ABSTRACT

Introduction: Ham Luong River is a branch of Mekong River located in Ben Tre Province, which has played a crucial role in supporting livelihoods of local residents and the province’s economic development. However, the saline intrusion has been expanding in Ham Luong River, which seriously affects the productive agriculture, aquaculture, and further causes tremendous difficulties for local people’s lives. Thus, it is crucial to have research for forecast the saline intrusion in Ham Luong River. Our aim was to develop mathematical models in order to forecast the saline intrusion in Ham Luong River, Ben Tre Province. Methods: The Auto regressive integrated moving average (ARIMA) model was built to forecast the weekly saline intrusion in Ham Luong River, which has been obtained from Ben Tre Province’s Hydro-Meteorological Forecasting Center over eight years (from 2012 to 2019). Results: The saline concentration increased from January to March and then decreased from April to June. The highest salinity occurred in February and March while the lowest salinity was observed in early June. Moreover, the ARIMA technique provided an adequate predictive model for a forecast of the saline intrusion in An Thuan, Son Doc, and An Hiep station. However, the ARIMA model in My Hoa and Vam Mon might be improved upon by other forecasting methods. Conclusion: Our study suggested that the nonseasonal/seasonal ARIMA is an easy-to-use modeling tool for a quick forecast of the saline intrusion.

Key words: ARIMA model, climate change, Mekong Delta, saline intrusion, time-series forecasts

INTRODUCTION

Ham Luong River (HLR) (in Vietnamese: Sông Ham Luồng) is a branch of the Mekong River in the Mekong Delta region that flows entirely within Ben Tre Province (BTP). HLR has played a crucial role in supporting the livelihoods of local residents, giving a productive environment for agriculture, aquaculture, capture fisheries, non-fish aquatic goods, and tourism revenue. However, saline intrusion (SI) has been expanding in Mekong Delta, especially in BTP in recent years, which seriously affect the productive agriculture, aquaculture, and also causes tremendous difficulties for local people’s lives. In the dry season, the saline water from the East Sea has intruded into HLR, and after that continued intrusion into complicated canal networks in BTP. SI is a complex phenomenon depending on a variety of variables include freshwater discharge from upstream, capacity, and morphology of the rivers/canals, a configuration of the drainage network, tidal conditions, and presence of control artificial structures such as dams, sluice gates. Moreover, the impacts of climate change and sea-level rise also exacerbate the damage of SI. However, SI might be predicted by using statistical models. Therefore, it is crucial to have research for forecast SI in HLR in order to give useful information that can be used in water resource management and saltwater monitoring as well.

Nowadays, capabilities to predict SI was a principle of interest in many studies. Various models have been developed to predict SI in main rivers. An artificial intelligence model, like an Artificial Neural Network (ANN) model, simulate SI using a trained neural network. Remote sensing techniques, like resolution applications of available satellite images for detecting SI. However, these methods mostly rely on complex statistics, artificial intelligence techniques, and large amounts of meteorological and topographic data. This leads to needing a model that is reliability, accurate, suitability whereas small amounts of hydrodynamic. The Auto regressive integrated moving average (ARIMA) model is regarded as a smooth method, and it is applicable when the data is reasonably long and the correlation between past observations is stable. ARIMA model, also known as the Box-Jenkins model or methodology, is commonly used in forecasting and analysis. Some significant advantages of ARIMA forecasting are: first, it only needs endogenous variables and does not need to use other exogenous variables. Second, the ARIMA technique only
requires the prior data of a time series to generalize the forecast. Hence, it can increase the forecast accuracy while keeping the number of parameters to a minimum. This lead to the ARIMA model has been applied to analyze hydrological time series, especially at the monthly scale. Several studies in the literature have used the ARMA model for saline intrusion prediction. Sun and Koch (2001) used ARIMA to analyze and forecast of salinity in Apalachicola Bay, Florida. The results show that ARIMA has been possible to statistically define the interaction of different parameters that affect the salinity change in Apalachicola Bay provided help understand the hydrodynamic circulation of the water body through the approach of data analysis. Felisa et al. (2015) applied the ARIMA model to forecast the groundwater salinization in Ravenna (Italy). The resulting predictive models were validated by comparison with data and demonstrated that data-driven approaches may provide useful information in situations where physics-based models have only limited success in characterizing the phenomenon of interest. As well as this, the ARIMA model is a major technique in hydrology and has been used extensively, mainly for the prediction of natural phenomena such as precipitation, streamflow events, solar radiation. Here, our primary objective was to develop the ARIMA model to forecast the weekly SI of HLR, BTP in consideration of the accuracy, suitability, adequacy, and timeliness of a collected data, which have been obtained from Ben Tre Province’s Hydro-Meteorological Forecasting Center (BTHMFC) over eight years (from 2012 to 2019). The reliability, accuracy, suitability, and performance of the model are investigated in comparison with those of established tests, such as standardized residuals.

