

FORECASTING OF CHANGES IN SALINITY INTRUSION IN THE VIETNAMESE MEKONG DELTA BY THE COMBINED MODEL OF LSTM (Long Short-Term Memory) AND SRM (Sinusoidal Regression Model)

Uyen T. Huynh

*University of Economics and Law (UEL), Ho Chi Minh City,
Vietnam National University, Ho Chi Minh City (VNU HCM), Vietnam
e-mail: uyenht@uel.edu.vn*

Abstract

The salinity intrusion in the Vietnamese Mekong Delta (VMD) has become more complex and temporally heterogeneous. This could seriously threaten the livelihoods of local residents and agricultural activities. Therefore, the research was conducted by using a combined model of LSTM (Long Short-Term Memory) and SRM (Sinusoidal Regression Model) to assess the trends and anomalies of salinity intrusion, with a series of data collected from main stations in the VMD in the year of 2021. The findings showed that the combined model exhibited high predictive ($R^2 = 0.9299$, $MSE = 2.0861$, and $MAPE = 0.1276$) in forecasting the increasing and decreasing trends of salinity intrusion and effectively detecting anomalous variations. Consequently, these results could be helpful to policymakers in predicting and responding to future salinity intrusion and to likely widespread implications for other regions impacted by saline intrusion.

1 Introduction

The Mekong Delta, situated in the southernmost region of Vietnam, stands out as one of Southeast Asia's largest and most fertile deltas. This delta is profoundly shaped

Key words: Mekong Delta, Salinity intrusion, LSTM and SRM models.

2020 AMS Classification: 62M10; 62P12; 68T07; 62J05.

by the ebb and flow of the Mekong River and the tides from the East and West Seas, resulting in a complex hydrological regime. However, being a low-lying flatland crisscrossed by an intricate network of rivers, the Mekong Delta is highly susceptible to the influences of both upstream flows from the Mekong River and tidal regimes from the East and West Seas. Consequently, the region consistently faces salinity intrusion, particularly during the dry season from January to May when upstream flows are minimal (Hung and Tuan (2019)). Recent seawater intrusion into rivers and canals has led to early-season salinity, adversely affecting the livelihoods and agricultural activities of residents in various coastal provinces of the Mekong Delta. Projections indicate a worsening trend in the coming years. The primary drivers of increased salinity intrusion are believed to be the impacts of climate change, sea-level rise, and reduced upstream flows, negatively affecting freshwater supply and agricultural production (Vu et al. (2018b)).

A significant body of research has been dedicated to computing, predicting, and early warning systems for salinity intrusion in the Mekong Delta (MD). These studies evaluate the integrated effects of climate change and the rapid development of hydropower systems upstream, taking into account variations in rainfall, sea level rise, and upstream flow. Various methods, such as hydrological models assessing salinity levels in river basins and hydrodynamic models combined with meteorological and tidal forecasting models, have been employed to reveal salinity intrusion and hydraulic regimes in the estuarine region (Bhattacharjya et al. (2009)). In recent years, hydrodynamic numerical models, including MIKE 11 and MIKE 21, have been successfully applied to simulate flow and salinity intrusion in the Hau River and other major rivers in the VMD region (Vu et al. (2020), Hochreiter and Schmidhuber (1997)). However, forecasting salinity intrusion for river systems using numerical models remains challenging due to the vast area, dense river networks, numerous hydraulic activities, and limited or infrequently updated datasets. On the other hand, artificial neural network (ANN) and autoregressive integrated moving average (ARIMA) models have seen increased use in modeling water resource variables, showing significant developments over the past decade. Nonetheless, these models have limitations, particularly in focusing on a single variable (Trung et al. (2021)).

Therefore, this research proposes an alternative method for predicting salinity intrusion in the MD: a combination of Long Short-Term Memory (LSTM) and Sinusoidal Regression Model (SRM). The combined model aims to identify trends and detect abnormal salinity intrusion variations. The input data for the model were collected as a time series from two different locations in the MD, namely Xeo Ro and Rach Gia, throughout 2021.

