TC-Diffuser: Bi-Condition Multi-Modal Diffusion for Tropical Cyclone Forecasting

Shiqi Zhang[1], **Pan Mu**[1], **Cheng Huang**[1], **Jinglin Zhang**[2], **Cong Bai**[1*]

[1]College of Computer Science, Zhejiang University of Technology
[2] School of Control Science and Engineering, Shangdong University
{201906062727, panmu, chenghuang}@zjut.edu.cn, jinglin.zhang@sdu.edu.cn, congbai@zjut.edu.cn

Abstract

Tropical cyclones (TCs) are complex weather systems with strong winds and heavy rainfall, causing substantial loss of life and property. Therefore, accurate TC forecasting is crucial for the effective prevention of disasters caused by TCs. TC forecasting can be regarded as a spatio-temporal prediction problem. It has been proven that using multi-modal data can effectively introduce atmospheric information to achieve better prediction results and higher interpretability. But it also introduces inevitably introduces noise into the prediction process. The diffusion model's unique noise modeling capability can reduce prediction noise when using multi-modal datasets. However, adapting it to TC forecasting has two main challenges: how to extract valuable information from multi-modal data, and how to utilize them to guide the generation process. For the first challenge, while recent methods can predict multiple TC attributes using multi-modal data, they often overlook the interdependence of multiple attributes and the semantic gap between modalities. Considering the interdependence of attributes, we propose two condition generators that capture the commonalities and characteristics of TC attributes, extracting spatio-temporal and environmental features and incorporating expert knowledge. To reduce the semantic gap between multi-modal data, we introduce the PGSA-LSTM module to map primary and auxiliary modalities. For the second challenge, we propose a novel Bi-condition diffusion model that sequentially processes conditions from the characteristics to commonalities of attributes, thereby expanding the guidance information that the diffusion model can accept. Our results surpass state-of-the-art deep learning models and outperform the numerical weather prediction model used by the China Central Meteorological Observatory. TC-Diffuser shows high generalizability across global ocean areas, strong robustness in handling missing data, and higher computational efficiency.

1 Introduction

Tropical cyclones (TCs) are complex weather systems with significant impacts on daily life, often bringing strong winds and heavy rainfall. Therefore, accurate forecasting is crucial to prevent TC-related disasters. TC forecasting uses historical TC attributes to predict future TC attributes, which typically encompass trajectory (longitude and latitude), pressure, and wind speed. Trajectory corresponds to the longitude and latitude of the TC center, pressure to the lowest atmospheric pressure (hPa) at the TC center, and wind speed to the maximum sustained wind speed (m/s) over two minutes near the TC center. These attributes are interdependent in physics (Callaghan and Smith 1998; Knaff and Zehr 2007; Yan et al. 2024). However, predicting TC attributes is difficult due to various factors, such as inherent interdependencies among TC attributes and atmospheric modalities. Despite the difficulty, there are still many methods trying to solve these problems. These methods are categorized into two main groups: traditional meteorological methods and deep learning methods.

In traditional meteorological methods, statistical dynamic forecasting methods employ regression techniques (DeMaria and Kaplan 1999), such as CLIPER (Neumann and Lawrence 1975). Numerical weather prediction models (NWPs) (Bauer, Thorpe, and Brunet 2015; Coiffier 2011; Kimura 2002), based on physical dynamic equations (Sanders, Pike, and Gaertner 1975; ECMWF 2011), are extensively used by official meteorological forecasting agencies, including the China Central Meteorological Observatory (CMO (CMO 2019)). These methods utilize multi-modal data but require substantial computational resources, such as supercomputers, to handle the numerous physical constraints involved (Chen et al. 2023).

In the field of deep learning, TC forecasting can be viewed as a spatio-temporal prediction problem; thus, several spatio-temporal prediction models are commonly employed. Notable examples include RNN (Moradi Kordmahalleh, Gorji Sefidmazgi, and Homaifar 2016; Alemany et al. 2019), LSTM (Gao et al. 2018), ConvLSTM (Kim et al. 2019), BiGRU (Song et al. 2022), and GAN (Rüttgers et al. 2019). Besides, it has been proven that using multi-modal data (Geng, Liu, and Shi 2023; Wu et al. 2021; Ma, Pan, and Bai 2024) can effectively introduce atmospheric information to achieve better prediction results and higher interpretability. However, the introduction of multi-modal data inevitably introduces noise in prediction process. The diffusion model's unique noise modeling capability can effectively reduce prediction noise.

There are two main challenges in adapting diffusion model to spatio-temporal prediction based on multi-modal data. First, how to extract valuable information from multi-

*Corresponding Author

modal data. Second, how to effectively utilize this information to guide the generation process.

For the first challenge, multi-modal data encompass multiple attributes of TCs. There are interdependent relationships among them, which can be categorized into common and characteristic relationships. Unlike single-task (single-attribute) prediction methods (Gao et al. 2018; Pan, Xu, and Shi 2019), predicting multiple TC attributes simultaneously (Huang et al. 2022, 2023) is advantageous for learning the commonalities among them (Zhao et al. 2023). However, different TC attributes have distinct characteristics, and focusing only on commonalities will lead to results that contradict physical laws (Baltrušaitis, Ahuja, and Morency 2018). So attribute-specific information must be extracted from the observed data, which has often been overlooked in previous work. For example, TC age is one of the factors influencing the wind-pressure relationship (Song, Duan, and Klotzbach 2020), making it essential to incorporate TC age into wind and pressure predictions. Moreover, previous methods are difficult to provide expert knowledge to explain the generated results, thus often lacking interpretability.

Besides, multi-modal data typically include historical TC attributes, geopotential height maps, and environmental data. Historical TC attributes are 1D data, while geopotential height maps are 2D data representing the atmospheric pressure structure. This creates a semantic gap between the 1D and 2D data. Previous methods have overlooked this gap, making accurate cross-modal mapping difficult to achieve.

