PIRT: A Physics-Informed Red Tide Deep Learning Forecast Model Considering Causal-Inferred Predictors Selection

Bin Mu, Bo Qin, Shijin Yuan, Xin Wang, and Yuxuan Chen

***Abstract*— In this letter, a Physics-Informed Red Tide (PIRT) forecast model considering causal-inferred predictors selection is proposed. Specifically, the directed acyclic graph-graph neural network (DAG-GNN) method is first applied to quantify the causality among multiple ocean-atmosphere-biology variables for selecting the most significant predictors of the red tides (or other chlorophyll variations). Then, the encoder-decoder model consisting of an Energy Attention Module (EAM) is built for daily red tide forecasting. The multisourced multivariate dataset during 2010–2020 covering the East China Sea serves to train and evaluate PIRT. The experimental results demonstrate that the predictors in the learned causal graph are closely related to the occurrence and decay of red tides, which exhibits high physical interpretability. PIRT has a superior forecasting skill, the predictions of which are with highly consistent spatial patterns, especially in extreme events. The seven-lead-day forecast errors for chlorophyll are within 0.9 mg·m^{-3}, which is much better than the other models. This also indicates that PIRT can be used as a reliable tool to study the ecology of the East China Sea.**

***Index Terms*— Causal inference, multivariate forecast, physics-informed neural network (PINN), red tide.**

I. Introduction

RED tide is a disastrous natural phenomenon involving harmful algal blooms, changing the sea color from normal to red or almost brown, which has a serious influence on sea ecosystems, coastal economy, and human sustainability [1]. Therefore, it is of great significance to accurately predict red tides. In general, chlorophyll concentration is the key indicator (red tides considered to occur when chlorophyll concentration is more than 10 mg·m^{-3}) for red tide forecasting [2], which represents the phytoplankton biomass [3] and eutrophication in water bodies [4]. Consequently, chlorophyll forecasting is considered as one of red tide warning techniques.

With the accumulation of marine ecological observation data, more and more studies tend to use deep learning techniques for red tide forecasting. However, existing deep learning models usually suffer from the following four problems: 1) limiting to single-grid forecasts rather than spatial oceans [5], [6], [7], [8]; 2) limiting to a single (or few) predictors, ignoring important multivariate predictors related to the complex ocean-atmosphere-biology coupled mechanisms during red tide evolutions [9], [10], [11]; 3) limiting to the fundamental and uncustomized model structures; and 4) the low credibility of the model [12], hardly involving the causal inference and physical interpretability among predictors [13]. These shortcomings confine performance improvement and mechanism cognition in red tide.

To tackle the above challenges, we construct a red tide forecast model for the East China Sea by thoroughly considering the dynamic mechanism of red tide (see Section II-A). When selecting predictors, we use directed acyclic graph-graph neural network (DAG-GNN) method [14] to quantify the causal relationship between ocean-atmosphere-biology multivariables and then select the most relevant predictors for red tide forecasts (see Section II-B). When constructing the model, we design a Physics-Informed Red Tide (PIRT) deep learning forecast model, which contains independent encoders/decoders for the chosen cause-and-effect predictors and our proposed Energy Attention Module (EAM) to simulate energy interaction between the causal chain to maintain the physical consistency (see Section II-C). PIRT is an end-to-end model and can predict the spatial-temporal distribution of these chosen multivariables in the East China Sea region by seven lead days, in which the CHL forecasting results can be used to determine the occurrence of red tides. Meanwhile, PIRT is based on the multisourced multivariate daily dataset of the East China Sea (see Section II-D), the loss function of which is MSE.

To the best of our knowledge, this is the first integration of causal inference and physical guidance for red tide forecasting. The main contributions of this letter are.

