A WEB-BASED WATER QUALITY PREDICTION AND DECISION SUPPORT SYSTEM FOR THE EARLY DEVELOPMENTAL STAGE OF *HOLUTHURIA SCABRA* (JAEGER, 1833) UTILIZING BAYESIAN NETWORKS

MARY JANE MAGNO-TAN[1], JAMES VINCENT O. LABNAO[1], JERICO M. TANA[1], MIKE LEMUEL B. BACAYO[1], AND ARMIN S. CORONADO[2]

[1]*Department of Information Technology, College of Computer and Information Sciences, Polytechnic University of the Philippines*
[2]*Institute for Science and Technology Research, Office of the Vice President for Research, Extension and Development, Polytechnic University of the Philippines*

Abstract: *Holuthuria scabra* is a high valued marine species in the Philippines. During its early developmental stage, the larvae depends on the water quality of the larval tank in order to survive. But as the quality of the larvae tank changes unpredictably, only 2% of the larvae survived into a juvenile. This research aims to develop a decision support system (DSS) using Bayesian Networks, which predicts the water quality of the larval tank during the early developmental stage of *H. scabra* and provides suggestion on the amount of water to be added to the larval tanks to maintain an optimum water quality. Using Mean Absolute Percentage Error (MAPE) to compute the forecast accuracy (FA), results show 84.37%, 88.44% and 88.44% accuracy in forecasting the salinity level, dissolved oxygen level and temperature, respectively and an overall FA of 87.08%. The FA in predicting the salinity level is significantly lower than the FA in predicting the dissolved oxygen level and temperature due to lack of variation on the training data. One-way Analysis of Variance (ANOVA) show that there is no significant difference ($\propto 0.05 < 0.2576$) between the advice of the DSS and human experts in the amount of water needed to be added to the larval tank.

Keywords: Bayesian networks, decision support system, *Holuthuria scabra*, water quality prediction

1. INTRODUCTION

The high valued marine species *Holuthuria scabra*, also known as sea cucumber or sandfish, has been an important part of the multispecies invertebrate fishery in the Indo-Pacific regions (Akamine, 2001). It also remains an important source of income in the coastal areas of the Philippines (Juinio-Meñez *et al.*, 2012). According to the outcome of multi-sectoral national forum, there have been no noteworthy efforts to effectively regulate and manage the sea cucumber fishery, either at the national or local government level which results to the depletion in the population of the *H. scabra* (Casilagan & Juinio-Meñez, 2007). There are institutions and individuals who cultivate *H. scabra* under controlled condition in their own premises like the Philippines' Bureau of Fisheries and Aquatic Resources (BFAR). According to the sandfish cultivators in BFAR, early developmental stage is the most critical part of the cultivation of sandfishes wherein large number of fertilized eggs collapse in this stage because of stress. Larvae gets stressed whenever there is an intervention in the development process like changing of larval tank water. Water from larval tanks are regularly changed to maintain the optimum quality of water that is essential in the development and survival of sandfish larvae. One of the factors that affect the quality of water is the large number of larvae that collapse during the developmental process. This leads cultivators to remove the dead larvae through the process of siphoning. This process is also one of the factors that cause stress to the larvae of sandfish.

In the study conducted by Palani & Liong (2008), Artificial Neural Network (ANN) was applied in forecasting water quality parameters such as salinity level, dissolved oxygen, temperature and chlorophyll-a (Chl-a). Though the system showed a promising accuracy in forecasting the water quality parameter, the researchers recommended that the future work should use new types of algorithms that are more appropriate for time series forecasting. ANN is known to be a slow learner because of its complexity.

Conservation of marine organism can be successfully done if their larvae must be well-protected. In the study of George (2013), decision support system (DSS) was developed and used to give more effective management measures like closed area and season of fishery. Study showed the influence of environmental parameters on the biology of a given ecosystem, to track the larval transport and biological abundance in relation to the environmental variables, and to compare the biological abundance and fish larval transport in the three marine ecosystems.

Based on the study conducted by Lou *et al.* (2015), the Philippines is one of the countries that will suffer high level of water shortage by the year 2040. This implies that the method of frequent changing of water of larval tanks to obtain optimum water quality is not advisable knowing the potential crisis on water supply. This requires a system that can accurately monitor and predict water quality and provides advice when to change water and how much water should be added to the larval tanks.

