DOI: 10.15625/vap.2019.000144

EVALUATING AND COMPARING TIME SERIES FORECASTING MODELS FOR WEEKLY FLUCTUATIONS OF SALINITY INTRUSION: THE CASE OF DAI ESTUARY, BEN TRE PROVINCE (SOUTHERN VIETNAM)

Thai Thanh Tran[1*], Ngo Xuan Quang[1, 2], Hoang Nghia Son[1,2]

[1]*Institute of Tropical Biology, Vietnam Academy of Science and Technology*

[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology*

ABSTRACT

Saltwater intrusion occurs naturally in a most coastal region; however it could harm a quality of local people. This phenomenon could be perceived with several models. Based on long-time database of salinity concentration collected in Dai estuary (belonging Mekong estuary systems), a number of models were comparated in order to selecting an adequate predictive model for prediction of salinity intrusion in the study area. Findings of this study demonstrated a nonseasonal/seasonal ARIMA (0,1,1)x(0,1,1)23 was a suitable model rather than the others. Outcomes of this study are useful for water resource operation, management in estuaries, and salinity monitoring as well. Our propose that ARIMA (0,1,1)x(0,1,1)23 can be applied in a part of early-warning of salinity intrusion systems.

Keywords: ARIMA, Ben Tre Province, Mekong Delta, salinity intrusion, time series forecasting.

1. INTRODUCTION

The Mekong River Delta also considered as the world's third-largest delta, high populated, known as Southeast Asia's most crucial food basket, and rich in biodiversity for all the world (Anthony et al., 2015). Covering approximately 12% of Vietnam's total land areas, it availability plays a crucial role in Vietnamese agriculture, provides 50% of the country's food and proud to be a "rice bowl" of Vietnam (Ikemoto et al., 2008). The delta now faces several significant sustainability challenges; notably salinity intrusion (SI) in main river is a major concern and frequently occurred in the Mekong Delta of Vietnam (SIWRR, 2015). The intrusion of saltwater is regarded as hugely productive agriculture (Kotera et al., 2008).

The coastal province Ben Tre is one of the provinces in the delta that is affected by SI. There are four main rivers in Ben Tre Province (BTP) including My Tho, Ba Lai, Ham Luong, and Co Chien River. My Tho River flows down from the north of Binh Dai District, the river widens from 50–60 m in Binh Dai up to 1 km at Dai Estuary (DS). Annual dry season, saline water from the DS-Tien River intrudes a upstream site of Ba Lai River (as a fresh water reservoir of BTP in the dry season) through An Hoa canal (particularly in the dry season) (Tran et al., 2019). To deal with the delta's annual SI, it is crucial to have a study for accurate predictions of SI in the main river or estuary.

In recent years some new models have been proposed to assist in the prediction of SI in rivers, and each one has its own advantages and disadvantages. There are a variety of models have been applicated such as random walk models, trend models, exponential smoothing, ARIMA models, etc.. However, a knowledge on what these models provide an adequate predictive model for forecast of salinity intrusion remains limited. Thus, the purpose of this paper presents some related works on the comparison of these models in order to select an adequate predictive model for prediction of salinity concentration in DS, BTP (Southern Vietnam).

2. MATERIALS AND METHOD

2.1. Study area and database collection

There is one salinity monitoring station in DS located in Binh Thang harbor, Binh Dai District. All of the models used the salinity monitoring data collected one time per week for a period of 23 weeks (from January to June). The river salinity monitoring data from 2012 to 2019 were available in the Center for Hydro-Meteorological forecasting of Ben Tre Province.

2.2. Predictive models description

The present study compared twenty models in order to find the most suitable model for the prediction of salinity intrusion in Dai estuary: Random walk models (1), Trend models (Constant mean (2), Linear (3), Quadratic (4), Exponential (5), S-Curve (6)), Exponential Smoothing (Simple exp. smoothing (7), Holt's linear exp. smoothing (8), Brown's linear exp. smoothing (9), Brown's quadratic exp. smoothing (10)), Nonseasonal ARIMA (ARIMA(2,1,1) (11), ARIMA(2,1,2) (12), ARIMA(1,1,0) (13), ARIMA(0,1,2) (14)), Nonseasonal/seasonal ARIMA(ARIMA(2,1,1)x(0,0,0)23 (15), ARIMA(2,1,2)x(0,0,0)23 (16), ARIMA(0,1,1)x(0,0,0)23 (17), ARIMA(1,1,0)x(0,0,0)23 (18), ARIMA(1,0,1)x(0,0,0)23 (19), ARIMA(0,1,1)x(0,1,1)23 (20)).