1) The learned causal graph identifies the predictors as NO_3, PO_4, SiO_3, SST, LH, OC, NPPV, Density, Wind, and CHL (more details in Table I). These variables are closely related to the occurrence and decay of red tide [15], [16], [17], providing guarantees of accurate red

Manuscript received 26 October 2022; revised 29 January 2023; accepted 25 February 2023. Date of publication 1 March 2023; date of current version 16 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42075141, in part by the first batch of Model Interdisciplinary Joint Research Projects of Tongji University in 2021 under Grant YB-21-202110, in part by the Key Project Fund of Shanghai 2020 "Science and Technology Innovation Action Plan" for Social Development under Grant 20dz1200702, in part by the Meteorological Joint Funds of the National Natural Science Foundation of China under Grant U2142211, and in part by the National Key Research and Development Program of China under Grant 2020YFA0608000. *(Bin Mu and Bo Qin contributed equally to this work.) (Corresponding authors: Bo Qin; Shijin Yuan.)*

The authors are with the School of Software Engineering, Tongji University, Shanghai 201804, China (e-mail: binmu@tongji.edu.cn; 1710966@tongji.edu.cn; yuanshijin@tongji.edu.cn; 2111284@tongji.edu.cn; 2211307@tongji.edu.cn).

The archive of dataset and source code can be found at doi.org/10.5281/zenodo.7550637

Digital Object Identifier 10.1109/LGRS.2023.3250642

TABLE I
TRAINING DATASET WITH SELECTED MULTIPLE VARIABLES RELATED TO RED TIDE

Name	Unit	Description
SST†	℃	Sea surface temperature
LH†	$W \cdot m^{-2}$	Sea water latent heat
Wind-u†	$m \cdot s^{-1}$	Eastward wind
Wind-v†	$m \cdot s^{-1}$	Northward wind
NO3ǂ	$mmol \cdot m^{-3}$	Mole concentration of nitrate
PO4ǂ	$mmol \cdot m^{-3}$	Mole concentration of phosphate
SiO3ǂ	$mmol \cdot m^{-3}$	Mole concentration of silicate
CHLǂ	$mg \cdot m^{-3}$	Mass concentration of chlorophyll A
Density†	m	Ocean mixed layer thickness defined by sigma theta
OC-u†	$m \cdot s^{-1}$	Eastward ocean current velocity
OC-v†	$m \cdot s^{-1}$	Northward ocean current velocity
NPPVǂ	$mg \cdot m^{-3}$	Net primary production of biomass per unit volume
Salinity†	$1e^{-3}$	Sea water salinity
Height†	m	Sea surface height above the geoid
O2ǂ	$mmol \cdot m^{-3}$	Mole concentration of dissolved molecular oxygen

Note: *According to the causal inference results (see Section III.A), the blue/green/orange shade covers the cause/CHL/effect variables. Meanwhile, the gray shade covers the variables filtered out directly. Superscript † and ǂ represent the oceanic variables (10.48670/moi-00168 and 10.48670/moi-00024), Superscript † represents the atmospheric variables (10.48670/moi-00185), Superscript ǂ represents the biological variables (10.48670/moi-00019).*

tide forecasting and effective latent mechanism exploring by PIRT.

2) PIRT has a very high forecasting skill, the seven-lead-day forecast errors within 0.9 $mg \cdot m^{-3}$ for CHL, and do not exceed 10% of the observed values for other variables as well, which is much better than the existing models.

II. METHODOLOGY

A. Problem Formalization

Compared with the traditional single-predictor-based series forecast model which can be formulated as $\hat{x}_{t+1} = \mathcal{F}(x_{\infty:t})$ (where $x_{\infty:t}$ is the series observation by time t, $\hat{x}_{t+1}$ is the prediction at time $t+1$, and $\mathcal{F}(\cdot)$ is the forecast system), our multivariate red tide forecast model can be formalized as follows:

$$\hat{P}_{t+1} = \mathcal{F}(P_{1:t}) \tag{1}$$

where $P_\tau = \{p_\tau^1, \ldots, p_\tau^M\}$ $(\tau = 1, 2, \ldots)$ is the input of selected predictors p_τ^m $(m = 1:M)$ at time step τ. For better forecasting skills and physical interpretability, we focus on physics-related predictors selection in p_τ^m and model structure customization in $\mathcal{F}(\cdot)$. We take advantage of causal inference and physics-informed neural network (PINN) [18] in the modeling.

