International Journal of Innovation Scientific Research and Review
Vol. 03, Issue, 06, pp.1312-1315, June, 2021
Available online at http://www.journalijisr.com

Research Article

ISSN: 2582-6131

FACTORS AFFECTING BIOCHEMICAL OXYGEN DEMAND CONCENTRATION IN RAYONG RIVER

Jatupat Mekparyup and * Kidakan Saithanu

Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.

Received 13th April 2021; Accepted 14th May 2021; Published online 12th June 2021

ABSTRACT

This research focused on BOD estimation in Rayong river, Rayong province, Thailand. Research methodology used in this study was regression analysis technique using 9 chemical factors, dissolved oxygen, temperature in water, pH, total phosphorus, nitrate-nitrogen, nitrite-nitrogen, ammonia, suspended solids and total dissolve solution, to generate models separated four periods, January to March, April to June, July to September and October to December. Data of 319 samples was divided into two groups; training and evaluation groups. The first group was to build the regression models and the other was to validate the achievement of forecasting BOD concentration in each period. After fitting the models, assumptions of regression analysis were detected by Anderson-Darling statistic, Durbin-Watson statistic, Breusch-Pagan statistic and Variance inflation factor. Finally, the performance of forecasting BOD values was calculated by mean absolute percentage error, mean absolute error, root mean squared error and mean squared error. The results found that the regression standard errors each period model wererespectively0.6761, 0.4446, 0.2932 and 0.4743with the adjusted coefficients of determination of 0.430, 0.517, 0.306 and 0.536. The best performances of fore casting accuracy were mean absolute percentage error of 2.4278 in the first period, mean absolute error of 3.7210 in the second period, mean absolute error of 4.4017 in the third period and mean absolute percentage error of 2.5738 in the fourth period.

Keywords: water quality index, biochemical oxygen demand.

INTRODUCTION

Rayong river, one of the important rivers of Rayong Province, consists of 6 water quality monitoring stations. In the Rayong river, water quality measured by the water quality index (WQI) are mainly in bad level since the year 2000 – 2018shown in Figure 1 (Water Quality Management Division, 2021).

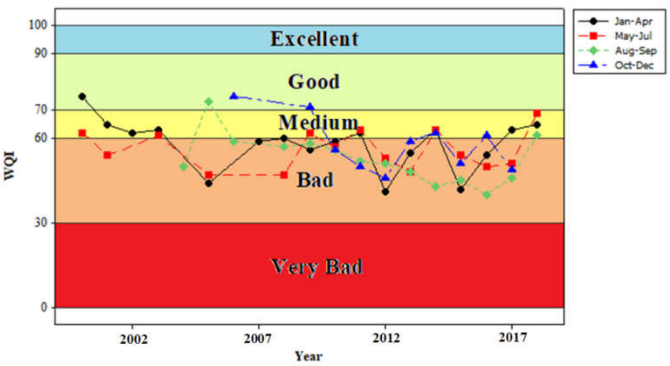

Figure 1:The WQI values in Rayong river of 4periods,January – March ,April–June, July–September and October – December.

Five factors used to calculate the WQI are dissolved oxygen (DO),total coliform bacteria (TCB), fecal coliform bacteria (FCB) ammonia (NH_3-N) and biochemical oxygen demand (BOD). BOD is one of the key factors that characterize water quality due to BOD indicates the amount of oxygen consumed by bacteria and other microorganisms while they decompose organic matter under aerobic (oxygen is present) conditions at a specified temperature (Office of the Environment Region 6, 2011; U.S. Geological Survey, 2021). Higher the BOD concentration, lower the WQI. Studying most of the research, it was found that DO,NH_3-N, temperature in water, pH, total phosphorus (TP), nitrate-nitrogen (NO_3-N), nitrite-nitrogen (NO_2-N), suspended solids (SS) and total dissolve solution (TDS) were used as factors to forecast BOD concentration (Dogan et al., 2009; Basant et al., 2010; Chen and Liu, 2015; Najafzadeh and Ghaemi, 2019; Tao et al., 2019; Salih et al., 2021). Main purpose of this research is to predict BOD concentration using 9 chemical factors mentioned above. Regression analysis technique was applied to generate models to prevision BOD concentration trend in Rayong river located the East of Thailand.

***Corresponding Author: Kidakan Saithanu,**
Burapha University, Department of Mathematics, Faculty of Science,169 Muang, Chonburi, Thailand.

MATERIALS AND METHODS

Secondary data was provided by Water Quality Management Division, Pollution Control Department, Ministry of Natural Resources and Environment since 2000 to 2018 divided into two sets, training and evaluation data sets. Total water sample was 319 gathered from 6 water quality monitoring stations in Rayong river. In each station, sample water was collected by 4 periods a year, January – March, April – June, July – September and October – December shown in Table 1.

