



# Enhancing Extended Weather Forecasts in the TCWAGFS Model Using Deep Learning Method for SST Bias Correction

Katherine Shu-Min Li<sup>1</sup>, Nadun Sinhabahu<sup>1</sup>, Ben-Jei Tsuang<sup>2</sup>, Fang-Chi Wu<sup>1</sup>, Wan-Ling Tseng<sup>3</sup>, Pei-Hsuan Kuo<sup>4,5</sup>, Sying-Jyan Wang<sup>6</sup>, Pang-Yen Liu<sup>4</sup>, Jen-Her Chen<sup>4</sup>, Bin-Ming Wang<sup>7</sup>, Yung-Yao Lan<sup>8</sup>, Sun-Yuan Kung<sup>9</sup>

<sup>1</sup>Department of Computer Science and Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan

<sup>2</sup>Department of Environmental Engineering, National Chung-Hsing University, Taichung, Taiwan.

<sup>3</sup>Ocean Center, National Taiwan University, Taipei, Taiwan.

<sup>4</sup>Central Weather Administration, Taipei, Taiwan

<sup>5</sup>Department of Atmospheric Sciences, National Central University, Taoyuan, Taiwan

<sup>6</sup>Department of Computer Science and Engineering, National Chung-Hsing University, Taichung, Taiwan

<sup>7</sup>Critical Infrastructure Research Center, National Taipei University of Technology, Taipei, Taiwan.

<sup>8</sup>Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan.

<sup>9</sup>Department of Electrical and Computer Engineering, Princeton University, New Jersey, U.S.A.

Correspondence to: Ben-Jei Tsuang, [btsuang@gmail.com](mailto:btsuang@gmail.com)

**Abstract.** The extended weather (> 7 days) and the seasonal climate predictions are highly dependent on the status of Madden Julian Oscillation (MJO) and El Niño-Southern Oscillation (ENSO). Both the evolutions of MJO and ENSO are found to be correlated to the anomalies of the global sea surface temperature (SST). To decrease the predicting SST bias (BiasSST) in a coupled ocean-atmosphere global model, we evaluate nine well-developed machine learning algorithms. By using the Bi-directional Long Short-Term Memory (Bi-LSTM) algorithm, it is found the bias can be reduced significantly. For example, the Root Mean Squared Error on Day 10 forecast (denoted as D10) is reduced to 0.01 K by the Bi-LSTM algorithm while the original bias is 0.38 K by the Taiwan Central Weather Administration Global Forecast System (TCWAGFS), of which the error is reduced by 97%.

## 1 Introduction

Extended weather prediction, which refers to potential weather conditions up to 46 days ahead (European Centre for Medium-Range Weather Forecasts; <https://www.ecmwf.int/en/forecasts/documentation-and-support>), requires an understanding of the complex interactions between ocean, land, and atmospheric processes. The improvement of accuracy in these predictions represents a fundamental objective within the field of meteorology, offering substantial benefits to both society and the economy (Haiden et al., 2018). To achieve this objective, various avenues must be explored, including increasing resolution (Demory et al., 2020; Gutjahr et al., 2016; Iles et al., 2020; Strandberg and Lind, 2021), enhancing parameterization schemes (Lee et al., 2020; Singh et al., 2022; Su et al., 2022), and advancing air-sea coupling processes (DeMott et al., 2014, 2015; Tseng et al., 2015, 2022; Stan, 2018; Gao et al., 2020; Lan et al., 2022, 2024). The advent of deep learning techniques signifies a considerable leap forward in this domain, exemplifying considerable promise and, in selected instances, rivaling the efficacy of conventional dynamic weather forecasting systems (Ji et al., 2019; Noack and Reyes, 2023;



Weyn et al., 2019). Nevertheless, despite the expansion and implementation of deep learning techniques that replace dynamic characteristics with statistical approaches, including neural network methods (Lopatka, 2019; Schultz et al., 2021), challenges remain, particularly in the context of extended weather forecasts. As comprehension and application of these techniques continue to expand, they are increasingly contributing to the enhancement of extended-range predictions. Consequently, the implementation of deep learning to enhance scheme performance or diminish model bias has emerged as a highly pragmatic and efficacious strategy within this domain (Vitart et al., 2022). These techniques have markedly enhanced forecasts of intricate climate phenomena such as the El Niño-Southern Oscillation (Nurdiati et al., 2021; Ham et al., 2019) and the Madden-Julian Oscillation (Kim et al. 2021; Silini et al. 2021).

Sea surface temperature (SST) is fundamentally important for climate research and accurate simulation in numerical weather prediction models. SST plays a crucial role in the earth's energy balance and the exchange of energy between the ocean and atmosphere, significantly modulating sensible and latent heat fluxes across the air-sea interface. For decades, coupled ocean-atmosphere integrations have been used for extended-range forecasts (Molteni et al., 2011; Zhou et al., 2022). These coupled models have produced more skillful and reliable forecasts of low-frequency intra-seasonal oscillations, such as the MJO, compared to atmosphere-only models (Kim et al., 2010). Despite the benefits of ocean-atmosphere coupling for extended forecasts, these models also exhibit large systematic errors in SSTs, which can exceed a few degrees after four weeks. Such errors affect atmospheric circulation and degrade forecast skill scores (Vitart & Balmaseda, 2018). Therefore, SST bias correction has been identified as a major issue to address in coupled prediction systems (Abhilash et al., 2014, Fei et al., 2022, Han et al., 2022).

In recent years, data-driven approaches have gained attention in geographical and oceanic research. Compared to physical-based numerical models, data-driven methods can automatically learn relationships with much less domain-specific knowledge and provide prediction results from trained algorithms in a short time (Xu et al., 2023). Xiao et al. (2019) used satellite data combined with LSTM and AdaBoost ensemble learning algorithms to predict short and mid-term daily SST. 1- to 7- day SSTs near the Korean Peninsula were predicted to designated high water temperature events by using an LSTM network with the European Centre for Medium-Range weather Forecasts (ECMWF) ERA5 SST time series data (Choi et al., 2023). Similarly, SST predictions for the next five days in the East and South China Seas were made using LSTM (Jia et al., 2022; Hao et al., 2023). However, a comprehensive evaluation of various data-driven methods for their SST bias correction performance is currently lacking.

This study aims to address this gap by evaluating nine well-developed machine learning algorithms for their performance in predicting SST within the TCWAGFS (Liou et al., 1997; Su et al., 2021, 2022). The proposed methodology provides a systematic and comprehensive approach to SST prediction using time series algorithms, considering the temporal dependencies and patterns inherent in time-series data. This approach enables the development of accurate prediction models that can effectively forecast future SST based on past data. The output from this methodology can serve as valuable input for climatology and oceanography studies, offering insights and immediate solutions for reducing SST bias and enhancing



71 prediction accuracy. The chapter is structured as follows: Section 2 describes the data, algorithms, and methodology, while  
72 Section 3 presents the results of SST bias correction. Finally, Section 4 provides a summary of the results.

## 73 **2 Data and methodology**

### 74 **2.1 Extended weather forecast model: TCWAGFS**

75 The TCWAGFS, Taiwan Central Weather Administration Global Forecast System (Liou et al., 1997; Su et al., 2021,  
76 2022), is a hydrostatic global spectral model using semi-Lagrangian advection scheme (Staniforth and Cote, 1991) with  
77 octahedral reduced Gaussian grids ( $T_{Co}$ ) (Hortal and Simmons, 1991; Malardel et al., 2016) in the horizontal and sigma-  
78 pressure hybrid in the vertical. The resolution of the model is 28 km in the horizontal ( $T_{Co}383$ ) and 72 layers with 0.1 hPa at  
79 the model top. The model is coupled with a one-column, turbulent, and kinetic-energy-type ocean mixed-layer model (snow-  
80 ice-thermocline, SIT) in 40°S-40°N sea grids. The 35-day SSTs from the daily runs of the TCWAGFS are analysed here. The  
81 SIT model provides accurate SST simulation and thus facilitates realistic air-sea interaction (Tseng et al., 2022). The daily  
82 mean SST tendencies in 40°S-60°S and 40°N-60°N sea grids are from TCWA1T (Juang et al., 2024). The bias of SST we use  
83 for later model training are collected from every 35-day forecasts in the period of 1 January 2021 to 30 June 2022, of which  
84 the data resolution is 2.5\*2.5 degrees in whole world. The observed SST used is obtained from TCWAGFS data assimilated  
85 systems (Su et al., 2021, 2022).

