CARBON-ECS: A BENCHMARK AND A PHYSICS-DECOUPLED MODEL FOR THE EAST CHINA SEA CARBON FLUX

Chengjie Fang[1], *Enyuan Zhao*[1*], *Hongming Yao*[3], *Ziyu Sun*[1], *Di Gao*[1], *Yanbiao Li*[1,2]

[1] Hangzhou Institute for Advanced Study,
University of Chinese Academy of Sciences, Hangzhou, China
[2] Computer Network Information Center, Chinese Academy of Sciences, Beijing, China
[3] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

ABSTRACT

Coastal ecosystems are critical zones in the global carbon cycle. In the East China Sea (ECS), the sea-air CO_2 flux (F_{CO_2}) is driven by complex biogeochemical and oceanic processes. Its forecasting faces two challenges: the **spatio-temporal discontinuity** of remote sensing and reanalysis data, and the difficulty for models to capture both **macro-scale transport** and **micro-scale structural evolution**. To address this, we construct **Carbon-ECS**, a 17-year, high-resolution, spatio-temporally complete benchmark for F_{CO_2} forecasting. We further propose a novel forecasting model, the **T**ransport and **I**terative **D**etail **E**nhancement **Net**work (**TIDE-Net**). The model features a physics-inspired decoupled architecture. It captures large-scale, ocean-current-driven carbon transport processes while simultaneously modeling the evolution and diffusion of fine-grained carbon sink structures. Our contributions enable a more physically consistent forecast of the spatio-temporal dynamics of F_{CO_2}, significantly enhancing the simulation capability for the complex ocean carbon cycle. The code and dataset are publicly available.

***Index Terms*—** Marine Carbon Sink, Flux Forecasting, Fourier Transform, Diffusion Model, Benchmark Dataset.

1. INTRODUCTION

Coastal ecosystems are critical global carbon sinks, yet accurately forecasting their sea-air CO_2 flux (F_{CO_2}) remains a significant challenge [1, 2, 3]. In the ECS, a typical marginal sea system shown in Fig. 1a, F_{CO_2} dynamics are driven by complex land-ocean interactions and ocean currents, exhibiting highly non-linear and multi-scale spatio-temporal characteristics [4, 5].

Accurately forecasting such complex dynamics essentially requires a model to decouple and capture two physical processes with significant scale differences: first, macro-scale transport dominated by large-scale advection, which determines the overall distribution and migration paths of carbon sink; and second, micro-scale structural evolution driven by local processes, which is responsible for the generation and dissipation of fine-grained structures.

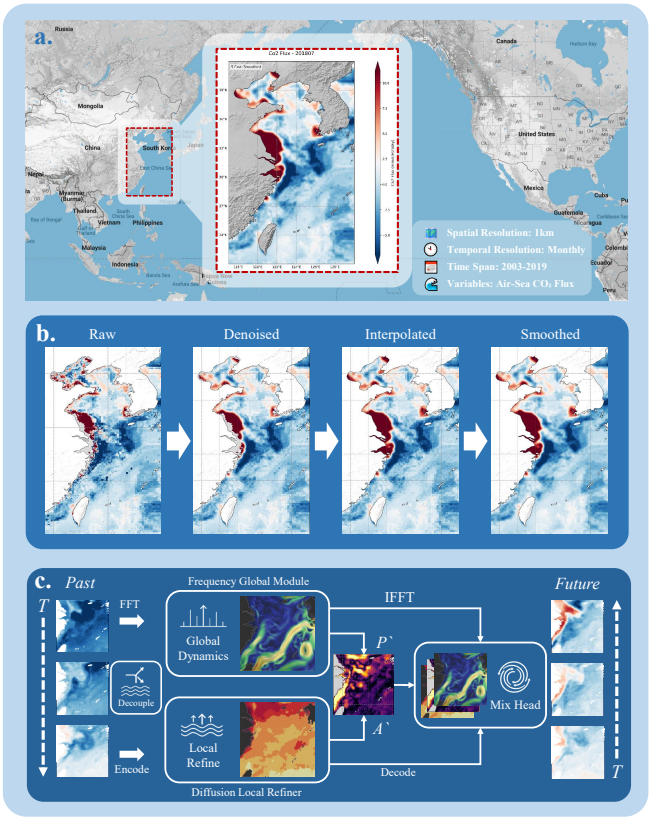


Fig. 1. Overview. (a) Study area & benchmark specifications. (b) Data completion pipeline. (c) TIDE-Net architecture.

