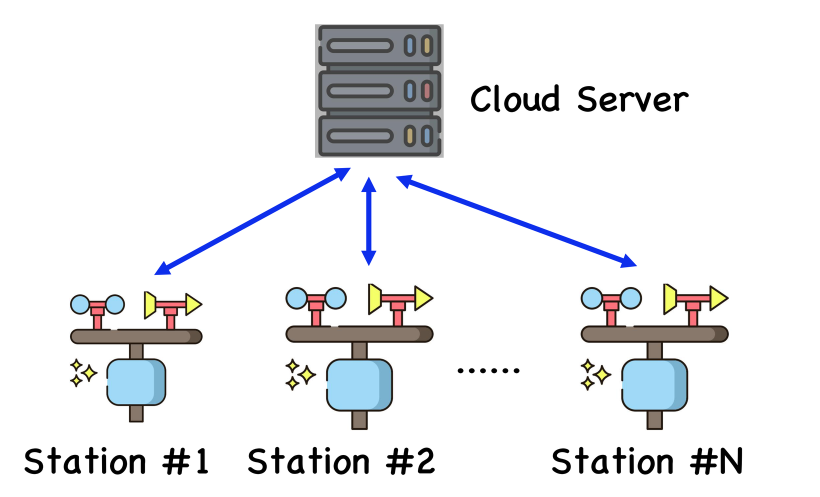
Federated Prompt Learning for Weather Foundation Models on Devices

Shengchao Chen, Guodong Long, Tao Shen, Jing Jiang, Chengqi Zhang
Australian Artificial Intelligence Institute, University of Technology Sydney