For the second challenge, a common idea is to input the feature of the multi-modal data as a condition into the diffusion model to guide the generation process. However, a single condition is not enough to fully represent the rich information of multiple attributes, that is, there are commonalities and characteristics among attributes. Thus, we need to design two conditions for commonalities and characteristics, respectively. Another question is how to effectively utilize these two conditions to guide our generation process. An unreasonable guidance order will lead to the missing of the information carried by a certain condition.

To address the above two challenges, we propose a novel framework, **TC-Diffuser**[1], for tropical cyclone forecasting. The main contributions are summarized as follows:

- We propose a novel **Bi-condition multi-modal diffusion model** for TC forecasting. This model sequentially processes conditions from the characteristics to commonalities of attributes, thereby extending the number of conditions and improving the range of guidance information that the diffusion model can accept.
- Considering the interdependence of multiple attributes, we propose two **condition generators** that consider the commonalities and characteristics of TC attributes, which extract the spatio-temporal and environmental features of TCs and incorporate expert knowledge.
- To reduce the semantic gap between multi-modal TC data, we introduce **PGSA-LSTM** module to discover mappings between primary and auxiliary modalities. This establishes the association between modalities.
- Comprehensive experiments were conducted using the China Meteorological Administration Tropical Cyclone Best Track Dataset (CMA-BST). Our method outperforms both the state-of-the-art deep learning model and the NWPs method used by the China Central Meteorological Observatory (CMO) across all metrics.

[1]Code: https://github.com/Zjut-MultimediaPlus/TC-Diffuser

2 Method

The objective of our model is to predict typical attributes of TCs, including trajectory, pressure, and wind speed. The trajectory includes longitude and latitude. The model input four parts of data: **1D** historical TC attribute $H = \{H_i^t\}$; **2D** geopotential height map $G = \{G_t\}$, $i \in \{longitude, latitude, pressure, wind\ speed\}$, $t \in \{T_{init}+1, T_{init}+2, ..., T_{init}+n\}$, where $T_{init}+1$ represents the first moment of historical data, n denotes the length of the historical data; **Env**ironmental data $E = \{E_{traj}, E_{pres}, E_{wind}\}$; and TC **Age** A. Consequently, the inputs are $X = \{H; G; E; A\}$.The outputs are predicted **1D** TC attributes $z = \{z_i^t\}$, where $t \in \{T_{init}+n+1, ..., T_{init}+n+m\}$, and m represents the length of the predicted data. Our overview is illustrated in Figure 1.

2.1 Bi-Condition Multi-Modal Diffusion Model

It has been proven that multi-modal data (Geng, Liu, and Shi 2023) can enhance prediction accuracy and interpretability by incorporating atmospheric information. However, it also introduces noise into the prediction process. The diffusion model's unique noise modeling capability can effectively reduce prediction noise. Thus, we propose Bi-condition multi-modal diffusion model tailored for TC forecasting.

We first define a diffusion sequence as $(z_0, z_1, ..., z_K)$, where K is the maximum diffusion step. The diffusion process adds noise into the ground truth region z_0. Correspondingly, the reverse diffusion sequence is $(z_K, z_{K-1}, ..., z_0)$. This process utilizes historical data as conditions, and gradually denoises the standard Gaussian z_K to obtain the future TC attributes z_0. This process reduces prediction noise.

The diffusion process is defined as $q(z_k|z_0) := \mathcal{N}(z_k; \sqrt{\bar{\alpha_k}}z_0, (1-\bar{\alpha_k})\mathrm{I})$, thus

$$z_k = \sqrt{\bar{\alpha_k}}z_0 + \sqrt{(1-\bar{\alpha_k})}\varepsilon \quad (1)$$

$\bar{\alpha_k} = \prod_{s=1}^{k} \alpha_s$, $\alpha_1, \alpha_2, ...\alpha_k$ are fixed variance schedulers, the noise variable $\varepsilon \sim \mathcal{N}(0, \mathrm{I})$. As the diffusion step k increases, $z_k \sim \mathcal{N}(0, \mathrm{I})$. This means z_0 is gradually destroyed into a Gaussian noise distribution. The training loss is:

$$L(\theta, \psi_{sh}, \psi_{sp}) = \mathbb{E}_{\varepsilon, z_0, k}||\varepsilon - \varepsilon_{(\theta, \psi_{sh}, \psi_{sp})}(z_k, k, X)|| \quad (2)$$

θ are parameters of Bi-condition multi-modal diffusion model, ψ_{sh} and ψ_{sp} are parameters of Shared and Specific Condition Generator, X are the input of the model.

We regard the process of gradually generating more reliable TC prediction results as the **reverse diffusion process**. We design the historical data as Bi-condition (in Section 2.2) to guide our prediction results from high uncertainty to low uncertainty. The Bi-condition (shared condition C^{share}, specific condition C^{spec}) are calculated as follows (see details in Section 2.2):