This web-based DSS for the early developmental stage of *H. scabra* could help sandfish cultivators and address the water supply shortage issue. It could be done by providing an optimized water change schedule. This includes the amount of water with saline solution. This should be added on the larval tank to obtain an optimum water quality value, which is based on the patterns of the water quality changes found in the model.

This research involves utilizing the Bayesian Networks algorithm in the development of a software tool that predicts the water quality of the larval tank during the early developmental stage of *H. scabra*. The software gives suggestion on the volume of water to be added to the larval tank in maintaining the optimum water quality. Bayesian networks were used from inference to prediction to modelling whereas artificial neural networks are used for purely predicting (Wu, 2015). The system can predict the future water quality as well as model the past, current and future water quality.

2. METHODOLOGY

In developing the web-based DSS, the researchers used programming language (i.e. Python, Hypertext Preprocessor (PHP), JavaScript, Hypertext Markup Language

(HTML) and Cascading Style Sheet (CSS)), utilized Python Libraries (i.e. Scikit-Learn and NumPy), JavaScript Libraries (i.e. Highcharts) for the visualizations and MySQL for the database.

Initially, the DSS was trained to learn using Bayesian networks algorithm by feeding it with the available three (3) months data provided by the Philippine's BFAR Regional Maritime Techno Demo Center Region 1 (RMTDC-Reg 1) in Alaminos, Pangasinan. The training data consists of the salinity level, dissolved oxygen level and temperature of the water in the larval tank during the months of March to May of the year 2015. The larval tank contains 0.75 millions larvae of *H. scabra* in 750 liters of water.

To test the system and measure its accuracy in forecasting the water quality parameters (salinity, dissolved oxygen and temperature), the actual values of the water quality parameters from January 23, 2016 until February 21, 2016 (30-day duration of the early developmental stage of *H. scabra*) were compared to the forecasted value of the DSS. Mean Absolute Percentage Error (MAPE) was used to determine the forecast accuracy of the system in predicting the salinity, dissolved oxygen and temperature following the equation:

$$FE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100 \qquad (1)$$

where: *A* is the actual value, *F* is the forecasted value of the water quality parameter, and FE is the forecast error.

On the other hand, the forecast accuracy was calculated using the equation:

$$FA = 100\% - FE \qquad (2)$$

where: FE is the computed forecast accuracy

To determine if there is a significant difference between the advice of the DSS and human expert in suggesting volume of water to be added to the larval tank to obtain optimum values for the water quality parameters, the researchers compared the suggested volume of water given by the DSS and by the human expert during the 30-day test. One-way Analysis of Variance (ANOVA) was used to determine if there is a significant difference between the suggested volume of water to be added to the larval tanks by the DSS and by the expert.

Figure 1 presents the architecture of DSS developed to predict the future water quality parameters suggesting the volume of water to be added to the tanks to obtain an optimum water quality. The figure details the process flow from the input to the presentation of output of the system.

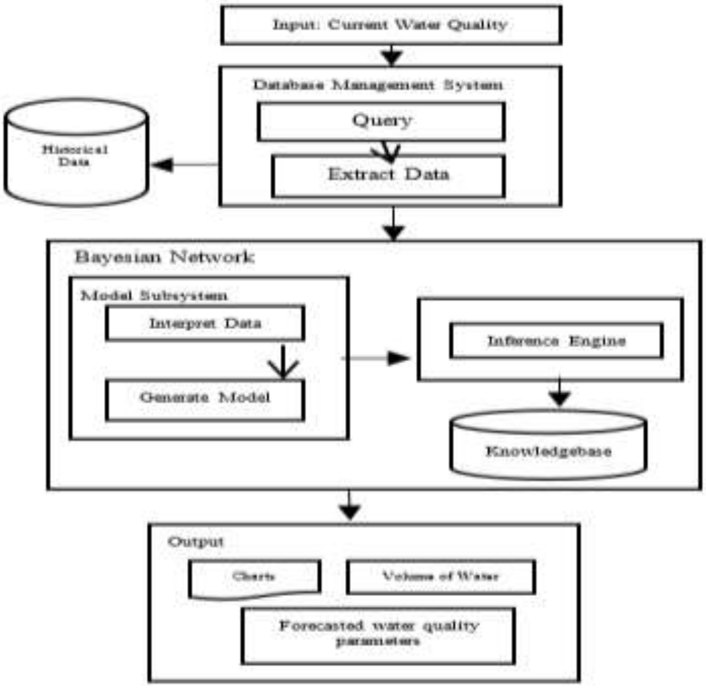

Figure 1. Architecture of the system.