**MATERIALS AND METHODS**

**Study area and dataset collection**

HLR is separated from Tien River in Tan Phu Commune, Chau Thanh District, BTP, creating a natural border between Bao and Minh islet. It has 72 km long, from 12 to 15 m in-depth, and from 1,200 to 1,500 m (over 3,000 m at estuary) in width. During the rainy season, average river flows are approximately 3,300–3,400 m³/s, while around 800–850 m³/s in the dry season. There are six saltwater monitoring stations (from estuary to upstream) situated in An Thuan-AT (Tien Tom harbor, Ba Tri District), Son Doc-SD (Hung Le Commune, Giong Trom District), Phu Khanh-PK (Phu Khanh Commune, Thanh Phu District), My Hoa-MH (Ben Tre city), An Hiep-AH (An Hiep Commune, Chau Thanh District), and Vam Mon-VM (Phu Son Commune, Cho Lach District) (Figure 1). In each station, the saltwater monitoring data were collected one time per week for a period of 23 weeks (from January to June that is the dry season in Mekong Delta). The river saltwater monitoring data from 2012 to 2019 were provided by BTHMFC (available at http://www.bentre.gov.vn/Lists/ThongTinCanBiet/TongQuat.aspx). The present study forecast the SI in HLR from Jan 1st–Jan 8th (week 1) to Jun 4th–Jun 11st (week 23) of 2020 based on saltwater monitoring data from 2012 to 2019 (Appendix 1).

![Figure 1: Map of Ham Luong River and its saltwater monitoring stations.](image)

**ARIMA models description and application**

ARIMA was first formed by Box and Jenkin in 1976. The general equation of successive differences at the $d$th difference of $X_t$ is briefly expressed as follows: $\Delta^d X_t = (1 - B)^d X_t$, where $d$ is the different order, and $B$ is the backshift operator.

The successive difference at one-time lag equals to:

$$\Delta^1 X_t = (1 - B)X_t = X_t - X_{t-1}$$

In this situation, the general non-seasonal ARIMA $(p, d, q)$ is as follows:

$$\Phi_p(B)W_t = \theta_q(B)v_t$$

where $\Phi_p(B)$ is an autoregressive operator of order $p$, $\theta_q(B)$ is a moving average operator of order $q$, and $W_t = \Delta^d X_t$.

A general nonseasonal/seasonal ARIMA $(p, d, q)x(P, D, Q)s$ model with nonseasonal parameters $p, d, q$, seasonal parameters $P, D, Q$, and seasonality $s$ that consists of several terms: A nonseasonal autoregressive term of order $p$, a one seasonal differencing of order $d$, a nonseasonal moving average term of order $q$, and a seasonal moving average term of order $Q$.
q, a seasonal autoregressive term of order P, a seasonal differencing of order D, a seasonal moving average term of order Q. ARIMA(0,1,1)x(0,1,1)s seasonal and nonseasonal MA terms of order 1 which was a common nonseasonal/seasonal ARIMA model. For a more detailed description of the terminology, see Box and Jenkins (1976), Bowerman and O’Connell (1987), and Pankraz (1991).