$$C^{share} = F_{\psi_{sh}}(H, G), C^{spec} = F_{\psi_{sp}}(E, A) \quad (3)$$

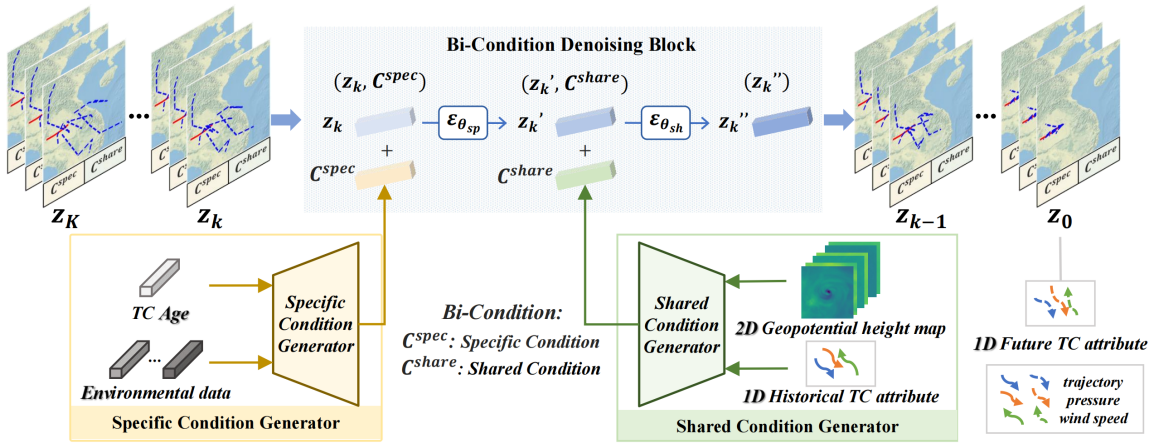

Figure 1: **Overview of TC-Diffuser inference.** The $z_K, z_{K-1}, ...z_0$ represents the TC prediction results under different diffusion steps k, the blue dashed lines inside are the prediction trajectory, and the red solid line is the ground truth. Only the future trajectory is visualized, but the wind speed and pressure prediction results are also included. From z_K to z_0, the blue dashed line gradually converges from noise to near ground truth. The single-step denoising process from z_k to z_{k-1} is implemented by **Bi-Condition Denoising block (in Section 2.1)**, which sequentially processes conditions from specific to shared. Specific and shared conditions are generated by the **two condition generators (in Section 2.2)**: (1) Specific Condition Generator encodes TC Age and Environmental data to incorporate attribute-specific information and expert knowledge. (2) Shared Condition Generator encodes 2D geopotential height map and 1D historical TC attribute data to incorporate spatio-temporal information. The **PGSA-LSTM module (in Section 2.3)** is part of Shared Condition Generator.

The reverse diffusion process with Bi-condition is defined as $p_\theta(z_{k-1}|z_k, C^{spec}, C^{share}) := \mathcal{N}(z_{k-1}; \mu_\theta(z_k, k, C^{spec}, C^{share}); \Sigma_\theta(z_k, k))$, where $\mu_\theta(\cdot) = \frac{1}{\sqrt{\alpha_k}}(z_k - \frac{\beta_k}{\sqrt{1-\bar{\alpha}_k}}\varepsilon_\theta(z_k, k, C^{spec}, C^{share}))$, $\Sigma_\theta(\cdot) = \sigma_k^2 \mathrm{I} = \beta_k \mathrm{I}$, thus:

$$z_{k-1} = \frac{1}{\sqrt{\alpha_k}}\left(z_k - \frac{\beta_k}{\sqrt{1-\bar{\alpha}_k}}\varepsilon_\theta(z_k, k, C^{spec}, C^{share})\right) + \sqrt{\beta_k}e \quad (4)$$

where e is a random variable in standard Gaussian Distribution, $\beta_k = 1 - \alpha_k$. ε_θ is implemented by the Bi-condition denoising block in Figure 1. The output of the block is:

$$z_k'' = \varepsilon_\theta(z_k, k, C^{spec}, C^{share}) \quad (5)$$

The Bi-condition denoising block is first guided by the specific condition C^{spec} to learn the attribute-specific information, and then by the shared condition C^{share} to learn the attribute-shared information. We discuss the guidance orders in Section 3.2. The parameters θ are consists of θ_{sp} and θ_{sh}:

$$z_k' = \varepsilon_{\theta_{sp}}(z_k, C^{spec}), z_k'' = \varepsilon_{\theta_{sh}}(z_k', k, C^{share}) \quad (6)$$

$\varepsilon_{\theta_{sp}}$ is implemented by three parallel gating units for 3 attributes respectively, which use condition to generate Gate and bias, and perform a gating transform for latent noise (z_k). $\varepsilon_{\theta_{sh}}$ is implemented by four serial gating units (without distinguishing 3 attributes) and a Transformer encoder.

2.2 Two Condition Generators

Considering the interdependence of TC attributes, we propose shared and specific condition generators for their commonalities and characteristics to extract spatio-temporal and environmental features and incorporate expert knowledge.

To explore characteristics of attributes, **specific condition generator** selects attribute-specific factors to generate a set of specific conditions for each attribute. **Env**ironmental data, such as the *future move direction* of TCs, which indicates the trajectory change direction, is attribute-specific and has minimal correlation with other attributes. Thus, we extract three sets of environmental factors from Env data to guide the prediction of three attributes: trajectory (E_{traj}: *future move direction*), pressure (E_{pres}: *intensity class*, *future intensity change direction*), and wind speed (E_{wind}: *wind*). *future move direction* and *future intensity change direction* are expert knowledge derived from 1D historical data. This means our model integrates expert knowledge. Obtaining this information may be challenging in practice, so we input wrong expert knowledge during testing, detailed further in Section 3.5. Besides, different TC **Age** A display unique patterns only in wind-pressure relationship (Song, Duan, and Klotzbach 2020), so the Age data is beneficial to produce predictions consistent with the wind-pressure relationship, a consideration often overlooked in prior work. Hence, TC age is incorporated into pressure and wind speed conditions, not in the trajectory condition. C^{spec} is calculated as follows:

$$C^{spec} = (F_{\psi_{sp}^{tr}}(E_{traj}), F_{\psi_{sp}^{pr}}(E_{pres}, A), F_{\psi_{sp}^{wi}}(E_{wind}, A)) \quad (7)$$

$F_{\psi_{sp}^{tr}}$ is a linear module, both $F_{\psi_{sp}^{pr}}$ and $F_{\psi_{sp}^{wi}}$ are two parallel linear modules, and then these two results are concatenated. In summary, these factors containing expert knowledge and environmental features guide the model to generate predictions in line with expert knowledge and atmospheric laws.