The numeric value of the current water quality parameters (salinity, dissolved oxygen and temperature) are the input to the system. After reading the input water quality parameters, the system saved the inputs in the database as an additional historical data. These historical data were used as training data set utilized by the system for forecasting the future water quality parameters. The database management system module is responsible in insertion of the data into and extraction of the data from the database to produce a query result. Utilizing the Bayesian networks algorithm, these data were transmitted to the Model subsystem module which interprets the data extracted from the database; and generated a prediction model based on interpreted data, and further fed into the inference engine. The data were interpreted based on the knowledge base that contains the rules for the decisions in the cultivation of *H. scabra* during the early developmental stage. Then the system were produced the output which are the advice on the volume of water to be added to the larval tanks and the predicted water quality parameters. The prediction was displayed in the web as charts and graphs displaying the forecasted water salinity, temperature and dissolved oxygen.

3. RESULTS AND DISCUSSION

3.1 Measurement of forecast accuracy

Figure 2 shows the difference between the actual data gathered and forecasted value of the salinity level during the 30-day test. It can be seen from the graph that the actual values of salinity level are almost constant and has no variation; the first half of the line is straight indicating the same value; the next half showing a little variation in the values.

These actual (historical) data are used to train the system in forecasting the future salinity levels. Data with no variation are not fit to train a prediction model. As seen in the figure, the forecasted values of salinity has variations but no trend. It is noticeable in the figure that the forecasted values for salinity level is significantly lower than the actual values. It implies that the forecast is far from the actual values hence, is not considered accurate. As prediction model is only as good as the data used to train the algorithm in the prediction system, the accuracy in forecasting the salinity is negatively affected; hence, interpreting the variations in the forecasted salinity makes no sense.

Figure 3 shows the difference between the actual data gathered and forecasted value of the dissolved oxygen level during the 30-day test. It can be seen from the graph that the actual values of dissolved oxygen has variations. These actual (historical) data were used to train the system in forecasting the future dissolved oxygen levels. Data with variations are fit to train a prediction model. As seen in the figure, the forecasted values of dissolved oxygen has variations too and overlap with the actual values of dissolved oxygen, indicating that the values are almost similar. It implies that the forecast is very near the actual values, hence, is considered accurate.

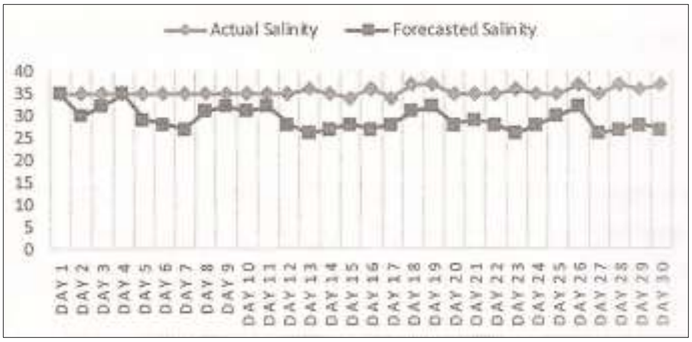

Figure 2. Comparison between the data of actual and forecasted salinity level.

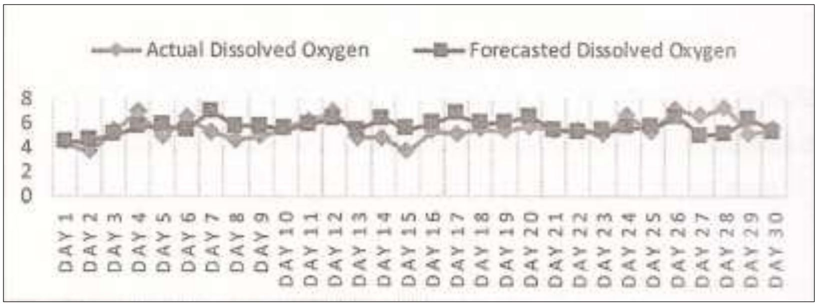

Figure 3. Comparison between the data of actual and forecasted dissolved oxygen level.