ARIMA modeling was developed using Statgraphics Centurion ver. 18 software. Model performance was evaluated using the root mean squared error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean error (ME), the mean percentage error (MPE).

Map visualizations
An Inverse Distance Weighting (IDW) method in ArcGIS 10.3 was used to interpolate forecast point data to create continuous surface maps.

\[ l_i = \frac{\sum_{j=1}^{G} \lambda_j / D_{ij}^p}{\sum_{j=1}^{G} 1/D_{ij}} \]

where \( l_i \) was the property at location i; \( \lambda_j \) was the property at location j; \( D_{ij} \) was the distance from i to j; and \( G \) was the number of sampled locations, and was the inverse-distance weighting power.

RESULTS
Long-term saline intrusion data in Ham Luong River from 2012 to 2019
The saline concentration data in HLR for eight years that is obtained from the BTHMFC and Figure 2 presented the basic trends of the collected data. Overall, the saltwater concentration in HLR increased from February to April. The maximum saltwater occurred at the end of March or the beginning of April in which was the driest months in the year. Subsequently, the saltwater concentration decreased slightly in late May, and fell rapidly in early June because of the seasonal change with rainfall in May. In early June, it is the beginning of the rainy season with much rainfall than those in May; therefore, the saline concentration decreased rapidly in the whole river. Notably, the highest saltwater concentration in HLR was observed in 2016 because of a severe El Niño, BTP experienced serious SI. The maximum saltwater concentration was 31.50‰ (05/02/2016), 26.01‰ (03/12/2016), 14.50‰ (03/12/2016), 12.40‰ (03/05/2016), 9.90‰ (03/12/2016), and 6.70‰ (03/05/2016) observed in AT, SD, PK, MH, AH, and VM, respectively. Saltwater (approximately 10‰) expanded through HLR by up to 50-60 km, considered to be the most extensive SI in the last 90 years.

The ARIMA model for the forecast of saline intrusion in Ham Luong River
In AT station, the highest saline concentration of 25.34‰ is observed in week 6, followed by 21.25‰ (week 10) and 21.16‰ (week 9). Furthermore, week 12 was expressed as the highest saltwater concentration (13.24‰), week 5 (8.95‰), week 12 (4.67‰), week 4 (1.68‰), week 11 (0.72‰). By contrast, the lowest saltwater concentration of 12.46‰ is observed in week 23. The saltwater concentration measured from 5.09 (week 22) to 13.24 (week 12), 4.31 (week 22)-9.40 (week 12), 1.61 (week 22) to 4.67 (week 12), 0.00 (week 22)-1.49 (week 12), and 0.00 (week 22)-0.72 (week 11) in SD, PL, MY, AH, and VM, respectively. Clearly, at the beginning of the rainy season (from May 28th to Jun 11th) observed with the lowest saltwater concentration. In turn, saline intrusion began in mid-March, saltwater entered deep to inland (Appendix 2). Table 1 showed an overview of the monthly average of the forecasted saltwater concentration for all stations in HLR from January to June 2020. Generally, the saltwater concentration increased from January to March and then decreased from April to June. The maximum saltwater occurred in February and March while the lowest saltwater was observed in early June. Figure 3 showed the historical data, the forecasts, and the forecast limits (95% P.I.)

Testing forecast models
A normal probability plot of the residuals can be displayed in Figure 4. If the residuals come from a normal distribution, they should fall close to the line. In fact, the residual plot in AT, SD, PK, AH showed some curvature away from the line while MH and VM did not. There are five tests have been run to determine whether or not the residuals form a random sequence of numbers. If a p-value for each test is greater than or equal to 0.05, we can not reject the hypothesis that the series is random at the 95.0% or higher confidence level. ARIMA forecasting model in AT, SD, PK, AH passed five tests while MH and VM did not (Table 2).

