fig. 2. ANN structure used in this work. The input is the time (min), and the output is the DO level (mg/L).

for 50 days. In other words, we natured the shrimps from the average weight of three grams to that of approximately 15 grams. Both ponds were in an open-air condition. The controlled, water-quality parameters used in our experiments are listed in Table I.

TABLE I
Controlled, water-quality parameters used in our experiments.

Water-quality parameters	Controlled value
Salinity	17 parts per thousand
pH	7.5 - 8.5
Alkalinity	130 - 150 mg/L
Total ammoniacal nitrogen	less than 1 mg/L

We installed some DO sensors, a temperature sensor, and a circulation pump at each pond. The pump was used to increase the DO content. Note that the BIOTEC team suggested the pump size and operation hours. Data read from sensors were transmitted to a database server via the UMTS network and the Internet, as described in Section II, and were used to construct the predictive models, as described in the previous section.

B. Experimental Results of the Regression-based Model

A set of thirty-day data obtained from the DO sensors, as listed in Table II, was used in the regression. We divided data into four segments (or four subsets), as shown in Fig. 4, and then determined four parabola equations that best fit these four subsets according to the quadratic regression. As a result, a model of the DO concentration d that changes with time t in the controlled, shrimp-farming environment is formulated as follows.

$$d(t) = \begin{cases} -\frac{1.21}{10^7}t^2 + \frac{33}{10^5}t + 6.34, & \text{for } 1 \leq t \leq 454, \\ -\frac{4.02}{10^6}t^2 + \frac{852}{10^5}t + 3.41, & \text{for } 455 \leq t \leq 677, \\ -\frac{8.82}{10^6}t^2 + \frac{1132}{10^5}t + 3.70, & \text{for } 678 \leq t \leq 951, \\ \frac{3.40}{10^6}t^2 - \frac{865}{10^5}t + 11.66, & \text{for } 952 \leq t \leq 1440, \end{cases} \tag{2}$$

where t is a positive integer that indicates the time sampled at a rate of one sample per minute. The plot of $d(t)$ is drawn by the blue line in Fig. 4. It should be noted that $d(t)$ is fitted to the average data points, which are marked by the red plus signs in Fig. 4, with the root-mean-square error (RMSE) of 0.0243. The coefficient of determination (R^2) is 0.9954.

The RMSE values between two data sets obtained from the DO sensors and the regression-based model are shown in Table II, and the average RMSE is 0.3580.

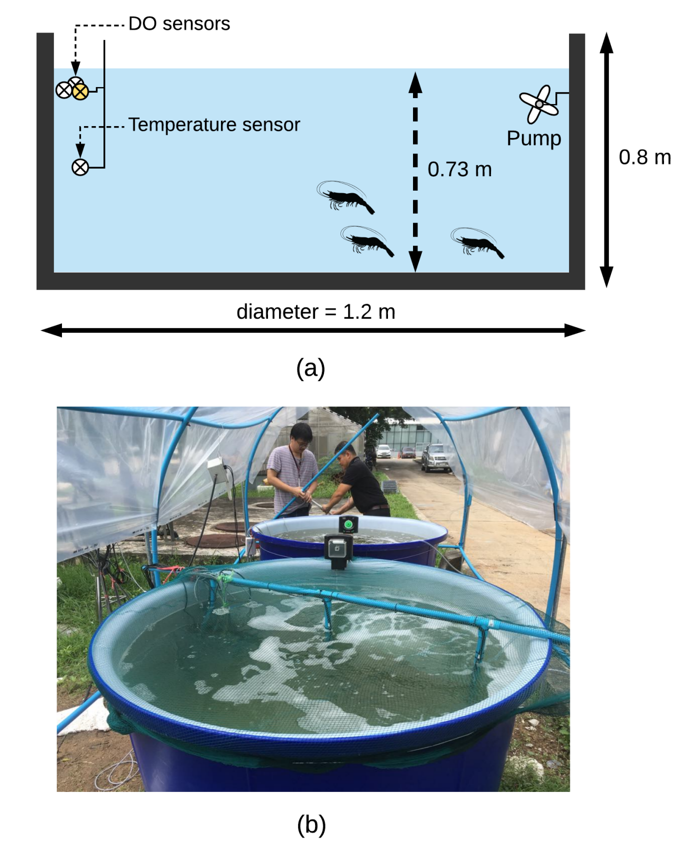

Fig. 3. Diagram of the shrimp pond together with sensors used in our experiment (top) and two experiment ponds in operation (bottom).

