Deep Learning-Based Forecasting of Oceanic pCO_2 at Station Papa: A Comparative Study of Time Series Models

Gwangun Yu, Gilhan Choi, Moonseung Choi, Sun-hong Min, Junu Jang, Yonggang Kim*
Kongju National University
Cheonan 31080, Korea
{ yugwangun, choikilhan12, cbhoo3, devminsh02, jwjang1102} @smail.kongju.ac.kr, ygkim@kongju.ac.kr

***Abstract*—Predicting oceanic pCO_2 is crucial for quantifying air sea carbon fluxes and enhancing projections of ocean acidification, which is essential for understanding the ocean response to ongoing climate change. This paper examines multistep forecasting of seawater pCO_2 using 3-hourly marine carbon observations obtained from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) Accession 0100074 dataset. A unified evaluation framework is developed to compare four representative time-series forecasting architectures: LSTM, PatchTST, iTransformer, and DLinear. All models are trained using the same multivariate inputs and preprocessing pipeline, and their performance is assessed across multiple prediction horizons. The results summarize the relative prediction behavior of recurrent, transformer based, and linear approaches, providing a consistent reference for applying modern deep learning methods to marine prediction pCO_2.**

***Index Terms*—Air–sea carbon fluxes, Marine carbon cycle, Nonlinear multiscale dynamics**

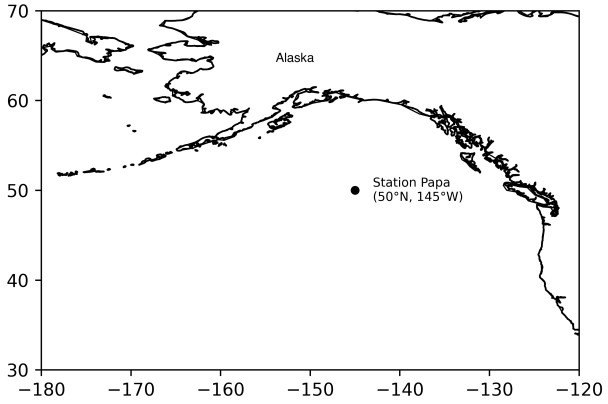


Fig. 1. Location of station Papa

I. Introduction

The global ocean plays a critical role in regulating atmospheric carbon dioxide CO_2 concentrations by absorbing about a quarter of all anthropogenic CO_2 emissions released since the industrial era. Variability in the partial pressure of CO_2 (pCO_2) at the sea surface directly influences the direction and magnitude of air-sea CO_2 fluxes, making high resolution monitoring and accurate predictions of oceanic pCO_2 essential for understanding marine carbon uptake, forecasting ocean acidification, and improving global carbon budget assessments [1], [2]. However, forecasting pCO_2 remains a significant challenge due to its nonlinear dependence on physical forcing, chemical equilibria, biological activity, and atmospheric variability, all of which interact across multiple temporal and spatial scales [3].

Station Papa (145°W, 50°N), located in the North Pacific Ocean as shown in figure1, is one of the longest running open ocean observational sites, providing continuous environmental measurements critical for advancing carbon cycle research. At this site, the National Oceanic and Atmospheric Administration (NOAA) Moored Autonomous pCO_2 ($MAPCO_2$) system provides high frequency measurements of seawater and atmospheric pCO_2, along with sea surface temperature, salinity, humidity, and other wind parameters [4]. These measurements are archived in the NOAA National Centers for Environmental Information (NCEI) Accession 0100074 dataset, creating a multiyear, quality controlled time series with approximately 3 hours of temporal resolution [5]. Such high resolution observations provide a valuable opportunity to examine short term fluctuations in oceanic pCO_2 while capturing broader seasonal and interannual variability.