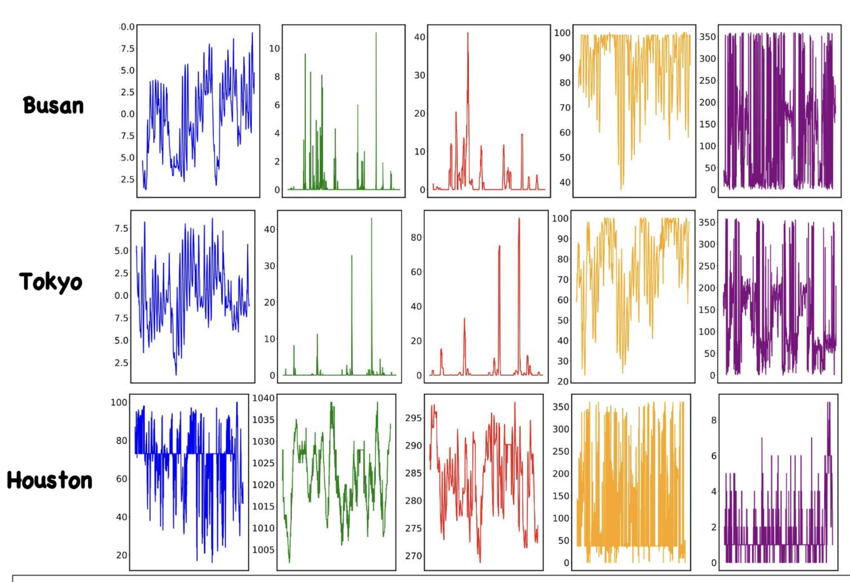
IJCAI JEJU 2024

INTRODUCTION

Background: On-device weather forecasting



Main Challenges:
(1) Non.iid affected by geographic location



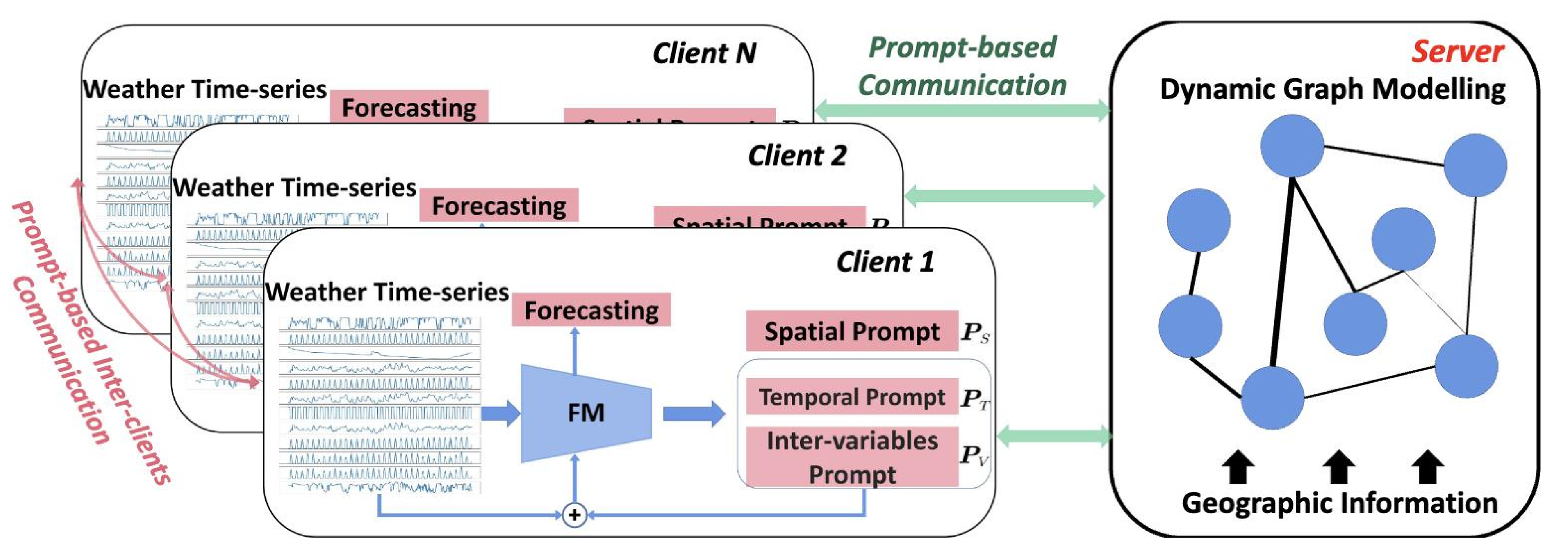
Weather data from different regions show significant non.iid. Additionally, devices on seashores and hilltops may collect different data even if they are geographically close.

(2) Communication Overhead

Low-resource weather edge devices cannot support large-scale parameter communications.

Main Contributions:
(1) FedPoD, an efficient FL framework for real-world on-device weather forecasting.
(2) Adaptive Prompt Tuning, to represent information and guide prediction through multi-level communication and knowledge sharing.
(3) Dynamic Graph Modeling, to enhance personalization by optimizing collaboration among clients with similar representations

MAIN ARCHITECTURE



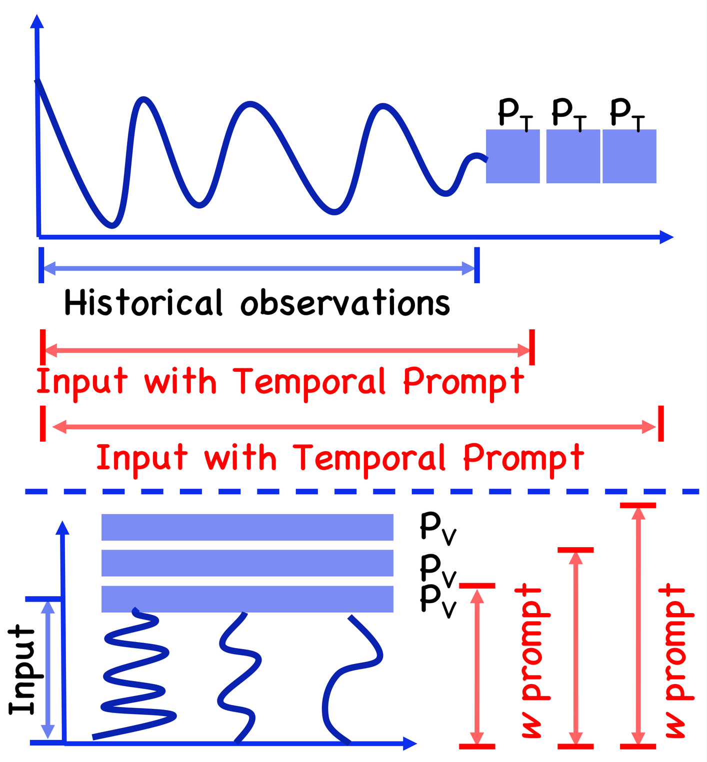
Architecture of FedPoD. Prompts comprise the **Spatial Prompt (P_S)**, **Temporal Prompt (P_T)**, and **Inter-variable Prompt (P_V)**, <-->: exchanges prompts among clients, <-->: exchange only prompts during client-server communication.