### 86 **2.2 Machine learning models**

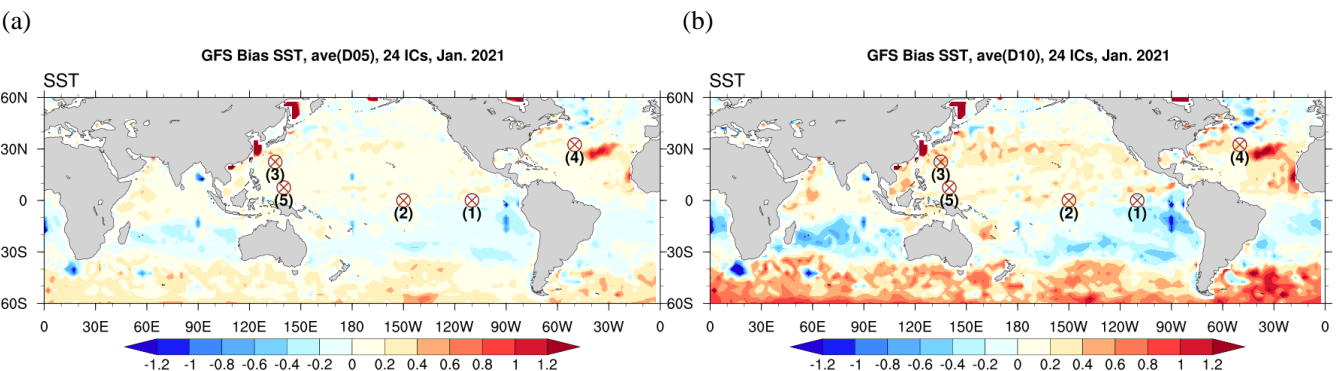
87 For the machine learning model selection, nine algorithms are selected to compare their performance in predicting SST  
88 with bias correction applied. The following machine learning methods were considered: (1) Recurrent Neural Network (RNN):  
89 Designed to handle sequence data by incorporating memory, RNNs can retain information across time steps. However, they  
90 often struggle with long-term dependencies due to vanishing and exploding gradient problems (Funahashi and Nakamura,  
91 1993). (2) Autoregressive Integrated Moving Average (ARIMA): A widely used statistical method for time series forecasting,  
92 ARIMA captures temporal dependencies through a linear combination of past values and errors (Box et al., 2015).  
93 (3) Seasonal Autoregressive Integrated Moving Average with eXogenous regressors (SARIMAX): This model is similar to  
94 ARIMA but includes seasonal and exogenous components, making it suitable for forecasting seasonal trends in time series  
95 data (Vagropoulos et al., 2016). (4) Long Short-Term Memory (LSTM): A type of RNN specifically designed to handle long-  
96 range dependencies in sequences of data, LSTMs use memory cells and gating mechanisms to learn and remember long  
97 sequences (Hochreiter and Schmidhuber, 1997). (5) Time Series Transformer I (Prophet): An additive model developed by  
98 Facebook for forecasting time series data, Prophet is robust to missing data and shifts in the trend, and handles outliers well  
99 (Taylor and Letham, 2018). (6) Time Series Transformer II and III (darts Transformer and darts Exponential Smoothing):  
100 Transformer-based models that allow parallel processing of sequence data, employing self-attention mechanisms to weigh the  
101 importance of sequence elements (Herzen et al., 2022). (7) Time Series Imaging: Transforms time series data into image data



and applies convolutional neural networks for analysis, combining strengths of time series analysis and image processing (Li et al., 2020). (8) Bi-directional Long Short-Term Memory (Bi-LSTM): A variant of LSTM that trains two LSTM networks on the input sequence in different temporal orders, capturing information from both past and future states for a richer representation of the sequence data (Schuster and Paliwal, 1997).

### 2.3 Data of SST bias preprocessing

The TCWAGFS forecast daily SST data from day 1 to day 40 is used to calculate the predicted SST bias. The predicted SST bias, referred to as BiasSST, is defined as the difference between the TCWAGFS predicted SST on the predicted day and the observed SST. Figure 1 shows the monthly averaged predicted SST biases in January 2021 for day 5 and day 10. The model configuration results in two large-scale circulation gyres separated by latitude, featuring a strong meandering zonal jet. This model emulates idealized versions of existing ocean current systems, such as the Gulf Stream in the North Atlantic and the Kuroshio Extension in the North Pacific, both characterized by dynamic eddies interacting with a robust mean flow. As illustrated in Figure 1, the process of predicting SST using a time series algorithm is comprehensive, beginning with rigorous data collection and preparation. Five major areas have been selected to showcase the enhancement of bias correction, as depicted in Fig. 1b: (a) Cold Tongue Equatorial Current, (b) El Niño Indicator Area, (c) Kuroshio Current, (d) North Atlantic Current, and (e) Pacific Ocean Warm Pool.



**Figure 1.** Averaged SST bias in the (a) day-5 and (b) day-10 in January 2021. The marks indicate the five selected areas. (1) Cold Tongue Equatorial Current, (2) El Niño Indicator Area, (3) Kuroshio Current, (4) North Atlantic Current, and (5) Pacific Ocean Warm Pool

In this study, data preprocessing is essential for maintaining the integrity of time-series patterns and ensuring that the algorithm can learn accurate representations of sea surface temperature trends. This involves handling missing data points, a common occurrence in real-world datasets. The technique used here is a forward-fill methodology, where null values are replaced with the last observed non-null value. This is widely accepted, especially for time series data, as it presumes that the absence of a data point is most closely related to the previous data point. Once the data is cleaned and complete, it undergoes



normalization using MinMaxScaler, a popular feature scaling technique. This scales all data values to a range between 0 and 1, ensuring no individual feature disproportionately influences the model due to its scale. Additionally, feature engineering is often applied to enhance model performance. This involves creating new features from existing data, such as polynomial features, to help uncover complex patterns within the data. Dimensionality reduction techniques like Principal Component Analysis (PCA) are also used to reduce the number of features in the dataset while retaining most of the important information, making the model more efficient and less prone to overfitting. The way data is represented or framed for a machine-learning algorithm can significantly impact model performance. With the data split, the next step is to preprocess it to fit the needs of the time series algorithm. Using a technique called the sliding window approach, sequences are generated from the data for the time series algorithm. The size of the window is defined by the 'look\_back' parameter, which determines how many previous time steps the algorithm should consider when making a prediction. Additionally, the 'time\_steps' parameter decides how many steps into the future the algorithm should predict.

## 2.4 Algorithm development

Following data collection and preprocessing, nine time-series algorithms (RNN, ARIMA, SARIMAX, LSTM, Prophet, darts Transformer, darts Exponential Smoothing, Time Series Imaging, Bi-LSTM) were trained using the Adam optimizer, with an initial learning rate of 0.001. Early stopping was applied based on validation loss, with a patience threshold of 250 epochs. Dropout layers and L2 regularization were added to prevent overfitting. A 5-fold cross-validation strategy was used to ensure robustness of the results. The predicted BiasSST was then applied to adjust the TCWAGFS SST forecasts. According to the results, the bidirectional LSTM algorithm is selected based on the algorithm performance to make predictions. LSTM is a type of recurrent neural network (RNN) that is particularly well-suited to time series prediction tasks like this one because of its ability to learn long-term dependencies and its capacity to consider both past and future data in the algorithm. The algorithm structure consists of an input layer, followed by a Bidirectional LSTM layer, and finally, a fully connected layer that maps the LSTM outputs to the final output size. The LSTM layer has a specific number of hidden units and LSTM layers, which are parameters that can be tuned to enhance the algorithm's performance. Training the algorithm involves optimizing its parameters to minimize a given loss function. The algorithm employs the Adam optimizer, an algorithm for first-order gradient-based optimization of stochastic objective functions. This is paired with the Mean Squared Error (MSE) loss function which calculates the quantity that the algorithm seeks to minimize during training.

One of the strengths of this algorithm is its capacity to look both backward and forward in time due to the use of a bidirectional LSTM. It uses past data (going back a specific number of steps defined by the 'look\_back' parameter, which is set as 1 in our work) and future data (defined by the 'time\_steps' parameter, which is set as 3 in our work) to make a prediction. A sensitivity analysis was conducted by varying the look-back window size from 1 to 5 days, and the Bi-LSTM performance was evaluated for each configuration. The model showed optimal performance with a look-back window of 3 days, beyond which performance gains were negligible. This analysis demonstrated the robustness of the Bi-LSTM model for time-series SST predictions.



161 For each data source (BiasSST), we calculated the following error metrics: Root Mean Square Error (RMSE), Mean  
162 Absolute Error (MAE), and Mean Bias Error (MBE). These three-error metrics can be calculated by following Eqs. (1)-(3):

$$163 \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2}, \quad (1)$$

$$164 \quad MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i|, \quad (2)$$

$$165 \quad MBE = \frac{1}{n} \sum_{i=1}^n (O_i - P_i), \quad (3)$$

166 where  $n$  is sample size,  $P_i$  is the prediction and  $O_i$  is the true value.

167 The algorithms were implemented using Python 3.11.2 and PyTorch 2.0.1. To enhance performance, noise removal and  
168 optimization techniques were applied, including data smoothing, the Huber loss function, overfitting compensation, and post-  
169 training fairness calculations. The training process consists of three main stages: pre-training, algorithm training, and post-  
170 training. During the pre-training stage, data preprocessing begins with applying a moving average to reduce noise. The data is  
171 then split into training, validation, and test sets, using a 75%/15%/10% ratio, respectively. In the training stage, MSE is used  
172 as the loss function, and the Adam optimizer is applied with a learning rate of 0.001. After training, the main algorithm  
173 generates an initial prediction, which is further refined by the secondary algorithm to produce the final, optimized prediction.

174

### 175 **3 Experimental Results**

#### 176 **3.1 Performance of Bi-LSTM for BiasSST Prediction**

177 Table 1 reports the performance of Bi-LSTM on day 10 with and without noise removal and optimization. Algorithms  
178 that incorporate noise removal and overfitting prevention generally show improved performance.

179



**Table 1.** Performance of Bi-LSTM for BiasSST prediction on day 10 (without noise removal and overfitting prevention) without and with noise removal. The 5 selected regions are marked in Fig. 1b.

