

# iTCTSL: Interpretable Tropical Cyclone Track and Intensity Forecasting via Task Sensitive Learning

## Supplementary Material

### 6. List of Notations

As shown in Tab. 4, we list the notation mentioned in the manuscript, and provide descriptions of them.

Notation	Description
$X_t^k$	historical TC attributes of sub-task $k$ at time $t$
$\mathcal{K}$	task set of TC multi-attribute forecasting: longitude, latitude, pressure, wind speed
$t$	time index for input or prediction sequence
$n$	length of historical input sequence
$\mathbf{Z}$	ERA5 meteorological reanalysis data at $t = 0$
$m$	length of predicted future sequence
$E_t$	environmental variables extracted from TC attributes at time $t$
$H_{wp}$	elevation data of the Western North Pacific region
$lon_t$	longitude of TC center at time $t$
$lat_t$	latitude of TC center at time $t$
$r_t$	elevation patch cropped at TC center position at time $t$
$s_t$	slope computed from $r_t$
$a_t$	aspect computed from $r_t$
$q_t$	roughness of $r_t$ , measured by standard deviation
$\text{Enc}_{\text{topo}}$	topographic feature encoder
$r'_t$	encoded elevation feature
$s'_t$	encoded slope feature
$a'_t$	encoded aspect feature
$q'_t$	encoded roughness feature
$\mathbf{R}'_k$	topographic feature combination for TC forecasting sub-task $k$
$\mathbf{R}_{lon,lat}$	topographic feature combination for longitude and latitude
$\mathbf{Q}$	query vector projected from $\mathbf{R}'_{k,lat}$
$\mathbf{C}$	learnable topographic codebook
$c_1, c_2, \dots, c_K$	topographic prototypes in the codebook
$\lambda$	number of prototypes in the codebook
$d$	dimension of latent representation of topographic prototypes
$\mathbf{K}$	key vectors projected from topographic prototypes
$\mathbf{V}$	value vectors projected from topographic prototypes to represent influence characteristics of TC track
$\mathbf{M}_a$	causal soft mask generated via Gumbel-Softmax attention over topographic prototypes
$\mathbf{M}_b$	confounding mask, complement of $\mathbf{M}_a$
$\mathbf{F}_{lon,lat}^k$	task-adaptive feature $\mathbf{F}_t^k$ for $k \in \{\text{lon}, \text{lat}\}$
$\mathbf{F}_{cau}^k$	causal branch feature extracted via the causal mask $\mathbf{M}_a$
$\mathbf{F}_{con}^k$	confounding branch feature extracted via the confounding mask $\mathbf{M}_b$
$\mathbf{F}_{cau}^{fd}$	predicted track direction from causal branch
$\mathbf{F}_{con}^{fd}$	predicted track direction from confounding branch
$\mathbf{F}_{int}^{fd}$	predicted intervened prediction from intervention branch
$\mathbf{F}_{gt}^{fd}$	ground truth label for future track direction
$\mathbf{F}_{dir}^{gt}$	uniform distribution over eight TC track development directions
$H(\cdot, \cdot)$	cross-entropy loss function
$\text{KL}(\cdot \  \cdot)$	Kullback-Leibler divergence function
$\mathbf{X}_t^k$	encoded historical TC attribute feature for sub-task $k$
$\mathbf{H}_k$	one-hot encoded task id for sub-task $k$