Machine learning (ML) and deep learning techniques have become widely used tools for environmental prediction tasks. In particular, recent advances in time series forecasting models, such as transformer based architectures, linear decomposition based models, and recurrent neural networks, have expanded the range of approaches available for analyzing complex environmental datasets. Among these are PatchTST and iTransformer, transformer based models designed to capture long term representations. DLinear is a linear forecasting model that identifies trends and seasonal components, while recurrent models like Long Short-Term Memory (LSTM) are widely used for nonlinear sequence modeling. In this study, we apply these time series forecasting models to the station Papa dataset and compare their predictive performance under a unified experimental setup. Instead of introducing a new modeling framework, our goal is to examine how different categories of existing time series models perform with oceanic data and provide a performance comparison across multiple prediction horizons.

II. Related work

AI and ML approaches have been applied to estimate or reconstruct ocean time series. One study compared MLR, CNN, XGBoost, SVM, and RF for global ocean surface pCO_2 and found that the optimized RF model achieved the best performance. The test results showed an MAE of 6.27 microatmospheres, an RMSE of 15.34 microatmospheres, and an R^2 of 0.92. The study produced global pCO_2 fields at 0.25 degree resolution [6].

Another study modeled the deviation between observed pCO_2 and model output in the Indian Ocean from 1980 to 2019 using XGBoost. The method improved RMSE by about 40 percent. The Bay of Bengal Ocean Acidification mooring was used as an independent dataset. It is the only point source surface pCO_2 record available in the region for the years 2014 to 2018 [7].

Deep learning methods have also been used for ocean time series prediction. One study evaluated ConvLSTM and ST ConvLSTM for sea surface temperature prediction in the South China Sea. The work used Copernicus reanalysis data from 2015 to 2019 and tested different input lengths, prediction lengths, and hidden sizes [8].

Another study constructed eight ML models for surface pCO_2 in global, offshore, and coastal regions. The models included MLR, GAM, CNN, GRU, LSTM, XGBoost, LSBoost, and RF. RF produced the highest accuracy in the global region, with an RMSE of 6.123 microatmospheres and an R^2 of 0.986 for the period 2000 to 2019 [9].

III. Methodology

A. Architectures

1) PatchTST: PatchTST is a Transformer based Long term Time Series Forecasting (LTSF) model that divides time series into fixed-length patches and treats them as input tokens while processing each channel of a multivariate sequence independently [10]. The patching mechanism preserves local semantic information and reduces the computational cost of attention, enabling the use of longer look-back windows. Channel independence further improves parameter efficiency and enhances generalization. In addition, masked patch prediction–based self-supervised pre-training provides stronger representation learning than supervised-only training. With this design, PatchTST consistently achieves lower forecasting errors than existing Transformer-based models such as FEDformer, Autoformer, and Informer while maintaining robustness to longer look-back windows and variations in patch length. A remaining limitation is that PatchTST does not explicitly model cross-channel dependencies.

2) iTransformer: iTransformer adopts a variate-centric design that inverts the conventional Transformer approach, which treats each time step as a token [11]. Instead, the model represents each variate (channel) by its entire time series as a single token. This variate tokenization enables self-attention to directly learn multivariate correlations, while the feed forward network extracts temporal representations for each variate. The architecture retains the Transformer's attention, feed-forward network (FFN), and layer normalization without modification and employs an encoder-only structure, achieving both simplicity and computational efficiency. With this design, iTransformer consistently attains lower forecasting errors than various Transformer variants (Transformer, Reformer, Informer, Flowformer, and FlashAttention-based models), and its performance improves as the look-back window increases. Variate-level generalization experiments further show that the model maintains reasonable predictive accuracy even for variates unseen during training. A remaining limitation is that the computational cost of self-attention increases with the number of variates, as attention is performed along the variate dimension.