Table 1: Data summary

Water monitoring period	Training data set		Validation data set		Total
	n	%	n	%	
First(January–March)	72	75.00	27	25.00	96
Second(April–June)	55	69.62	24	30.38	79
Third(July–September)	60	71.43	24	28.57	84
Fourth(October–December)	42	70.00	18	30.00	60

The forecasting model consists of 10 chemical factors which were BOD, DO, water temperature (Temp), pH, TP, NO_3-N, NO_2-N,NH_3-N,SS and TDS. A response variable was BOD and the remaining factors were the predictor variables as of equation (1)

$$\mathrm{BOD}_i = \beta + \beta_{\mathrm{DO}}\mathrm{DO}_i + \beta_{\mathrm{Temp}}\mathrm{Temp}_i + \beta_{\mathrm{pH}}\mathrm{pH}_i + \beta_{\mathrm{TP}}\mathrm{TP}_i + \beta_{\mathrm{NO_3\text{-}N}}\mathrm{NO_3\text{ - }N}_i + \beta_{\mathrm{NO_2\text{-}N}}\mathrm{NO_2\text{ - }N}_i + \beta_{\mathrm{NH_3\text{-}N}}\mathrm{NH_3\text{ - }N}_i + \beta_{\mathrm{SS}}\mathrm{SS}_i + \beta_{\mathrm{TDS}}\mathrm{TDS}_i + \varepsilon_i \quad (1)$$

where $i = 1,\ 2,\ 3,\ 4$ (1 for the first period, 2 for the second period, 3 for the third period and 4 for the fourth period), β isa *y*-intercept of the regression line, $\beta_{\mathrm{DO}},\ \beta_{\mathrm{Temp}},\ \ldots,\ \beta_{\mathrm{TDS}}$ represent for the regression coefficients and εstands for error term consequently.

Firstly, the linear relationship between BOD and 9 chemical factors were checked by Pearson correlation. After detected correlation, four models were fitted by regression analysis technique then the assumptions were validated with Anderson-Darling statistic (AD) (Anderson and Darling, 1952), Durbin-Watson statistic (DW) (Durbin and Watson, 1950), Breusch-Pagan statistic (BP) (Breusch and Pagan, 1979) and Variance inflation factor (VIF) for detecting normality, auto-correlation, homoscedasticity and multi collinearity respectively. Finally, forecasting accuracy was assessed using 4 key performance indexes which were mean absolute percentage error (MAPE), mean absolute error (MAE), root mean squared error (RMSE) and mean squared error (MSE).

RESULTS

Pearson correlation coefficients between BOD and 9 chemical factors (DO, Temp, pH, TP, NO_3-N, NO_2-N, NH_3-N, SS and TDS) were calculated and illustrated in Table 2.The highest positive value was significantly detected in the fourth period between BOD and SS with 0.434 (p-value 0.004). In contrast, the highest negative value was remarkably found in the fourth period between BOD and TDS -0.315 (p-value = 0.042).

Table 2: Pearson correlation coefficients between BOD and the predictor variables.

Period	Predictor variables								
	DO	Temp	pH	TP	NO_3-N	NO_2-N	NH_3-N	SS	TDS
First	-0.159	0.063	0.054	-0.101	-0.005	0.264	0.167	0.238	0.166
	(0.181)	(0.600)	(0.654)	(0.399)	(0.965)	(0.025)	(0.162)	(0.080)	(0.163)
Second	-0.289	0.223	0.002	0.331	0.080	-0.163	0.241	0.238	0.175
	(0.032)	(0.101)	(0.991)	(0.014)	(0.562)	(0.234)	(0.076)	(0.080)	(0.202)
Third	-0.264	0.215	-0.112	0.294	0.229	0.425	0.249	0.257	-0.198
	(0.042)	(0.099)	(0.393)	(0.023)	(0.078)	(0.001)	(0.055)	(0.048)	(0.129)
Fourth	0.073	0.327	-0.201	0.310	0.352	0.390	0.429	0.434	-0.315
	(0.645)	(0.035)	(0.202)	(0.046)	(0.022)	(0.011)	(0.005)	(0.004)	(0.042)

p-values in parenthesis

Therefore, the equations in each period were generated by multiple regression analysis with stepwise method. The estimators of regression coefficients $(\hat{\beta}_{\mathrm{DO}},\ \hat{\beta}_{\mathrm{Temp}},\ \ldots,\ \hat{\beta}_{\mathrm{TDS}})$ and the regression equations of forecasting BOD in each period were exhibited as equation (2) – (5).