Region	Metrics (unit:K)	w/o noise removal	w/ noise removal
(1) Cold Tongue Equatorial Current	RMSE	0.044	0.008
	MAE	0.040	0.006
	MBE	0.040	0.006
(2) El Niño Indicator Area	RMSE	0.009	0.002
	MAE	0.007	0.002
	MBE	-0.007	-0.001
(3) Kuroshio Current	RMSE	0.035	0.013
	MAE	0.031	0.006
	MBE	0.027	-0.004
(4) North Atlantic Current	RMSE	0.014	0.006
	MAE	0.013	0.005
	MBE	0.001	0.004
(5) Pacific Ocean Warm Pool	RMSE	0.019	0.004
	MAE	0.017	0.003
	MBE	-0.017	-0.002
Average	RMSE	0.024	0.007
	MAE	0.022	0.004
	MBE	0.009	0.001

### 3.2 Identify the improvement of Bi-LSTM

The TCWAGFS is the foundational model used for predicting sea data statistics. It is developed based on established physical laws and relationships governing oceanic and atmospheric conditions. This model relies on equations derived from physics to forecast future states of the sea, including factors such as temperature, currents, and sea level pressure. However, it is limited by its simplified assumptions and inability to fully capture the complexities and variabilities of natural processes. This limitation becomes evident in the model's progressively increasing RMSE, MAE, and MBE values over the course of 30 days. While the Bi-LSTM is suggested as the best algorithm to correct the SST bias, the further evaluation is presented here. The performance metrics provide key insights into the magnitude, direction, and percentage of error relative to the actual data, helping to evaluate the model's accuracy. As shown in Figure 2, different algorithms demonstrate varying levels of effectiveness in predicting SST in the El Niño indicator area and the Western Pacific warm pool. These predictions are evaluated based on the bias temperature gradient (BiasSST) data collected over 30 days. This information is critical for understanding and forecasting climate phenomena such as El Niño.

Focusing first on the TCWAGFS model, the RMSE values for the El Niño indicator area increase from 0.129 to 0.591 over 30 days, while in the Western Pacific warm pool, the RMSE values rise from 0.169 to 0.453. This upward trend indicates a growing magnitude of prediction errors over time, suggesting that the model's performance deteriorates as the forecast

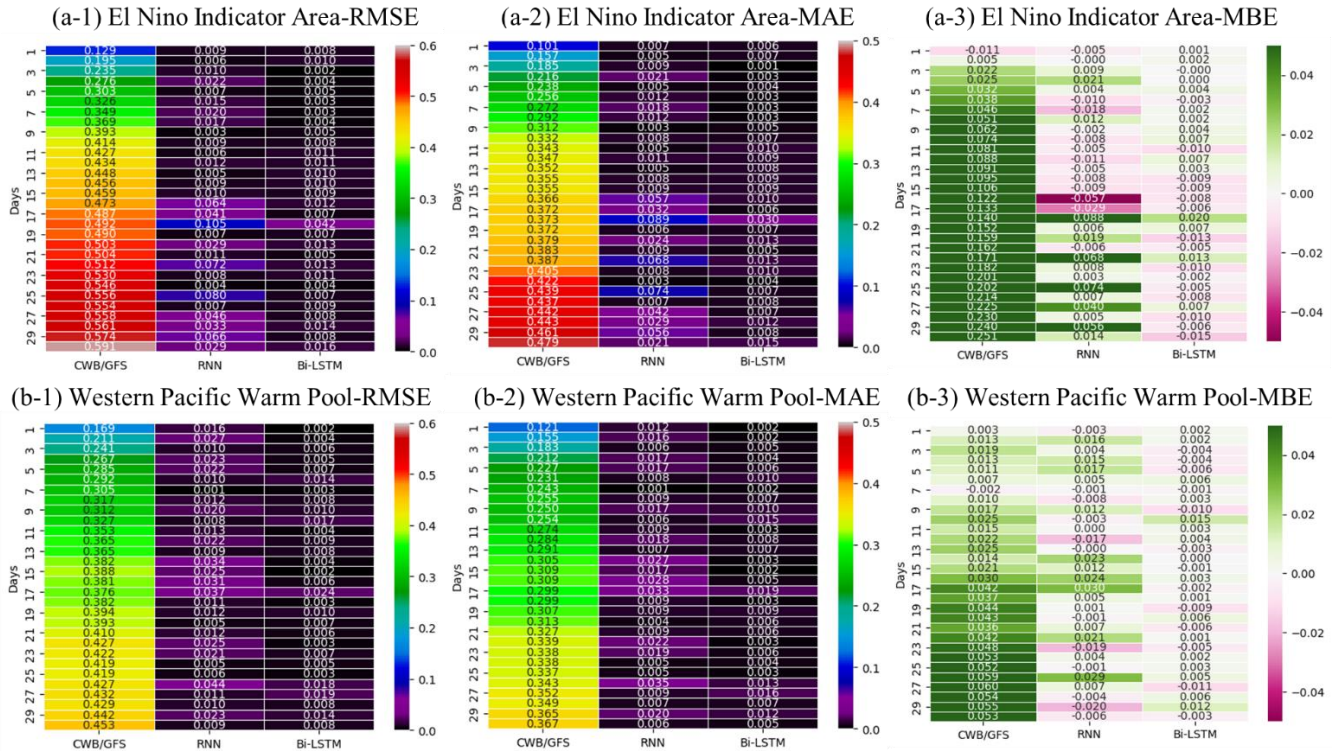




198 horizon extends. A similar pattern is observed with the MAE and MBE metrics, further supporting this conclusion. In contrast,  
199 the Bi-LSTM algorithm, a recurrent neural network variant, consistently achieves lower RMSE, MAE, and MBE values  
200 compared to the TCWAGFS model. Bi-LSTM is particularly well-suited for tasks involving sequential data because it  
201 processes information in both directions—backwards and forwards—allowing it to understand the context of data more  
202 effectively. This ability makes Bi-LSTM especially effective in time-series predictions where surrounding context is crucial  
203 for improving accuracy. Similar performance improvements are observed in other locations shown in Figure 1b (not depicted  
204 here). Several factors explain why the Bi-LSTM algorithm outperforms traditional models in sea data prediction, a time-series  
205 problem: (1) Sea data often has long-term dependencies. Traditional methods may struggle to capture these over extended  
206 periods. For example, sea currents or temperatures may be influenced by factors from days, weeks, or even months earlier.  
207 LSTM’s ability to learn and retain such long-term dependencies is advantageous, and Bi-LSTM enhances this by analyzing  
208 data in both directions. (2) Unlike standard LSTMs, which process data from past to future, Bi-LSTMs look at data in both  
209 directions. This is particularly beneficial in cases where future data provides additional context to past states, leading to more  
210 accurate predictions. For instance, knowing about a future storm can inform the prediction of current sea conditions. (3) Noise  
211 Reduction: Sea data often contains noise due to environmental variability, sensor errors, or other factors. Bi-LSTM’s  
212 bidirectional processing helps it manage this noise more effectively, improving prediction accuracy. (4) Complex Patterns:  
213 Sea data can exhibit intricate temporal patterns due to factors like tidal movements, seasonal shifts, and long-term climate  
214 trends. Bi-LSTM’s architecture, which captures both linear and non-linear relationships, is well-suited for learning and  
215 predicting such patterns. (5) Robustness: Sea data can be volatile, with sudden changes due to extreme weather events, such  
216 as storms or tsunamis. Bi-LSTMs are robust enough to handle these sudden shifts, maintaining strong prediction performance  
217 even under such conditions.

218       The key reason Bi-LSTMs perform better in this context is their ability to incorporate both past and future information  
219 into predictions, giving the model a more holistic view of the data. This leads to more accurate and informed predictions,  
220 particularly when dealing with complex, temporal, and noisy datasets like sea data statistics. Bi-LSTM’s performance suggests  
221 that machine learning methods, particularly those capable of handling long-range dependencies and complex patterns, have  
222 significant potential to outperform traditional physical models in capturing the complexities of environmental systems.