To explore commonalities of attributes, **shared condition generator** selects attribute-shared modalities to generate a shared condition. What modality is attribute-shared? The ridges of 2D geopotential height maps may guide TC trajectory changes (Wu and Wang 2004). Lower geopotential

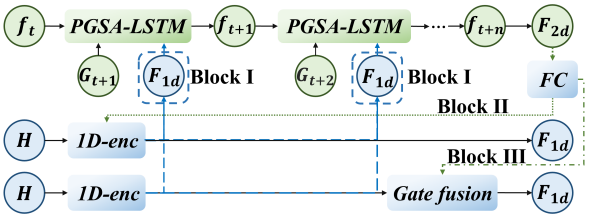

Figure 2: **Four guidance strategies between 1D and 2D modalities**: no guidance between modalities (none), 1D guide 2D (Block I), 2D guide 1D (Block II), mutual guidance (Block I and III). Modules connected by dotted lines are optional. The brackets signify the selection of blocks. The hidden feature f_t is initialized to 0, $G_{t+1}, ..., G_{t+n}$ represents historical 2D data at historical n moments. H represents 1D data. The final f_{t+n} is taken as 2D feature F_{2d}.

height leads to higher wind speed and lower pressure (Gray 1968). Due to their high relevance with all three attributes, 2D data are divided into shared condition. Besides, we follow the previous multi-task prediction methods and divide 1D data as attribute-shared modality. Thus, attribute-shared modalities consist of **1D** historical TC attribute representing temporal aspects and **2D** historical geopotential height map representing spatial aspects, collectively termed spatio-temporal historical information. However, there are two issues between the 1D and 2D data: (a) Different dimensions. The 2D data have one more dimension than 1D data; (b) Semantic gap. 1D TC attribute represent the values of TC attributes. However, the 2D geopotential height map represents the atmospheric pressure distribution information, and thus there is an obvious semantic gap between them. To address these two issues, we propose the PGSA-LSTM (Prompt-Guided Self-Attention ConvLSTM) module.

2.3 PGSA-LSTM Module: Semantic Gap between Primary and Auxiliary Modalities

To reduce the semantic gap between 1D and 2D data, PGSA-LSTM module discovers mappings between primary and auxiliary modalities in multi-modal data. Because the outputs of model are 1D future data, we first define the 1D data as primary modalities, and 2D data is treated as auxiliary modalities. As shown in Figure 2, we explore four distinct guidance strategies between the primary (1D) and auxiliary (2D) modalities. In the *"1D guide 2D"* strategy, which is inspired by prompt learning principles, the primary modality's feature F_{1d} acts as a guiding prompt to orchestrate the temporal fusion of auxiliary 2D data. To match the dimensions of F_{1d} and 2D data G_{t+i}, we reshape ($\boldsymbol{\mathcal{R}}$) the last dimension (256) of F_{1d} into two dimensions (16, 16). Notably, at the beginning of temporal encoding of the 2D data, guidance is provided by F_{1d}, which contains information from all historical moments. This approach enriches the temporal fusion features with global historical information. The F_{2d} obtained through this method fully incorporates the information from 1D data. After *"1D guide 2D"*, the similarity score of F_{1d} and F_{2d} has increased from 115.66 to 136.26 (See the visualization results in the supplementary material), thus the semantic gap between 1D and 2D data has been reduced. Experimental results in the supplementary material demonstrate that *"1D guides 2D"* outperforms other strategies. Here, F_{1d} is encoded by the *1D-Encoder*, which is designed following the Trajectron++'s encoder (Salzmann et al. 2020). As shown in Figure 2, each historical moment has a PGSA-LSTM block, which contains *2D-Encoder* ($\mathbf{En_{2D}}$, primarily consists of CNN and spatial attention layers; encode 2D data G_{t+i}) and **S**elf-**A**ttention ConvLSTM (Lin et al. 2020) (**SA**, temporal fusion of 2D data under the prompt from reshaped F_{1d}: $\boldsymbol{\mathcal{R}}(F_{1d})$, self-attention captures the semantic interactions between 1D and 2D data). Finally, we use Transformer encoder ($\boldsymbol{\mathcal{T}}$) and residual to obtain the shared condition:

$$\begin{aligned} f_{t+i} &= \mathbf{SA}(cat(f_{t+i-1}, \mathbf{En_{2D}}(G_{t+i}), \boldsymbol{\mathcal{R}}(F_{1d}))) \\ F_{2d} &= f_{t+n} \\ C^{share} &= F_{1d} + \alpha * \boldsymbol{\mathcal{T}}(concat(F_{1d}, F_{2d})) \end{aligned} \tag{8}$$

where the hidden feature f_t is 0, α is a hyperparameter.

3 Experiments

We first describe our experimental setup, conduct ablation studies, and compare our method with SOTA methods. Visualization results for trajectory are then presented (pressure and wind speed results are in the supplementary material), followed by a discussion of expert knowledge in application.

3.1 Experimental Setup

For fair comparison, following MGTCF (Huang et al. 2023), the sampling number is 6, which means that the model generates 6 possible tendencies. Similarly, our model inputs 48 hours (n = 8, the time resolution is 6 hours) of TC historical data and outputs 24 hours (m = 4) of 1D future TC attribute data. We deployed TC-Diffuser on the PyTorch framework. Training was performed using Adam Optimizer with a learning rate of 0.001, a batch size of 256, and a duration of 10 hours. All experiments, including other deep learning methods used for comparison, were conducted on an NVIDIA RTX A6000 GPU, and the seeds in training and testing were fixed. We set the epoch to 270 based on model convergence criteria. For α, we followed the common practice in deep learning, tested four orders of magnitude: 0.1, 0.01, 0.001, 0.0001, and selected the best-performing, 0.001.

Datasets. We employed the dataset introduced by MGTCF (Huang et al. 2023), encompassing all the 1722 TCs data from 1950 to 2021 over the Western North Pacific (WP). So the datasets contain sufficient diverse conditions of TCs. 80% of the TC data from 1950 to 2016 was allocated for training, 20% for validation, and the data from 2017 to 2021 were reserved for testing. The dataset contains 1D historical TC attribute (**1D**), 2D geopotential height map (**2D**), and environmental data (**Env**). The 1D data include historical values of longitude, latitude, pressure, and wind speed. The 2D data depict the current pressure structure. We enhanced the central information of 2D data, see more in the supplementary material. The environmental data describes

Settings	Distance (km)				Pressure error (hPa)				Wind speed error (m/s)			
	6h	12h	18h	24h	6h	12h	18h	24h	6h	12h	18h	24h
C^{share} and C^{spec}	**19.56**	22.86	47.29	84.56	1.16	0.74	1.71	2.79	**0.68**	0.40	1.01	1.61
C^{share} then C^{spec}	20.90	23.26	44.75	81.13	1.26	0.83	1.82	2.93	0.79	0.43	1.00	1.61
C^{spec} then C^{share}	20.50	**22.63**	**40.85**	**75.15**	**1.12**	**0.60**	**1.59**	**2.67**	0.69	**0.34**	**0.88**	**1.50**

Table 1: **Ablation experimental results on three condition guidance orders**. Lower value indicates better performance. "Distance" refers to the error in trajectory task. "6h" represents 6-hour result. C^{share} represents shared condition, C^{spec} represents specific condition. Underlined numbers indicate suboptimal performance.