Against this backdrop, existing forecasting paradigms suffer from inherent limitations. Traditional statistical and semi-mechanistic models struggle to capture non-linear dynamics [6, 7].While general-purpose machine learning methods improve fitting accuracy, their ”black-box” nature and static regression formulation result in poor physical inter-

*Corresponding author: zhaoenyuan@ucas.ac.cn

pretability and an inability to explicitly model temporal evolution [8, 9]. Emerging spatio-temporal dynamic models face a dilemma: deterministic models (e.g., ConvLSTM [10]) often yield blurry forecasts with intensity decay, whereas probabilistic models (e.g., GANs [11]) suffer from structural drift despite generating high-fidelity details.

Furthermore, a more fundamental bottleneck stems from the data itself: due to cloud cover in satellite remote sensing, existing datasets widely suffer from spatio-temporal discontinuity [12, 13]. This fundamentally constrains the upper limit of forecasting accuracy, hinders the development and training of advanced models, and makes fair comparisons between different methods exceptionally difficult.

To address these challenges, we propose a novel solution from both data and model perspectives. First, we construct **Carbon-ECS**, the first long-term (17-year), high-resolution, and spatio-temporally complete benchmark dataset (Fig. 1b), to resolve the data bottleneck. Second, building on this dataset, we propose a physics-decoupled predictive model, **TIDE-Net** (Fig. 1c). The model's core innovation lies in its decoupled architecture: one branch models the large-scale, ocean-current-driven advection-transport in the frequency domain to accurately predict the macro-scale position and intensity of carbon sink, while the other branch iteratively refines local detail structures. Finally, a fusion module combines their strengths to generate more physically consistent predictions. Our contributions are as follows:

- **Dataset:** We construct and release **Carbon-ECS**, the first spatio-temporally complete, high-resolution benchmark for F_{CO_2} forecasting in the ECS region.
- **Model:** The proposal of **TIDE-Net**, a novel physics-inspired decoupled framework that explicitly models advection-transport and local evolution.
- **Performance:** Experiments demonstrate that TIDE-Net outperforms state-of-the-art models in the complex task of ocean carbon flux forecasting.

2. METHODOLOGY

2.1. Benchmark Dataset: Carbon-ECS

A primary bottleneck for training deep learning models in oceanography is the significant data loss in satellite remote sensing products due to cloud occlusion [12]. To overcome this, we constructed Carbon-ECS (Table 1), the first long-term, high-resolution, and spatio-temporally complete benchmark dataset for F_{CO_2} forecasting in the East China Sea. Based on the work of MeSAA-ML[1], our dataset was systematically repaired through a rigorous multi-stage process—integrating regional anomaly removal, wavelet denoising, and K-nearest neighbors interpolation—to fix noise, biases, and missing values. Carbon-ECS establishes a solid foundation and fair benchmark for future research.

[1]https://zenodo.org/records/8042265

Table 1. Core specifications of Carbon-ECS.

SPECIFICATION	VALUE
TEMPORAL COVERAGE	2003–2019 (204 MONTHS)
TEMPORAL RESOLUTION	MONTHLY
SPATIAL COVERAGE	YANGTZE RIVER ESTUARY & ECS SHELF
SPATIAL RESOLUTION	$\sim$1 KM (0.01°)
AREA	$\sim 8.5 \times 10^5$ KM^2
DATA QUALITY	COMPLETE, DENOISED, BIAS-CORRECTED

2.2. TIDE-Net Architecture

The spatio-temporal evolution of F_{CO_2} is driven by two key physical processes: large-scale **advection-transport** and local **evolution-diffusion**. Existing models that attempt to address these in a coupled manner often struggle to capture both effectively.