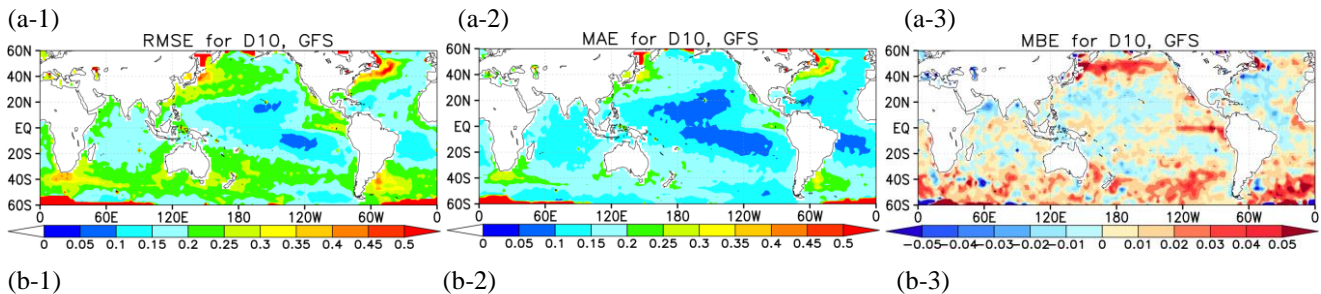


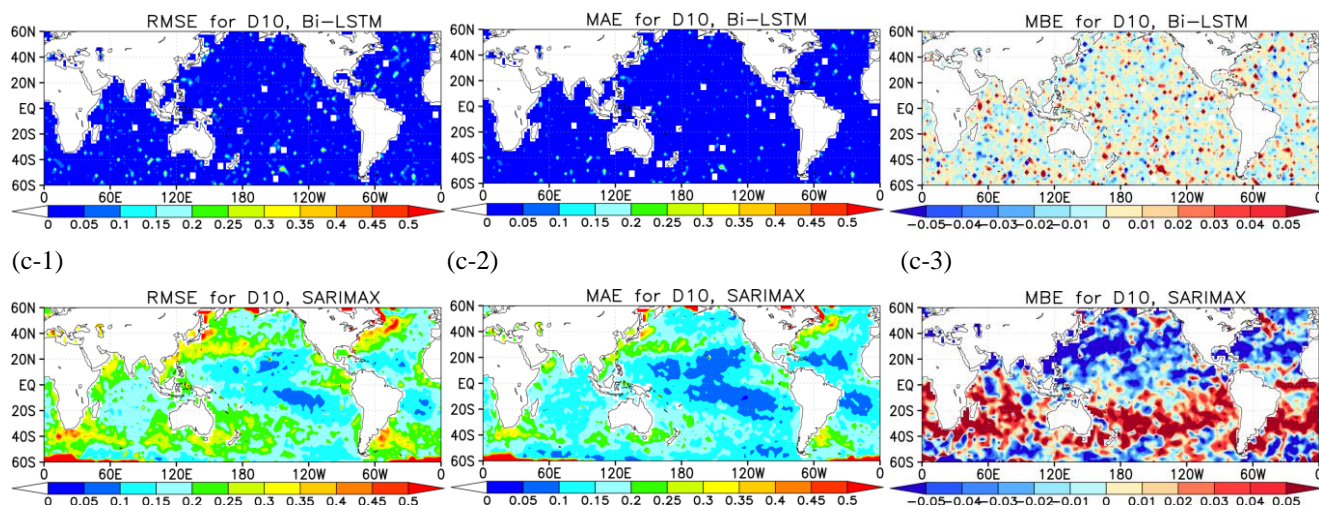
**Figure 2.** Variation of performance parameters over 30 days for BiasSST in two selected areas: (a) El Niño indicator area (Mark 2 in Figure 1b) and (b) Western Pacific warm pool (Mark 5 in Figure 1b).

The simple RNN method generally produces RMSE, MAE, and MBE values that are higher than those of the Bi-LSTM algorithm but still lower than those of the TCWAGFS model. This suggests that while the RNN can process time-series data effectively, it is not as sophisticated or efficient as the Bi-LSTM. The performance of the three algorithms from day 1 to day 5, day 10, day 15, day 20, day 25, and day 30 is summarized in Figure 2. Both RNN and Bi-LSTM consistently outperform the TCWAGFS model. For day 5, the Bi-LSTM algorithm achieves the best results with the lowest RMSE (0.005 K) and MAE (0.004 K), while RNN also outperforms the TCWAGFS model, though not as well as Bi-LSTM. As the forecast horizon extends to day 10, Bi-LSTM remains the best performer, with an RMSE of 0.008 K and an MAE of 0.007 K. The RNN algorithm shows a slight increase in errors compared to day 5 but continues to perform better than the TCWAGFS model. By day 15, Bi-LSTM still leads with an RMSE of 0.009 K and an MAE of 0.009 K, although it slightly underestimates forecasts, as indicated by its negative MBE (-0.009 K). On day 20, Bi-LSTM experiences a moderate increase in RMSE (0.013 K), though it remains the top performer. By day 25, Bi-LSTM improves again with an RMSE of 0.007 K and an MAE of 0.007 K, while RNN shows increased errors. Finally, by day 30, Bi-LSTM still leads with an RMSE of 0.016 K and an MAE of 0.015 K, outperforming the other algorithms. RNN sees a slight reduction in RMSE and MAE compared to day 25 but is still inferior to Bi-LSTM.



The TCWAGFS model struggles to predict complex and volatile sea data statistics over extended periods. Meanwhile, both the RNN and Bi-LSTM algorithms provide promising alternatives due to their ability to process sequential data. RNN offers improvements over traditional mathematical models, as recurrent neural networks are designed for sequential data and can recognize patterns over time. However, RNN's performance is limited by issues like vanishing and exploding gradients, which make it difficult for the model to capture long-term dependencies. In contrast, Bi-LSTM outperforms both the TCWAGFS and RNN algorithms. Its ability to process data in two directions—past to future and future to past—provides additional context, improving prediction performance. Bi-LSTM also proves to be more robust against the inherent noise and volatility in sea data, making it a more reliable choice for such predictions. However, while Bi-LSTM demonstrates strong results, it is not without limitations. Its performance can be highly dependent on the selection of hyperparameters, and finding the optimal set can be computationally intensive. Additionally, Bi-LSTM, like other deep learning algorithms, requires a large amount of training data, which may not always be available. The BiasSST prediction performance of GFS, SARIMAX, and Bi-LSTM is shown in Figure 3. As illustrated, the Bi-LSTM method outperforms the others, as it successfully predicts the systematic error (Column: System Error), leaving only random noise (Column: Random Noise). The selected-days forecasting performance for BiasSST of GFS and Bi-LSTM is also summarized in Table 2, where Bi-LSTM consistently achieves the best performance, with lower RMSE, MAE, and MBE values. The average RMSE for BiasSST over days 5, 10, 15, 20, and 30 in the sea grids (0E-360E, 60S-60N) for TCWAGFS and Bi-LSTM is 0.666 K and 0.009 K, respectively. Under the same conditions, the MAE is 0.559 K for TCWAGFS and 0.033 K for Bi-LSTM, while the MBE is 0.096 K and 0.0005 K, respectively. Figure 4 highlights the bias correction effect from Day 5 to Day 30. The left panels display the BiasSST of the TCWAGFS model, and the right panels show the bias after correction, which is the BiasSST of the TCWAGFS model minus the bias predicted by Bi-LSTM. For example, on Day 5, the average bias for the TCWAGFS model is 0.0261 K, and after bias correction, it is reduced to -0.009 K. In this case, the bias is reduced to just 3.4% of the original value, the highest reduction observed. On average, the bias shown in Figure 4 is reduced to around 2%. This means that, after bias correction, only minimal residual bias remains, resulting in more accurate SST predictions. The improvements in SST bias reduction, as measured by MAE (Figure A1) and RMSE (Figure A2), are provided in the Appendix, further demonstrating the effectiveness of the BiasSST correction.





**Figure 3** BiasSST prediction performance for validation data (15%) of (a) GFS, (b) Bi-LSTM, and (c) SARIMAX on day 10 for RMSE (the left column), MAE (the middle column), and MBE (the right column).

**Table 2.** Selected-days forecasting performance for BiasSST of selected algorithms in sea grids of (0°E-360°E, 60°S-60°N).

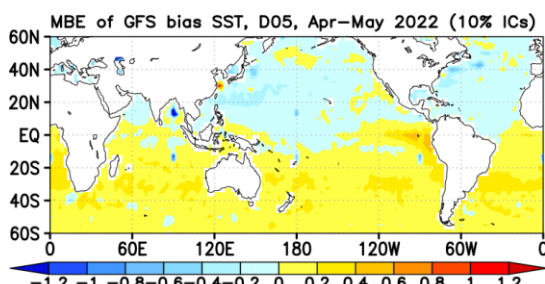
Day	Model	RMSE (K)	MAE (K)	MBE (K)
5	TCWAGFS	0.223	0.307	0.031
	Bi-LSTM	0.009	0.015	-0.001
10	TCWAGFS	0.379	0.414	0.057
	Bi-LSTM	0.010	0.023	-0.001
15	TCWAGFS	0.547	0.508	0.082
	Bi-LSTM	0.007	0.030	-0.001
20	TCWAGFS	0.739	0.607	0.108
	Bi-LSTM	0.009	0.037	0.002
25	TCWAGFS	0.928	0.706	0.137
	Bi-LSTM	0.010	0.043	0.003
30	TCWAGFS	1.182	0.814	0.158
	Bi-LSTM	0.011	0.048	0.001
Average	TCWAGFS	0.666	0.559	0.096
	Bi-LSTM	0.009	0.033	0.0005