$$\widehat{\mathrm{BOD}}_{\mathrm{Jan-Mar}} = 1.8136 - 0.16047\mathrm{DO}_{\mathrm{Jan-Mar}} + 10.743\mathrm{NO_2\text{ - }N}_{\mathrm{Jan-Mar}} + 0.007509\mathrm{SS}_{\mathrm{Jan-Mar}} \quad (2)$$

$$\widehat{\mathrm{BOD}}_{\mathrm{Apr-Jun}} = 0.6234 + 1.3177\mathrm{NH_3\text{ - }N}_{\mathrm{Apr-Jun}} + 0.0034149\mathrm{SS}_{\mathrm{Apr-Jun}} \quad (3)$$

$$\widehat{\mathrm{BOD}}_{\mathrm{Jul-Sep}} = 2.2568 - 0.16596\mathrm{DO}_{\mathrm{Jul-Sep}} \quad (4)$$

$$\widehat{\mathrm{BOD}}_{\mathrm{Oct-Dec}} = -6.887 + 0.27838\mathrm{Temp}_{\mathrm{Oct-Dec}} + 3.799\mathrm{TP}_{\mathrm{Oct-Dec}} + 0.4996\mathrm{NH_3\text{ - }N} + 0.005182\mathrm{SS}_{\mathrm{Oct-Dec}} - 0.00001548\mathrm{TDS} \quad (5)$$

All models were tested by analysis of variance (ANOVA) after fitting regression models together with the standard error of regression (*S*) and the adjusted coefficients of determination (r^2_{adj}) displayed in Table 3.It was illustrated that four models were respectively appropriated with *F*-statistic values 17.12, 26.69, 19.09 and 10.46.The highest r^2_{adj} of 0.536 was found in the fourth period model with *s* 0.4743 while the lowest r^2_{adj} of 0.306 was established in the third period with *s* 0.2932.

Table 3: ANOVA

Period	Sum of square		Mean of square		*F*-statistic	*S*	r_{adj}^2
	Regression	Error	Regression	Error			
First	23.4817	27.8839	7.8272	0.4571	17.12 (0.000)	0.6761	0.430
Second	10.5507	9.0917	5.2754	0.1976	26.69 (0.000)	0.4446	0.517
Third	1.6417	3.4394	1.6417	0.0860	19.09 (0.000)	0.2932	0.306
Fourth	11.7708	8.0989	2.3542	0.2250	10.46 (0.000)	0.4743	0.536

p-values in parenthesis

Furthermore, all regression assumptions, normality, auto-correlation, homoscedasticity and multicollinearity, were validated by AD, DW, BP in Table 4 and VIF in Table 5. The results indicated that all assumptions were satisfied.

Table 4: Assumption testing

Period	Test statistics		
	AD	DW	BP
First	0.400 (0.354)	1.787 [1.696]	0.144 (0.986)
Second	0.294 (0.587)	1.801 [1.623]	1.141 (0.565)
Third	0.238 (0.769)	1.587 [1.553]	0.176 (0.675)
Fourth	0.493 (0.206)	2.294 [1.781]	1.075 (0.956)

p-values in parenthesis
upper critical values in brackets

Table 5: Variance Inflation Factor (VIF)

Period	Predictor variables								
	DO	Temp	pH	TP	NO_3-N	NO_2-N	NH_3-N	SS	TDS
First	1.0					1.0		1.1	
Second							1.0	1.0	
Third									
Fourth		1.0		1.0			1.1	1.2	1.1

Finally, the accuracy of prediction was presented with numerical display in Table 6 following and graphical display as of Figure 2. In Table 6, the most performance of forecasting BOD concentration was found in the first period with MAPE=2.4278, the second period with MAE=3.7210, the third period with MAE=4.4017 and the fourth period with RMSE=9.8642. Moreover, time series were plotted to compare between the real and forecasted data.