DISCUSSION
The perspective view of the saline intrusion in Ham Luong River in 2020 is predicted by the ARIMA model
At the beginning of the dry season (January), the saltwater levels of 10‰ will have occurred in a location where between Mo Cay Nam and Thanh Phu District, over 50 km away from Ham Luong estuary.
Figure 2: The trend of saline intrusion in Ham Luong River from 2012 to 2019.

Table 1: Monthly average saltwater concentration (‰) in Ham Luong River from January to June of 2020. For: Forecast, 95% (L/H): the 95% prediction interval (low/high)

| Month   | AT     | SD    | PK    |
|---------|--------|-------|-------|
| January | 19.54  | 10.90/28.18 | 10.96 | 1.38/20.54 | 8.38 | 1.99/14.77 |
| February | 20.98   | 10.14/31.83 | 12.29 | 0.07/26.24 | 8.61 | 0.14/17.47 |
| March   | 20.50   | 8.03/32.97  | 12.99 | 0.00/29.85 | 8.96 | 0.00/19.52 |
| April   | 17.71   | 3.64/31.78  | 10.79 | 0.00/30.41 | 7.67 | 0.00/19.87 |
| May     | 13.51   | 0.00/29.02  | 6.45  | 0.00/28.49 | 5.49 | 0.00/19.12 |
| June    | 12.46   | 0.00/28.72  | 5.17  | 0.00/28.45 | 5.50 | 0.00/19.88 |

| Month   | MH     | AH    | VM    |
|---------|--------|-------|-------|
| January | 3.92   | 0.00/8.10 | 0.85 | 0.00/3.84 | 0.29 | 0.00/2.28 |
| February | 3.97    | 0.00/10.30 | 1.05 | 0.00/5.61 | 0.54 | 0.00/3.53 |
| March   | 4.53   | 0.00/12.26 | 1.40 | 0.00/6.97 | 0.60 | 0.00/4.24 |
| April   | 3.62   | 0.00/12.66 | 0.67 | 0.00/7.20 | 0.17 | 0.00/4.43 |
| May     | 1.94   | 0.00/12.14 | 0.09 | 0.00/7.45 | 0.08 | 0.00/4.88 |
| June    | 2.40   | 0.00/13.18 | 0.00 | 0.00/7.78 | 0.08 | 0.00/5.15 |

Also, the saltwater levels from 5-10‰ will cover almost all of Giong Trom and half of Mo Cay Nam District. These districts in upstream such as Chau Thanh and Cho Lach District will be covered by under 2‰ (Figure 5A). Subsequently, at the driest month (February and March), saltwater will be intruded into an area within 60-70 km from the mouth of HLR; therefore all of Giong Trom and Mo Cay Nam District will be affected with the saltwater rate 10‰. Ben Tre City and a small part of Chau Thanh District will be covered by under 5‰ (Figure 5B, C). Finally, at the beginning of the rainy season (early June), saltwater will be pushed away from the inland. The saltwater levels of 10‰ will be observed in Ba Tri District, approximately 10km away from the estuary (Figure 5F). Based on the forecasting results of the ARIMA model, saltwater with 5‰ will be entered up to 60-70 km deep inland that means Ben Tre city (areas with the highest population) and Chau Thanh District (areas with large-scale fruit production) seems to be affected by SI. Outcomes of this study are useful for reducing damages caused by the saline intrusion in the Mekong.
Figure 3: Time sequence plot displays for saltwater concentration in Ham Luong River include the forecasts and the forecast limits.

Figure 4: Residual normal probability plot.