Row	Settings			Distance (km)				Pressure error (hPa)				Wind speed error (m/s)			
				6h	12h	18h	24h	6h	12h	18h	24h	6h	12h	18h	24h
1	w/o *1D guide 2D*			21.17	26.83	48.05	84.31	1.01	1.06	1.94	2.98	0.63	0.58	1.03	1.60
2	Bi-condition (**Shared, Specific**)	-	-	81.28	163.12	245.31	329.25	1.91	3.29	4.47	5.42	1.12	1.95	2.72	3.37
3		-	Env, Age	77.98	153.99	230.07	306.49	1.67	2.94	3.97	4.89	0.98	1.67	2.21	2.63
4		1D, 2D	-	21.43	23.68	44.72	81.18	**0.99**	0.73	1.68	2.79	**0.59**	0.36	0.89	1.51
5		2D	1D, Env, Age	26.49	46.70	77.81	117.96	1.17	1.65	2.59	3.38	0.72	0.93	1.39	1.74
6	TC-Diffuser	**1D, 2D**	**Env, Age**	**20.50**	**22.63**	**40.85**	**75.15**	1.12	**0.60**	**1.59**	**2.67**	0.69	**0.34**	**0.88**	**1.50**

Table 2: **Ablation experiments on proposed modules.** Row2: w/o Bi-condition. Row3: w/o C^{share}. Row4: w/o C^{spec}. Row5: 1D in C^{spec}. Bi-condition are necessary (Row 2&3&4). 1D data must be considered as shared condition (Row 5&6).

the environmental features during the TC development process. The TC age data (**Age**) is computed from the 1D data.

Metrics. We compute the absolute error between the predicted results and ground truth, including trajectory (distance, km), pressure (hPa), and wind speed (m/s).

3.2 Ablation Studies

Analyzing three condition guidance orders. For the second challenge, we discuss three condition guidance orders in the Bi-condition multi-modal diffusion model: combined shared and specific (C^{share} and C^{spec}), shared then specific (C^{share} then C^{spec}), and specific then shared (C^{spec} then C^{share}, TC-Diffuser). In Table 1, "C^{spec} then C^{share}" performs best on long-term metrics and second best on short-term metrics. This suggests that guidance from the attribute-specific condition first, followed by the attribute-shared condition, benefits long-term learning. In "C^{share} then C^{spec}," multiple attributes in the latent noise are already fused, making it difficult to learn a single attribute with specific condition guidance. Besides, accepting the specific condition first enables the latent noise to acquire attribute-specific domain knowledge. Thus, in "C^{spec} then C^{share}", both specific and shared information are learned, not just one or the other.

Analyzing the effectiveness of proposed modules. We conducted ablation experiments on two groups: ***"1D guide 2D"*** strategy (the core component of PGSA-LSTM), which reduces the semantic gap between 1D and 2D data; **Bi-condition** obtained from two condition generators. As shown in Table 2, *"1D guide 2D"* significantly improves performance (13.34%, 16.87%, 12.62% in attribute predictions), demonstrating that reducing the semantic gap enhances accuracy. In Row 2, a standard diffusion model without Bi-condition performs poorly. A well-designed Bi-condition is essential for adapting the diffusion model to TC forecasting. For example, 1D data must be divided into shared conditions to learn attribute correlations (Row 5&6). In other words, the Bi-condition design must follow attribute-specific and shared constraints. These results show that classifying shared and specific conditions by analyzing modality-attribute correlations is effective and reasonable.

In summary, Bi-condition multi-modal diffusion model, through well-designed Bi-condition and conditions guidance orders, extends the number of conditions and the range of guidance information that the diffusion model can accept.

Discussion on improvement in long-term prediction. Unlike methods predicting TC attributes every six hours, we predict four future moments simultaneously, achieving substantial improvements, especially in long-term prediction, and mitigating error accumulation. As shown in Table 2, suboptimal results occurred in the 6-hour predictions. This is because future moments were predicted together, sharing a common Bi-condition. In fact, the 6-hour results are not significantly different from the final historical moment data. In other words, our Bi-condition overemphasizes global historical information, disrupting the final moment's feature. For 6-hour prediction, *"w/o C^{spec}"* (Row4) performs best because expert knowledge in C^{spec}, such as *future move direction*, benefits long-term results. Regarding the 12-hour error being smaller than the 6-hour error, *"w/o 1D guide 2D"* basically does not conform to this pattern. In other words, adding *"1D guide 2D"* improves long-term prediction more than short-term prediction. Because F_{1d} contains 1D information from all historical moments, aiding 2D data fusion with global context at earlier moments. This favors long-term predictions but may cause overemphasis on global historical features in 6-hour predictions.