B. Causal-Inferred Predictors Selection

We use DAG-GNN [14] as the causal inference model to quantitatively mine ocean–atmosphere–biology relationships related to red tides, which are usually represented by a DAG [19] from the data to identify dominant predictors and filter irrelevant variables. Specifically, the DAG-GNN [right side of (2)] extends the linear structural equation model [SEM, left side of (2)], where $A \in R^{M\times M}$ is the weighted adjacency matrix of DAG, $X \in R^{M\times N}$ is the joint distribution of data samples and $Z \in R^{M\times N}$ is the bias, M, N denote the nodes in DAG and the number of samples. DAG-GNN performs nonlinear transformations of X and Z by introducing reversible $f(\cdot)$ and $g(\cdot)$ operators to tap nonlinear causality, respectively,

$$X = A^T X + Z \Longleftrightarrow f(X) = A^T f(X) + g(Z). \tag{2}$$

Equation (2) can be further disassembled into an autoencoder-based model via a special graph convolution operator $I - A^T$ as shown in (3). This model can be trained by variational inference [20] for matching the actual scenes better

$$\begin{cases} \text{encoder: } Z = g^{-1}\big((I - A^T) f(X)\big) \\ \text{decoder: } X = f^{-1}\Big(\big(I - A^T\big)^{-1} g(Z)\Big). \end{cases} \tag{3}$$

Besides, to filter irrelevant variables, we set the weights in A to 0 when they are less than 0.03 (an effective threshold confirmed by repeated experiments) to ignore their effects.

C. PIRT Forecast Model

We propose the PIRT Forecast Model for predicting the spatial-temporal distribution of multiple variables closely related to red tides as shown in Fig. 1. Structurally, it can be divided into three parts: encoder, EAM, and decoder.

According to the learned DAG, the key variable CHL will be affected by some related variables (denoted as cause) and act on some others (denoted as effect) as energy transfer of cause → CHL → effect. Thus, we first divide the input of PIRT into three groups (P^t_{cause}, P^t_{CHL}, P^t_{effect}), which represent the cause variables, CHL, and effect variables, respectively. Then the three groups of variables (each stacked along channel) are separately fed into their individual encoders and mapped to hidden states (h^t_{cause}, h^t_{CHL}, h^t_{effect}). Finally, the hidden states perform energy transfers in EAM [as shown in blue boxes of Fig. 1(a)] and are decoded separately to (P^{t+1}_{cause}, P^{t+1}_{CHL}, P^{t+1}_{effect}), which is the prediction of the next time step.

For specific structures of PIRT, we use ConvLSTM [21] as the skeleton structure of encoder and decoder. At each time step t, ConvLSTM cells first compute the interactive feature among current input, historical memory (c^{t-1}_{cause}, c^{t-1}_{CHL}, c^{t-1}_{effect}) and hidden state (h^{t-1}_{cause}, h^{t-1}_{CHL}, h^{t-1}_{effect}) by convolution operation, and then use these feature information to conduct "forget" and "obtain" operation under gating mechanism. Subsequently, these cells output the updated (c^t_{cause}, c^t_{CHL}, c^t_{effect}) and (h^t_{cause}, h^t_{CHL}, h^t_{effect}) for the subsequent network layers. Between encoder and decoder, we propose EAM to simulate real-world multivariate energy transfer in red tide as introducing inductive biases in PINN [22]. In detail, EAM incorporates two vanilla attention modules, one is $E^t_{\text{cause}\to\text{CHL}}$ which represents the energy transferred by process cause → CHL and the other is $E^t_{\text{CHL}\to\text{effect}}$ which represents the energy