3) DLinear: DLinear is a one-layer linear Direct Multi-Step (DMS) model proposed as a critique of the architectural complexity of Transformer-based LTSF methods [12]. For each variate X_i, which represents an individual input feature in the multivariate time series, the model applies a one-layer linear projection of the form $\hat{X}_i = WX_i$ to directly predict future values, where W is the weight matrix. The decomposition-based variant, Decomposition-Linear, further improves performance on data with clear trend components by applying a moving-average trend–seasonality decomposition. Experimental results show that DLinear consistently achieves lower forecasting errors than sparse attention Transformer based models across all benchmarks. Its performance continues to improve as the look back window increases, whereas Transformer models tend to overfit temporal noise when the input length becomes large. In terms of efficiency, DLinear surpasses Transformer variants in parameter count, multiply–accumulate operations (MACs), inference time, and GPU memory usage. A noted limitation is that the one-layer linear structure struggles to capture complex temporal dynamics arising from change points.

4) LSTM: The LSTM architecture was designed to address the vanishing and exploding gradient problems observed in recurrent neural networks (RNNs) [13]. It employs a Constant Error Carousel (CEC), a self-connected linear unit that preserves gradients over long time spans to maintain a constant error flow. The input and output gates, implemented as multiplicative units, protect the CEC from irrelevant inputs and regulate when its internal state is exposed. Outside the CEC, truncated gradients are used to ensure computational stability. Experimental results show that LSTM reliably solves long time lag problems spanning hundreds of steps, significantly outperforming conventional RNNs, Backpropagation Through Time (BPTT), and Real-Time Recurrent Learning (RTRL), and exhibiting strong performance on noisy, distributed, and real-valued inputs, as well as on synthetic tasks such as the Reber grammar and temporal order tests. The paper also notes that during early training, the memory cell may be abused by behaving like a constant, which can temporarily distort its intended functionality.

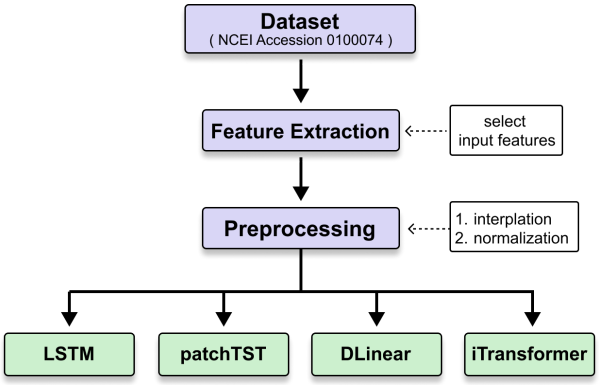


Fig. 2. Overall workflow diagram

B. Forecasting Framework

We approach the pCO_2 forecasting task as a supervised time-series prediction problem. All models share the same dataset, feature set, and preprocessing pipeline to ensure a fair comparison. Four representative architectures are evaluated: LSTM, PatchTST, iTransformer, and DLinear. The overall workflow is shown in Figure 2. Each model is trained using uniform optimization settings and early stopping based on validation loss. Model performance is assessed on the test set using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). MAE and RMSE quantify the magnitude of prediction errors, while R^2 measures the proportion of variance in the target that is explained by the model.

IV. Dataset

The measurement data span approximately three years, from 16 May 2022, 06:17 UTC to 30 May 2025, 21:17 UTC, with a temporal resolution of three hours The raw dataset includes 21 observational variables and a timestamp column. The original dataset includes oceanic and atmospheric CO_2 variables, meteorological and oceanographic conditions, and several biogeochemical parameters. Among these, we set seawater saturated pCO_2 as the prediction target, while raw sensor measurements and environmental variables serve as input features.