3.3 Comparison with State-Of-The-Art Methods

We first compared our method with single-task methods using only historical trajectory or wind speed data: GRU (Cho et al. 2014), NMPT (Gao et al. 2018), DLM (Pan, Xu, and Shi 2019), and TCIF-fusion (Wang, Li, and

Methods	Distance (km)				Pressure error (hPa)				Wind speed error (m/s)				Model size	Training/ inference time
	6h	12h	18h	24h	6h	12h	18h	24h	6h	12h	18h	24h		
GRU (2014)	45.85	104.07	180.29	275.77	-	-	-	-	-	-	-	-	-	-
NMPT(2018)	44.10	101.72	177.06	270.91	-	-	-	-	-	-	-	-	-	-
DLM (2019)	-	-	-	-	-	-	-	-	1.09	1.85	2.48	3.04	-	-
TCIF-fusion (2024)	-	-	-	-	-	-	-	-	-	-	-	3.56	-	-
SGAN (2018)	28.88	61.75	98.74	140.61	1.91	3.12	4.20	5.12	1.05	1.69	2.28	2.81	3.1M	16h/0.12s
GBRNN (2019)	29.93	65.06	105.74	152.06	-	-	-	-	1.16	1.89	2.52	3.10	1.5M	16h/0.55s
MMSTN (2022)	27.57	59.09	96.54	139.19	1.69	2.86	3.94	4.74	0.95	1.52	2.10	2.55	4.8M	18h/0.16s
MGTCF (2023)	23.14	43.37	67.09	93.08	1.37	2.04	2.66	3.29	0.73	1.17	1.55	1.86	3.6M	20h/0.18s
TAM-CL (2024)	-	-	-	155.04	-	-	-	-	-	-	-	4.12	22M	26h/1.11s
CMO (2019)	37.08	52.93	60.69	75.49	2.67	4.30	5.04	6.31	2.29	3.45	2.75	5.00	-	-/-
Pangu (2023)	42.80	44.75	50.85	**65.68**	16.04	16.51	16.70	16.92	-	-	-	-	64M	16days/4min
TC-Diffuser	**20.50**	**22.63**	**40.85**	75.15	**1.12**	**0.60**	**1.59**	**2.67**	**0.69**	**0.34**	**0.88**	**1.50**	9.5M	10h/1.04s

Table 3: **Comparative analysis on performance and efficiency of different methods.** In last column, "s" means seconds, "min" means minutes, "h" means hours. The last two column for CMO are blank because there is no literature record.

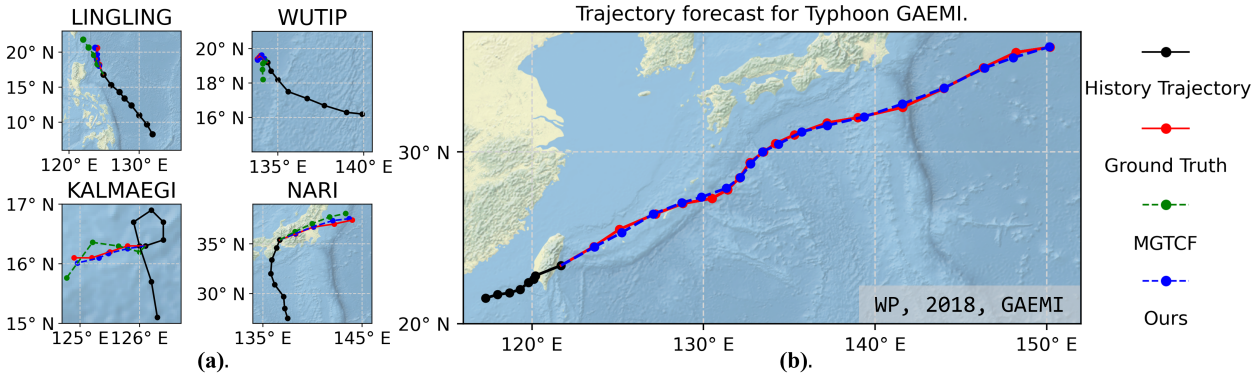

Figure 3: **Visualization of our trajectory forecasting results and comparison with the SOTA deep learning method MGTCF.** (a) The four small figures visualize the ground truth (red solid line), our trajectory forecasting results (blue dashed line), and MGTCF's trajectory forecasting results (green dashed line). Four representative trajectories were selected: natural linear variation, sudden turning, spiral variation, and turning upon encountering land. (b) The trajectory forecasting results during the entire development stages of tropical cyclone GAEMI, which occurred in the western North Pacific (WP) in 2018.

Zheng 2024). Then, multi-task methods were compared: SGAN (Gupta et al. 2018), GBRNN (Alemany et al. 2019), MMSTN (Huang et al. 2022), MGTCF (Huang et al. 2023), and TAM-CL (Li et al. 2024). Next, NWPs method (CMO 2019), used by official meteorological agency, China Central Meteorological Observatory (CMO), was compared. Last, the large-scale model Pangu (Bi et al. 2023) was compared. As shown in Table 3, our approach outperforms the state-of-the-art deep learning model MGTCF and the NWPs method used by CMO across all metrics. TC-Diffuser demonstrates superior performance over MGTCF by 11%∼48% in trajectory, 18%∼71% in pressure, and 6%∼71% in wind speed forecasting. Additionally, TC-Diffuser shows significant improvements in long-term predictions (12, 18, 24-hour), with smaller gains at 6-hour. This is due to MGTCF predicting TC attributes every six hours. In contrast, we predict four future time steps simultaneously, reducing error accumulation. TC-Diffuser achieves these significant improvements with less training time but a bit longer inference time than other deep learning methods. This is due to the unique multi-step denoising mechanism of the diffusion model. Given that the task is to predict the next 6 hours' TC attributes, this is acceptable. The only one outperforming TC-Diffuser is Pangu's 24-hour trajectory forecasting. However, TC-Diffuser achieve higher or equivalent accuracy compared to large-scale models with much less model parameters and time cost, demonstrating its efficiency.

Remarkably, our method for the first time surpasses the typical NWPs model used by CMO in all metrics with only one A6000. This is a critical point, which proves that deep learning models can outperform NWPs methods with much smaller computational resources.