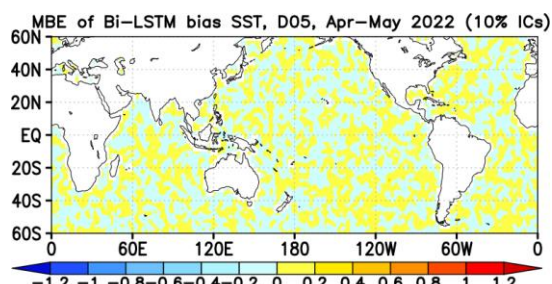
279

280 (a-1) Average bias = 0.031 K

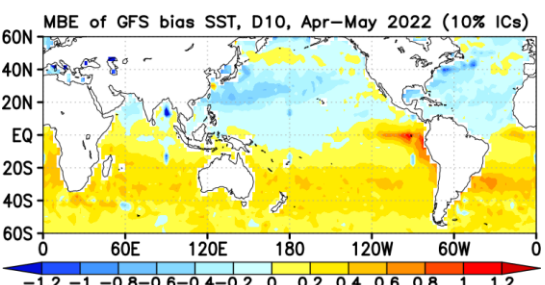


281

(a-2) Average bias = -0.001 K

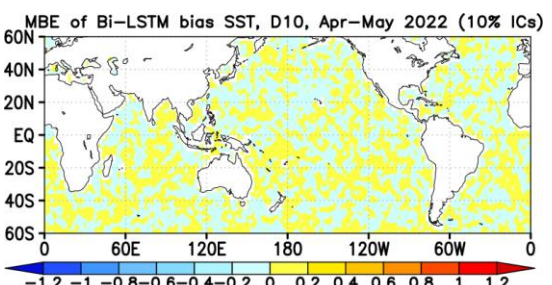


282 (b-1) Average bias = 0.057 K

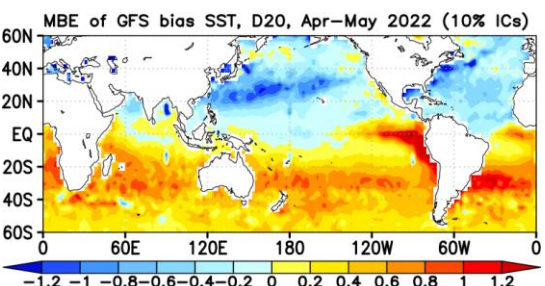


283

(b-2) Average bias = -0.001 K

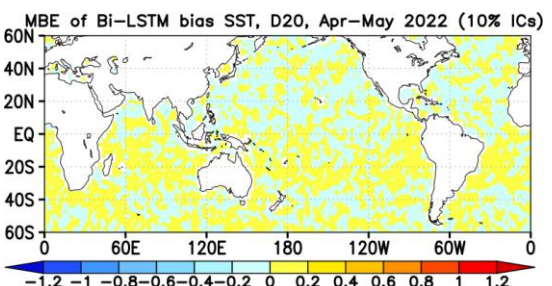


284 (c-1) Average bias = 0.108 K

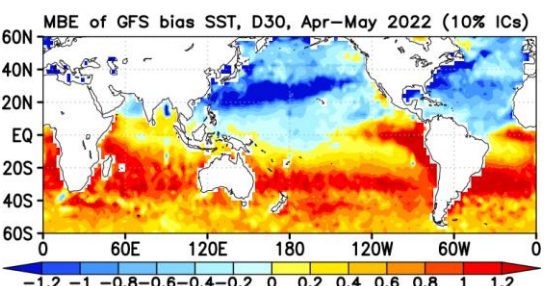


285

(c-2) Average bias = 0.002 K

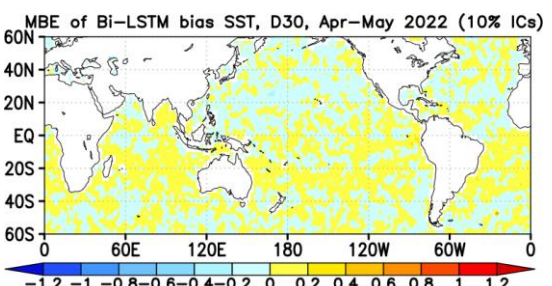


286 (d-1) Average bias = 0.158 K



287

(d-2) Average bias = 0.001 K



288 **Figure 4.** The bias correction effect of the model results. The left column represents the original bias from GFS, while the  
289 right column displays the bias correction after applying the Bi-LSTM method. The figures are arranged in chronological order  
290 from top to bottom, covering Day 5 (a-1, a-2), Day 10 (b-1, b-2), Day 20 (c-1, c-2), and Day 30 (d-1, d-2).



## 291 4 Conclusion

292 This study evaluates the performance of nine well-established machine learning algorithms for predicting sea surface  
293 temperature (SST) in the TCWAGFS model, offering a systematic and comprehensive approach to time-series-based SST  
294 forecasting. The algorithms considered include Recurrent Neural Network (RNN), Autoregressive Integrated Moving Average  
295 (ARIMA), Seasonal ARIMA with exogenous variables (SARIMAX), Long Short-Term Memory (LSTM), Prophet, Darts  
296 Transformer, Darts Exponential Smoothing, Time Series Imaging, and Bidirectional Long Short-Term Memory (Bi-LSTM).  
297 Compared to traditional dynamical models like TCWAGFS, these machine learning approaches excel due to their ability to  
298 handle sequential data, capture complex temporal dependencies, and learn patterns from past time steps, which is crucial for  
299 accurate weather and climate predictions. Among the algorithms tested, Bi-LSTM stands out as the most effective, consistently  
300 outperforming the others, including conventional methods like RNN and the baseline TCWAGFS model. Bi-LSTM achieved  
301 significantly lower RMSE, MAE, and MBE values, demonstrating its superior ability to predict SST bias. The bidirectional  
302 structure of the Bi-LSTM model, which processes information in both forward and backward directions, allows it to capture  
303 long-term dependencies and contextual information critical to the accurate prediction of time-series data, such as SST statistics.  
304 This capability positions Bi-LSTM as a powerful tool for improving the accuracy of SST bias correction in the TCWAGFS  
305 model.

306 A critical takeaway from this study is the importance of accounting for temporal dependencies in SST prediction.  
307 SSTs are influenced by complex, historical factors that have long-lasting effects on present conditions. Conventional methods  
308 often struggle to retain and leverage these long-term dependencies, which limits their predictive accuracy. Bi-LSTM, by  
309 contrast, excels in learning and retaining these intricate relationships, providing a more holistic and accurate view of SST  
310 evolution. The comparison between the original bias from TCWAGFS and the bias correction achieved through the Bi-LSTM  
311 method illustrates the algorithm's ability to reduce the original bias by an average of 98%. This substantial improvement  
312 highlights the potential for machine learning methods to enhance the precision and reliability of environmental forecasts. While  
313 the offline bias correction achieved by Bi-LSTM in this study is significant, there is even greater potential in integrating  
314 dynamical and statistical models for real-time applications. A hybrid approach that combines the strength of dynamical models,  
315 like TCWAGFS, with the advanced learning capabilities of machine learning models could further enhance extended weather  
316 forecasts. Such an integration could address not only SST bias but also other biases within global forecasting systems, offering  
317 a more comprehensive solution for long-range and sub-seasonal weather prediction. Looking ahead, the next step will involve  
318 transitioning from offline bias correction to the full integration of the Bi-LSTM model with the TCWAGFS system. This  
319 integration will aim to reduce SST bias in real-time forecasting, thereby improving the overall accuracy of the TCWAGFS  
320 model in predicting weather and climate conditions. Additionally, this approach could be extended to other climatological  
321 variables that exhibit similar temporal dependencies, such as atmospheric temperature, humidity, and precipitation patterns.  
322 The broader implications of this study extend beyond SST bias correction. The machine learning algorithms explored here,  
323 particularly Bi-LSTM, demonstrate a significant potential for addressing challenges in various domains of environmental and



climate modeling. Their capacity to handle sequential data and capture long-term dependencies offers valuable insights not only for improving numerical weather prediction models but also for enhancing the accuracy of sub-seasonal and seasonal climate predictions. Moreover, the methodologies developed in this study could be adapted to other domains of Earth system science, where bias correction and long-range forecasting are essential for policy-making, disaster management, and sustainability planning.

In summary, this study underscores the transformative potential of machine learning algorithms in environmental forecasting. The superior performance of Bi-LSTM in reducing SST bias, coupled with its broader applicability in time-series analysis, makes it a valuable tool for improving the accuracy and reliability of numerical weather prediction models. By incorporating advanced statistical techniques alongside traditional dynamical models, we can enhance the precision of future climate and weather forecasts, ultimately contributing to more informed decision-making in the face of environmental challenges.

#### **Code and data availability**

All model codes and data presented in this study are archived at Zenodo: <https://doi.org/10.5281/zenodo.15400607> (Li et al., 2025). The development repository is also publicly accessible via GitHub: <https://github.com/btsuang/CWA-biastg-Train>. All experiments were conducted at the National Center for High-Performance Computing.

#### **Acknowledgements**

This work was supported by the Taiwan Ministry of Science and Technology (grant nos. MOST 111-2111-M-002-015-, MOST 112-2121-M-005-001- and MOST 113-2121-M-005-006-). We are grateful to the National Center for High-Performance Computing for providing computer facilities.

#### **Author contributions**

SML, NS, FCW, WLT, PHS and SJW conceptualized the study, performed data analysis, and wrote the manuscript. BJT, PYL, JHC, YYL, and SYK developed the model and conducted the simulations. All authors reviewed and approved the final manuscript.

#### **Competing interests**

The contact author has declared that none of the authors has any competing interests.

#### **Financial support**

This research has been supported by the Ministry of Science and Technology, Taiwan (grant nos. MOST 111-2111-M-002-015-, MOST 112-2121-M-005-001- and MOST 113-2121-M-005-006-).