2 Materials

2.1 Study area and data collection

The study area is a coastal area in the Mekong Delta. This study used salinity data at two monitoring hydrological stations (Xeo Ro and Rach Gia station) in Kien Giang province. Xeo Ro hydrological station is located on the Cai Lon River, at coordinates $09^{\circ}52'N$, $105^{\circ}06'E$, about 8,000 meters from the mouth of the river into Rach Gia Bay. Rach Gia monitoring hydrological station is located on the Kien River at coordinates $10^{\circ}00'N$, $105^{\circ}05'E$, not far from the mouth of the river into Rach Gia Bay, about 800 meters. The study area has natural rivers, including Giang Thanh River, Cai Lon River, and Cai Be River, which are large rivers emptying into the West Vietnam Sea. It significantly drains floods from the inland to the West Vietnam Sea. However, it is also susceptible to salinity intrusion from the West Vietnam Sea into the inland during the dry season. Salinity intrusion, which is caused by seawater flowing from the sea to inland when not enough fresh water flows to the estuaries, also causes problems for production and human health. The series of data used is hourly salinity value (from 1 to 23 hours daily) at Rach Gia and Xeo Ro salinity monitoring stations in the dry season in the year of 2021 (Fig.1). Hourly salinity data were also collected from the Hydro-Meteorological Centre of Southern Vietnam.

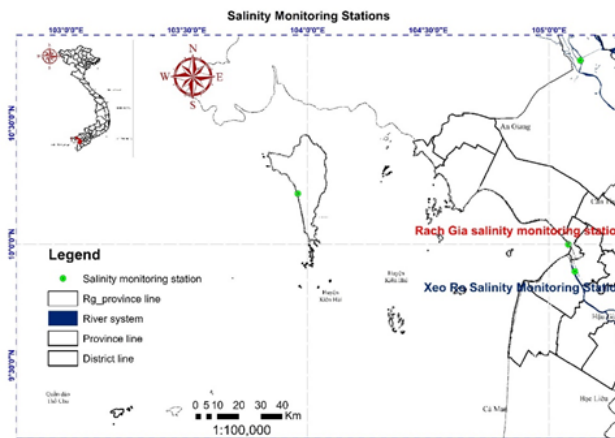


Figure 1: Salinity Monitoring Stations

2.2 Long Short-Term Memory Model (LSTM)

2.2.1 Model principle

Long Short-Term Memory (LSTM) plays a crucial role in time series forecasting across various domains (Gers and Schmidhuber (2000)). This model is founded on the architecture of Recurrent Neural Networks (RNN). Its power is a ability to comprehend and learn intricate relationships within sequential data. To overcome this challenge, the LSTM model incorporates gating mechanisms, allowing it to regulate the information flow from the past to the future (Kantoush et al. (2017)). This adaptive approach aids the model in capturing accurate trends, significantly enhancing the accuracy of salinity data predictions (Thai et al. (2021)).

The structure of each LSTM cell is comprised of three essential gates: the forget gate (f_t), the input gate (i_t), and the output gate (o_t) (Fig.2). The model uses these gates to control which information is retained, discarded, or transmitted to the next layer. This capability enables LSTM to retain and learn crucial details during sequence processing (Hoai et al. (2022)).