Table 2: Tests for the randomness of residuals. 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 2nd half, VAR = Test for difference in variance 1st half to 2nd half

| Test types | AT   | SD   | PK   | MH  | AH  | VM  |
|------------|------|------|------|-----|-----|-----|
| RUNS       | N.S. | N.S. | N.S. | N.S.| N.S.| N.S.|
| RUNM       | N.S. | N.S. | N.S. | N.S.| N.S.| N.S.|
| AUTO       | N.S. | N.S. | N.S. | *   | N.S.| *   |
| MEAN       | N.S. | N.S. | N.S. | N.S.| N.S.| N.S.|
| VAR        | N.S. | N.S. | N.S. | N.S.| N.S.| *   |

N.S. = not significant (p ≥ 0.05), * = marginally significant (0.01 < p ≤ 0.05)
The ARIMA model: advantages and disadvantages

Forecast is an activity to calculate or predict future events or situations, usually as a result of rational study or analysis of suitable data. The accurate information for saline forecast will become more and more difficult to predict due to climate change and extreme weather. In recent years, there are several quantitative forecast techniques available such as ARIMA models, Random walk models, Trend models, or Exponential Smoothing. Generally, ARIMA models are considered as statistical theory and mathematically complex techniques while the others are defined as simple prediction techniques. Therefore, the ARIMA model has been regarded as the most efficient prediction technique in hydrology. In the empirical research, many advantages of the ARIMA model were found and support it as a proper way in especially short-term time series forecasting. The ARIMA model requires fewer the prior data inputs to generalize the forecast, only needs endogenous variables and does not need to use other exogenous variables. Basically, this model is relatively more robust and efficient than other complex structural models in relation to short-run predictions. However, the main limitation of ARIMA is the lack of a deterministic cause. In addition, many traditional techniques for time series forecast, such as ARIMA, which assume that the series is generated from linear processes.
and as a result might be inappropriate for most real-world problems that are nonlinear. This problem has now been circumvented through large numbers of past data inputs, stochastic events, and the accuracy of past data inputs that must be enhanced.

CONCLUSION
This paper presents a new approach to forecasting the SI in HLR of the Mekong River systems based on ARIMA forecasting model. Our result showed that the nonseasonal/seasonal ARIMA (0,1,1)x(0,1,1)23 model has been applied successfully for the forecasting of SI in HLR. However, the ARIMA forecasting model in AH and VM could be improved upon by other forecasting methods or still ARIMA with other parameters. ARIMA model with its convenience, accurate forecasting, low data input requirement, and simple computational process, it is bound to obtain a good picture of the prediction of SI over the main river. This makes the analytical model a powerful tool to guide future adaptation management on climate change and also SI in the Mekong Delta.

LIST OF ABBREVIATIONS
AH: An Hiep
ANN: Artificial Neural Network
ARIMA: Auto regressive integrated moving average
AT: An Thuan
AUTO: Ljung-Box test for excessive autocorrelation
BTHMFC: Ben Tre Province’s Hydro-Meteorological Forecasting Center
BTP: Ben Tre Province
CI: confidence interval
HLR: Ham Luong River
IDW: Inverse Distance Weighting
MAE: Mean absolute error
MAPE: Mean absolute percentage error
ME: Mean error
MEAN: Test for difference in mean 1st half to 2nd half
MH: My Hoa
MPE: Mean percentage error
PK: Phu Khanh
RMSE: Root mean squared error
RUNM: Test for excessive runs above and below median
RUNS: Test for excessive runs up and down
SD: Son Doc
SI: Saline intrusion
VAR: Test for difference in variance 1st half to 2nd half
VM: Vam Mon

COMPETING INTERESTS
The authors declare that they have no conflicts of interest.

AUTHORS’ CONTRIBUTIONS
Thai Thanh Tran has contributed to collections, analyses, interpretation of data, and writing the manuscript. Luong Duc Thien has contributed to mapping visualizations and interpolation techniques. Ngo Xuan Quang and Lam Van Tan have supported data analyses and revising the manuscript.

ACKNOWLEDGEMENTS
This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.06-2019.51. Moreover, we are particularly grateful to editors and anonymous referees, who kindly provided the constructive and critical reviews of our manuscript. If you have any questions or if we can help you with more information, please feel free to let us know. We are here to assist you.