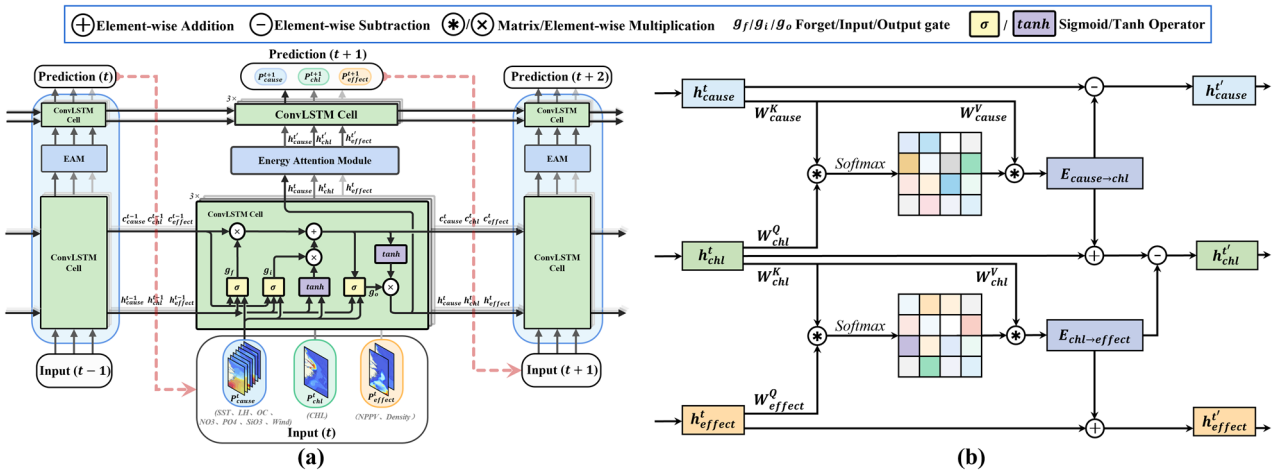

Fig. 1. (a) Model architecture of our proposed PIRT, which contains the individual encoders and decoders (green boxes) for the cause variables, CHL, and effect variables. The designed EAM, (blue boxes) between encoder and decoder is used to simulate the flow of energy interaction by cause → CHL → effect. The encoder and decoder in each time step adopt the ConvLSTM structure as shown in the time step t. (b) Detailed structure of our proposed EAM, which is composed of two attention modules. (Best viewed when zoomed in-view.)

in CHL → effect part. Quantitatively, we use these three parallel hidden states to calculate the energy flux of the above two processes respectively as shown as follows:

$$\begin{cases} E^t_{\text{cause}\to\text{CHL}} \\ = \text{softmax}\left(W^Q_{\text{CHL}} h^t_{\text{CHL}} \cdot W^K_{\text{cause}} h^t_{\text{cause}}\right) \cdot W^V_{\text{cause}} h^t_{\text{cause}} \end{cases} \tag{4}$$

$$\begin{cases} E^t_{\text{CHL}\to\text{effect}} \\ = \text{softmax}\left(W^Q_{\text{effect}} h^t_{\text{effect}} \cdot W^K_{\text{CHL}} h^t_{\text{CHL}}\right) \cdot W^V_{\text{CHL}} h^t_{\text{CHL}} \end{cases} \tag{5}$$

where W^Q_{CHL}, W^K_{cause}, W^V_{cause}, W^Q_{effect}, W^K_{CHL}, and W^V_{CHL} are all the trainable transformation matrixes (Q, K, V are the implications of query, key, and value similar in the original attention mechanism, respectively) used to unify the dimensions of the three hidden states. The final hidden states after energy transfer are calculated by (6) using the extracted energy above and then passed to the decoder. This process is detailed in Fig. 1(b)

$$\begin{cases} h^{t'}_{\text{cause}} = h^t_{\text{cause}} - E^t_{\text{cause}\to\text{CHL}} \\ h^{t'}_{\text{CHL}} = h^t_{\text{CHL}} + E^t_{\text{cause}\to\text{CHL}} - E^t_{\text{CHL}\to\text{effect}} \\ h^{t'}_{\text{effect}} = h^t_{\text{effect}} + E^t_{\text{CHL}\to\text{effect}}. \end{cases} \tag{6}$$

The detailed hyperparameter setting including training strategy of PIRT is shown in Table II.