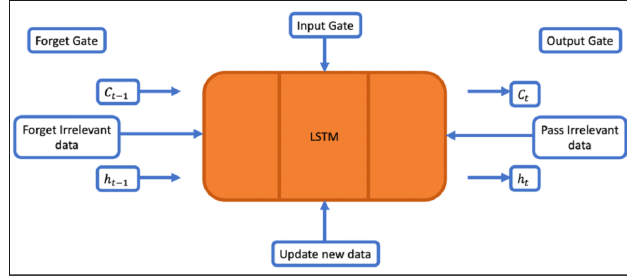


Figure 2

$$h_t = \sigma(W_x x_t + W_h \times h_{t-1} + b) \quad (2.1)$$

$$t_t = \sigma(W_{tx} x_t + W_{th} h_{t-1} + b_t) \quad (2.2)$$

$$o_t = \sigma(W_{ox} x_t + W_{oh} h_{t-1} + b_o) \quad (2.3)$$

The gate f_t Acts like a conveyor belt in the RNN model. However, the LSTM model possesses the capability to preserve crucial information as needed. This attribute is referred to as Long-Term Memory and is absent in the RNN model. Thus, the LSTM model encompasses both short-term and long-term memory properties.

2.2.2 Model Evaluation Metrics

In this research, the forecasting quality of the LSTM model is accessed by the following metrics (Hoai et al. (2022)):

- Pearson Correlation Coefficient (R^2): when this coefficient reaches 1, the LSTM model accurately reflects the data trend; otherwise, it fails. The formula for the correlation coefficient is as follows:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2.4)$$

- Mean Squared Error (MSE): this commonly used performance measure in prediction and regression tasks approaches 0 when the model's predictions are more accurate. An increasing value indicates a greater deviation from the actual data:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.5)$$

- Mean Absolute Error (MAE): Similar to MSE, MAE is a performance measure with properties resembling MSE, but outliers do not influence it:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

- Mean Absolute Percentage Error MAPE: This metric differs from MSE and MAE in that it does not consider the units of measurement of the models because the error is presented in the form of a percentage. When MAPE is closer to 0%, the model is better, and as it increases, the model deviates further from the actual data:

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2.6)$$

2.2.3 The Algorithm steps

- Step 1: Forget Gate (f_t):

The forget gate decides what information to discard from the cell state. It uses the current input x_t and the previous hidden state h_{t-1} :

$$f_t = \sigma(W_t[h_{t-1}, x_t] + b_f)$$

- Step 2: Input Gate (i_t)

The input gate decides what new information to store in the cell state. It consists of two parts: the gate layer that decides which values will be updated and the $\tilde{C}_t$ layer that creates a vector of new candidate values:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c). \end{aligned}$$

Step 3: Updating Cell State (C_t)

The new cell state C_t is a combination of the old cell state C_{t-1} , updated by the forget gate, and the new candidate values, scaled by the input gate:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Step 4: Output Gate (o_t)

The output gate decides what part of the cell state to output. It uses the current input x_t and the previous hidden state h_{t-1} :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

The final output h_t is based on the cell state passed through a tanh function, scaled by the output gate:

$$h_t = o_t * \tanh(C_t)$$

where σ : Sigmoid activation function, outputs between 0 and 1, $\tanh$: Hyperbolic tangent function, outputs between -1 and 1; W_f, W_i, W_C, W_o : Weight matrices for the forget, input, cell, and output gates; b_f, b_i, b_C, b_o : Bias vectors for the forget, input, cell, and output gates; h_{t-1} : Hidden state from the previous time step; x_t : Current input; C_{t-1} : Cell state from the previous time step; f_t : Forget gate output; i_t : Input gate output; $\tilde{C}_t$: Candidate cell state; C_t : Cell state; o_t : Output gate; h_t : Hidden state (output) at the current time step.

2.2.4. Model Setup

The parameters of the LSTM model in Xeo Ro and Rach Gia are discussed in Table 1. These parameters are based on the characteristics of the trends and seasons in each respective region:

The calculations were performed using Python 3.7 with the following libraries: Pandas and NumPy for data formatting; Keras, TensorFlow, and Scikit-learn for training and parameter computation. Graphs were plotted using R 4.2 via the `ggplot2` package. Input data consisted of hourly time series from 2021 at the Rach Gia and Xeo Ro sites as training data.