We propose **TIDE-Net** (Fig. 2), a dual-branch architecture that explicitly decouples these two processes. Given historical inputs $X \in \mathbb{R}^{T_{\text{in}} \times C \times H \times W}$, TIDE-Net forecasts future states $Y \in \mathbb{R}^{T_{\text{out}} \times C \times H \times W}$ through three core modules:

- **Frequency Global Module (FGM):** Predicts large-scale dynamics in the frequency domain by decoupling the signal's amplitude and phase.
- **Diffusion Local Refiner (DLR):** Generates high-fidelity local details through an iterative refinement process that enhances fine-grained spatial structures.
- **Synergistic Fusion Head (SFH):** Adaptively fuses the complementary global forecast from FGM and local details from DLR to generate the final prediction.

2.3. Frequency Global Module (FGM)

The Frequency Global Module (FGM) models large-scale structures by operating in the frequency domain. It transforms the input sequence X_k using a 2D Discrete Fourier Transform (DFT), decomposes the spectrum into amplitude and phase, extrapolates these components, reconstructs the future spectrum, and applies an inverse DFT to obtain the forecast.

The input map X_k is first transformed into its complex spectrum $\hat{X}_k(u, v)$, where (h, w) and (u, v) are spatial and frequency coordinates, respectively:

$$\begin{aligned}\hat{X}_k(u, v) &= \mathcal{F}(X_k)(u, v) \\ &= \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} X_k(h, w)\, e^{-i2\pi\left(\frac{uh}{H} + \frac{vw}{W}\right)}. \end{aligned} \tag{1}$$

The spectrum is then decomposed into amplitude $\mathcal{A}(X_k)$ and phase $\mathcal{P}(X_k)$:

$$\begin{aligned}\mathcal{A}(X_k)(u, v) &= \sqrt{R^2(\hat{X}_k(u, v)) + I^2(\hat{X}_k(u, v))}, \\ \mathcal{P}(X_k)(u, v) &= \arctan\left(\frac{I(\hat{X}_k(u, v))}{R(\hat{X}_k(u, v))}\right). \end{aligned} \tag{2}$$

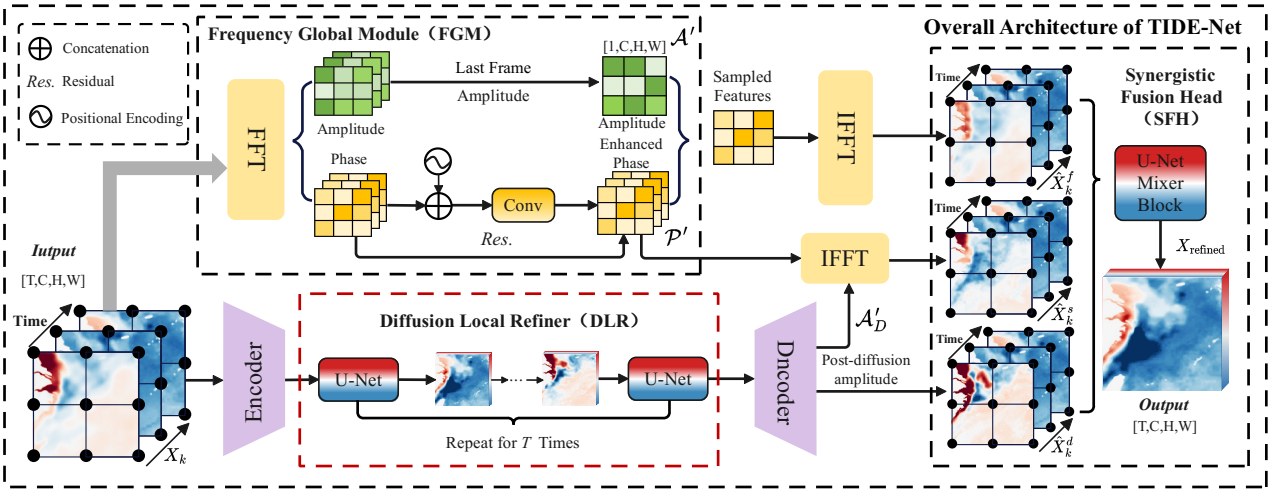


Fig. 2. The architecture of TIDE-Net. The model first decouples the input into two parallel streams: the **Frequency Global Module (FGM)** predicts large-scale dynamics via spectral analysis, while the **Diffusion Local Refiner (DLR)** iteratively enhances fine-grained spatial details. Finally, the **Synergistic Fusion Head (SFH)** intelligently fuses the outputs from both modules to generate the final refined prediction.