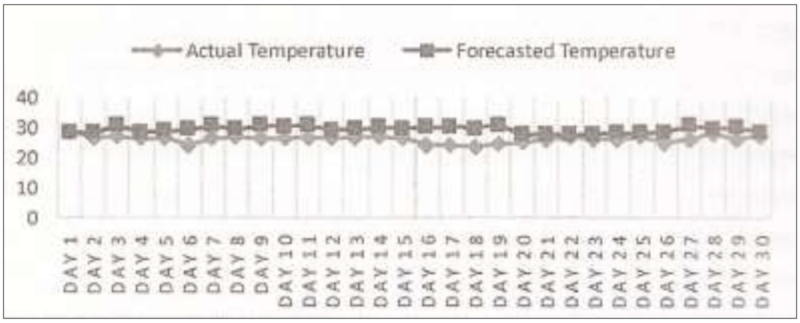

Figure 4. Comparison between the data of actual and forecasted temperature level.

Table 1. Forecast Percentage Error (FPE) and Forecast Accuracy (FA) of the Decision Support System (DSS) for *H. scabra*.

Water Parameters	FPE	FA
Salinity	15.63%	84.37%
Temperature	11.56%	88.44%
Dissolved Oxygen	14.59%	88.44%
Overall		**87.08%**

Figure 4 shows the difference between the actual data gathered and forecasted value of the temperature level during the 30-day test. It can be seen from the graph that the actual values of temperature has very little variation but in contrast to Figure 1, the straight line is only 10 % of the entire line and was observed in between variations, indicating that there were very little occurrence of same values in the actual (historical) data used to train the prediction model. Data with variations were fitted to train a prediction model. As seen in the figure, the forecasted values of temperature has very little variations too but they overlap with the actual values of temperature, indicating that the values are almost similar. It implies that the forecast is very near the actual values, hence, is considered accurate.

The figures show that the forecasted values for salinity level is significantly lower than the actual values. The forecasted values for dissolved oxygen level is closer to the actual data. Moreso, the forecasted values for temperature level is slightly higher than the actual values.

Using the Forecast Percentage Error (FPE) to determine the forecast accuracy, Table 1 shows that the accuracy of the system in forecasting the salinity level is significantly lower than that of temperature and dissolved oxygen. This is because the training data for the salinity level is less diverse than the other water quality parameters. Similar to the actual data where the values for salinity are almost the same and has no variation over the 30 days, the historical training data for salinity has no variation over the 3-month period. This affects the performance of the prediction model generated from the training data. Nevertheless, the overall forecast accuracy of the system is 87.08%.

Table 2. The absolute percentage error (APE) of the water quality parameters based on the advice of the expert and decision support system (DSS).

Day	Salinity (ppt)			Temperature (°C)			Dissolved Oxygen (mg/L)		
	Actual	**Forecast**	**APE (%)**	**Actual**	**Forecast**	**APE (%)**	**Actual**	**Forecast**	**APE (%)**
1	35	35	0	28.5	28.5	0	4.40	4.40	0
2	35	30	0.14	26.8	28.3	0.06	3.63	4.57	0.26
3	35	32	0.09	26.9	30.6	0.14	5.36	5.05	0.06
4	35	35	0	26.4	28.5	0.08	6.93	5.68	0.18
5	35	29	0.17	26.5	28.7	0.08	4.89	5.81	0.19
6	35	28	0.20	24.3	29.4	0.21	6.44	5.45	0.15
7	35	27	0.23	26.5	30.6	0.15	5.28	6.98	0.32
8	35	31	0.11	26.9	29.5	0.10	4.67	5.75	0.23
9	35	32	0.09	26.5	30.6	0.15	4.88	5.75	0.18
10	35	31	0.11	26.4	30.4	0.15	5.57	5.48	0.02
11	35	32	0.09	26.6	30.8	0.16	6.18	5.88	0.05
12	35	28	0.20	26.4	29.1	0.10	6.93	6.41	0.08
13	36	26	0.28	26.7	29.4	0.10	4.80	5.45	0.14
14	35	27	0.23	27.0	30.0	0.11	4.81	6.38	0.33
15	34	28	0.18	26.5	29.7	0.12	3.63	5.55	0.53
16	36	27	0.25	24.3	29.9	0.23	5.36	5.95	0.11
17	34	28	0.18	24.1	30.4	0.25	5.12	6.78	0.32
18	37	31	0.16	23.8	29.7	0.24	5.50	5.95	0.08
19	37	32	0.14	24.6	30.9	0.26	5.40	6.05	0.12
20	35	28	0.20	25.1	27.5	0.10	5.79	6.51	0.12
21	35	29	0.17	26.0	27.9	0.10	5.41	5.45	0.01
22	35	28	0.20	27.1	27.7	0.02	5.41	5.25	0.03
23	35	26	0.28	26.2	27.8	0.05	5.14	5.35	0.04
24	35	28	0.20	26.0	28.1	0.08	6.59	5.65	0.14
25	35	30	0.14	27.0	28.1	0.04	5.32	5.65	0.06
26	337	32	0.14	25.0	28.6	0.14	7.16	6.51	0.09
27	35	26	0.26	26.0	30.5	0.17	6.59	4.99	0.24
28	37	27	0.27	28.0	29.8	0.10	7.34	5.15	0.30
29	36	28	0.22	26.2	30.0	0.15	5.13	6.38	0.24
30	37	27	0.28	27.2	28.4	0.04	5.51	5.25	0.05
TAPE (%)			**15.63**			**11.56**			**14.58**