Table 1: LSTM model parameters

Parameter	Xeo Ro	Rach Gia
Number of hidden layers	128	64
Dropout	0.2	0.2
Learning rate	0.005	0.005
Batch size	16	32
Epochs	200	200

2.3. Sinusoidal Regression Model (SRM)

2.3.1. Model Principle

Data collected at Xeo Ro and Rach Gia were measured every 2 hours from 1 to 23 hours daily. Since the data exhibit time-dependent oscillations and wave-like patterns, a Sinusoidal Regression Model (SRM) over time ($t = 1, 3, 5, \dots, 23$) is used to predict salinity for new observations. The SRM is a trigonometric model that best fits a sinusoidal curve. Rather than fitting a straight line, this approach aims to enhance accuracy in modeling naturally occurring cyclic phenomena. Figures 3 (not shown here) depict daily salinity cycles at Xeo Ro and Rach Gia, highlighting daily fluctuations.

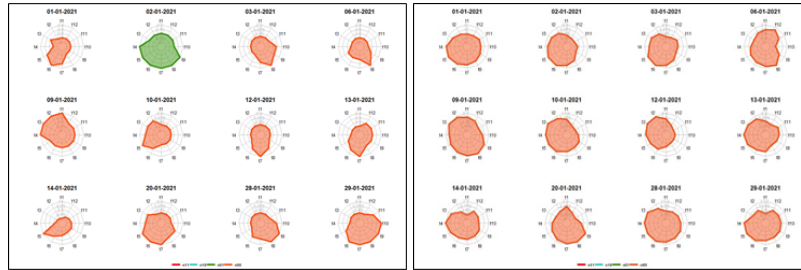


Figure 3: The daily salinity cycle at the Rach Gia and Xeo Ro site measured in January 2021

where $t = \{1, 3, \dots, 23\}$ represents the time of measurement during the day, and 12 measurement times during the day are selected sequentially as $t = \{t_1, t_2, \dots, t_{12}\}$ that why our Fig.3 have only 12 days

2.2.4 Model Setup

The Sinusoidal Regression Model (SRM) with salinity data measured at Xeo Ro and Rach Gia is set up as follows:

$$Y(m, t) = A \cdot \sin(\omega t + \phi) + C + \varepsilon(m, t)$$

where: $t = \{1, 3, \dots, 23\}$ represents the time of measurement during the day, and 12 measurement times during the day are selected sequentially as $t = \{t_1, t_2, \dots, t_{12}\}$, $m = \{1, 2, 3, \dots, 21\}$ denotes the measurement day in the month; A : amplitude; ω : angular frequency; ϕ : phase shift; C : vertical shift (baseline); $\varepsilon(m, t)$: error term.

The residual is determined as follows:

$$e(m, t) = Y(m, t) - \hat{Y}(m, t) \quad (2.7)$$

Since this is a nonlinear model, R^2 is not employed to evaluate the fit. Instead, the residuals are modified using the following transformation:

$$e^*(m, t) = [e(m, t) - \min(e(m, t))] \geq 0 \quad (2.8)$$

with the formulas for $e(m, t)$ and $e^*(m, t)$, the residuals are estimated by the model at Xeo Ro and Rach Gia. These are illustrated in Fig. 4.

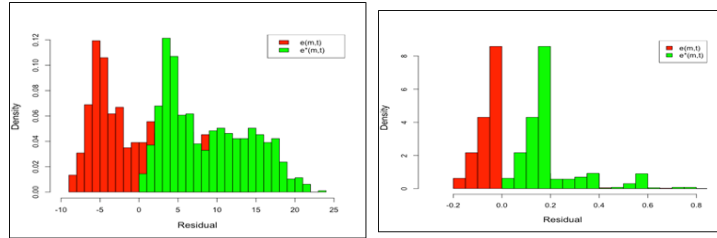


Figure 4: Histograms of $e(t, k)$ and $e^*(t, k)$ at Rach Gia

The Weibull distribution is used for testing and comparison that examines the suitability of $e^*(t, k)$ with a particular distribution. The results are demonstrated in Fig. 5.