$\mathbf{Z}_t^k$	encoded ERA5 meteorological reanalysis feature for sub-task $k$
$\mathbf{W}_k^{(i)}$	$i$ -th group convolution kernel
$C$	number of channels in meteorological feature $\mathbf{Z}_t^k$
$H$	spatial height of meteorological feature
$W$	spatial width of meteorological feature
$\mathbf{Z}_t^k$	channel-expanded meteorological features for sub-task $k$
$s_k$	channel relevance scores for sub-task $k$
$\mathbf{g}_k$	Gumbel noise for sub-task $k$ sampled from $\text{Gumbel}(0, 1)$
$\sigma_g$	scaling factor for Gumbel noise
$\tilde{s}_k$	noisy channel scores for sub-task $k$ after Gumbel perturbation
$\mathbf{D}_k$	channel assignment cost matrix for sub-task $k$
$\tau$	temperature parameter for Sinkhorn algorithm
$T_{\text{sink}}$	number of iterations in Sinkhorn normalization process
$\mathbf{P}_k$	soft assignment matrix from Sinkhorn for sub-task $k$
$\nu$	right marginal constraint for Sinkhorn output (number of selected channels)
$\mathbf{m}_k$	soft selection probabilities for expanded channels in sub-task $k$
$\mathbf{Z}_k$	task-adaptive meteorological feature for sub-task $k$
$\tilde{\mathbf{Z}}_k$	channel-reduced task-adaptive meteorological feature derived from $\mathbf{Z}_k$ for sub-task $k$
$\mathbf{Q}_k$	query projected from $\tilde{\mathbf{Z}}_k$ for sub-task $k$
$\mathbf{K}_k$	key projected from original meteorological feature $\mathbf{Z}_t^k$ for sub-task $k$
$\mathbf{V}_k$	value projected from original meteorological feature $\mathbf{Z}_t^k$ for sub-task $k$
$\gamma$	channel dropout rate
$\mathbf{M}_k$	binary mask for channel dropout in sub-task $k$
$d_k$	feature dimension of projected queries and keys in sub-task $k$
$\mathbf{G}_k$	cross-attention output from $\mathbf{Z}_t^k$ attending to $\mathbf{Z}_t^k$ in sub-task $k$
$\mathbf{G}'_k$	FFN-transformed feature from $\mathbf{G}_k$ in sub-task $k$
$\mathbf{F}_k$	fusion feature of $\mathbf{X}_t^k$ (TC attributes feature) and $\mathbf{Z}_t^k$ (task-adaptive meteorological feature) for sub-task $k$
$\mathbf{F}_k'$	fusion feature of $\mathbf{F}_k$ (task-adaptive met+attr feature) and $\mathbf{R}'_k$ (topographic feature) for sub-task $k$
$\mathbf{F}_k''$	fusion feature of $\mathbf{F}_k'$ and $\mathbf{F}_{cau}^k$ for sub-task $k$
$\mathcal{T}_k$	forecasting sub-task $k$
$\theta$	shared parameter subspace
$\nabla_{\theta} \mathcal{L}_{\mathcal{T}_k}$	gradient of sub-task $\mathcal{T}_k$ 's loss w.r.t. shared parameters
$\mathbf{A}$	asymmetric task affinity matrix between sub-tasks
$\epsilon$	stability constant for normalization
$w_k^{\text{raw}}$	raw influence score of sub-task $\mathcal{T}_k$ from affinity matrix
$\alpha_k$	learnable directional scaling factor for sub-task $\mathcal{T}_k$
$w_k$	positive weight for sub-task $\mathcal{T}_k$ after softplus transformation
$\tilde{w}_k$	normalized loss weight for sub-task $\mathcal{T}_k$ in current batch