## References

- Abhilash, S., Sahai, A. K., Borah, N., Chattopadhyay, R., Joseph, S., Sharmila, S., De, S., and Goswami, B. N.: Does bias correction in the forecasted SST improve the extended range prediction skill of active-break spells of Indian summer monsoon rainfall?, *Atmos. Sci. Lett.*, 15(2), 114–119, <https://doi.org/10.1002/asl2.477>, 2014.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M.: *Time Series Analysis: Forecasting and Control*, 5th Edition ch4, John Wiley, 2015.
- Choi, H. M., Kim, M. K., and Yang, H.: Deep-learning model for sea surface temperature prediction near the Korean Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, 208, 105262, <https://doi.org/10.1016/j.dsr2.2023.105262>, 2023.
- Demory, M.-E., Berthou, S., Fernández, J., Sørland, S. L., Brogli, R., Roberts, M. J., Beyerle, U., Seddon, J., Haarsma, R., Schär, C., Buonomo, E., Christensen, O. B., Ciarlo, J. M., Fealy, R., Nikulin, G., Peano, D., Putrasahan, D., Roberts, C. D., Senan, R., Steger, C., Teichmann, C., and Vautard, R.: European daily precipitation according to EURO-CORDEX regional climate models (RCMs) and high-resolution global climate models (GCMs) from the High-Resolution Model Intercomparison Project (HighResMIP), *Geosci. Model Dev.*, 13, 5485–5506, <https://doi.org/10.5194/gmd-13-5485-2020>, 2020.
- DeMott, C. A., Klingaman, N. P., and Woolnough, S. J.: Atmosphere-ocean coupled processes in the Madden-Julian oscillation, *Rev. of Geophysics*, 53, 1099–1154, <https://doi.org/10.1002/2014RG000478>, 2015.
- DeMott, C. A., Stan, C., Randall, D. A., and Branson, M. D.: Intraseasonal variability in coupled GCMs: The roles of ocean feedbacks and model physics, *J. Climate*, 27(13), 4970–4995. <https://doi.org/10.1175/JCLI-D-13-00760.1>, 2014.
- Fei, T., Huang, B., Wang, X., Zhu, J., Chen, Y., Wang, H., and Zhang, W.: A Hybrid Deep Learning Model for the Bias Correction of SST Numerical Forecast Products Using Satellite Data, *Remote Sens.*, 14, 1339, <https://doi.org/10.3390/rs14061339>, 2022.
- Funahashi, K.-i., and Nakamura Y.: Approximation of dynamical systems by continuous time recurrent neural networks, *Neural Netw.*, 6(6): 801–806, [https://doi.org/10.1016/S0893-6080\(05\)80125-X](https://doi.org/10.1016/S0893-6080(05)80125-X), 1993.
- Gao, Y., Klingaman, N. P., DeMott, C. A., and Hsu, P.-C.: Boreal summer intraseasonal oscillation in a superparameterized general circulation model: effects of air–sea coupling and ocean mean state, *Geosci. Model Dev.*, 13, 5191–5209, <https://doi.org/10.5194/gmd-13-5191-2020>, 2020.
- Gutjahr, O., Schefczyk, L., Reiter, P., and Heinemann, G.: Impact of the horizontal resolution on the simulation of extremes in COSMO-CLM, *Meteorol. Z.*, 25, 543–562, <https://doi.org/10.1127/metz/2016/0638>, 2016.
- Haiden, T., Janousek, M., Vitart, F., Bouallègue, Z. B., Ferranti, L., Prates, F., & Richardson, D. (2018). Evaluation of ECMWF forecasts, including the 2018 upgrade. ECMWF Technical Memorandum No. 831, <https://doi.org/10.21957/ldw15ckqi>, 2018.





- 390 Ham, Y. G., Kim, J. H., and Luo, J. J.: Deep learning for multi-year ENSO forecasts, *Nature*, 573(7775), 568-572,  
391 <https://doi.org/10.1038/s41586-019-1559-7>, 2019.
- 392 Han, G., Zhou, J., Shao, Q., Li, W., Li, C., Wu, X., Cao, L., Wu, H., Li, Y., and Zhou, G.: Bias correction of sea surface  
393 temperature retrospective forecasts in the South China Sea, *Acta Oceanologica Sinica*, 41, 41-50,  
394 <https://doi.org/10.1007/s13131-021-1880-5>, 2022.
- 395 Hao, P., Li, S., Song, J., and Gao, Y.: Prediction of Sea Surface Temperature in the South China Sea Based on Deep  
396 Learning, *Remote Sens*, 15, 1656, <https://doi.org/10.3390/rs15061656>, 2023.
- 397 Herzen, J., Lässig, F., Piazzetta, S.G., Neuer, T., Tafti, L., Raille, G., Van Pottelbergh, T., Pasieka, M., Skrodzki, A.,  
398 Huguenin, N., Dumonal, M., Kościsz, J., Bader, D., Gusset, F., Benheddi, M., Williamson, C., Kosinski, M., Petrik, M.,  
399 Grosch, G.: Darts: User-friendly modern machine learning for time series, *J. Mach. Learn. Res.*, 23.1 (2022): 5442-  
400 5447, <https://doi.org/10.48550/arXiv.2110.03224>, 2022.
- 401 Hochreiter, S., and Schmidhuber, J.: Long short-term memory. *Neural comput.*, 9(8), 1735-1780,  
402 <https://doi.org/10.1162/neco.1997.9.8.1735>, 1997.
- 403 Hortal M, Simmons A. J.: Use of reduced Gaussian grids in spectral models, *Mon. Weather. Rev.* 119(4):1057–1074,  
404 [https://doi.org/10.1175/1520-0493\(1991\)119<1057:UORGGI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2), 1991.
- 405 Iles, C. E., Vautard, R., Strachan, J., Joussaume, S., Eggen, B. R., and Hewitt, C. D.: The benefits of increasing resolution in  
406 global and regional climate simulations for European climate extremes, *Geosci. Model Dev.*, 13, 5583–5607,  
407 <https://doi.org/10.5194/gmd-13-5583-2020>, 2020.
- 408 Ji, L., Wang, Z., Chen, M., Fan, S., Wang, Y., and Shen Z.: How Much Can AI Techniques Improve Surface Air  
409 Temperature Forecast? —A Report from AI Challenger 2018 Global Weather Forecast Contest, *J Meteorol Res* 33,  
410 989–992, <https://doi.org/10.1007/s13351-019-9601-0>, 2019.
- 411 Jia, X., Ji, Q., Han, L., Liu, Y., Han, G., and Lin, X.: Prediction of Sea Surface Temperature in the East China Sea Based on  
412 LSTM Neural Network. *Remote Sens*, 14, 3300, <https://doi.org/10.3390/rs14143300>, 2022.
- 413 Juang, H. M. H., Wu, T. Y., Liu, P. Y. B., Lin, H. Y., Lee, C. T., Kueh, M. T., Fan, J. F., Chen, J. H. R., Lu, M. M., and Lin,  
414 P. L.: The First Operational Version of Taiwan Central Weather Bureau’s One-Tier Global Atmosphere–Ocean  
415 Coupled Forecast System for Seasonal Prediction. *Weather and Forecasting*, 39(5), 745-763, 2024.
- 416 Kim, H., Ham, Y., Joo, Y., and Son, S.: Deep learning for bias correction of MJO prediction, *Nat. Commun.*, 12, 3087,  
417 <https://doi.org/10.1038/s41467-021-23406-3>, 2021.
- 418 Kim, H.-M., Hoyos, C. D., Webster, P. J., and Kang, I.-S.: Ocean–atmosphere coupling and the boreal winter MJO, *Clim.*  
419 *Dynam.*, 35, 771-784, <https://doi.org/10.1007/s00382-009-0612-x>, 2010.
- 420 Lan, Y. Y., Hsu, H. H., & Tseng, W. L. To quantify the impact of SST feedback periodicity on atmospheric intraseasonal  
421 variability in the tropical regions. *Geosci. Model Dev. Discuss.*, 17, 3897–3918, <https://doi.org/10.5194/gmd-17-3897-2024>,  
422 2024.