Table 2 shows the actual value and forecasted value of the water quality parameters (salinity, temperature and dissolved oxygen), the computed absolute percentage error (APE) and the total absolute percentage error (TAPE). It can be observed that the APE for both salinity and temperature ranges from 0 to .25 while the dissolved oxygen ranges from 0 to .52; all of which are considered very low. The TAPE for salinity, temperature and dissolved oxygen were 15.63%, 11.55% and 14.58%, respectively. Temperature has the lowest TAPE since the individual predicted values and actual values have almost no variation.

1.2 Measurement of significant difference

Table 3 shows the system advice (DSS) and the expert advice on the volume of water to be added to the larval tank. Analysis of Variance (ANOVA) revealed that there is no significant difference ($\propto_{0.05} < 0.2576$) between the advice of the DSS and of the human expert in the suggested volume of water to be added to the larval tanks per day to maintain an optimum water quality. Thus, the system advice can be used in the adjustment of water volume in the tank.

Table 3. The advice of the expert and of the decision support system (DSS).

Day	**Water Volume (L)**	
	Expert Advice	**System Advice**
1	167	167
2	167	0
3	167	67
4	167	167
5	167	200
6	167	400
7	167	600
8	167	34
9	167	67
10	167	34
11	167	67
12	167	400
13	200	800
14	167	600
15	134	400
16	200	600
17	134	400
18	234	34
19	234	67
20	167	400
21	167	200
22	167	400
23	200	800
24	167	400
25	167	0
26	234	67
27	167	800
28	234	600
29	200	400
30	234	600

4. CONCLUSIONS

It is concluded that the advice of the DSS on the volume of water to be added per day is accurate and the overall forecast accuracy of the system is considered high. The Bayesian networks algorithm used is effective in time-series forecasting (i.e. predicting future water quality parameters based on historical values) and also in developing a decision support system that does not only forecast future values of water salinity, temperature and dissolved oxygen, but also models the prediction through graphs, and suggests volumes of water to be added to the larval tanks.

5. REFERENCES

Akamine, J. (2001). Holothurian exploitation in the Philippines: continuities and discontinuities. *Tropics,* 10 (4), 591– 607.

Bureau of Reclamation. (2013). Information required for short-term water management decisions outlined. *ScienceDaily*. Retrieved September 29, 2015 from www.sciencedaily.com/releases/2013/01/130109185924.htm.

Casilagan, I.L. & Juinio-Meñez, M.A. (2007). Documentation report of the National Forum on Sea Cucumber Fisheries Management, Dagupan, Pangasinan, Philippines. 6-9.

Juinio-Meñez M., Paña, M., de Peralta, G., Catbagan, T., Olavides, R., Edullantes, C & Rodriguez B. (2012). Establishment and management of communal sandfish (Holothuria scabra) sea ranching in the Philippines. *Marine Science Institute, University of the Philippines.*

Lou, T., Young, R., & Reig, P. (2015). Aqueduct projected water stress world ranking. *World Resources Institute.*

Palani S. & Liong S. (2008). An ANN Application for Water Quality Forecasting. *Marine Pollution Bulletin*, 56 (9),1586-97, doi: 10.1016/j.marpolbul.2008.05.021.

George, G. (2013). Fish larval transport in the coastal waters through ecological modelling. N*ational Institute of Oceanography.*

Wu, E. (2015). What is the difference between a Bayesian network and an artificial neural network? https://www.quora.com/What-is-the-difference-between-a-Bayesian-network-and-an-artificial-neural-network