Table 4. List of notations and their descriptions.

### 7. Experimental Details

#### 7.1. Datasets

The tropical cyclone (TC) attribute data used in this study are derived from the CMA-BST dataset [36], and environmental factors are extracted from ERA5 reanalysis data [13]. The full dataset covers 1861 TC events from 1950 to 2024 over the Western North Pacific (WP). TCs from 1950 to 2016 are split into 80% for training and 20% for validation, while those from 2023 to 2024 are reserved for testing. The study region spans from 100°E to 179.75°E in longitude and 0° to 59.75°N in latitude, encompassing the majority of TC activities in the WP [30]. For environmental variables, we extract six key meteorological variables within a 25°×25° square region centered on the TC center: geopotential height (gph), temperature (u), meridional wind (u), and zonal wind (v) at 200 hPa, 500 hPa, 700 hPa, and 850 hPa levels, as well as sea surface temperature (sst) and total column water vapor (tcwv). These variables are provided at a spatial resolution of 0.25°, offering a comprehensive representation of TC-related environmental conditions. In addition, we include topographic elevation data with a spatial resolution of 0.01°, covering the region from 5°S to 64.75°N and 95°E to 184.75°E to account for topographic influence.

In addition to the above data sources, we also incorporate the Env-Data proposed in MGTCF [15], which provides a sequence of environmental variables  $E_t$ , extracted from TC attributes at each time step  $t \in \{-n+1, -n+2, \dots, 0\}$ . These variables include a normalized continuous variable representing the TC's moving velocity, as well as several one-hot encoded variables. Specifically, the one-hot encoded variables include the month, the TC's longitude and latitude (discretized within the domain 80°E–160°W and 0°N–60°N, respectively), the movement direction over the past 24 hours, and the intensity change over the past 24 hours. Among them, History Direction (24 h) and History Intensity Change (24 h) are categorical variables. The former is categorized into eight compass directions: east, southeast, south, southwest, west, northwest, north, and northeast. Notably, the same directional categorization is adopted in our TDG module for predicting the future movement direction of TC. The latter is classified into four intensity evolution types: intensifying, intensifying-then-weakening, weakening, and stable. As  $E_t$  is adopted from previous work and not directly related to our core contributions, we do not elaborate on it in the manuscript. It

is used as shared input features and concatenated with the task-specific features  $\mathbf{F}_k$ , which are produced by the Task-sensitive Environmental Routing Module (TERM), to guide the forecasting of TC attributes.

## 7.2. Metrics

To evaluate the performance of TC track forecasting models, a commonly used metric is the geographical distance between the forecasted and ground-truth latitude and longitude coordinates. In this study, the following distance formula is adopted:

$$d = \sqrt{\left(\frac{\Delta\text{lon}}{10} \cdot 111 \cos\left(\frac{\text{lat}_{\text{gt}}}{10} \frac{\pi}{180}\right)\right)^2 + \left(\frac{\Delta\text{lat}}{10} \cdot 111\right)^2} \quad (15)$$

Here,  $\Delta\text{lat}$  and  $\Delta\text{lon}$  represent the differences between the forecasted and ground-truth latitude and longitude, respectively. This formula accounts for the Earth's curvature by applying a cosine correction to the longitudinal component, using the ground-truth latitude as reference. This adjustment improves the accuracy of the Euclidean distance computation between two geographic coordinates.

The primary evaluation metric adopted in this work is the Mean Absolute Error (MAE), which quantifies the average magnitude of prediction errors between the forecasted and ground-truth values. In particular, MAE is computed for the TC track error, defined as the spatial distance between the forecasted and actual TC centers, the minimum central pressure error and the maximum wind speed error near the TC center. The MAE is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

where  $y_i$  is the  $i$ -th ground-truth value,  $\hat{y}_i$  is the corresponding forecasted value, and  $n$  is the total number of samples.

## 7.3. Hyperparameter Selection

We conduct a sensitivity study on two key hyperparameters of the Sinkhorn algorithm within the core module, Task-sensitive Environmental Routing Module (TERM): the temperature  $\tau$  and the iteration count  $T_{\text{sink}}$ . We investigate how different combinations of these parameters affect model performance on both track and intensity forecasting tasks. As shown in Fig. 9, lower temperature values generally yield sharper soft masks but may sacrifice prediction accuracy due to overly discrete assignments. On the other hand, increasing  $T_{\text{sink}}$  improves marginal alignment but may introduce noise when overly large. These findings suggest that an appropriate balance between selection sharpness and marginal refinement is necessary.

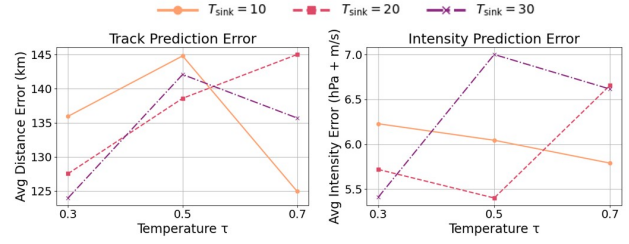


Figure 9. Effects of temperature  $\tau$  and Sinkhorn iteration count  $T_{\text{sink}}$  on forecasting performance.

## 8. Detailed Discussion of Experimental Results

### 8.1. Comparative Forecast Visualizations for Coastal TCs

Fig. 10 provides a visual comparison of TC track forecasts by different models for two representative coastal TCs: GAEMI (nearshore) and HAIKUI (landfalling). In these cases, iTCTSL consistently achieves higher track accuracy than other methods. The zoomed-in visualizations further highlight this advantage by clearly showing iTCTSL's closer alignment with the actual TC track under complex coastal conditions, reflecting its ability to capture the evolving development patterns of tropical cyclones in such challenging environments.