where $R(\cdot)$ and $I(\cdot)$ are the real and imaginary parts. We forecast the future amplitude $\mathcal{A}'$ and phase $\mathcal{P}'$ by modulating the amplitude with the last frame X_t and modeling phase evolution with a learnable kernel $Conv_{pm}$:

$$\begin{aligned} \mathcal{A}'(X_k)(u,v) &= \mathcal{A}(X_k)(u,v) \otimes X_t, \\ \mathcal{P}'(X_k)(u,v) &= \mathcal{P}(X_k)(u,v) \otimes Conv_{pm}, \end{aligned} \tag{3}$$

The forecast spectrum is reconstructed by combining the forecasted components:

$$\hat{X}_k(u,v) = \mathcal{A}'(X_k)(u,v) \cdot e^{i\mathcal{P}'(X_k)(u,v)}. \tag{4}$$

Finally, the inverse DFT ($\mathcal{F}^{-1}$) transforms the spectrum back to the spatial domain, yielding the forecast $\hat{X}_k^f$ for future frames $k = t+1, \ldots, 2t$:

$$\begin{aligned} \hat{X}_k^f(h,w) &= \mathcal{F}^{-1}(\hat{X}_k) \\ &= \frac{1}{HW} \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} \hat{X}_k(u,v)\, e^{i2\pi\left(\frac{uh}{H}+\frac{vw}{W}\right)}. \end{aligned} \tag{5}$$

2.4. Diffusion Local Refiner (DLR)

To restore fine details to the FGM predictions, the Diffusion Local Refiner (DLR) iteratively refines an initial forecast. First, an encoder f_θ generates an initial estimate $\hat{X}_k^d$ from the input sequence X_k:

$$\hat{X}_k^d = f_\theta(X_k). \tag{6}$$

Then, the prediction is updated over T refinement steps by adding residual corrections from a UNet, where α_t is a time-dependent scaling factor that stabilizes the refinement and balances the contribution of the residual:

$$X_{k+t} = \frac{1}{\sqrt{\alpha_t}} \left(X_k + UNet_\phi\left(X_k, t\right) \right). \tag{7}$$

The final output $\hat{X}_k^d$ provides the detailed component for the fusion stage.

2.5. Synergistic Fusion Head (SFH)

The Synergistic Fusion Head (SFH) combines the outputs from FGM and DLR. It first constructs a synergistic spectrum $X_k^s(u,v)$ using the phase from FGM's prediction ($\mathcal{P}'$) and the amplitude from DLR's output ($\mathcal{A}'_D$):

$$X_k^s(u,v) = \mathcal{A}'_D(X_k)(u,v) \cdot e^{i\mathcal{P}'(X_k)(u,v)}, \tag{8}$$

This spectrum is transformed back to the spatial domain through the same inverse DFT process as in Eq. (5), yielding a synergistic field $\hat{X}_k^s$. This field is then concatenated with the FGM output $\hat{X}_k^f$ and the DLR output $\hat{X}_k^d$, and processed by a U-Net mixer to learn multi-scale features and produce the final prediction X_{refined}:

$$X_{\text{refined}} = UNet_{mixer}(\hat{X}_k^f \oplus \hat{X}_k^s \oplus \hat{X}_k^d). \tag{9}$$

where $\oplus$ denotes channel-wise concatenation. The model is trained end-to-end with mean squared error (MSE) loss against ground truth $\hat{X}_k^{(gt)}$:

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left\| X_{\text{refined}} - \hat{X}_k^{(gt)} \right\|_2^2. \tag{10}$$

3. EXPERIMENTS

3.1. Experimental Setup

We evaluate on Carbon-ECS under two forecasting settings: **short-term** ($T_{in} = 3, T_{out} = 3$) and **medium-term** ($T_{in} = 6, T_{out} = 6$). Performance is assessed using MSE, MAE, RMSE, PSNR, SSIM, CRPS, and LPIPS. All models are implemented in PyTorch and trained with AdamW on two NVIDIA A100 GPUs.

Table 2. Comparison on two forecasting tasks. Metrics with arrows indicate improvement direction (↓: lower is better, ↑: higher is better). The best results are in **bold** and the second-best are underlined. Note that MSE, MAE, RMSE, and CRPS are scaled by 10^2 for readability.