In this case, the predictive model is estimated based on the formula:

$$\hat{Y}(m, t) - \min(e(m, t)) = \hat{A} \cdot \sin(\hat{\omega}t + \hat{\phi}) + \hat{C} + e^*(m, t) \quad (11)$$

In the context provided, Extreme Value Distributions are utilized to characterize the extreme values of salinity peaks observed during monitoring. Extreme value distributions are the limiting distributions for the minimum or the maximum of a very large collection of random observations from the same arbitrary distribution. In this study, focus is placed on the maximum or minimum salinity peak values during monitoring.

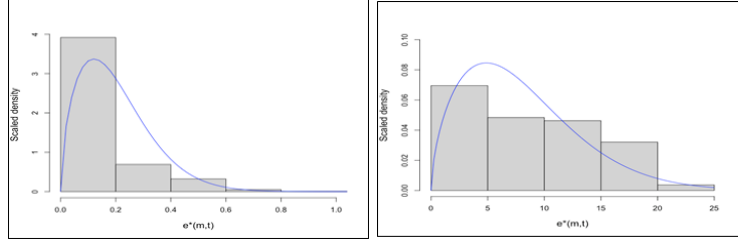


Figure 5: The suitability $e^*(t, k)$ of the Weibull Distribution at Rach Gia and Xeo Ro

The transformation of $e(t, k)$ into $e^*(t, k)$, where $e^*(t, k) \geq 0$ satisfies the condition of the Weibull distribution and applies it to find the estimated parameters. After estimating the parameters for formula (11), the SRM model's result is used to estimate the time intervals for the LSTM model.

3 Results and Discussion

3.1 LSTM Model

The errors shown in Table 2 resulted from the LSTM model using data collected at the Xeo Ro and Rach Gia sites with trained parameters.

Table 2: LSTM Model Training Results

Training Errors	Xeo Ro	Rach Gia
MSE	2.0861	0.1367
MAE	0.9555	0.0667
R^2	0.9299	0.6218
MAPE	0.1276	0.1681

Table 2 demonstrates the error metrics of the LSTM model for Xeo Ro and Rach Gia. The MSE, MAE, and MAPE values are impressively low, indicating excellent training performance. However, the R^2 for Rach Gia is 0.6218, lower than Xeo Ro's 0.9299, suggesting that the trend-fitting capability of the LSTM model in Rach Gia is less optimal compared to Xeo Ro.

Fig.6 illustrates forecasts of salinity fluctuations for the next year at the Xeo Ro and Rach Gia sites. The simulated and measured results showed a similar trend. However, there was a difference in the predicted salinity values between the Xeo Ro and Rach Gia sites. At the Xeo Ro site, the forecasting results aligned well with the measured data, indicating that the model obtained a fitting estimate of the cyclic variations in salinity. On the other hand, at the Rach Gia site, due to the high randomness in