A. Feature Extraction

The input features consist of four categories. First, raw sensor variables such as the mole fraction of CO_2 (xCO_2), H_2O, temperature within the LI-COR analyzer, atmospheric pressure, and oxygen concentration are considered. Second, physical environmental variables like sea surface temperature and salinity are included. Third, autoregressive terms are included by using previous pCO_2 values to capture the autocorrelation structure of the time series. Finally, time embeddings were added to explicitly represent seasonal and diurnal periodicity. The day of year (DOY), ranging from 1 to 365 and indicating the position in the annual cycle, and the hour of day (HOD), ranging from 0 to 23 and capturing the diurnal cycle, are both circular variables. Because these variables exhibit periodic boundaries, they were encoded using sinusoidal transformations that map each value onto a unit circle:

$$\sin_{\text{doy}} = \sin\left(2\pi\frac{\text{DOY}}{365.25}\right), \qquad \cos_{\text{doy}} = \cos\left(2\pi\frac{\text{DOY}}{365.25}\right),$$

$$\sin_{\text{hod}} = \sin\left(2\pi\frac{\text{HOD}}{24}\right), \qquad \cos_{\text{hod}} = \cos\left(2\pi\frac{\text{HOD}}{24}\right).$$

We use 365.25 days to incorporate the leap-year correction, which prevents gradual phase drift and provides a more accurate representation of the seasonal cycle without increasing model complexity.

Several variables were excluded due to high missing rates, deterministic dependence on other features, or direct functional dependence on the target variable. For example, the fugacity of CO_2 (fCO_2) is nearly a linear transformation of pCO_2 because the fugacity coefficient slightly deviates from unity. In seawater applications, fCO_2 is commonly approximated following [14]

$$fCO_2 \approx pCO_2\left[1 - 10^{-4}(44.6 - 0.227T)\right].$$

where the adjustment factor is

$$c(T) = 1 - 10^{-4}(44.6 - 0.227T).$$

It is a linear function of temperature. Therefore,

$$fCO_2 \approx pCO_2\, c(T).$$

This shows that fCO_2 differs from pCO_2 only by a small, temperature(T) dependent scaling factor, making it effectively redundant as a model input and preventing potential target leakage that could arise from including variables that are deterministic functions of the target itself. Furthermore, the difference variables compare seawater (sw) and atmospheric (air) measurements,

$$\text{dpCO}_2 = \text{pCO}_2^{SW} - \text{pCO}_2^{Air},$$
$$\text{dpCO}_2 = \text{fCO}_2^{SW} - \text{fCO}_2^{Air}.$$

are linear combinations of the target and atmospheric CO_2, directly encoding target information and therefore must be excluded to avoid leakage. Dry mole-fraction variables also exhibit deterministic dependence since

$$\text{xCO}_2^{\text{dry}} = \frac{\text{xCO}_2^{\text{wet}}}{1 - \text{H}_2\text{O}}.$$

rendering xCO_2 (dry) unnecessary due to the included wet mole-fraction and humidity measurements. Lastly, atmospheric pCO_2 can be derived from existing sensor inputs through

$$\text{pCO}_2^{air} = \text{xCO}_2^{air(dry)}\,(P - e_{air}).$$

Where P denotes the total atmospheric pressure measured by the LI-COR analyzer, and e_{air} is the water vapor partial pressure computed from atmospheric humidity and temperature, providing no additional independent information while increasing the risk of multicollinearity and target leakage. For these reasons, all such deterministically derived variables were omitted from the model inputs.

B. Preprocessing

Interpolation was performed independently for each variable under the assumption that no abrupt physical changes occur within a three-hour interval and that the underlying time grid is uniformly sampled. After gap filling, all numerical features were standardized using z-score normalization. The mean μ_{train} and standard deviation σ_{train} were computed from the training subset and used to standardize each feature according to

$$x_{\text{norm}} = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}}.$$

These statistics ensure that all features share a comparable numerical scale during training, improving optimization stability. The target variable, pCO_2, was normalized using the same procedure

$$y_{\text{norm}} = \frac{y - \mu_{\text{train,target}}}{\sigma_{\text{train,target}}},$$

to maintain consistent scaling between inputs and outputs, ensuring that the input and output share a comparable scale during model optimization. Time-embedding features (sinusoidal DOY/HOD components) inherently lie within the range $[-1, 1]$ and therefore require no additional normalization. To prevent data leakage, statistics from the training period were consistently applied when transforming the validation and test sets. Model predictions were subsequently mapped back to physical units via inverse normalization to enable direct interpretation in terms of pCO_2. This preprocessing procedure ensures a consistent pipeline across training and inference, enabling stable model convergence and physically meaningful evaluation.