Model	Medium-term Forecasting (Tin=6, Tout=6)							Short-term Forecasting (Tin=3, Tout=3)						
	MSE↓	MAE↓	RMSE↓	PSNR↑	SSIM↑	CRPS↓	LPIPS↓	MSE↓	MAE↓	RMSE↓	PSNR↑	SSIM↑	CRPS↓	LPIPS↓
SimVP (CVPR'22) [14]	0.2564	3.4999	4.8492	26.708	0.8182	3.4999	0.3044	0.2756	3.6412	4.9398	26.738	0.8161	3.6412	0.2700
TAU (CVPR'23) [15]	0.2489	3.4198	4.7729	26.824	0.8153	3.4198	0.2815	**0.2370**	3.4080	4.6608	27.071	0.8185	3.4080	0.2809
WaST (AAAI'24) [16]	0.2768	3.5524	5.0355	26.396	0.8202	3.5524	0.2158	0.2494	3.5117	4.7867	26.819	0.8223	3.5117	0.2219
SmaAtUnet (PRL'21) [17]	0.2815	3.6429	5.0679	26.362	0.8122	3.6429	0.2274	0.2911	3.6917	5.1210	26.406	0.8142	3.6917	0.2219
PhyDNet (CVPR'20) [18]	0.2550	3.3489	4.7565	27.006	0.8138	3.3489	0.3274	0.2387	3.2958	**4.5930**	27.380	0.8176	3.2958	0.3373
MotionRNN (CVPR'21) [19]	0.5459	5.4703	6.8396	24.145	0.7096	5.4703	0.4560	0.4449	4.7460	6.1847	24.978	0.7109	4.7460	0.4802
Earthformer (NIPS'22) [20]	0.3253	3.8520	5.3804	25.995	0.8205	3.8520	0.3728	0.4514	4.9365	6.3003	24.661	0.7998	4.9365	0.4170
AlphaPre (CVPR'25) [21]	0.3285	3.8863	5.4281	25.799	0.8336	3.8863	0.2505	0.3584	3.9900	5.6479	25.506	0.8388	3.9900	0.2414
+FGM	0.3263	3.9326	5.4447	25.716	0.8310	3.9326	0.2493	0.3610	4.0307	5.6882	25.403	0.8377	4.0307	0.2476
w/o SFH	0.3102	3.7568	5.2606	26.113	0.8377	3.7568	0.2714	0.3382	3.8345	5.4693	25.811	0.8395	3.8345	0.2364
w/o DLR	0.2637	3.3645	4.8602	26.756	0.8691	3.3645	**0.1728**	0.2673	3.3331	4.8387	26.935	0.8715	3.3331	0.2218
TIDE-Net (Full, Ours)	**0.2433**	**3.1800**	**4.6176**	**27.285**	**0.8796**	**3.1800**	0.1746	0.2419	**3.1051**	4.6587	**27.395**	**0.8723**	**3.1051**	**0.2075**

Table 3. Pearson Correlation Coefficient (PCC) for medium-term (6→6) and short-term (3→3) forecasting. All reported correlations are statistically significant ($p < 0.001$).

Model	6→6	3→3	Model	6→6	3→3
SimVP	0.8266	0.8479	TAU	0.8284	0.8685
SmaAtUnet	0.8102	0.8673	PhyDNet	0.8229	0.8744
MotionRNN	0.4629	0.7958	Earthformer	0.7839	0.6736
AlphaPre	0.5922	0.8418	**TIDE-Net**	**0.8330**	**0.8828**

3.2. Comparison & Ablation Experiment

The primary comparative and ablation results are presented in Table 2 and Fig 3. Our model outperforms eight mainstream baselines, achieving state-of-the-art results in both forecasting settings and generating predictions with clearer, more physically reasonable structures. The ablation study confirms the necessity of each component. As shown in Table 2, the full model outperforms its variants: FGM alone (+FGM), w/o DLR, and w/o SFH. Removing DLR or SFH, or relying solely on FGM, leads to notable performance drops, underscoring their synergistic roles in capturing global trends, refining local details, and integrating multi-scale information.

3.3. Quantitative Analysis

PCC Analysis. We use the Pearson Correlation Coefficient (PCC) to evaluate forecast accuracy. As detailed in Table 3, our model achieves the highest PCC among all baselines in both forecasting settings, indicating a stronger ability to capture spatio-temporal dynamics.