LOCAL PROMPTS UPDATING

Temporal Prompt & Inter-variable Prompt Updating

Algorithm 1 Implementation of P_T and P_V Updating

Initialize Original input series X_{ipt}, frozen PFM F_M, Temporal/Variable updating steps K_t and K_s.
for time forecasting step $q = 1, 2, \ldots$ **do**
  $\texttt{Updating}(F_M(\|X_{\text{ipt}}, P_T\|^T)), P_T \in \mathbb{R}^{q \cdot K_t \times n}$
    ▷ $\|.\|^T$: *concat* along temporal dimension
  $P_T \leftarrow \|P_T, P'_T \in \mathbb{R}^{K_t \times n}\|^T$
    ▷ P'_T: Next temporal prompt block
end for
for variable forecasting step $p = 1, 2, \ldots$ **do**
  $\texttt{Updating}(F_M(\|X_{\text{ipt}}, P_V\|^V)), P_V \in \mathbb{R}^{m \times p \cdot K_v}$
    ▷ $\|.\|^V$: *concat* along variable dimension
  $P_V \leftarrow \|P_V, P'_V \in \mathbb{R}^{m \times K_v}\|^V$
    ▷ P'_V: Next inter-variable prompt block
end for



Upper: Temporal Prompt
Bottom: Inter-variable Prompt

Spatial Prompt Updating [X_{geo}: geographic information]

$$P_S, X \leftarrow \texttt{LayerNorm}(\|X_{\text{ipt}}, X_{\text{geo}}\|, \|X, P_S\|)$$

$$\overline{X} = \texttt{FFN}(F(X_{\text{ipt}} + X))$$

→ **Final forecasting results**

DYNAMIC GRAPH MODELING

Construct Prompt Groups:
(1) Temporal & Inter-variable Prompts [P_T, P_V]
(2) Spatial Prompts [P_S]
(3) Full Prompts [P_T, P_V, P_S]

→ Graph Attention Network for dynamic graph modeling.

Haversine Formula: $D\tan^{-1}\left(\sqrt{\frac{\sin^2(\frac{\Delta\phi}{2}) + \cos(\phi_i)\cdot\cos(\phi_j)\cdot\sin^2(\frac{\Delta\lambda}{2})}{1-(\sin^2(\frac{\Delta\phi}{2}) + \cos(\phi_i)\cdot\cos(\phi_j)\cdot\sin^2(\frac{\Delta\lambda}{2}))}}\right)$ → A_{Geo}

Adjacent Matrix: $A_{i,j} = \frac{e_{i,j}}{1+e^{-W[W_iP_i - W_jP_j]}}$ → A_S, A, A_{TV}

Correlation Reconstruction:

$$A' \leftarrow \texttt{Softmax}\left(\frac{(A_{Geo} - A_S)A_{TV}^\top}{\sqrt{d_k}}\right)A,$$

$$\{P_i\}_{i=1}^{l,N} \leftarrow \alpha A\{P_i\}_{i=1}^N + (1-\alpha)A'\{P_i\}_{i=1}^N.$$

$[A_{\text{Geo}} - A_{\text{S}}]$ **(1)** Highlights the discrepancy between the actual geographic correlation and the encoded spatial correlation.
(2) Dynamically adjusts the spatio-temporal correlation among devices to achieve precise potential correlation graph modeling.