Table 6: Accuracy of prediction

Period	Key performance			
	MAPE	MAE	RMSE	MSE
First	2.4278	4.1223	10.1340	102.6984
Second	6.2667	3.7210	10.4025	108.2115
Third	11.1559	4.4017	10.5251	110.7774
Fourth	2.5738	4.3891	9.8642	97.3016

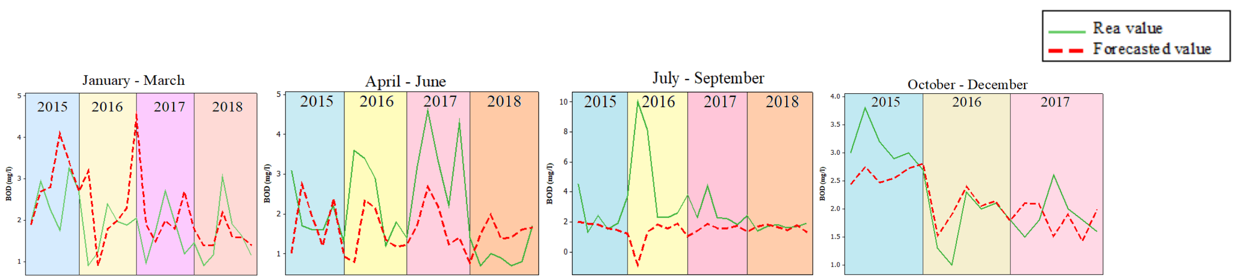

Figure 2: Time series plot between the real and forecasted values.

CONCLUSION AND DISCUSSION

Multiple regression models were generated to forecast BOD concentration in Rayong river in fourth periods, the first, the second, the third and the fourth. DO was the essential chemical factor to forecast BOD which found in equation (2) and (4) whileNH3-N was act in equation (3) and (5). Prediction BOD in the third period, July – September, was the least forecasting efficiency with MAPE=11.1559, MAE=4.4017, RMSE=10.5251 and MSE =110.7774.

ACKNOWLEDGEMENTS

Specially, the authors were appreciated to provide the secondary data from Water Quality Management Division, Pollution Control Department, Ministry of Natural Resources and Environment for this research.

REFERENCES

Abyaneh, H. Z. (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. Journal of Environmental Health Science and Engineering, 12(1), 40.

Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. The annals of mathematical statistics, 23(2),193212.

Basant, N., Gupta, S., Malik, A., & Singh, K. P. (2010). Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water—a case study. Chemometrics and Intelligent Laboratory Systems, 104 (2), 172-180.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological), 26 (2), 211-243.

Breusch, T. S.; Pagan, A. R. (1979). "A Simple Test for Heteroskedasticity and Random Coefficient Variation". Econometrica, 47 (5), 1287-1294.

Dogan, E., Sengorur, B., & Koklu, R. (2009). Modeling biological oxygen demand of the MelenRiver in Turkey using an artificial neural network technique. Journal of Environmental Management, 90(2), 1229-1235.

Durbin, J.; Watson, G. S. (1950). "Testing for Serial Correlation in Least Squares Regression, I". Biometrika, 37 (3–4), 409-428.

Chen, W. B., & Liu, W. C. (2015). Water quality modeling in reservoirs using multivariate linear regression and two neural network models. Advances in Artificial Neural Systems, 2015, 6.

Montgomery, D. C., & A skin, R. G. (1981). Problems of no normality and multicollinearity for forecasting methods based on least squares. AIIE Transactions, 13 (2), 102-115.

Najafzadeh, M., & Ghaemi, A. (2019). Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. Environmental monitoring and assessment, 191(6), 1-21.

Office of the Environment Region 6. (2011). It's about "water" in environment view, Available: http://reo06.mnre.go.th/newweb/index.php/2011-07-27-08-44-12/2011-08-04-07-38%2041/2011-08-04-08-02-46/730-2013-04-11-03-45-18. [Accessed February 5, 2019].

Salih, S. Q., Alakili, I., Beyaztas, U., Shahid, S., & Yaseen, Z. M. (2021). Prediction of dissolved oxygen, biochemical oxygen demand, and chemical oxygen demand using hydrometeoro logical variables: case study of Selangor River, Malaysia. Environment, Development and Sustainability, 23(5), 8027-8046.

Tao, H., Bobaker, A. M., Ramal, M. M., Yaseen, Z. M., Hossain, M. S., & Shahid, S. (2019). Determination of biochemical oxygen demand and dissolved oxygen for semi-arid river environment: application of soft computing models. Environmental Science and Pollution Research, 26(1), 923-937.

U.S. Geological Survey. (2021). Biological Oxygen Demand (BOD) and Water, Available: https://www.usgs.gov/special-topic/water-science-school/science/biological-oxygen-demand-bod-and-water?qt-science_center_objects=0#qt-science_center_objects. [Accessed June 1, 2021].

Water Quality Management Division, Pollution Control Department, Available:http://iwis.pcd.go.th/index.php?method=waterquality&etc=1550340427183.[AccessedApril 1, 2021].

Welsch, R. E., & Kuh, E. (1977). Technical Report 923-77: Linear Regression Diagnostics. Cambridge, MA: Sloan School of Management, Massachusetts Institute of Technology, 775.

****