3.4 Qualitative Evaluation

Qualitative analysis of the trajectory forecasting. We visualize and compare our trajectory forecasting results with those of the SOTA deep learning method MGTCF. Figure 3(a) demonstrates our model's ability to accurately predict diverse trajectory variations, which showcases the versatility of our method. Besides, we depict the entire developmental stages of GAEMI occurred in the WP in 2018. It is important to note that our task primarily focuses on short-term forecasting. However, through multiple tests, the results in Figure 3(b) can be effectively achieved. In summary,

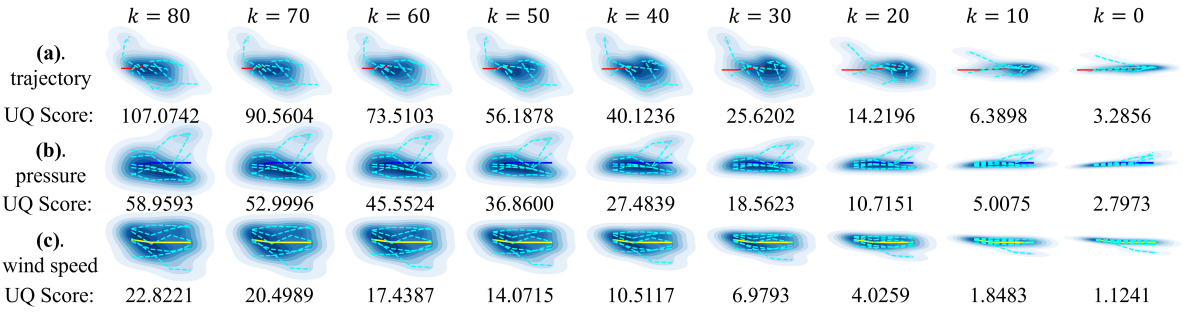

Figure 4: **Visualization of the uncertainty reduction process** of predicted TC attributes for three attributes (trajectory, pressure, wind speed), sampled at diffusion step k. Cyan dashed lines represent the predictions of 6 tendencies, solid lines represent GT. The prediction results are represented as a Gaussian distribution with uncertainty, illustrated in the blue area. Simultaneously, the corresponding uncertainty quantification scores are given.

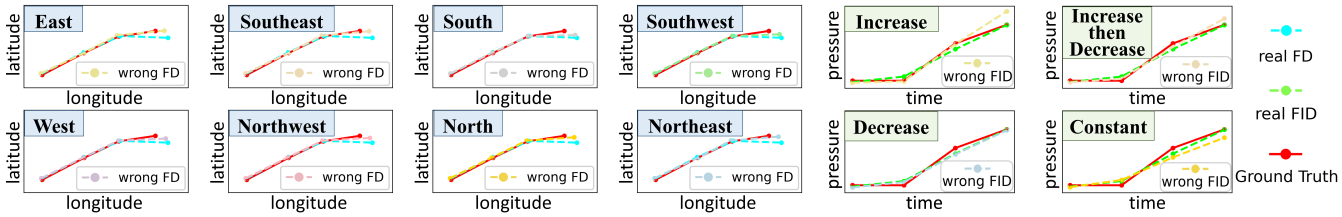

Figure 5: **Visualization of prediction results when using the wrong FD or FID in testing**. The left eight figures: the trajectory prediction results using the real and wrong FD (8 directions, as detailed within blue rectangular boxes). The right four figures: the pressure prediction results using the real and wrong FID (4 situations, as detailed within green rectangular boxes).

TC-Diffuser performs well in trajectory forecasting.

Visualization of the uncertainty reduction process. Our testing process, leveraging the denoising mechanism of the diffusion model, is a reverse diffusion process, which transforms a high-uncertainty prediction distribution into one focused on the low-uncertainty ground truth (GT). As shown in Figure 4, we demonstrate how our predictions align with the ground truth, including 3 attributes. To objectively observe uncertainty reduction, we also provide an uncertainty quantification method. The uncertainty quantification (UQ) score is the MSE loss between six prediction results and GT. The results show that our model successfully reduces uncertainty and achieves accurate, reliable predictions.

3.5 Discussion about Expert Knowledge

Future move Direction (FD) refers to the future move direction of TC trajectory, while ***Future Intensity change Direction*** (FID) indicates pressure change trend. FD&FID are expert knowledge derived from historical 1D data. However, real-time expert knowledge acquisition may not be feasible. Thus, although trained with real FD&FID, our model was tested using the wrong FD&FID. As shown in Figure 5, comparing predictions with real and incorrect FD&FID, the impact on results is negligible. In summary, TC-Diffuser takes into account expert knowledge successfully by training and still keeps accuracy in applications, even with wrong expert knowledge. Thus, random FD&FID values can be used in application. (See quantitative experiments in supplementary material.) This showcases our model's high robustness.

4 Conclusion & Discussion

In this paper, we proposed a novel Bi-condition multi-modal diffusion model for TC forecasting. It improves the range of guidance information that the diffusion model can accept through well-designed Bi-condition, thus achieving accurate predictions with low uncertainty. Considering the interdependence of multiple attributes, we propose two condition generators to obtain shared and specific conditions (Bi-condition) that consider the commonalities and characteristics. To reduce the semantic gap between modalities, we introduce the PGSA-LSTM module to discover mappings between them. Our performance significantly outperforms the current state-of-the-art deep learning methods and the NWPs method used by China Central Meteorological Observatory.

Generalizability. Test results from five other oceans are in supplementary materials. The prediction errors decreased or remained equivalent. This suggests our architecture is applicable to TCs in various oceans and is generalizable.

Robustness.TC-Diffuser maintains high accuracy with most input missing. The original input includes 48 hours' data of 1D, 2D, Env, and Age. With only 24 hours of 1D data, prediction errors decrease by 3% for trajectory and 1% for pressure and wind speed.

High computational efficiency. Compared to NWP models and large models, TC-Diffuser achieves higher or equivalent accuracy with faster inference and much less training time. TC-Diffuser achieves higher accuracy with inference times comparable to other deep learning methods.

Acknowledgments

This work is partially supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LRG25F020002 and LR21F020002, as well as the Natural Science Foundation of China under Grant No. 61976192 and 62202429.

References

Alemany, S.; Beltran, J.; Perez, A.; and Ganzfried, S. 2019. Predicting hurricane trajectories using a recurrent neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 468–475.

Baltrušaitis, T.; Ahuja, C.; and Morency, L.-P. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2): 423–443.