Convergence Analysis. The training curves in Fig 4 confirm the DLR module's role in enhancing training stability and accelerating convergence. The full model reaches a lower and more stable amplitude loss, whereas the variant without DLR exhibits significant oscillations, underscoring DLR's effectiveness in the training process.

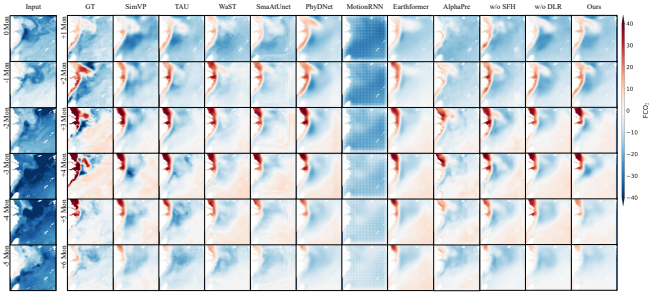


Fig. 3. Qualitative comparison of 6-month-ahead F_{CO_2} forecasts. Our model's predictions are visually most consistent with the Ground Truth (GT), exhibiting clearer physical structures and fewer artifacts.

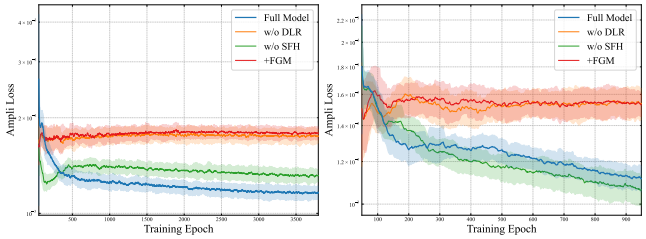


Fig. 4. Amplitude loss curves for the full model and its variants. Models equipped with the DLR module exhibit more stable convergence, highlighting its role in stabilizing amplitude learning.

4. CONCLUSION

Our work introduced Carbon-ECS, a spatio-temporally complete benchmark, and TIDE-Net, a novel physics-decoupled model, to advance the forecasting of F_{CO_2}.

Looking ahead, future efforts will focus on three key directions: first, incorporating multi-modal variables to strengthen the model's physical constraints; second, extending the TIDE-Net framework to other critical ocean regions; and third, integrating uncertainty quantification to enhance its reliability for scientific and policy-making applications.

5. ACKNOWLEDGMENTS

This research was supported in part by the Science and Technology Program of Zhejiang Province (Grant No. 2025C02038), the Natural Science Foundation of Zhejiang Province (Grant No. LTGG23F020004), the Hangzhou Qianjiang Distinguished Expert Program (Grant No. HZ20251217834), the National Natural Science Foundation of China (Grant No. 42206035), and the Shenzhen Science and Technology Program (Grant No. JCYJ20250604183032041).

6. REFERENCES

[1] Nadine Schubert, Fernando Tuya, Viviana Peña, Paulo A Horta, Vinícius W Salazar, Pedro Neves, Cláudia Ribeiro, Francisco Otero-Ferrer, Fernando Espino, Kathryn Schoenrock, et al., ""pink power"—the importance of coralline algal beds in the oceanic carbon cycle," *Nature Communications*, vol. 15, no. 1, pp. 8282, 2024.

[2] Wolfgang Koeve, Angela Landolfi, Andreas Oschlies, and Ivy Frenger, "Marine carbon sink dominated by biological pump after temperature overshoot," *Nature Geoscience*, vol. 17, no. 11, pp. 1093–1099, 2024.

[3] Moritz Mathis, Fabrice Lacroix, Stefan Hagemann, David Marcolino Nielsen, Tatiana Ilyina, and Corinna Schrum, "Enhanced co2 uptake of the coastal ocean is dominated by biological carbon fixation," *Nature Climate Change*, vol. 14, no. 4, pp. 373–379, 2024.

[4] Yuchen Zhang, Mingsheng Long, Kaiyuan Chen, Lanxiang Xing, Ronghua Jin, Michael I Jordan, and Jianmin Wang, "Skilful nowcasting of extreme precipitation with nowcastnet," *Nature*, vol. 619, no. 7970, pp. 526–532, 2023.