D. Dataset and Loss Function

In this study, we use the ocean-atmosphere-biology daily dataset from 2010 to 2020 to train and evaluate PIRT, provided by CMEMS (https://www.copernicus.eu), which contains remote sensing data (atmospheric variables), reanalysis data (oceanic variables), and hindcast data (biological variables). We select 15 variables (see Table I), and determine the training set (2010–2016), validation set (2017), and testing set (2018–2020) by 7 : 1 : 3. The spatial resolution of all variables is 0.25° × 0.25° in (120° − 125°E, 25° − 35°N), completely covering the high incidence area of red tide in the East China Sea. We use mean square errors (MSE) as the loss function of PIRT.

TABLE II
DETAILED HYPERPARAMETER SETTING OF PIRT

Module	Component	Description
Encoder	*cause*	3-layer ConvLSTM@{16, 64, 256}@3×3
	chl	3-layer ConvLSTM@{16, 64, 256}@3×3
	effect	3-layer ConvLSTM@{16, 64, 256}@3×3
EAM	$cause \to chl$	Equation (4) with transformation dimension 64 for W^Q_{chl}, W^K_{cause}, and W^V_{cause}
	$chl \to effect$	Equation (5) with transformation dimension 64 for W^Q_{effect}, W^K_{chl}, and W^V_{chl}
Decoder	*cause*	3-layer ConvLSTM@{256, 64, 9}@3×3
	chl	3-layer ConvLSTM@{256, 64, 1}@3×3
	effect	3-layer ConvLSTM@{256, 64, 2}@3×3

Note: *The initial learning rate is 0.001, and is reduced with a scale factor 0.1 after every 50 epochs in which the evaluation metrics do not get better. 3-layer ConvLSTM@{16, 64, 256}@3×3 means the layers, filter number, and kernel size of the convolution, and so are others.*

III. EXPERIMENT RESULTS AND ANALYSES

A. Determinations of Predictors and Optimal Warm-Up Length

We infer the causal graph based on DAG-GNN to select the predictors (as shown in Fig. 2(a), where we filter out Height, Salinity, and O_2 with weak causalities. Among the variables pointing to CHL, NO_3, PO_4, and SiO_3 are essential nutrients for phytoplankton growth during red tide occurrences and maintain positive correlations. The increased nutrients in the sea will inevitably reduce phytoplankton foraging competition pressure and cause increased CHL [15]. Wind affects upwelling, which brings nutrients from the seafloor to the surface and causes an increase in phytoplankton growth, thereby increasing CHL [16]. LH is negatively correlated with CHL. When its value is low, less heat is supplied to the atmosphere through evaporation over the sea surface and more heat is retained by seawater, which is beneficial to the growth of phytoplankton. The SST and CHL are negatively correlated in the East China Sea. The higher the SST, the higher the stratification of seawater, the weaker the mixing of the upper and lower layers of seawater, leading to the lack of