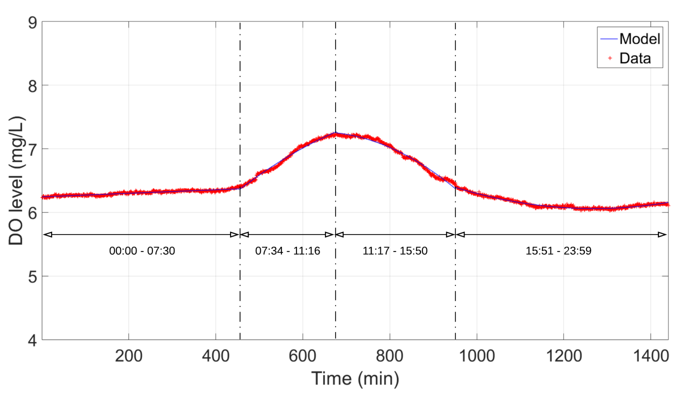

Fig. 4. Data are segmented into 4 pieces for the quadratic regression.

C. Experimental Results of the ANN-based Model

The same set of data used in the regression was used in training the network described in Section III-B. Also, they were shuffled and divided into two subsets; the first one (80%) is for training, and the second one (20%) is for testing. The network was trained for 500 iterations.

The plot of estimated DO concentration obtained from the ANN-based model is illustrated in Fig. 5.

The RMSE values between two data sets obtained from the DO sensors and the model are shown in Table II, and the average RMSE is 0.3593. It can be seen that, given the limitation of the number of data points, the regression-based model is slightly better than the ANN-based model in the RMSE comparison.

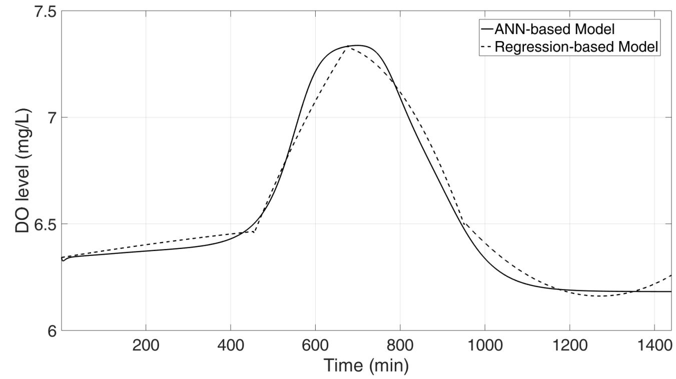

Fig. 5. Estimated DO content comparison between the ANN-based model (solid line) and the regression-based model (dashed line).

TABLE II
Root-mean-square-error (RMSE) comparison between the regression-based model and the ANN-based model.

Date	Sensor ID	Regression-based model	ANN-based model
18 Aug 17	0x10	0.4773	0.4908
18 Aug 17	0x01	0.1694	0.1892
18 Aug 17	0x02	0.3307	0.3696
18 Aug 17	0x03	0.3791	0.3640
18 Aug 17	0x04	0.4841	0.4879
18 Aug 17	0x06	0.3274	0.3651
19 Aug 17	0x10	0.3453	0.3285
19 Aug 17	0x01	0.3603	0.4058
19 Aug 17	0x02	0.2397	0.2853
19 Aug 17	0x03	0.8043	0.7554
19 Aug 17	0x04	0.1193	0.1253
19 Aug 17	0x06	0.1743	0.1988
20 Aug 17	0x10	0.2974	0.2623
20 Aug 17	0x01	0.3263	0.2818
20 Aug 17	0x02	0.1829	0.2096
20 Aug 17	0x03	1.3414	1.3017
20 Aug 17	0x04	0.2385	0.2321
20 Aug 17	0x06	0.1610	0.1673
21 Aug 17	0x10	0.3300	0.2995
21 Aug 17	0x01	0.2345	0.2363
21 Aug 17	0x02	0.1669	0.1602
21 Aug 17	0x03	0.9441	0.9078
21 Aug 17	0x08	0.3216	0.3675
21 Aug 17	0x06	0.1704	0.1539
23 Aug 17	0x10	0.2114	0.2301
23 Aug 17	0x08	0.3501	0.3858
23 Aug 17	0x02	0.2151	0.1885
23 Aug 17	0x03	0.5041	0.4648
23 Aug 17	0x07	0.2352	0.2176
23 Aug 17	0x06	0.2972	0.3004

V. Discussion

Based on the experimental results, there are three issues concerning the performance of both predictive models to be discussed in this section. First, the average RMSE of 0.3580 for the regression-based model is acceptable for many applications (e.g., using the model as a source of