V. Simulation

Time series training employed a sliding-window approach. We defined one time step as 3 hours. The input sequence lengths were set to 8, 16, 24, 32, and 40 steps (approximately 1–5 days), and the output sequence lengths were set to 1, 2, 4, 8, and 16 steps (approximately 3–48 hours), resulting in 25 (input length, output length) combinations in total. For each combination, contiguous segments of length (input window length + output window length) were extracted from the full time series, where the first part was used as the input sequence and the immediately following part as the future target sequence.

For all generated windows, we grouped the samples by the year and month corresponding to the last time step of the input sequence. Within each month, the windows were sorted in chronological order, and a 9:1 split was applied, assigning the first 90% to the training set and the last 10% to the validation set. The loss function was the MSE computed on the normalized targets, and the Adam optimizer was used with a learning rate of 1×10^{-3}. The batch size was set to 128, and training was conducted for up to 500 epochs, with early stopping applied if the validation MSE did not improve for 5 consecutive epochs. During validation, predictions and ground truth were de-normalized, and the MSE, RMSE, MAE, and coefficient of determination (R^2) were computed in the original units. The model weights at the epoch with the minimum validation MSE were saved as the final model for each (input, output) configuration.

TABLE I
Forecasting performance of the best (input, output) configuration for each model on the April 2025 test period.

Model (in, out)	MSE	RMSE	MAE	R^2
PatchTST (16, 1)	**1.1974**	**1.0942**	0.7417	0.8727
DLinear (8, 1)	1.1985	1.0948	0.7048	**0.8943**
LSTM (8, 1)	1.4393	1.1997	0.8689	0.8730
iTransformer (8, 1)	1.2016	1.0962	**0.6881**	0.8941

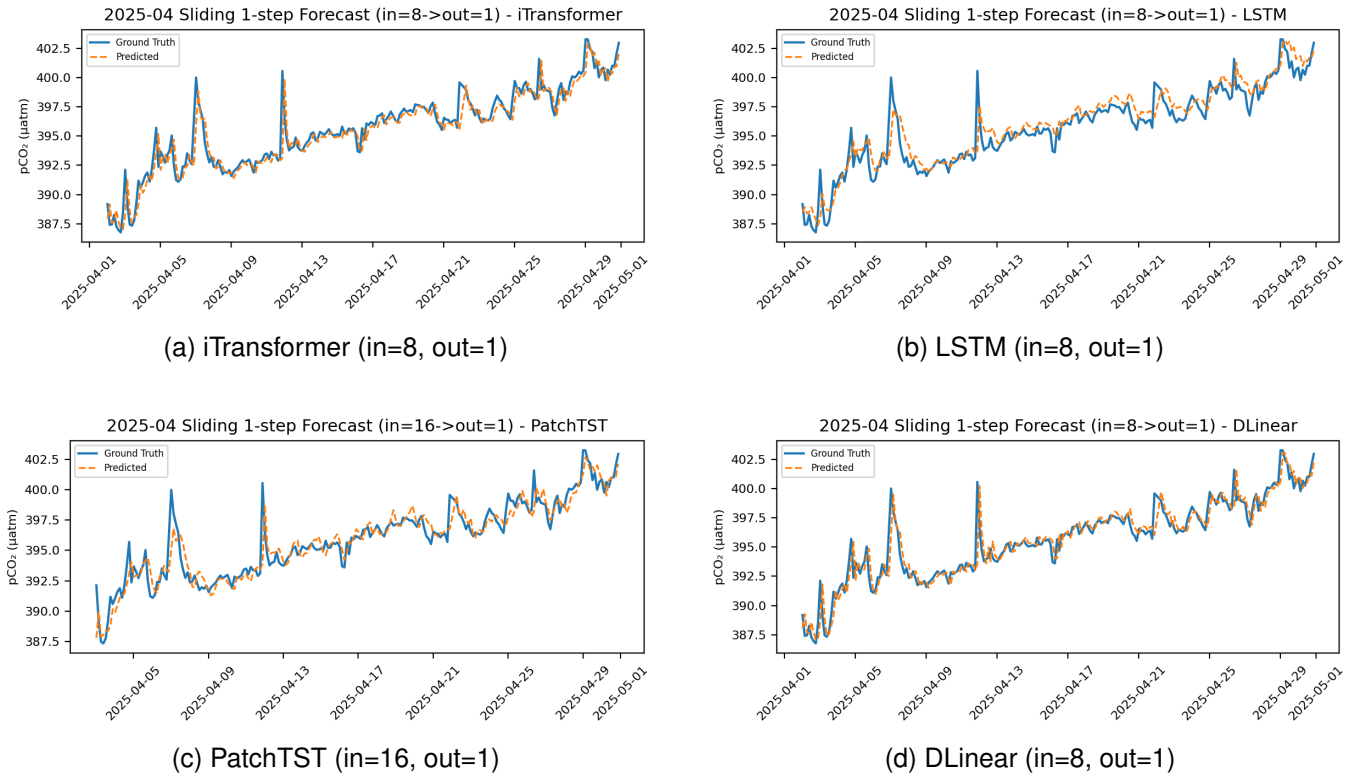
For LSTM, DLinear, PatchTST, and iTransformer, we employed the same data preprocessing and window configuration, loading model checkpoints corresponding to the minimum validation MSE obtained during training to evaluate multi-step forecasting performance on an independent test period. The test data were organized as a 3-hourly time series sorted in ascending order by datetime, consistent with the training stage. The testing input features were created to match those used during training.