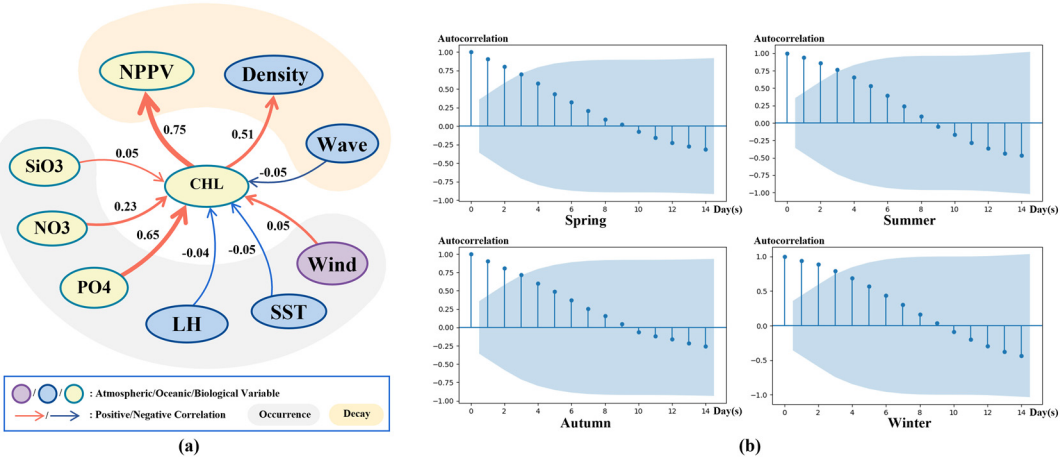

Fig. 2. (a) Learned causal graph, where the arrows indicate the direction of causality, red/blue arrows indicate positive/negative feedback relationships between connected variables, and the thickness of the arrows represents the strength of quantified causality. Gray/Orange shades mean the causal relationships are majorly modulated during the occurrence/decay phase of red tides. (b) Contains the seasonal-average autocorrelation tests for daily CHL observations, where the horizontal axis represents the lagged days and the vertical axis represents the autocorrelation coefficient. The blue shade is the 95% confidence interval. When the autocorrelation coefficient exceeds/does not exceed the confidence interval, it indicates that there is a strong/weak relationship.

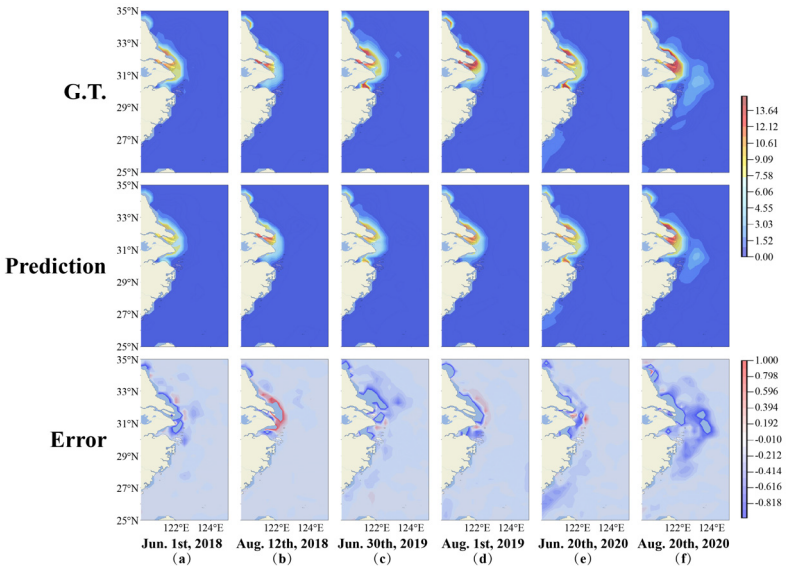

Fig. 3. Prediction results of CHL (unit: mg·m^{-3}) with seven forecast lead days for (a)–(f) six typical red tide events in the East China Sea during the testing period by PIRT. The first row is the ground truths (G.T.), the second row is predictions, and the third row represents the prediction errors. (Best viewed when zoomed in-view.)

nutrients and thus lower chlorophyll values [23]. During the decay of red tides, OC causes the migration and dispersal of phytoplankton to new seawater environments where nutrients are reduced, making CHL decrease. Moreover, NPPV and Density are significantly correlated with CHL. Chlorophyll, the main pigment in phytoplankton cells, converts inorganic matter into organic matter using solar energy. When CHL increases, phytoplankton aggregation and photosynthetic rate both increase, so NPPV and density subsequently increase. According to the above analyses, the variables in Fig. 2(a) are chosen for red tide forecasts, the physical significance of which can be comprehensively tracked.

Furthermore, we use a seasonal-average autocorrelation test [24] for all chosen variables to determine the optimal warm-up length of the model. Fig. 2(b) shows that the autocorrelation coefficients of CHL for 1–3 lagged days are all outside or at the edge of the blue shade for all seasons, which implies the strong relationships within historical three days. Other predictors have the same characteristics. Therefore, we choose three as the optimal warm-up length for predicting the next seven days, which covers the whole process of red tide from occurrence to decay.