Notably, Fig. 10(b) reveals a significant track deviation in Pangu's forecast. As discussed in the manuscript, large-scale AI-based weather models such as Pangu-Weather treat TC forecasting as a downstream task, with their performance reliant on the effectiveness of postprocessing TC tracking algorithms. This dependence can lead to substantial errors in complex scenarios, particularly for landfalling TCs, where the rapid weakening of TC intensity after landfall presents a major challenge for accurate localization by tracking algorithms.

### 8.2. Understanding the Unusual Error Pattern in Intensity Forecasting of Coastal TCs

For coastal TC samples, many methods exhibit lower long-term intensity forecast errors compared to short-term ones. This is because rapid intensity changes during the early stage of landfall are difficult to predict, whereas TC intensity typically weakens and stabilizes in the later stage, making it easier to forecast. Notably, this phenomenon is especially evident for the Pangu and ECMWF numerical prediction methods. Since ERA5 data itself tends to underestimate TC intensity [5, 26], both Pangu and ECMWF are biased toward lower TC intensities in both short-term and long-term forecasts. After landfall, the actual TC intensity decreases rapidly, causing the underestimated long-term forecasts to become closer to the actual values.

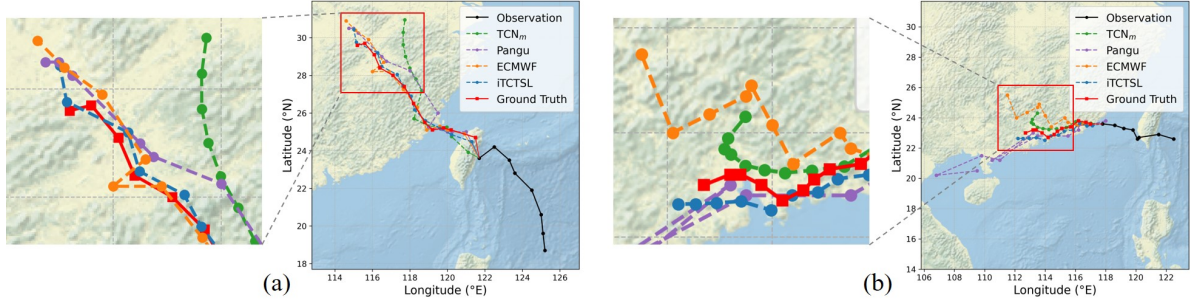


Figure 10. Compares the TC track forecasts of different models in nearshore and landfall regions. (a) A nearshore case for TC GAEMI (No. 3, 2024) at 12:00 UTC on July 24. (b) A landfall case for TC HAIKUI (No. 11, 2023) at 18:00 UTC on September 4.

### 8.3. Visualization and Interpretation of Learned Topographic Prototypes

To understand what the topography-prototype codebook has learned, we visualize the prediction samples and their corresponding topographic prototype classes within a complex region of the WP, as illustrated in Fig. 11. The results show that areas with similar topographic characteristics are consistently mapped to the same prototype. Specifically, samples located near the Central Mountain Range of Taiwan and the mountains of Luzon are predominantly assigned to high-impact prototypes, while those over oceanic regions are grouped into low-impact categories. Prior research has confirmed that both the Central Mountain Range of Taiwan and the mountains of Luzon exert considerable influence on TC track [21, 38]. These findings indicate that the learned prototypes effectively capture the causal effects of major topographic barriers on TC movement.

### 8.4. Intuitive Visualization of Task-Specific Meteorological Variable Contributions

We further investigated the reasons behind the performance improvement observed after task decoupling. Specifically, we conducted an interpretability analysis of how each meteorological variable contributes to different forecasting tasks during training, aiming to understand the model’s decision-making process. As shown in Fig. 12, each subplot corresponds to a specific task—longitude, latitude, pressure, and wind speed—and presents a line chart depicting the changing contribution rates of six meteorological variables over the course of training. Additionally, the accompanying bar charts provide a more intuitive visualization of the average contribution of each variable after training, highlighting their overall importance to the respective task.