Table II reports the test-set R^2 scores obtained by each model for all 25 (input, output) configurations in the April 2025 test period. From this table, we observe that, across most (input, output) combinations, iTransformer attains the highest R^2 among the four models, indicating comparatively robust performance with respect to the choice of window lengths. For each model, we then selected a single representative configuration for detailed comparison. The primary selection criterion was the highest coefficient of determination R^2 in the test set, as R^2 is one of the most widely used goodness-of-fit measures to assess the agreement between observed and predicted values in environmental and geophysical applications [15]. When two configurations exhibited nearly identical values R^2, we further preferred the configuration with the smaller MAE, following the recommendation that MAE provides a natural and interpretable measure of average model error and is less sensitive to outliers than RMSE [16]. This multi-metric selection strategy is consistent with previous work suggesting that forecast models should be evaluated using complementary accuracy measures rather than a single criterion [17]. According to this rule, the best-performing settings were PatchTST (16, 1), DLinear (8, 1), LSTM (8, 1) and iTransformer (8, 1).

Figure 3 illustrates sliding 1-step forecasts for the April 2025 test period using the best configuration of each model and Table I summarizes the quantitative performance of these best configurations on the test set. Table I summarizes the performance of the best (input, output) configuration for each model on the April 2025 test period. Overall, all four models achieve relatively low prediction errors with MSE values around 1.2, and among them DLinear (8, 1) and iTransformer (8, 1) obtain the highest coefficients of determination, with $R^2 = 0.8943$ and $R^2 = 0.8941$, respectively. iTransformer attains the smallest MAE (0.6881), providing the best pointwise absolute

TABLE II
TEST-SET R^2 SCORES FOR ALL (INPUT, OUTPUT) CONFIGURATIONS ON THE APRIL 2025 TEST PERIOD.