B. Evaluations of Red Tide Forecast Skill

We evaluate PIRT during the testing period. Fig. 3 exhibits the forecast results (and errors) of six extreme red tides (usually from June to August each year) in the East China Sea. The prediction of PIRT has quite consistent numerical values with highly similar spatial patterns, and the systematic forecast errors are stably controlled within 0.9 mg·m^{-3}. Numerically, these results fully reflect the outstanding performance of PIRT in red tide forecasting. It is worth noting that PIRT tends to slightly underestimate the intensity of red tides. We think this is due to our data processing (filling 0 on land grids) and

TABLE III
FORECAST SKILL COMPARISONS WITH OTHER METHODS (VALUES ARE ALL MSE)

Model	Forecast-1d	Forecast-3d	Forecast-7d
(1) CNN	0.22	0.60	1.13
(2) FC-LSTM	0.22	0.68	1.23
(3) ConvLSTM	0.16	0.44	0.93
(4) PredRNN	0.15	0.36	0.90
(5) PhyDNet	0.11	0.36	0.89
(6) PIRT (ours)	**0.09**	**0.35**	**0.87**

Note: *Forecast-1d represents making forecasting in 1 lead day, and so are others. Bold numbers indicate the best performance.*

imbalance of training set (fewer red tide samples). Meanwhile, this is also because the vertical distributions of oceanic and biological variables are non-homogeneous. PIRT covers limited ocean depth, which ignores the rise of surface nutrient concentrations caused by upwelling ocean waves, resulting in an underestimation.

Additionally, to further demonstrate the superiority of PIRT, we compare the performance of PIRT with universal advanced spatial-temporal sequence prediction models in Table III, including 1) purely convolutional model (CNN); 2) FC-LSTM model [25]; 3) ConvLSTM [21]; 4) PredRNN [26]; and 5) PhyDNet [27]. All these models use the same training set and loss function, and we tune them to their optimal respectively. PIRT is obviously superior to other models, especially in long-term (seven days) predictions. Besides, the prediction errors of all other variables of PIRT are also less than 10% of the observation, indicating that PIRT can be used as a reliable tool for ecological prediction in the East China Sea.

IV. CONCLUSION

Red tide is of significant influence on coastal economic development and human sustainability, indicating that more accurate red tide forecasting is always the pursuit for the sake of formulating ecological protection measures and disaster prevention policies in advance. In this letter, we construct a PIRT multivariate deep learning forecast model for the East China Sea by thoroughly considering its dynamic mechanisms. The experiments show that PIRT can accurately forecast red tide in seven lead days, which exhibits low prediction errors and highly similar spatial patterns. The quantitative comparison in long-term forecasts also shows the superiority of PIRT to other advanced methods. In the future, we will use remotely sensed reflectance data to distinguish different red tide species and thus achieve finer prediction results.

REFERENCES

[1] S. Park and S. R. Lee, "Red tides prediction system using fuzzy reasoning and the ensemble method," *Appl. Intell.*, vol. 40, no. 2, pp. 244–255, Mar. 2014, doi: 10.1007/s10489-013-0457-1.

[2] A. P. Yunus, J. Dou, and N. Sravanthi, "Remote sensing of chlorophyll-a as a measure of red tide in Tokyo Bay using hotspot analysis," *Remote Sens. Appl., Soc. Environ.*, vol. 2, pp. 11–25, Dec. 2015.

[3] S. A. Henson et al., "Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity," *Biogeosciences*, vol. 7, no. 2, pp. 621–640, Feb. 2010.

[4] W. Zou et al., "Relationships between nutrient, chlorophyll *a* and Secchi depth in lakes of the Chinese eastern plains ecoregion: Implications for eutrophication management," *J. Environ. Manage.*, vol. 260, Apr. 2020, Art. no. 109923.

[5] X. He, S. Shi, X. Geng, and L. Xu, "Hierarchical attention-based context-aware network for long-term forecasting of chlorophyll," *Appl. Intell.*, pp. 1–16, Aug. 2022, doi: 10.1007/s10489-022-03242-8.