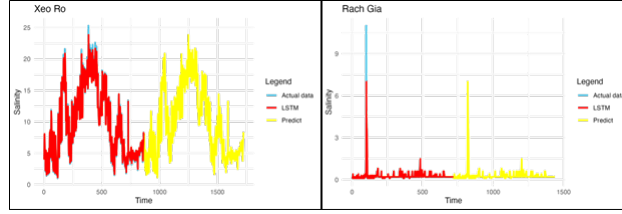


Figure 6: Training and Forecasting Results of the LSTM Model

the input data, the simulated and measured results agreed regarding the trend but did not match precisely regarding values. Thus, the model only produced predictions without accurately reproducing the salinity peaks observed in the previous year. The LSTM model showed training results, which were relatively good performance (Table 2 and Fig. 9). Particularly, considering the low values of MSE (2.0861) and MAE (0.9555) at the Xeo Ro site indicated that there was a relatively low deviation between the modeled and measured values. This could reflect the complexity of the data, which tended to change continuously but still followed the tidal cycle. Thus, the model could be able to provide reasonably accurate predictions. The high R^2 value at the Xeo Ro site (0.9299) illustrated that the LSTM model's training results conformed to the cyclic changes in salinity over time. Finally, the relatively low value of MAPE (0.1276) showed that the training results of the model had a significantly slight deviation compared to the measured data. In Rach Gia, the model results reveal discrepancies in the error metrics, as illustrated in Table 1 and Figure 9. The MSE (0.1367), MAE (0.0667), and MAPE (0.1681) errors in Rach Gia are relatively low, indicating a favorable performance, but the R^2 (0.6218) stands at a moderate level. The root cause lies in the time series data for Rach Gia, which fails to exhibit cyclic patterns and includes some unusual salinity peaks. Consequently, the time series of Rach Gia demonstrates a random nature. Therefore, the LSTM model has not yet achieved accurate forecasting of these salinity peaks (see in Fig. 9). Therefore, the model accurately simulated both the trend and the values, as indicated by a low MAPE of 0.1276, at the Xeo Ro site. While the Rach Gia site model simulated the cyclic pattern well, the measured and simulated results did not agree totally at the peak with an estimated value of around 11 %. Nevertheless, the model error remained small, with a MAPE of 0.1681. Therefore, this result could provide the basis for forecasting future changes in saline intrusion. In this study, the LSTM model was utilized for forecasting the upcoming year. The results are illustrated in Fig.9. The forecasted line demonstrates the trend and cyclic variations in salinity for the next year based on the input data and the LSTM model. The obtained results in Xeo Ro exhibit salinity variations that align with the input data characteristics. For Rach Gia, the forecasting results also indicate the presence of salinity peaks, but the cyclic variations are not clearly evident, differing from the input data.

2.4.2. SRM Model

To estimate these parameters, you can use a combination of linear and non-linear regression techniques. Then, we can get the estimate Vertical Shift (C), Amplitude (A), Angular Frequency (ω), and Phase Shift (φ) as follows:

$$\hat{C} = \frac{1}{n} \sum_{t=1}^n y(t), \hat{\alpha} = A \cos(\phi), \hat{\beta} = A \sin(\phi), \hat{A} = \sqrt{\alpha^2 + \beta^2}, \hat{\phi} = \arctan\left(\frac{\beta}{\alpha}\right)$$

The parameter estimation for the SRM model (equation 2.2) at the Xeo Ro and Rach Gia sites is presented in Table 3.

Table 3: Estimated Parameters of the SRM Model.

Estimated Parameters of the SRM Model	Xeo Ro	Rach Gia
$\hat{A}$	-0.6492*	0.0133*
$\hat{\omega}$	5.7697*	0.6549*
$\hat{\phi}$	-6.3647*	4.1319*
$\hat{C}$	9.7502*	0.2448*

*Statistically significant at the 0.05 significance level

All parameters in both measured salinity sites are statistically significant, indicating that the SRM model was well-suited for estimating salinity values in both Xeo Ro and Rach Gia sites. This could be because salinity varied cyclically with tidal patterns; the SRM model was built as a sinusoidal function that obtained these cyclical changes. Determining the $Y(m, t)$ cycle is crucial to finding precise parameters. Using the Weibull distribution for error instead of the Normal distribution corrected for the asymmetric error in the model.

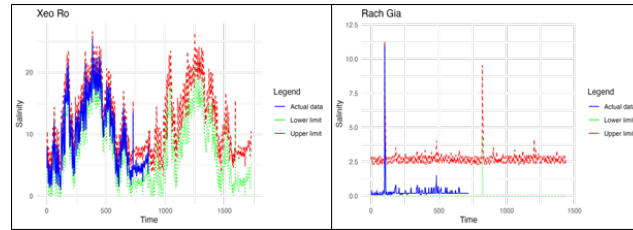


Figure 7: Combined Results of the SRM and LSTM Model

Using the Weibull distribution for error terms—rather than the normal distribution—corrects for asymmetry in model residuals.