In particular, for longitude forecasting (Fig. 12a), the meridional wind ( $u$ ) and geopotential height contribute the most, while for latitude forecasting (Fig. 12b), the zonal wind ( $v$ ) shows the highest contribution, followed by geopotential height. Notably, previous studies in meteorology have shown that large-scale circulation variables play a cru-

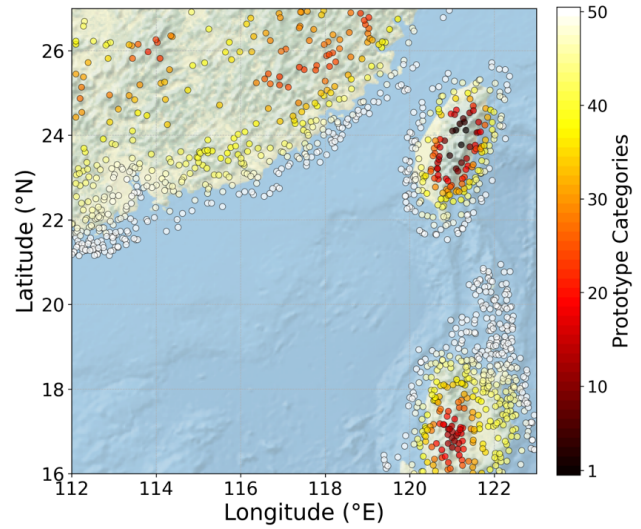


Figure 11. Visualization of the learned topographic prototypes (codebook) in a complex topographic sector of the WP. Each circle denotes the current TC center for a prediction sample, and the color indicates the corresponding topography-prototype class assigned by iTCTSL.

cial role in TC track forecasting [6], and iTCTSL successfully captures this characteristic. In the pressure forecasting task (Fig. 12c), sea surface temperature and total column water vapor gradually become the dominant variables as training progresses, indicating that iTCTSL effectively learns the strong relationship between thermal conditions and TC intensity. The contribution of temperature also increases, while wind and geopotential height play relatively minor roles. Prior research has demonstrated that TC intensity is highly sensitive to the surrounding thermal environment [29], with factors such as sea surface temperature and water vapor being key drivers of intensity changes, which aligns well with the behavior learned by iTCTSL. For the wind speed forecasting task (Fig. 12d), iTCTSL similarly relies on thermal variables that provide energy for convec-





Figure 12. Contribution rate of six meteorological variables for four TC forecasting tasks (longitude, latitude, central pressure, and maximum wind speed). The line chart for each task illustrates how the contribution rate of six meteorological variables change over the course of training. The bar chart shows the average contribution rate of each variable after training, indicating their overall importance to the corresponding task.

tion and rotation. Meanwhile, the wind field, which reflects pressure gradients, makes a moderate contribution.

Overall, these results indicate that iTCTSL can adaptively identify and leverage the most relevant meteorological variables for each forecasting task. This finding is consistent with established meteorological knowledge, demonstrating that iTCTSL has learned meaningful physical relationships for TC forecasting and exhibits strong interpretability.

## 8.5. Comparative Analysis of Model Performance and Efficiency

Tab. 5 summarizes the model size and training/inference time of Pangu and our proposed method iTCTSL. Compared with the large-scale model Pangu (256M parameters, trained on 192 V100 GPUs over 15 days), iTCTSL is significantly more lightweight, with only 36.04M parameters, trained on a single NVIDIA RTX A6000 GPU within 30 hours, and capable of single-sample inference within 0.31 seconds. Despite the substantially reduced model size and training cost, iTCTSL demonstrates better performance than Pangu in several important aspects.

For coastal TCs, iTCTSL yields lower track and intensity forecast errors across all lead times. For all TCs, iTCTSL outperforms Pangu in intensity forecasting across all lead times and achieves more accurate track forecasts within 24 hours. These results demonstrate the superior efficiency of iTCTSL in both computational and predictive aspects.

Methods	Model Size	Training / Inference Time
Pangu	256.00M	15 days / 4 min
iTCTSL	36.04M	30 h / 0.31 s

Table 5. Comparison of model size and training/inference time between Pangu and iTCTSL. Here, “s” means seconds, “min” means minutes, and “h” means hours.