2.3. Criteria for choosing the adequate predictive models

According to Goh and Case (2016) and Stat (2017) fit models should be meet three general criteria: (i) small RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and AIC (Akaike Information Criterion) values; (ii) ME (Mean Error) and MPE (Mean Percentage Error) close to 0; (iii) a p-value for five test (RUNS-Test for excessive runs up and down, RUNM-Test for excessive runs above and below median, AUTO-Ljung-Box test for excessive autocorrelation, MEAN-Test for difference in mean 1st half to 2^{nd} half, VAR-Test for difference in variance 1^{st} half to 2^{nd} half) greater than or equal to 0.05, three (RUNS, RUNM, and, AUTO) of which were the most importance.

3. RESULTS AND DISCUSSION

3.1. Salinity intrusion in DS for the last 8 years (2012-2019)

The highest average salt concentrations (SC) (‰) of 22.46 is observed in 2016; the lower adjacent was 16.00 and the upper adjacent 27.16. The lowest SC, 18.89, was observed in 2012. The lower adjacent was 16.40 in week 1, and the upper adjacent was 23.50 in week 7 (Figure 1A). Due to a severe El Niño in 2016, SC in Dai estuary was considerably higher than the remaining period. Related to SC by week, the highest average SC of 24.22 is observed in week 9; the lower adjacent was 17.70 and the upper adjacent 25.58. The black square represented outliers, which mostly belonged to the lower adjacent. There were two lower and one upper outliers were 15.60, 11.13, and 22.57, which occurred in week 17, 20, and 21, respectively. The largest variation in SC was observed in week 2, 3, 4, and 11, which may be attributed to seasonal weather fluctuations in this week of the year (Figure 1B).

3.3. Testing forecast models

Model No. (15), (16), (20), and (15) had the smallest value of RMSE, MAE, MAPE, AIC. Menawhile, the value of ME and MPE in model No. (3) and (11) close to 0 as well (Table 1). Clearly, models of the ARIMA group had the statistical index smaller than the others. Related to the five forecast error test, it determined whether the residuals form a random time series. Among them, RUNS, RUNM, and AUTO test. Runs up and down/ above and below median (RUNS/RUNM) count the number of times the series goes up or down/ above or below its median. This number is compared to the expected value for a random time series. Small p-values indicate that the time series is not purely random. Box-Pierce Test or Ljung-Box test (AUTO) constructs a test statistic based on the first k residual autocorrelations. For either test, the test statistic is compared to a chi-squared distribution with k degrees of freedom. As with the other two tests, small p-values indicate that the residuals are not purely random. If the p-values for all three tests are well above 0.05, there will be no reason to doubt that the residuals are white noise (Stat, 2017).

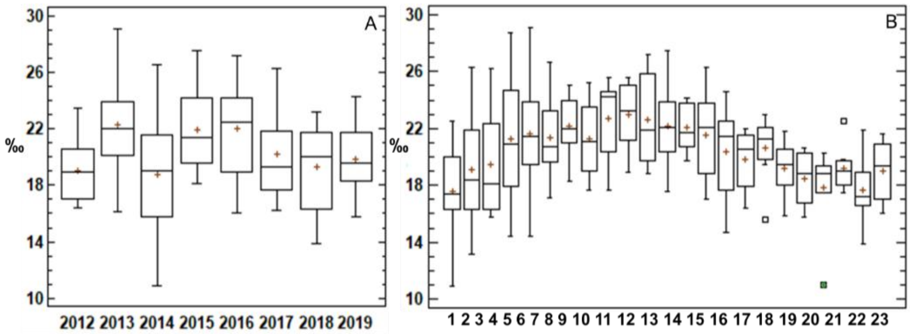

Figure 1. The salinity concentration in Dai estuary from 2012 to 2019. (A) Year, (B) Week.

Table 1 also presented that models of the Exponential Smoothing group (7-9, except for Brown's Quadratic Exp. Smoothing-10) and ARIMA group (11-20) passed three tests (mentioned above) while the others did not.

Table 1. Statistics is based on the one-ahead forecast errors and Tests for randomness of residuals. Bold indicated the highest value