- 423 Lan, Y.-Y., Hsu, H.-H., Tseng, W.-L., and Jiang, L.-C.: Embedding a one-column ocean model in the Community  
424 Atmosphere Model 5.3 to improve Madden–Julian Oscillation simulation in boreal winter, *Geosci. Model Dev.*, 15,  
425 5689–5712, <https://doi.org/10.5194/gmd-15-5689-2022>, 2022.
- 426 Lee, W.-L., Wang, Y.-C., Shiu, C.-J., Tsai, I., Tu, C.-Y., Lan, Y.-Y., Chen, J.-P., Pan, H.-L., and Hsu, H.-H.: Taiwan Earth  
427 System Model Version 1: description and evaluation of mean state, *Geosci. Model Dev.*, 13, 3887–3904,  
428 <https://doi.org/10.5194/gmd-13-3887-2020>, 2020.
- 429 Li, K. S.-M., Sinhabahu, N., Tsuang, B.-J., Wu, F.-C., Tseng, W.-L., Kuo, P.-H., Wang, S.-J., Liu, P.-Y., Chen, J.-H., Lan,  
430 Y.-Y., and Kung, S.-Y.: CWA-biastg-Train – Model code and data for SST bias correction in TCWAGFS, Zenodo [data  
431 set], <https://doi.org/10.5281/zenodo.15400607>, 2025.
- 432 Li, X., Kang, Y., and Li, F.: Forecasting with time series imaging, *Expert Systems with Applications*, 160, 113680,  
433 <https://doi.org/10.1016/j.eswa.2020.113680>, 2020.
- 434 Liou C.-S., Chen J.-H., Terng C.-T., Wang F.-J., Fong C.-T., Rosmond T. E., Kuo H.-C., Shiao C.-H., Cheng M.-D.: The  
435 second-generation global forecast system at the central weather bureau in Taiwan, *Wea. Forecasting* 12(3):653–663,  
436 [https://doi.org/10.1175/1520-0434\(1997\)012<0653:TSGGFS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0653:TSGGFS>2.0.CO;2), 1997.
- 437 Lopatka, A.: Meteorologists predict better weather forecasting with AI, *Physics Today*, 72(5), 32–34,  
438 <https://doi.org/10.1063/PT.3.4201>, 2019.
- 439 Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kühnlein, C., Mozdzyński, G., Hamrud, M., Smolarkiewicz, P.: A  
440 new grid for the IFS. ECMWF Newsl 146:23–28, <https://doi.org/doi:10.21957/zwdu9u5i>, 2016.
- 441 Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T.,  
442 and Vitart, F.: The new ECMWF seasonal forecast system (System 4), ECMWF Technical Memorandum, 656, 49 pp.,  
443 2011.
- 444 Noack, M. M., and Reyes, K. G.: Mathematical nuances of Gaussian process-driven autonomous experimentation, *MRS*  
445 *Bulletin* 48, 153–163, <https://doi.org/10.1557/s43577-023-00478-8>, 2023.
- 446 Nurdianti, S., Sopaheluwakan, A., Pratama, Y. A., and Najib, M. K.: Statistical bias correction on the climate model for el  
447 nino index prediction. *Al-Jabar: j. pendidik. mat.*, 12, 273–282, <http://dx.doi.org/10.24042/ajpm.v12i2.8884>, 2021.
- 448 Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A. and Stadtler, S.: Can deep  
449 learning beat numerical weather prediction? *Phil. Trans. R. Soc. A*.3792020009720200097,  
450 <http://doi.org/10.1098/rsta.2020.0097>, 2021.
- 451 Schuster, M., and Paliwal, K. K.: Bidirectional recurrent neural networks, *IEEE trans. Signal Process.*, 45(11), 2673–2681,  
452 <https://doi.org/10.1109/78.650093>, 1997.
- 453 Silini, R., Barreiro, M., and Masoller, C.: Machine learning prediction of the Madden-Julian oscillation, *Npj. Clim. Atmos.*  
454 *Sci.* 4, 57, <https://doi.org/10.1038/s41612-021-00214-6>, 2021.

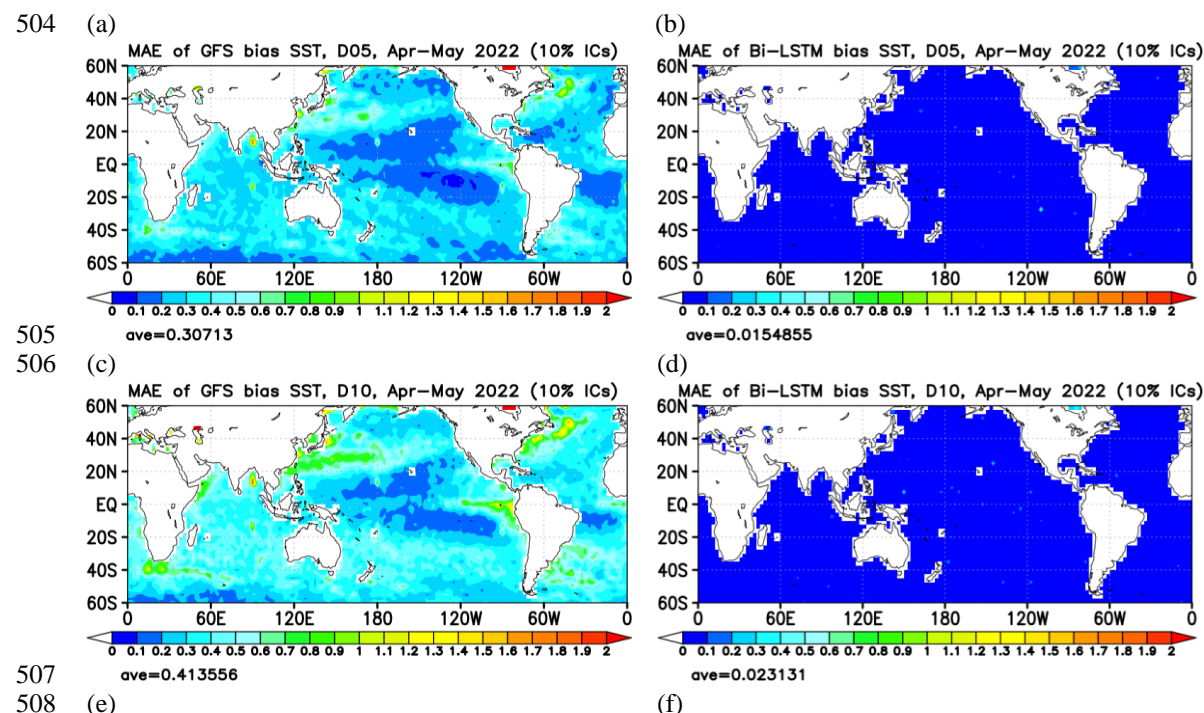


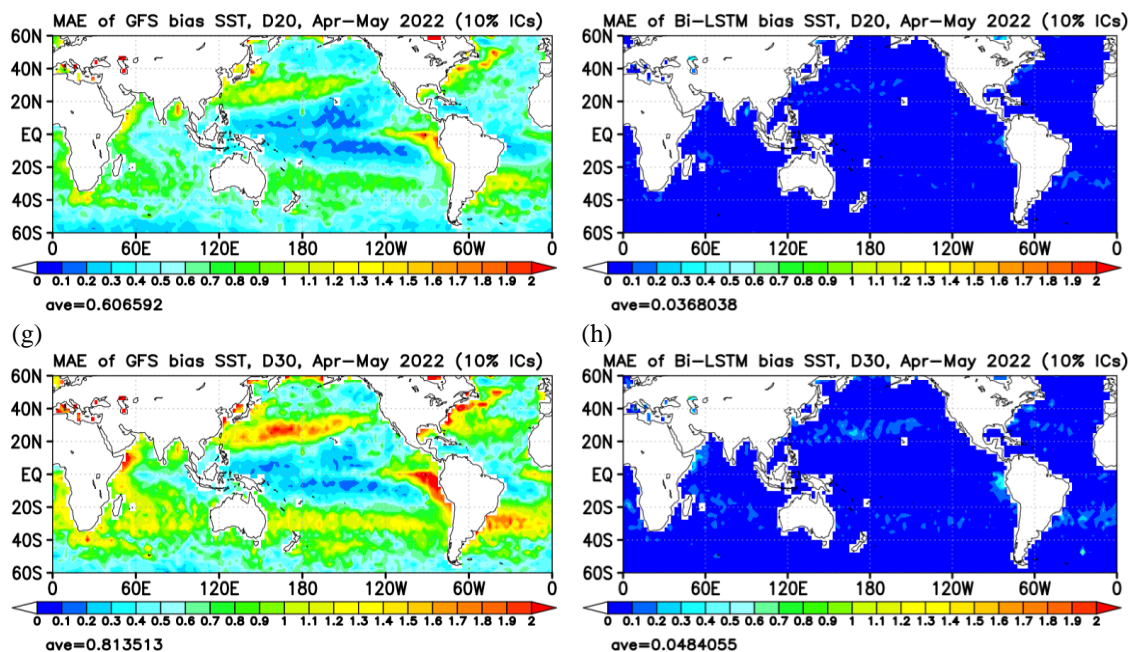
- 455 Singh, V. K., Mohan, M. and Bhati, S.: Evaluation of different parameterization schemes in the WRF model for assessment  
456 of meteorological conditions over an industrial region in South-East India, *Theor Appl Climatol* 150, 1045–1066,  
457 <https://doi.org/10.1007/s00704-022-04196-0>, 2022.
- 458 Staniforth, A., and J. Côté.: Semi-Lagrangian integration schemes for atmospheric models—A review, *Mon. Wea. Rev.*, 119,  
459 2206–2223, [https://doi.org/10.1175/1520-0493\(1991\)119<2206:SLISFA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<2206:SLISFA>2.0.CO;2), 1991.
- 460 Strandberg, G. and Lind, P.: The importance of horizontal model resolution on simulated precipitation in Europe – from  
461 global to regional models, *Weather Clim. Dynam.*, 2, 181–204, <https://doi.org/10.5194/wcd-2-181-2021>, 2021.
- 462 Su, C.-Y., Wu, C.-M., Chen, W.-T., and Chen J.-H.: The effects of the unified parameterization in the CWBGFS: the diurnal  
463 cycle of precipitation over land in the Maritime Continent, *Clim Dyn* 58, 223–233, [https://doi.org/10.1007/s00382-021-](https://doi.org/10.1007/s00382-021-05899-2)  
464 05899-2, 2022.
- 465 Su, C.-Y., Wu, C.-M., Chen, W.-T., and Chen, J.-H.: Implementation of the Unified Representation of deep moist convection  
466 in the CWBGFS, *Mon. Weather Rev.*, 149(10), 3525–3539. <https://doi.org/10.1175/mwr-d-21-0067.1>, 2021.
- 467 Taylor, S. J., and Letham, B.: Forecasting at scale, *Am. Stat.*, 72(1), 37–45, <https://doi.org/10.1080/00031305.2017.1380080>,  
468 2018.
- 469 Tseng, W.-L., Hsu, H.-H., Lan, Y.-Y., Lee, W.-L., Tu, C.-Y., Kuo, P.-H., Tsuang, B.-J., and Liang, H.-C.: Improving  
470 Madden–Julian oscillation simulation in atmospheric general circulation models by coupling with a one-dimensional  
471 snow–ice–thermocline ocean model, *Geosci. Model Dev.*, 15, 5529–5546, <https://doi.org/10.5194/gmd-15-5529-2022>,  
472 2022.
- 473 Tseng, W.-L., Tsuang, B.-J., Keenlyside, N. S., Hsu, H.-H. and Tu, C.-Y.: Resolving the upper-ocean warm layer improves  
474 the simulation of the Madden-Julian oscillation, *Clim. Dynam.*, 44, 1487–1503, [https://doi.org/10.1007/s00382-014-](https://doi.org/10.1007/s00382-014-2315-1)  
475 2315-1, 2015.
- 476 Vagropoulos, S. I., Chouliaras, G. I., Kardakos, E. G., Simoglou, C. K., and Bakirtzis, A. G.: Comparison of SARIMAX,  
477 SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. In 2016 IEEE  
478 international energy conference (ENERGYCON), pp. 1–6, Leuven, Belgium,  
479 <https://doi.org/10.1109/ENERGYCON.2016.7514029>, 2016.
- 480 Vitart, F., and Balmaseda, M.: Impact of sea surface temperature biases on extended-range forecasts. ECMWF Tech. Memo.  
481 830, 21 pp., <https://doi.org/10.21957/xvbl914ps>, 2018.
- 482 Vitart, F., Robertson, A. W., Spring, A., Pinault, F., Roškar, R., Cao, W., Bech, S., Bienkowski, A., Caltabiano, N., De  
483 Coning, E., Denis, B., Dirkson, A., Dramsch, J., Dueben, P., Gierschendorf, J., Kim, H. S., Nowak, K., Landry, D.,  
484 Lledó, L., Palma, L., Rasp, S., and Zhou, S.: Outcomes of the WMO Prize Challenge to Improve Subseasonal to  
485 Seasonal Predictions Using Artificial Intelligence, *Bull. Am. Meteorol. Soc.*, 103, E2878–E2886,  
486 <https://doi.org/10.1175/BAMS-D-22-0046.1>, 2022.