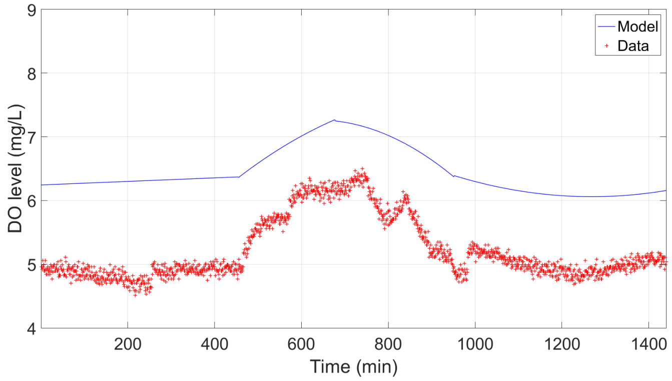

Fig. 6. Data obtained from the sensor 0x03 (20 Aug 2017).

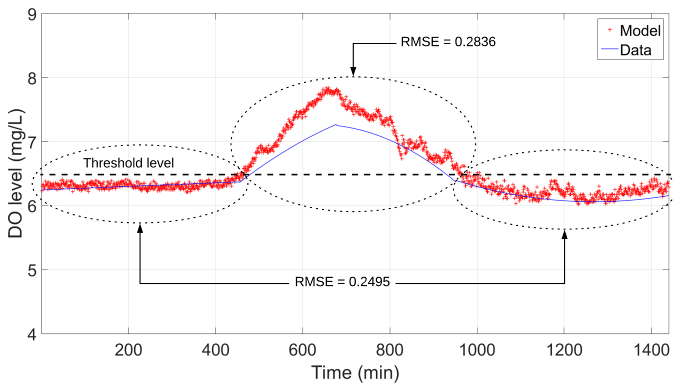

Fig. 7. Data obtained from the sensor 0x01 (18 Aug 2017).

information in data fusion techniques), but this number is, in a sense, overestimated. It is overestimated because sensor errors are not taken into consideration in this work. For example, let us examine data obtained from the sensor 0x03 on the 20th of August, 2017, as plotted in Fig. 6. It can be seen that the actual data are largely dropped, compared with those expected from the model. Many factors affect sensor efficiency, for instance, dirt or biofilms that form on membranes of the sensor. Thus, when we had cleaned our data by removing ones that might be caused by such a failure, the average RMSE decreased from 0.3580 to 0.2833.

Also, we found that the RMSE values increase when the DO concentration increases. For example, if we set a threshold level of 6.5 mg/L, as illustrated in Fig. 7, the average RMSE of the lower part is smaller than that of the upper part. This finding implies that it somehow boosts a level of confidentiality in using the model since we practically more concern when the DO concentration is low.

Second, under this experimental condition, we believed that what limits the performance of the ANN-based model is the limitation on the number of training data, not the complexity of the neural network. For example, when we have trained another ANN consisting of three hidden layers, of which there were 128, 64, and 16 nodes,

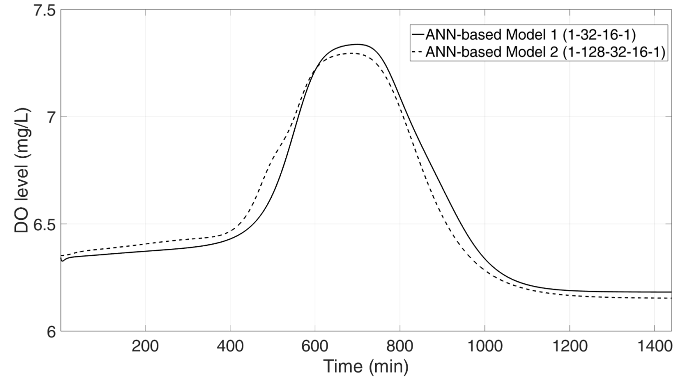

Fig. 8. Estimated DO concentration comparison between two different ANN-based models.

respectively, and have compared its result with that of the proposed model, we found that the difference between the estimated DO level was of no significance, as shown in Fig. 8. However, the performance of the model might depend upon network types. In future work, applying other types of the network, such as recurrent neural networks, to the application of DO estimation will be investigated.

Third, regarding the aim of this work, which is the cost reduction, the proposed model could be integrated with the system as follows. As detailed in Section III-B, our previously proposed system deploys a few DO sensors to crosscheck DO levels read from those sensors since relying on data obtained from one sensor is risky. When we have a suitable model that can describe the dynamics of the pond, the DO levels estimated by it then can be used in cross-checking. Thus, the number of required DO sensors in the system is reduced; hence, the cost is reduced.