	Statistics is based on the one-ahead forecast errors						Tests for randomness of residuals				
Model	RMSE	MAE	MAPE	ME	MPE	AIC	RUNS	RUNM	AUTO	MEAN	VAR
1	3.00266	2.22117	11.2545	0.005109	-1.20105	2.199	N.S	*	**	N.S	*
2	3.43892	2.75929	14.1258	8.60E-15	-3.02602	2.48481	***	***	***	*	N.S
3	3.44218	2.74911	14.0729	**9.83E-15**	-3.00724	2.5012	N.S	***	***	N.S	N.S
4	3.41578	2.67974	13.7708	1.28E-14	-2.96437	2.50029	*	***	***	N.S	N.S
5	3.4552	2.76354	13.9429	0.293635	-1.54518	2.50874	N.S	***	***	N.S	N.S
6	3.43609	2.72678	13.7605	0.291065	-1.53459	2.49766	N.S	***	***	*	N.S
7	2.80772	2.11626	10.6826	0.015198	-1.37038	2.07924	N.S	N.S	N.S	N.S	*
8	2.84377	2.13694	10.9318	-0.32070	-3.02893	2.11925	N.S	N.S	N.S	N.S	*
9	3.01711	2.25784	11.4083	0.00633	-1.08099	2.22309	N.S	N.S	N.S	N.S	**
10	3.18673	2.39256	12.0852	-0.00182	-0.96328	2.33248	N.S	*	*	N.S	**
11	2.71336	2.09942	10.5503	0.209404	**-0.74844**	2.03986	N.S	N.S	N.S	N.S	*
12	2.7125	2.09592	10.5386	0.199849	-0.80028	2.05371	N.S	N.S	N.S	N.S	*
13	2.83195	2.13191	10.7643	0.014574	-1.25583	2.09642	N.S	N.S	N.S	N.S	*
14	2.82711	2.12777	10.7345	0.018261	-1.37275	2.10749	N.S	N.S	N.S	N.S	*
15	**2.69949**	2.09056	10.5577	0.114544	-1.22044	**2.0296**	N.S	N.S	N.S	N.S	*
16	2.69883	**2.08578**	10.5300	0.123199	-1.18277	2.04361	N.S	N.S	N.S	N.S	*
17	2.81797	2.12352	10.7161	0.017581	-1.34985	2.08653	N.S	N.S	N.S	N.S	*
18	2.83195	2.13191	10.7643	0.014574	-1.25583	2.09642	N.S	N.S	N.S	N.S	*
19	2.81618	2.11131	10.6142	0.113454	-0.87648	2.09975	N.S	N.S	N.S	N.S	*
20	2.85485	2.26617	**11.4161**	-0.06658	-1.86624	2.12702	**N.S**	**N.S**	**N.S**	**N.S**	**N.S**

*N.S = not significant ($p \geq 0.05$), * = marginally significant ($0.01 < p \leq 0.05$), ** = significant ($0.001 < p \leq 0.01$), *** = highly significant ($p \leq 0.001$).*

Therefore, the Exponential Smoothing and ARIMA might be quite models to forecast the SI in DS. Specially, there was only model No. (20) nonseasonal/seasonal ARIMA(0,1,1)x(0,1,1)23 passed five tests. Combinedly, this model had the statistical index (RMSE, MAE, MAPE, AIC) with quite small and ME, MPE close to 0 as well. Although model No. (20) had not the smallest these statistic indices, but it passed five forecast error test. This lead to the nonseasonal/seasonal ARIMA(0,1,1)x(0,1,1)23 model considered a accuracy, suitability, adequacy, and timeliness of the collected data. Therefore, this model might be the adequate predictive model for prediction of salinity intrusion in DS.

4. CONCLUSION

The result showed that nonseasonal/seasonal ARIMA (0,1,1)x(0,1,1)23 provided an adequate predictive model rather than the others. Because of little data requirements, simplicity, computational efficiency, etc., this model might be used as the part of the early-warning of salinity intrusion systems.

REFERENCES

[1]. Alsharif, M. H., Younes, M. K., & Kim, J. (2019). Time series arima model for prediction of daily and monthly average global solar radiation: The case study of seoul, south korea. *Symmetry*, *11*(2), 240.

[2]. Goh, Y. M., & Case, K. (Eds.) 2016. Advances in Manufacturing Technology XXX: Proceedings of the 14th International Conference on Manufacturing Research, Incorporating the 31st National Conference on Manufacturing Research, September 6–8, 2016, Loughborough University, UK (Vol. 3). IOS Press.

[3]. Ikemoto, T., Tu, N. P. C., Okuda, N., Iwata, A., Omori, K., Tanabe, S.,... & Takeuchi, I., 2008. Biomagnification of trace elements in the aquatic food web in the Mekong Delta, South Vietnam using stable carbon and nitrogen isotope analysis. Archives of environmental contamination and toxicology, 54(3), 504-515.

[4]. Kotera, A., Sakamoto, T., Nguyen, D. K., & Yokozawa, M., 2008. Regional consequences of seawater intrusion on rice productivity and land use in coastal area of the Mekong River Delta. Japan Agricultural Research Quarterly: JARQ, 42(4), 267-274.

[5]. SIWRR (Southern Institute of Water Resources Research), 2015. Saline intrusion status in the Mekong Delta report. Ho Chi Minh Southern Institute of Water Resources Research.

[6]. Statgraphics Centurion 18, 2017. The User's Guide to STATGRAPHICS® Centurion 18, StatPoint Technologies.

[7]. Tran, T. T., Le, Q. L. N., Le, H. D., Nguyen, T. M. Y., & Ngo, X. Q (2019). Intertidal meiofaunal communities in relation to salinity gradients in the Ba Lai river, Vietnam. *Journal of Vietnamese Environment*, *10*(2), 138-150.