Figure 6 illustrates forecasts of salinity fluctuations for the next year at the Xeo Ro and Rach Gia sites. The simulated and measured results showed a similar trend. However, differences in the predicted salinity values were observed:

- **Xeo Ro:** The forecasting results closely matched measured data, indicating that the model accurately captured cyclic salinity variation.
- **Rach Gia:** Due to high randomness in the input data, the predicted values aligned with the trend but did not precisely match the observed salinity peaks.

The LSTM model showed relatively good performance based on Table 2 and Figure 6:

These metrics reflect:

- Low deviation between modeled and measured values at Xeo Ro.
- Strong agreement with cyclic tidal patterns at Xeo Ro.
- Moderate agreement at Rach Gia due to irregular, non-cyclic salinity peaks.

While the LSTM model simulated the general trend well at both sites, it was more accurate at Xeo Ro. For Rach Gia, salinity peak predictions were less reliable, though errors remained relatively small.

2.4.3 Performance Comparison

Algorithmic Steps of Leave One Out of Bootstrap (LOOB)

Apply the Leave One Out of Bootstrap (LOOB) algorithm with $n = 67$ days. In LOOB, remove Day $-d$ and use the remaining 66 days to estimate A, w, ϕ, C and the parameters for the distribution of ε . Call these as $\hat{A}^{-d}, \hat{\omega}^{-d}, \hat{\phi}^{-d}, \hat{C}^{-d}$

$$\hat{Y}^{(-d)}(1, 1) = \hat{A}^{-d} \cdot \sin(\hat{\omega}^{-d} \cdot 1 + \hat{\phi}^{-d}) + \hat{C}^{-d} \quad (3.1)$$

$$\vdots$$

$$\hat{Y}^{(-d)}(1, 12) = \hat{A}^{-d} \cdot \sin(\hat{\omega}^{-d} \cdot 12 + \hat{\phi}^{-d}) + \hat{C}^{-d} \quad (3.2)$$

Obtain the observed residual with the first day (Day $-d$) removed as

$$e^{(-d)}(d, 1) = Y(d, 1) - \hat{Y}(d, 1) \quad (3.3)$$

$$\vdots$$

$$e^{(-d)}(d, 12) = Y(d, 12) - \hat{Y}(d, 12) \quad (3.4)$$

In this step, we can find the modified residuals and estimated parameter using MLE method, $d = 1, 2, \dots, D (= 67)$.

Then we find

$$\text{PRMSE} = \sqrt{\frac{1}{n} \sum_{d=1}^D \sum_{k=1}^{12} e^{(-d)}(d, k)^2} \quad (3.5)$$

$$\text{PMAE} = \frac{1}{D} \sum_{d=1}^D \sum_{k=1}^{12} |e^{(-d)}(d, k)| \quad (3.6)$$

where: $D = 67$ (days); $k = 1, 2, \dots, 12$ (hours); PRMSE: Prediction Root Mean Squared Error; PMAE: Prediction Mean Absolute Error.

Model Evaluation Results

Rach Gia			Xeo Ro		
Model	PRMSE	PMAE	Model	PRMSE	PMAE
LSTM	0.36	0.25	LSTM	12.05	8.14
SRM	0.14	0.09	SRM	5.64	4.86

Table 4: PRMSE and PMAE for Rach Gia and Xeo Ro using LSTM and SRM models

The research results revealed that the lower and upper bounds of the estimates included the abnormal salinity peaks that the LSTM model initially could not accurately predict (Fig.6). However, the current model did not provide individual point forecasts; instead, it generated ranges that consisted of lower and upper bounds, which still followed the pattern of measured values. The size of the forecast intervals varied depending on the cyclical variation in the trained data.