OPTIMIZATION OBJECTIVE

$$\mathcal{L}_{ap} = \texttt{MSE}(y', y) + \mathcal{R}(\{P_i\}; \{P_j\}^l; \{P_i\}^l; \{P\}^*)$$ **Prompt-based Regularization Term**

$$\mathcal{L}_{ap} = \texttt{MSE}(y', y) + \frac{1}{\xi^2}L^2(\{P_i\}, \{P\}^*) + \frac{1}{\xi^2}L^2(\{P_i\}, \{P_i\}^l) + \frac{1}{\tau^2}\cdot\frac{1}{(|\mathcal{N}|/S_G)-1}\sum_{j\in\mathcal{N}}L^2(\{P_i\}, \{P_j\}^l) + 4\{\log_2(\xi) + \log_2(\tau)\}.$$

Expanding from a multi-task learning perspective [Kendall et al. 2018]
P_i :Prompts from i-th client
P_j :Prompts from j-th client
P^* :Global prompts

Global Optimization Objective

$$\arg\min_{\{P_i\};A} \sum_{i=1}^N [\frac{n_i}{n}F_i(\{P_i\}; D_i) + \mathcal{R}(\{P_i\}; \{P_j\}^l; \{P_i\}^l; \{P\}^*)] + \tau\mathcal{G}(A),$$

$$s.t. \quad \{P\}^* \in \arg\min_{\{P_1\},\ldots,\{P_N\}} \sum_{i=1}^N \frac{n_i}{n}F_i(\{P_i\}),$$

$$\{P\}^l \in \arg\min_{\{P_i\}^l} \sum_{j\in\mathcal{N}} A_{j,i}S(\{P_i\}^l, \{P_j\}^l)$$

G(A) is a graph-based constraint that ensures each client aggregates with similar neighboring node, Adjacent Matrix from Dynamic Graph Modeling

MAIN RESULTS

Fine-Tuning Strategy	Method	AvePRE Task1	AvePRE Task2	SurTEMP Task1	SurTEMP Task2	SurUPS Task1	SurUPS Task2
Conventional Fine-tuning	FedAvg [McMahan *et al.*, 2017]	34.6/44.8	56.0/90.1	47.6/64.4	56.5/78.3	53.5/74.2	54.1/74.6
	FedProx [Li *et al.*, 2020]	31.7/42.1	54.4/87.2	44.4/62.7	52.9/76.4	51.2/69.5	52.3/72.4
	Per-FedAvg [Fallah *et al.*, 2020]	30.9/40.7	54.3/71.5	41.4/60.9	51.8/73.3	50.2/69.7	51.7/71.8
	APFL [Deng *et al.*, 2020]	32.5/43.8	56.1/84.9	46.2/63.1	59.4/77.3	54.3/73.7	53.8/73.4
	FedAMP [Huang *et al.*, 2021]	31.9/41.3	54.7/84.2	43.8/62.9	52.3/73.7	51.5/70.0	53.2/73.4
	FedATT [Jiang *et al.*, 2020]	34.5/44.7	63.2/89.8	48.7/63.1	61.0/79.4	58.8/73.6	64.6/82/0
	pFedMe [T Dinh *et al.*, 2020]	32.2/42.7	64.0/85.2	42.9/61.8	50.7/74.6	51.7/70.1	52.5/72.0
	SFL [Chen *et al.*, 2022]	30.0/40.2	53.1/81.2	39.9/62.6	51.7/76.1	48.0/69.1	51.0/70.4
Adaptive Prompt Tuning (Ours)	FedAvg [McMahan *et al.*, 2017]	32.4/42.8	51.0/76.3	41.2/61.7	54.4/76.8	52.1/72.2	53.2/73.8
	FedProx [Li *et al.*, 2020]	27.1/38.0	47.1/70.2	39.7/61.5	51.7/75.2	48.1/67.1	51.0/67.6
	Per-FedAvg [Fallah *et al.*, 2020]	29.3/37.9	45.3/67.4	37.8/60.0	51.3/72.2	47.6/68.2	50.1/69.5
	APFL [Deng *et al.*, 2020]	29.5/38.7	46.0/67.7	38.6/64.2	55.7/75.7	56.2/67.1	59.7/68.2
	FedAMP [Huang *et al.*, 2021]	27.1/37.4	46.7/69.7	39.2/61.0	51.2/73.1	51.5/67.9	52.1/69.3
	FedATT [Jiang *et al.*, 2020]	30.5/40.8	58.7/79.7	38.4/63.7	52.4/79.1	50.9/70.0	53.5/72.6
	pFedMe [T Dinh *et al.*, 2020]	28.2/39.7	47.5/69.9	38.5/61.4	**50.5**/74.1	48.4/66.9	51.2/68.8
	SFL [Chen *et al.*, 2022]	31.1/39.2	46.4/68.8	37.6/59.3	54.2/73.7	47.2/66.0	49.8/67.2
	FedPoD (Ours)	**23.7/32.9**	**44.3/65.5**	**35.7/55.0**	51.4/**71.2**	**43.9/62.5**	**45.2/63.9**
Other Prompt Tuning	PromptFL [Guo *et al.*, 2023]	33.8/42.7	49.2/70.0	44.1/63.2	59.7/78.9	51.1/73.7	58.2/69.2
	MetePFL [Chen *et al.*, 2023b]	29.9/37.2	46.1/68.0	40.1/58.6	51.3/73.0	48.4/67.7	52.4/67.6