## 8.6. Rationale for Test Period Selection

The 2023–2024 test period was chosen to align with Pangu’s prediction results that we obtained from authoritative sources, ensuring all models are evaluated against the same test data, thereby supporting a fair and transparent comparison.

## 8.7. Additional Coastal TC Case Studies with Baseline Comparisons

Fig. 13 presents extended versions of the two coastal case studies shown in the main paper, incorporating predictions from multiple baseline models (e.g., TCN<sub>M</sub>, Pangu, ECMWF) alongside iTCTSL and its ablation variant. These results provide a more comprehensive comparison of how different forecasting methods respond to complex coastal topography and abrupt changes in TC motion. Across both time snapshots—GAEMI 2024/07/24 06:00 UTC and GAEMI 2024/07/24 12:00 UTC—the additional baselines reveal varying sensitivities to the topographic barriers imposed by Taiwan’s Central Mountain Range. While several models tend to project unrealistic direct crossing or insuf-

Table 6. Forecasting robustness under meteorological variable perturbations. Perturbations are applied by injecting independent Gaussian noise into each meteorological variable, where the noise level is defined as the ratio between the noise standard deviation and the standard deviation of the original variable.

Noise level	Variable	Distance Error (km)						Pressure Error (hPa)						Wind Speed Error (m/s)					
		12h	24h	36h	48h	60h	72h	12h	24h	36h	48h	60h	72h	12h	24h	36h	48h	60h	72h
0%	/	<b>42.64</b>	<b>72.06</b>	<b>102.69</b>	<b>134.79</b>	<b>174.93</b>	<b>217.74</b>	<b>2.76</b>	<b>3.97</b>	<b>4.07</b>	<b>2.98</b>	<b>2.36</b>	<b>2.64</b>	<b>1.36</b>	<b>2.01</b>	<b>3.06</b>	<b>2.67</b>	<b>2.10</b>	<b>1.70</b>
1%	GPH	51.52	92.94	135.50	158.94	184.63	246.49	2.96	4.12	4.17	3.18	2.42	2.68	2.14	2.73	3.26	2.78	2.76	2.16
	WIND	58.80	94.06	145.41	165.02	185.65	248.84	3.01	4.14	4.36	3.61	2.52	2.95	2.65	3.01	3.39	3.16	2.84	2.53
	TEMP	49.72	74.27	111.41	134.47	174.83	232.15	3.26	4.31	4.73	3.81	2.82	3.05	2.31	2.83	3.14	2.85	2.74	2.26
	SST	44.48	75.82	110.14	144.81	171.19	221.44	3.52	5.09	5.36	4.47	4.38	3.69	2.48	3.12	3.74	3.31	3.09	2.61
	TCWV	47.80	76.61	116.51	148.61	168.43	225.66	3.49	4.93	5.25	4.32	4.18	3.40	2.60	2.99	3.58	2.93	2.79	2.88
5%	GPH	57.36	100.83	137.84	165.66	189.53	263.74	3.37	4.25	4.38	3.00	2.82	2.99	2.36	2.87	3.97	3.56	3.45	3.17
	WIND	62.37	108.54	148.27	171.70	193.14	272.82	3.65	4.17	4.41	3.19	3.03	3.19	3.62	3.05	4.24	3.73	3.56	3.26
	TEMP	49.43	80.25	116.82	145.55	178.80	241.47	3.99	4.56	5.04	4.59	4.13	4.44	2.46	2.93	4.03	3.48	3.36	3.24
	SST	53.15	86.62	109.05	158.61	171.45	220.61	4.58	5.25	6.11	5.61	5.08	5.41	2.89	3.67	4.65	3.86	3.20	3.69
	TCWV	56.17	90.80	111.77	152.29	174.51	228.96	4.28	5.18	5.78	5.38	4.37	4.89	2.78	3.59	4.32	3.79	3.00	3.53
50%	GPH	62.71	114.39	152.68	185.86	233.94	346.92	3.68	4.23	4.68	3.75	3.66	4.17	2.30	3.30	4.85	4.74	4.42	4.37
	WIND	68.62	127.30	156.01	189.98	240.27	358.79	3.66	4.67	4.81	3.90	3.71	4.38	2.47	3.26	5.09	4.89	4.50	4.40
	TEMP	53.32	107.83	129.54	149.01	200.86	247.02	4.16	5.07	5.31	4.94	4.77	5.21	2.51	3.34	4.88	4.86	4.49	4.24
	SST	50.84	107.72	134.99	157.29	197.37	225.61	5.19	5.43	7.06	6.84	6.23	6.29	3.23	3.68	5.79	4.94	4.76	4.43
	TCWV	54.66	110.93	141.52	164.46	184.65	219.16	4.91	4.78	6.26	6.14	5.56	5.55	3.17	3.70	6.01	4.90	4.64	4.45