REFERENCES
1. Thach P, Doan T. Ben Tre Geography: Social Sciences Publishing House (in Vietnamese); 2001.
2. Tran TT, Ngo QX, Ha HH, Nguyen NP. Short-term forecasting of saline intrusion in Ham Luong river, Ben Tre province using Simple Exponential Smoothing method. Journal of Vietnamese Environment. 2019;11(2):43–50.
3. Hashimoto TR. Environmental Issues and Recent Infrastructure Development in the Mekong Delta: Review, Analysis and Recommendations with Particular Reference to LargeScale Water Control Projects and the Development of Coastal Areas. Sydney, Australia; 2001.
4. Nguyen AD, Savenjie H, Pham DN, Tang DT. Using Salt Intrusion Measurements to Determine the Freshwater Discharge Distribution over the Branches of a Multi-Channel Estuary: The Mekong Delta Case. Estuarine, Coastal and Shelf Science. 2008;77:633–645.
5. Nguyen PT, Koedsin W, McNeil D, Van TP. Remote sensing techniques to predict SI: application for a data-poor area of the coastal Mekong Delta, Vietnam. International journal of remote sensing. 2018;39(20):6676–6691.
6. Bhattacharjya RK, Datta B, Satish MG. Artificial Neural Networks: Approximation of Density Dependent Saltwater Intrusion Process in Coastal Aquifers. Journal of Hydrologic Engineering. 2007;12(3):273–282.
7. Yadav AK, Chandel SS. Solar radiation prediction using Artificial Neural Network techniques: A review. Renewable and Sustainable Energy Reviews. 2014;33:772–781.
8. Farhath ZA, Arputhamary B, Arockiam DL. A Survey on ARIMA Forecasting Using Time Series Model. Int J Comput Sci Mobile Comput. 2016;5:104–109.
9. Box GEP, Jenkins GM. Time series analysis: Forecasting and Control. San Francisco: Holden-Day; 1976.
10. Liu X, Zhang C, Liu P, Yan M, Wang B, Zhang J, et al. Application of Temperature Prediction Based on Neural Network in Intrusion Detection of IoT. Security and Communication Networks. 2018;Article ID 1635081:10.
11. Wang HR, Wang C, Lin X, Kang J. An improved ARIMA model for precipitation simulations. Nonlinear Processes in Geophysics. 2014;21(6):1159–1168.
12. Sun H, Koch M. Case Study: Analysis and Forecasting of Salinity in Apalachicola Bay, Florida, Using Box-Jenkins ARIMA Models. Journal of Hydraulic Engineering. 2001;127(9):718–727.
13. Felisa G, Ciriello V, Antonellini M, Federico VO, Tartakovsky DM. Data-driven models of groundwater salinization in coastal plains. Journal of Hydrology. 2015;531:187–197.
14. Adebiyi AA, Adewumi AO, Ayo CK. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. Journal of Applied Mathematics. 2014;Article ID 614342:10.
15. Alsharif M, Younes M, Kim J. Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea. Symmetry. 2019;11(2):240.
16. Rapid Integrated & Ecosystem-Based Assessment of Climate Change Vulnerability & Adaptation for Ben Tre Province, Vietnam. In: Tuan LA, Du V, Skinner T, editors. Completed under the ‘Global Cooperation on Water Resource Management’ (WWF and Coca-Cola) and the ‘Capacity building and sustainable production’ programme (WWF – DANIDA) by World Wildlife Fund for Nature (WWF); 2012.
17. Bowerman BL, Connell O, T R. Time series forecasting. Boston: Duxbury Press; 1987.
18. Pankraz A. Forecasting with dynamic regression models. New York: WileyInterscience; 1991.
19. Statgraphics Centurion XVI. The User’s Guide to STATGRAPHICS® Centurion XV, StatPoint Technologies; 2005.