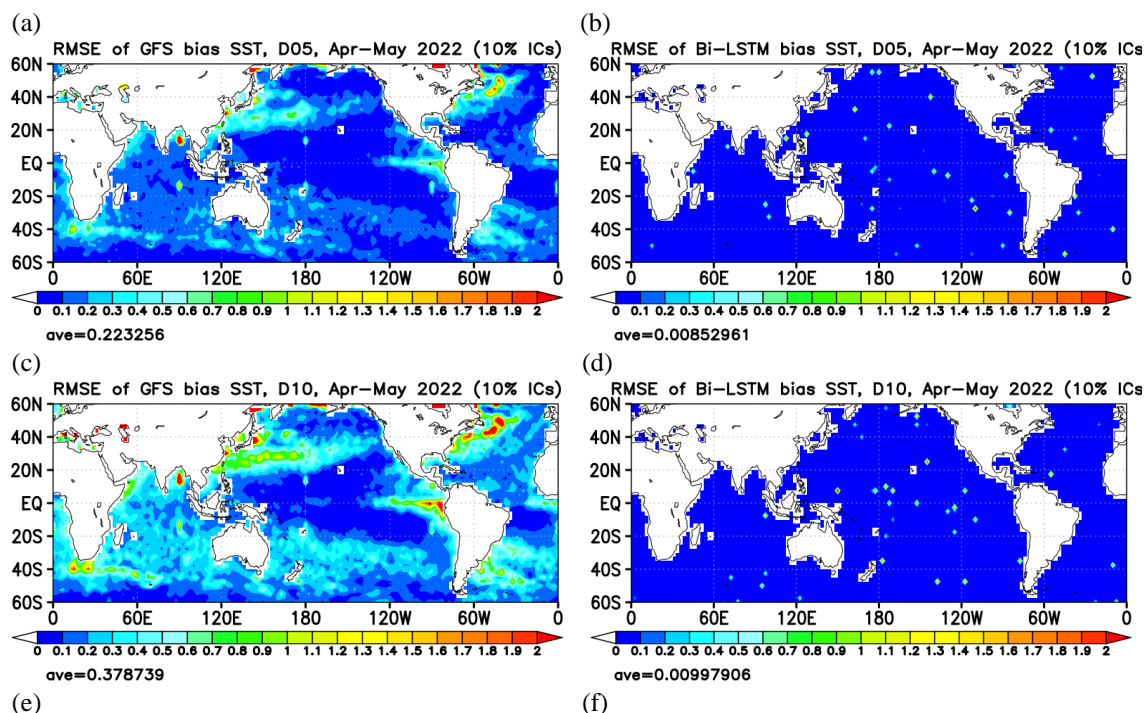
- 487 Weyn, J. A., Durran, D. R., and Caruana, R.: Can machines learn to predict weather? Using deep learning to predict gridded  
488 500-hPa geopotential height from historical weather data, *J. Adv. Model. Earth Syst.* 11, 2680–2693,  
489 <https://doi.org/10.1029/2019MS001705>, 2019.
- 490 Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., and Chen, Z.: Short and mid-term sea surface temperature prediction using  
491 time-series satellite data and LSTM-AdaBoost combination approach, *Remote Sens. Environ.*, 233, 111358,  
492 <https://doi.org/10.1016/j.rse.2019.111358>. 2019.
- 493 Xu, S., Dai, D., Cui, X., Yin, X., Jiang, S., Pan, H., Wang, G.: A deep learning approach to predict sea surface temperature  
494 based on multiple modes, *Ocean Model.*, 181: 102158, <https://doi.org/10.1016/j.ocemod.2022.102158>, 2023.
- 495 Zhou, X., Zhu, Y., Hou, D., Fu, B., Li, W., Guan, H., Sinsky, E., Kolczynski, W., Xue, X., Luo, Y., Peng, J., Yang, B.,  
496 Tallapragada, V., and Pegion, P.: The Development of the NCEP Global Ensemble Forecast System Version 12,  
497 *Weather Forecast.*, 37, 1069–1084, <https://doi.org/10.1175/waf-d-21-0112.1>, 2022.

## 503 Appendix

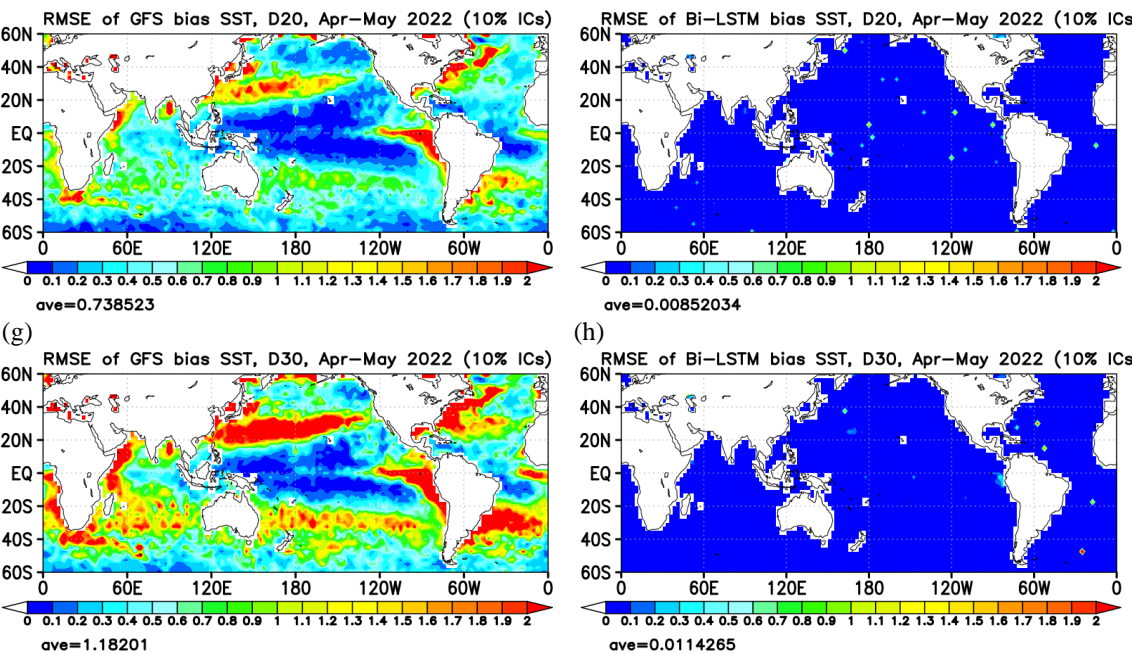




**Figure A1.** The forecasting performance (MAE) for BiasSST of the model results. The left column represents the MAE results from GFS, while the right column displays the MAE results from Bi-LSTM. The figures are arranged in chronological order from top to bottom, covering Day 5 (a, b), Day 10 (c, d), Day 20 (e, f), and Day 30 (g, h).







**Figure A2.** The forecasting performance (RMSE) for BiasSST of the model results. The left column represents the RMSE results from GFS, while the right column displays the RMSE results from Bi-LSTM. The figures are arranged in chronological order from top to bottom, covering Day 5 (a, b), Day 10 (c, d), Day 20 (e, f), and Day 30 (g, h).