VI. Conclusion

In this paper, we analyzed and reported the performance of two DO content prediction models under the specific controlled environment for shrimp farming. The first model is based on the quadratic regression, and the second one is based on the ANN. We found that, under the limitation of the number of data used in the analysis, the regression-based model performed better in the RMSE comparison. Also, the estimated values obtained from both models more fitted to observed data when the DO concentration was low. This finding implies that the average RMSE is overestimated to some degree since in practice we more concern with the efficiency of the model in the case of low DO content.

Acknowledgment

This work was collaborative research between the National Electronics and Computer Technology Center (NECTEC) and the Aquaculture Product Development and Services (AAPS) laboratory of the National Center of Genetic Engineering and Biotechnology (BIOTEC), Thailand. The authors would like to express their sincere gratitude to Dr. Sage Chaiyapechara and his colleagues for domesticating the whiteleg shrimps in the experiments.

References

[1] L. Lebel, S.H. Gheewala, and P. Lebel, "Innovation cycles, niches and sustainability in the shrimp aquaculture industry in Thailand," Environmental Science & Policy, vol. 13, pp 291–302, 2010.

[2] Office of Agricultural Economics [Accessed 16 August 2018], http://oldweb.oae.go.th/oae_report/export_import/export.php

[3] UN-DESA [Accessed 16 August 2018], http://www.un.org/en/development/desa/news/population/2015-report.html

[4] P. G. Lee, "A review of automated control systems for aquaculture and design criteria for their implementation," Aquacultural engineering, vol. 14(3), pp. 205–227, 1995

[5] Y. Shifeng, K. Jing, and Z. Jimin, "Wireless monitoring system for aquiculture environment," in Proc. IEEE International Workshop on Radio-Frequency Integration Technology, pp. 274–277, December 2007.

[6] D. S. Simbeye, J. Zhao, and S. Yang, "Design and deployment of wireless sensor networks for aquaculture monitoring and control based on virtual instruments," Computers and Electronics in Agriculture, vol. 102, pp. 31–42, 2014.

[7] N. T. K. Duy, N. D. Tu, T. H. Son, and L. H. D. Khanh, "Automated monitoring and control system for shrimp farms based on embedded system and wireless sensor network," in Proc. IEEE International Conference on Electrical, Computer and Communication Technologies, pp. 1–5, March 2015.

[8] T. W. Flegel, "Major viral diseases of the black tiger prawn (Penaeus monodon) in Thailand," World Journal of Microbiology and Biotechnology, vol. 13(4), pp. 433–442, 1997.

[9] J. Rodr?guez and G. Le Moullac, "State of the art of immunological tools and health control of penaeid shrimp," Aquaculture, vol. 191(1–3), pp. 109–119, 2000.

[10] Y. Li, J. Li, and Q. Wang, "The effects of dissolved oxygen concentration and stocking density on growth and non-specific immunity factors in Chinese shrimp, Fenneropenaeus chinensis," Aquaculture, vol. 256(1–4), pp. 608–616, 2006.

[11] W. Wiyoto, S. Sukenda, E. Harris, K. Nirmala, and D. Djokosetiyanto, "Water Quality and Sediment Profile in Shrimp Culture with Different Sediment Redox Potential and Stocking Densities Under Laboratory Condition," Ilmu Kelautan, vol. 21(2), pp. 65–76, 2016.

[12] K. Galajit, T. Duangtanoo, K. Rungprateepthaworn, S. Sartsatit, P. Dangsakul, and J. Karnjana, "Flexible and Automatic Aerator-control System for Shrimp Farming in Thailand," in Proc. Advanced Research in Electrical and Electronic Engineering Technology, pp. 1–6, 2017.

[13] Y. M. Svirezhev, V. P. Krysanova, and A. A. Voinov, "Mathematical modelling of a fish pond ecosystem," Ecological modelling, vol. 21(4), pp. 315–337, 1984.

[14] W. J. S. Mwegoha, M. E. Kaseva, and S. M. M. Sabai, "Mathematical modeling of dissolved oxygen in fish ponds," African Journal of Environmental Science and Technology, vol. 4(9), pp. 625–638, 2010.

[15] N. Kaushik, B. Tyagi, and G. Jayaraman, "Modeling of the Dissolved Oxygen in a River with Storage Zone on the Banks," Applied Mathematics, vol. 3(07), pp. 699, 2012.