Model	(8,1)	(8,2)	(8,4)	(8,8)	(8,16)	(16,1)	(16,2)	(16,4)	(16,8)	(16,16)	(24,1)	(24,2)	(24,4)	(24,8)	(24,16)	(32,1)	(32,2)	(32,4)	(32,8)	(32,16)	(40,1)	(40,2)	(40,4)	(40,8)	(40,16)
iTransformer	0.8941	**0.8904**	**0.8770**	**0.8844**	**0.8586**	0.8640	**0.8246**	**0.8485**	**0.8444**	**0.7472**	0.8416	**0.8278**	**0.8074**	**0.8229**	0.5302	0.8377	**0.8235**	**0.8070**	**0.7474**	0.4036	0.8394	**0.8072**	**0.7976**	**0.6506**	0.3573
PatchTST	0.6110	0.8236	0.7849	0.7111	0.6393	**0.8727**	0.7326	0.6402	0.6621	0.5736	**0.8614**	0.7916	0.7265	0.5546	0.3849	**0.8558**	0.8232	0.7180	0.6650	0.5666	0.8446	0.7391	0.3914	0.6074	-0.0780
LSTM	0.8730	0.8375	0.8057	0.7325	0.4840	0.6471	0.8168	0.6947	0.3932	0.3207	0.7915	0.6980	0.4111	0.5516	**0.6140**	0.6486	0.7105	0.7071	0.5873	**0.6349**	0.8213	0.3449	0.5608	0.0995	**0.6377**
DLinear	**0.8943**	0.8501	0.7966	0.7167	0.6027	0.8589	0.8203	0.7455	0.6478	0.5473	0.8467	0.7877	0.7110	0.6153	0.5934	0.8489	0.7850	0.7159	0.6316	0.5995	**0.8528**	0.7897	0.7035	0.6015	0.6103



(a) iTransformer (in=8, out=1)

(b) LSTM (in=8, out=1)

(c) PatchTST (in=16, out=1)

(d) DLinear (in=8, out=1)

Fig. 3. Sliding 1-step (3 h) forecasts of $p\mathrm{CO}_{2,\mathrm{SW}}(\mathrm{sat})$ for the April 2025 test period using the best configuration of each model. Blue lines denote ground truth and orange dashed lines denote model predictions.

accuracy, whereas DLinear achieves the second-smallest MAE together with the highest R^2 and is therefore selected as the overall best configuration according to the selection rule used for Table II (primary criterion R^2, secondary criterion MAE).

PatchTST (16, 1) exhibits the smallest MSE and RMSE among the four models, but at the same time yields the lowest R^2 and the second-largest MAE; under our criteria, it is therefore regarded as the weakest configuration. LSTM (8, 1) shows the largest MSE and MAE and an R^2 only slightly higher than that of PatchTST, indicating that it also performs clearly worse than the transformer-based models.

VI. CONCLUSION

In this paper, we investigated deep learning–based forecasting of seawater saturated partial pressure of CO_2 at Station Papa using high-frequency MAPCO$_2$ observations. Building on a unified preprocessing and sliding-window forecasting framework, we compared four representative time-series models—LSTM, PatchTST, iTransformer, and DLinear—across 25 (input, output) configurations and multiple evaluation metrics on a recent three-year time period. The results show that all models achieve reasonably accurate short-term forecasts, with test-set MSE values of approximately 1.2 $\mu\mathrm{atm}^2$ for the best configuration of each model. Among them, iTransformer yields the smallest MAE, while DLinear achieves the highest R^2 with the second-smallest MAE and is therefore identified as the overall best configuration under our multi-metric selection rule. These findings highlight that, for one-dimensional oceanic pCO$_2$ forecasting at a single station, a carefully tuned linear baseline can match or even surpass more complex deep architectures in terms of deterministic accuracy.

Despite these encouraging results, several limitations remain. First, our analysis is restricted to a single open-ocean site, so the robustness of the models across different biogeochemical regimes (e.g., coastal, upwelling, or high-latitude regions) has yet to be assessed. Second, we focused on point predictions and deterministic error metrics, without explicitly

quantifying predictive uncertainty or evaluating performance under extreme or rapidly fluctuating events. In future work, we plan to extend this framework to multi-site and possibly spatially explicit settings, incorporate additional physical drivers (such as mixed-layer depth and large-scale climate indices), and explore probabilistic and physics-informed forecast models. Such extensions would help bridge purely data-driven approaches and process-based understanding, ultimately improving our ability to monitor and predict changes in the ocean carbon system.