Bauer, P.; Thorpe, A.; and Brunet, G. 2015. The quiet revolution of numerical weather prediction. *Nature*, 525(7567): 47–55.

Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; and Tian, Q. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970): 533–538.

Callaghan, J.; and Smith, R. 1998. The relationship between maximum surface wind speeds and central pressure in tropical cyclones. *Australian Meteorological Magazine*, 47(3): 191–202.

Chen, S.; Long, G.; Shen, T.; and Jiang, J. 2023. Prompt Federated Learning for Weather Forecasting: Toward Foundation Models on Meteorological Data. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23. ISBN 978-1-956792-03-4.

Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Moschitti, A.; Pang, B.; and Daelemans, W., eds., *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734. Doha, Qatar: Association for Computational Linguistics.

CMO, C. M. O. 2019. Typhoon network of Central Meteorological Observatory. http://typhoon.nmc.cn/web.html. Accessed: 2003-03-10.

Coiffier, J. 2011. *Fundamentals of numerical weather prediction*. Cambridge University Press.

DeMaria, M.; and Kaplan, J. 1999. An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Weather and Forecasting*, 14(3): 326–337.

ECMWF. 2011. IFS Documentation-Cy37r2, operational implementation.

Gao, S.; Zhao, P.; Pan, B.; Li, Y.; Zhou, M.; Xu, J.; Zhong, S.; and Shi, Z. 2018. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. *Acta Oceanologica Sinica*, 37: 8–12.

Geng, X.; Liu, Z.; and Shi, Z. 2023. Spatio-Temporal Alignment and Track-To-Velocity Module for Tropical Cyclone Forecast. *Remote Sensing*, 15(20): 4938.

Gray, W. M. 1968. Global view of the origin of tropical disturbances and storms. *Monthly Weather Review*, 96(10): 669–700.

Gupta, A.; Johnson, J.; Fei-Fei, L.; Savarese, S.; and Alahi, A. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2255–2264.

Huang, C.; Bai, C.; Chan, S.; and Zhang, J. 2022. MMSTN: A Multi-Modal Spatial-Temporal Network for Tropical Cyclone Short-Term Prediction. *Geophysical Research Letters*, 49(4): e2021GL096898.

Huang, C.; Bai, C.; Chan, S.; Zhang, J.; and Wu, Y. 2023. MGTCF: multi-generator tropical cyclone forecasting with heterogeneous meteorological data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 5096–5104.

Kim, S.; Kim, H.; Lee, J.; Yoon, S.; Kahou, S. E.; Kashinath, K.; and Prabhat, M. 2019. Deep-hurricane-tracker: Tracking and forecasting extreme climate events. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1761–1769. IEEE.

Kimura, R. 2002. Numerical weather prediction. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(12-15): 1403–1414.

Knaff, J. A.; and Zehr, R. M. 2007. Reexamination of tropical cyclone wind–pressure relationships. *Weather and Forecasting*, 22(1): 71–88.

Li, T.; Lai, M.; Nie, S.; Liu, H.; Liang, Z.; and Lv, W. 2024. Tropical cyclone trajectory based on satellite remote sensing prediction and time attention mechanism ConvLSTM model. *Big Data Research*, 36: 100439.

Lin, Z.; Li, M.; Zheng, Z.; Cheng, Y.; and Yuan, C. 2020. Self-attention convlstm for spatiotemporal prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 11531–11538.

Ma, Q.; Pan, J.; and Bai, C. 2024. Direction-Oriented Visual–Semantic Embedding Model for Remote Sensing Image–Text Retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–14.

Moradi Kordmahalleh, M.; Gorji Sefidmazgi, M.; and Homaifar, A. 2016. A sparse recurrent neural network for trajectory prediction of atlantic hurricanes. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, 957–964.

Neumann, C. J.; and Lawrence, M. B. 1975. An operational experiment in the statistical-dynamical prediction of tropical cyclone motion. *Monthly Weather Review*, 103(8): 665–673.

Pan, B.; Xu, X.; and Shi, Z. 2019. Tropical cyclone intensity prediction based on recurrent neural networks. *Electronics Letters*, 55(7): 413–415.

Rüttgers, M.; Lee, S.; Jeon, S.; and You, D. 2019. Prediction of a typhoon track using a generative adversarial network and satellite images. *Scientific reports*, 9(1): 6057.

Salzmann, T.; Ivanovic, B.; Chakravarty, P.; and Pavone, M. 2020. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, 683–700. Springer.

Sanders, F.; Pike, A. C.; and Gaertner, J. P. 1975. A barotropic model for operational prediction of tracks of tropical storms. *Journal of Applied Meteorology and Climatology*, 14(3): 265–280.

Song, J.; Duan, Y.; and Klotzbach, P. J. 2020. Revisiting the relationship between tropical cyclone size and intensity over the western North Pacific. *Geophysical Research Letters*, 47(13): e2020GL088217.

Song, T.; Li, Y.; Meng, F.; Xie, P.; and Xu, D. 2022. A novel deep learning model by Bigru with attention mechanism for tropical cyclone track prediction in the Northwest Pacific. *Journal of Applied Meteorology and Climatology*, 61(1): 3–12.

Wang, C.; Li, X.; and Zheng, G. 2024. Tropical cyclone intensity forecasting using model knowledge guided deep learning model. *Environmental Research Letters*, 19(2): 024006.

Wu, L.; and Wang, B. 2004. Assessing impacts of global warming on tropical cyclone tracks. *Journal of climate*, 17(8): 1686–1698.

Wu, Y.; Geng, X.; Liu, Z.; and Shi, Z. 2021. Tropical cyclone forecast using multitask deep learning framework. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Yan, H.; Mu, P.; Huang, C.; Zhang, J.; and Bai, C. 2024. Phy-CoCo: Physical Constraint-Based Correlation Learning for Tropical Cyclone Intensity and Size Estimation. In *European Conference on Artificial Intelligence*, 2226–2233. IOS Press.

Zhao, D.; Wang, Q.; Zhang, J.; and Bai, C. 2023. Mine diversified contents of multispectral cloud images along with geographical information for multilabel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–15.