[6] Z. Yu, K. Yang, Y. Luo, and C. Shang, "Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi lake based on wavelet analysis and long-short term memory network," *J. Hydrol.*, vol. 582, Mar. 2020, Art. no. 124488.

[7] H. Cho, U. J. Choi, and H. Park, "Deep learning application to time series prediction of daily chlorophyll-a concentration," *WIT Trans. Ecol. Environ.*, vol. 215, pp. 157–163, Oct. 2018.

[8] D. Wenxiang, Z. Caiyun, S. Shaoping, and L. Xueding, "Optimization of deep learning model for coastal chlorophyll a dynamic forecast," *Ecol. Model.*, vol. 467, May 2022, Art. no. 109913.

[9] M. Liu, J. He, Y. Huang, T. Tang, J. Hu, and X. Xiao, "Algal Bloom forecasting with time-frequency analysis: A hybrid deep learning approach," *Water Res.*, vol. 219, Jul. 2022, Art. no. 118591.

[10] A. Marndi and G. K. Patra, "Chlorophyll prediction using ensemble deep learning technique," in *Progress in Computing, Analytics and Networking*. Singapore: Springer, 2020, doi: 10.1007/978-981-15-2414-1_34.

[11] S. Shamshirband, E. J. Nodoushan, J. E. Adolf, A. A. Manaf, A. Mosavi, and K.-W. Chau, "Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters," *Eng. Appl. Comput. Fluid Mech.*, vol. 13, no. 1, pp. 91–101, Jan. 2019.

[12] J.-H. Choi, J. Kim, J. Won, and O. Min, "Modelling chlorophyll-a concentration using deep neural networks considering extreme data imbalance and skewness," in *Proc. 21st Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2019, pp. 631–634.

[13] Z. Du, M. Qin, F. Zhang, and R. Liu, "Multistep-ahead forecasting of chlorophyll a using a wavelet nonlinear autoregressive network," *Knowl.-Based Syst.*, vol. 160, pp. 61–70, Nov. 2018.

[14] Y. Yu, J. Chen, T. Gao, and M. Yu, "DAG-GNN: DAG structure learning with graph neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7154–7163.

[15] D. J. Jiang and H. Zhang, "Analysis of spatial and temporal characteristics of chlorophyll-a concentration and red tide monitoring in Bohai Sea," *Mar. Sci.*, vol. 42, no. 5, pp. 23–31, 2018.

[16] Y. Fu, S. Xu, and J. Liu, "Temporal-spatial variations and developing trends of chlorophyll-a in the Bohai Sea, China," *Estuarine, Coastal Shelf Sci.*, vol. 173, pp. 49–56, May 2016.

[17] G. Kim, Y.-W. Lee, D.-J. Joung, K.-R. Kim, and K. Kim, "Real-time monitoring of nutrient concentrations and red-tide outbreaks in the southern sea of Korea," *Geophys. Res. Lett.*, vol. 33, no. 13, 2006, Art. no. L13607.

[18] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.

[19] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA, USA: MIT Press, 2017.

[20] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2014, *arXiv:1312.6114*.

[21] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[22] G. E. Karniadakis et al., "Physics-informed machine learning," *Nature Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021.

[23] J. Shan, W. Yicheng, and M. Renfeng, "Change conalysis of chlorophyll concentration in the East China Sea and its response to seawater temperature," *Bull. Surv. Mapping*, no. 6, pp. 39–44, 2020, doi: 10.13474/j.cnki.11-2246.2020.0177.

[24] J. H. F. Flores, P. M. Engel, and R. C. Pinto, "Autocorrelation and partial autocorrelation functions to improve neural networks models on univariate time series forecasting," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–8.

[25] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1745–1749, Oct. 2017, doi: 10.1109/LGRS.2017.2733548.

[26] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 879–888.

[27] V. Le Guen and N. Thome, "Disentangling physical dynamics from unknown factors for unsupervised video prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11474–11484.