ACKNOWLEDGMENT

This work was supported by the Regional Innovation System & Education(RISE) program through the Chungnam RISE center, funded by the Ministry of Education(MOE) and the Chungcheongnam-do, Republic of Korea (2025-RISE-12-003).

REFERENCES

[1] J. G. Canadell, P. M. Monteiro, M. H. Costa, L. C. Da Cunha, P. M. Cox, A. V. Eliseev, S. Henson, M. Ishii, S. Jaccard, C. Koven, *et al.*, "Global carbon and other biogeochemical cycles and feedbacks," *IPCC AR6 WGI, final government distribution*, pp. chapter–5, 2021.

[2] R. Wanninkhof and W. R. McGillis, "A cubic relationship between air-sea co2 exchange and wind speed," *Geophysical Research Letters*, vol. 26, no. 13, pp. 1889–1892, 1999.

[3] P. Rustogi, P. Landschützer, S. Brune, and J. Baehr, "The impact of seasonality on the annual air-sea carbon flux and its interannual variability," *npj Climate and Atmospheric Science*, vol. 6, no. 1, p. 66, 2023.

[4] A. J. Sutton, C. L. Sabine, S. Maenner-Jones, N. Lawrence-Slavas, C. Meinig, R. Feely, J. Mathis, S. Musielewicz, R. Bott, P. McLain, *et al.*, "A high-frequency atmospheric and seawater pco 2 data set from 14 open-ocean sites using a moored autonomous system," *Earth System Science Data*, vol. 6, no. 2, pp. 353–366, 2014.

[5] A. J. Sutton, C. L. Sabine, C. Dietrich, S. M. Jones, S. Musielewicz, R. Bott, and J. Osborne, ""high-resolution ocean and atmosphere pco2 time-series measurements from mooring papa_145w_50n in the north pacific ocean"." NOAA National Centers for Environmental Information, Dataset, 2012. Accessed on: DEC. 9, 2025.

[6] H. Wu, L. Wang, X. Ling, L. Cui, R. Sun, and N. Jiang, "Spatiotemporal reconstruction of global ocean surface pco2 based on optimized random forest," *Science of the Total Environment*, vol. 912, p. 169209, 2024.

[7] P. K. Ghoshal, A. Joshi, and K. Chakraborty, "An improved long-term high-resolution surface p co2 data product for the indian ocean using machine learning," *Scientific Data*, vol. 12, no. 1, p. 577, 2025.

[8] P. Hao, S. Li, J. Song, and Y. Gao, "Prediction of sea surface temperature in the south china sea based on deep learning," *Remote Sensing*, vol. 15, no. 6, p. 1656, 2023.

[9] H. Wu, Y. Ji, L. Wang, X. Liu, W. Zhou, L. Cui, Y. Wang, M. Liu, and Z. Li, "Reconstruction and spatiotemporal analysis of global surface ocean pco_2 considering sea area characteristics," *EGUsphere*, vol. 2025, pp. 1–32, 2025.

[10] Y. Nie, "A time series is worth 64words: Long-term forecasting with transformers," *arXiv preprint arXiv:2211.14730*, 2022.

[11] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "itransformer: Inverted transformers are effective for time series forecasting," *arXiv preprint arXiv:2310.06625*, 2023.

[12] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, pp. 11121–11128, 2023.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] A. G. Dickson and C. Goyet, "Handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water. version 2," tech. rep., Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 1994.

[15] D. R. Legates and G. J. McCabe Jr, "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation," *Water resources research*, vol. 35, no. 1, pp. 233–241, 1999.

[16] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.

[17] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.