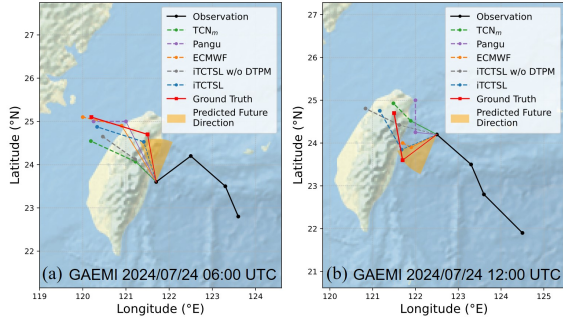


Figure 13. Predictions from multiple forecasting methods for two representative coastal TC cases. The baseline approaches (TCN<sub>M</sub>, Pangu, ECMWF) exhibit different responses to the terrain-induced steering near Taiwan’s Central Mountain Range, with some yielding implausible direct-crossing tracks or insufficient deflection. In contrast, iTCTSL provides more physically consistent forecasts that better match the observed motion, highlighting the advantage of incorporating topographic guidance.

icient deflection, iTCTSL consistently produces physically plausible tracks that better align with the true track.

## 8.8. Robustness to Meteorological Variable Perturbations

To further evaluate the robustness of iTCTSL under realistic forecasting conditions, Tab. 6 reports perturbation experiments in which independent Gaussian noise is injected into each meteorological variable at various noise levels (i.e., 1%, 5%, 50%). The noise level is defined as the ratio between the noise standard deviation and the original data standard deviation for the perturbed variable.

Across all meteorological variables, iTCTSL exhibits

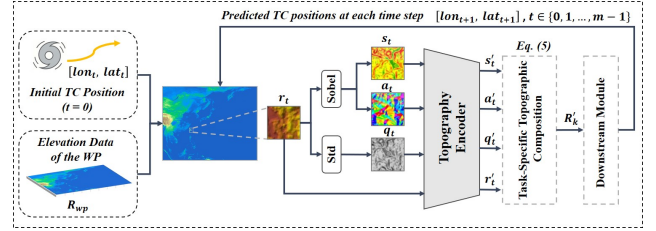


Figure 14. The detailed structure of DTA.

only slight performance degradation when the noise level is within 5%, demonstrating strong resilience to moderate perturbations in environmental inputs. More noticeable degradation emerges only when the noise level reaches 50%; however, even under such severe disturbances, the model maintains stable prediction behavior without exhibiting exponential error escalation. This indicates that iTCTSL is robust across a wide range of input uncertainties.

Moreover, the degradation patterns observed in these perturbation experiments align closely with the task-specific variable dependencies identified in our interpretability analysis. Perturbing large-scale circulation variables (e.g., wind, gph) disproportionately affects track forecasting, whereas disturbances applied to thermal variables (e.g., sst, tcwv) primarily influence intensity predictions. This consistency provides further empirical support that iTCTSL not only captures physically meaningful relationships between environmental factors and forecasting tasks but also leverages them robustly under noisy conditions.

## 9. More Details about Model

### 9.1. Dynamic Topography Awareness(DTA)

Fig. 14 illustrates the architecture of the DTA module, which continuously updates topographic features based on the predicted TC locations at each time step.