The rising salinity measurements were within the forecast range at the Rach Gia site. This range represented the highest variability in the forecasting results. In the final phase, as the level of variability decreased, the field became more consistent. The forecasting intervals continuously change. This could be because the data measured at the Xeo Ro site have fewer abnormal peaks; however, the measured values still exhibited continuous variations with small amplitudes. Therefore, the forecast intervals also had to follow these changing patterns. Thus, the SRM model coupled with the LSTM model has generated results that accurately illustrate the nature of the measured values. These results conformed to the cyclical patterns of the measurements and stabilized the value of sudden fluctuations.

The forecasting results in Fig.7, the prediction intervals provide salinity results within the next cycle. This allows the forecasting results to have higher reliability compared to using only the LSTM model. When the data lacks cyclicity, as in Rach

Gia, interval forecasting results indicate the range of fluctuations in the next step rather than a single value. Thus, the SRM model has addressed the previous limitation of the LSTM model when predicting for Rach Gia.

4 Conclusion

A salinity forecasting model has been developed based on 2021 monitoring data at the Xeo Ro and Rach sites using the Long Short-Term Memory (LSTM) model coupling with the Sinusoidal Regression Model (SRM). The computed results demonstrated that the developed model could simulate cyclical patterns and abnormal values well. By combining these two models, the research has addressed the limitation of being unable to model abnormal values in time series forecasting using only the LSTM model. This advantage allows the model to be widely applied to various environmental management-related data. Furthermore, the study proposed that the input data should be measured in a more extended time series, covering several years, which creates a more generalizable model suitable for multiple regions. In addition, the model can enhance accuracy by considering additional input variables related to hydrological factors.

Acknowledgement

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number C2025-34-06.

References

- [1] Bengio, S., & Bengio, Y. (2000). On the curse of dimensionality in joint distributions using neural networks. *IEEE Transactions on Neural Networks, Special issue on Data Mining and Knowledge Discovery*, 11(3), 550–557.
- [2] Gers, F. A., & Schmidhuber, J. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471.
- [3] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
- [4] Hoai, P. N., Quoc, P. B., & Thanh, T. T. (2022). Apply machine learning to predict saltwater intrusion in the Ham Luong River, Ben Tre Province. *VNU Journal of Science: Earth and Environmental Sciences*, 38(3). Retrieved from <https://doi.org/10.25073/2588-1094/vnuees.4852>

- [5] Hung, H. V., & Tuan, H. V. (2019). Using artificial neural networks to forecast tide-affected river water levels. *Journal of Water Resources Science and Technology*, 52.
- [6] Kantoush, S., Binh, D., Sumi, T., & La, T. (2017). Impact of upstream hydropower dams and climate change on hydrodynamics of the Vietnamese Mekong Delta. *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)*, 73(1), 109–114. https://doi.org/10.2208/jscejhe.73.I_109
- [7] Thai, T. T., Liem, N. D., Luu, P. T., Yen, N. T. M., Yen, T. T. H., Quang, N. X., Tan, L. V., & Hoai, P. N. (2022). Performance evaluation of Auto-Regressive Integrated Moving Average models for forecasting saltwater intrusion into Mekong river estuaries of Vietnam. *Vietnam Journal of Earth Sciences*, 44(1), 18–32. Retrieved from <https://doi.org/10.15625/2615-9783/16440>
- [8] Trung, T. D., Vinh, T. N., & Kim, J. (2021). Improving the accuracy of dam in-flow predictions using a long short-term memory network coupled with wavelet transform and predictor selection. *Mathematics*, 9(5), 551.
- [9] Vu, D. T., Yamada, T., & Ishidaira, H. (2018). Assessing the impact of sea level rise due to climate change on seawater intrusion in Mekong Delta, Vietnam. *Water Science and Technology*, 77(6), 1632–1639. <https://doi.org/10.2166/wst.2018.038>
- [10] Vu, P., Vu, P., Pham, C., & Can, N. (2020). The dynamics of saltwater intrusion and its impacts on the territory of Ben Tre.