[5] Demin Yu, Xutao Li, Yunming Ye, Baoquan Zhang, Chuyao Luo, Kuai Dai, Rui Wang, and Xunlai Chen, "Diffcast: A unified framework via residual diffusion for precipitation nowcasting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 27758–27767.

[6] Christian Rödenbeck, Dorothee CE Bakker, Nicolas Gruber, Yosuke Iida, Andrew R Jacobson, S Jones, Peter Landschützer, Nicolas Metzl, Shinichiro Nakaoka, Are Olsen, et al., "Data-based estimates of the ocean carbon sink variability–first results of the surface ocean pco 2 mapping intercomparison (socom)," *Biogeosciences*, vol. 12, no. 23, pp. 7251–7278, 2015.

[7] Yan Bai, Wei-Jun Cai, Xianqiang He, Weidong Zhai, Delu Pan, Minhan Dai, and Peisong Yu, "A mechanistic semi-analytical method for remotely sensing sea surface p co 2 in river-dominated coastal oceans: A case study from the e ast c hina s ea," *Journal of Geophysical Research: Oceans*, vol. 120, no. 3, pp. 2331–2349, 2015.

[8] Jiye Zeng, Yosuke Iida, Tsuneo Matsunaga, and Tomoko Shirai, "Surface ocean co2 concentration and air-sea flux estimate by machine learning with modelled variable trends," *Frontiers in Marine Science*, vol. 9, pp. 989233, 2022.

[9] Varvara E Zemskova, Tai-Long He, Zirui Wan, and Nicolas Grisouard, "A deep-learning estimate of the decadal trends in the southern ocean carbon storage," *Nature communications*, vol. 13, no. 1, pp. 4056, 2022.

[10] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.

[11] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[12] Shujie Yu, Zigeng Song, Yan Bai, Xianghui Guo, Xianqiang He, Weidong Zhai, Huade Zhao, and Minhan Dai, "Satellite-estimated air-sea co2 fluxes in the bohai sea, yellow sea, and east china sea: Patterns and variations during 2003–2019," *Science of The Total Environment*, vol. 904, pp. 166804, 2023.

[13] Zigeng Song, Shujie Yu, Yan Bai, Xianghui Guo, Xianqiang He, Weidong Zhai, and Minhan Dai, "Construction of a high spatiotemporal resolution dataset of satellite-derived pco2 and air–sea co2 flux in the south china sea (2003–2019)," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[14] Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z Li, "Simvp: Simpler yet better video prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 3170–3180.

[15] Cheng Tan, Zhangyang Gao, Lirong Wu, Yongjie Xu, Jun Xia, Siyuan Li, and Stan Z Li, "Temporal attention unit: Towards efficient spatiotemporal predictive learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 18770–18782.

[16] Xuesong Nie, Yunfeng Yan, Siyuan Li, Cheng Tan, Xi Chen, Haoyuan Jin, Zhihang Zhu, Stan Z Li, and Donglian Qi, "Wavelet-driven spatiotemporal predictive learning: Bridging frequency and time variations," in *Proceedings of the AAAI conference on artificial intelligence*, 2024, vol. 38, pp. 4334–4342.

[17] Kevin Trebing, Tomasz Staczyk, and Siamak Mehrkanoon, "Smaat-unet: Precipitation nowcasting using a small attention-unet architecture," *Pattern Recognition Letters*, vol. 145, pp. 178–186, 2021.

[18] Vincent Le Guen and Nicolas Thome, "Disentangling physical dynamics from unknown factors for unsupervised video prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11474–11484.

[19] Haixu Wu, Zhiyu Yao, Jianmin Wang, and Mingsheng Long, "Motionrnn: A flexible model for video prediction with spacetime-varying motions," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15435–15444.

[20] Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Bernie Wang, Mu Li, and Dit-Yan Yeung, "Earthformer: Exploring space-time transformers for earth system forecasting," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25390–25403, 2022.

[21] Kenghong Lin, Baoquan Zhang, Demin Yu, Wenzhi Feng, Shidong Chen, Feifan Gao, Xutao Li, and Yunming Ye, "Alphapre: Amplitude-phase disentanglement model for precipitation nowcasting," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 17841–17850.