The full elevation data of the WP (denoted as  $R_{wp}$ ) is preloaded into memory. At each time step  $t \in \{0, 1, \dots, m-1\}$ , a square elevation patch  $r_t$  centered at the predicted TC location  $(lon_t, lat_t)$  is extracted by cropping a  $512 \text{ km} \times 512 \text{ km}$  region from  $R_{wp}$ . Since different forecasting sub-tasks are affected by topography in different ways, we apply a set of transformations to  $r_t$  to extract multiple topographic descriptors. The elevation gradient is first computed by applying the Sobel operator, which produces spatial derivatives along the longitudinal and latitudinal directions, denoted as  $\nabla_x r_t$  and  $\nabla_y r_t$ . The slope  $s_t$  and aspect  $a_t$  are then calculated as:

$$s_t = \sqrt{(\nabla_x r_t)^2 + (\nabla_y r_t)^2}, a_t = \arctan 2(\nabla_y r_t, \nabla_x r_t) \quad (17)$$

where  $\arctan 2(\cdot, \cdot)$  computes the four-quadrant inverse tangent, returning the angle of the gradient vector with respect to the positive  $x$ -axis. Additionally, the terrain roughness  $q_t$  is computed as the standard deviation of elevation values within  $r_t$ . The extracted topographic features  $(r_t, s_t, a_t, q_t)$  are encoded by a topography feature encoder (i.e.,  $\text{Enc}_{\text{topo}}$ ):

$$[r'_t, s'_t, a'_t, q'_t] = \text{Enc}_{\text{topo}}(r_t, s_t, a_t, q_t) \quad (18)$$

To accommodate the varying relevance of topographic cues across forecasting sub-tasks, a task-specific composition step is performed. The final topographic representation  $\mathbf{R}'_k$  for each sub-task is defined as:

$$\mathbf{R}'_k = \begin{cases} r'_t + s'_t + a'_t, & k \in \{\text{lon}, \text{lat}\} \\ r'_t, & k = \text{pres} \\ r'_t + q'_t, & k = \text{wind} \end{cases} \quad (19)$$

The resulting  $\mathbf{R}'_k$  is then passed to the downstream module for predicting the TC attributes at time step  $t+1$ .

### 9.2. Sinkhorn Algorithm

As described in the manuscript, we adopt a differentiable approximation of top- $C$  channel selection by formulating it as an entropy-regularized optimal transport problem between  $2C$  expanded channels and two selection states: “selected” and “unselected.” The transport cost matrix  $\mathbf{D}_k \in \mathbb{R}^{2C \times 2}$  is constructed from perturbed channel scores  $\tilde{s}_k$  (with Gumbel noise), and converted into log-probabilities:

$$\mathbf{A}_k = \exp(-\mathbf{D}_k/\tau) \quad (20)$$

We apply the Sinkhorn algorithm to project  $\mathbf{A}_k$  into a soft transport matrix  $\mathbf{P}_k \in \mathbb{R}^{2C \times 2}$  with prescribed marginal constraints: uniform row sums (i.e.,  $\sum_j \mathbf{P}_{ij} = 1$ ) and column sums  $(2C - C, C)$ , corresponding to “unselected” and “selected” counts, respectively. Note that the column marginals are defined in terms of mass (counts) rather than normalized probabilities, consistent with the channel count constraint in implementation. The iterative normalization process begins with:

Repeat for  $T_{\text{sink}}$  steps:

$$\mathbf{P} \leftarrow \begin{cases} \mathbf{P} \oslash \left( \sum_j \mathbf{P}_{:,j} + \varepsilon \right) & \text{(row norm)} \\ \mathbf{P} \oslash \left( \left( \sum_i \mathbf{P}_{i,:} \right) \oslash \boldsymbol{\nu} + \varepsilon \right) & \text{(column norm)} \end{cases} \quad (21)$$

where  $\oslash$  denotes element-wise division (i.e., Hadamard division),  $\varepsilon$  is a small constant for numerical stability, and  $\boldsymbol{\nu} = [2C - C, C]$  specifies the target column marginal (i.e., the total counts for “unselected” and “selected” states, respectively).

The final soft selection vector  $\mathbf{m}_k \in [0, 1]^{2C}$  is obtained from the second column of  $\mathbf{P}_k$  and used for channel-wise gating. This mechanism enables differentiable channel routing, allowing the model to learn interpretable and task-sensitive environmental dependencies without sacrificing gradient flow or flexibility.