Main results with different fine-tuning strategy (MAE/RMSE report). Conventional fine-tuning: update local FM with a FFN head. **Other Prompt Tuning:** add parameters to input to update local FM. AvePRE/SurTEMP/SurUPS: weather forecasting dataset from NASA.

Variant	P_V	P_T	W_{bv}	W_{bt}	P_S	Federated Aggregation Strategy	Local Loss	Task 1	Task 2
FedPoD-A	*w/o*	*w*	-	-	*w*	$\{P_i\}_{i=1}^{l,N} \leftarrow A_{\text{T}}\{P_i\}_{i=1}^N + (1-\alpha)A_{\text{S}}\{P_i\}_{i=1}^N$	Ours MSE	29.9/40.4 31.7/42.4	53.7/78.4 54.4/80.0
FedPoD-B	*w*	*w/o*	-	-	*w*	$\{P_i\}_{i=1}^{l,N} \leftarrow A_{\text{S}}\{P_i\}_{i=1}^N + (1-\alpha)A_{\text{V}}\{P_i\}_{i=1}^N$	Ours MSE	28.2/37.2 29.2/39.0	57.1/85.0 58.2/85.9
FedPoD-C	*w/o*	*w/o*	-	-	*w*	$\{P_i\}_{i=1}^{l,N} \leftarrow A_{\text{S}}\{P_i\}_{i=1}^N$	Ours MSE	30.8/41.2 31.8/42.4	52.0/77.7 54.8/78.9
FedPoD-D	*w*	*w*	*w*	*w*	*w/o*	$\{P_i\}_{i=1}^{l,N} \leftarrow A_{\text{TV}}\{P_i\}_{i=1}^N$	Ours MSE	30.1/40.9 31.6/42.1	48.7/74.7 50.9/76.0
FedPoD-D	*w*	*w/o*	-	-	*w/o*	$\{P_i\}_{i=1}^{l,N} \leftarrow A_{\text{V}}\{P_i\}_{i=1}^N$	Ours MSE	29.4/39.8 31.1/40.8	56.2/84.7 59.0/87.8
FedPoD-E	*w/o*	*w*	-	-	*w/o*	$\{P_i\}_{i=1}^{l,N} \leftarrow A_{\text{T}}\{P_i\}_{i=1}^N$	Ours MSE	30.1/40.6 31.7/43.5	53.7/79.0 54.2/80.5
FedPoD (Ori.)	*w*	*w*	*w*	*w*	*w*	$\{P_i\}_{i=1}^{l,N} \leftarrow \alpha A\{P_i\}_{i=1}^N + (1-\alpha)A'\{P_i\}_{i=1}^N$	Ours MSE	**23.7/32.9** **25.0/34.4**	**44.3/65.5** **47.7/68.0**

Ablation results (MAE/RMSE report). For (1) Local Adaptive Prompts; (2) Local Optimization.

Term 1	Term 2	Term 3	Task 1	Task 2
w	*wo*	*wo*	29.1/36.9	47.1/70.1
w	*wo*	*w*	**27.3/36.3**	**46.0/69.9**
w	*w*	*wo*	29.1/34.3	46.6/72.5
wo	*w*	*w*	29.0/34.6	47.9/74.8
wo	*wo*	*w*	28.2/37.0	49.2/74.4

Ablation results about the optimization objective.

Updating step of P_T	Updating step of P_V	MAE	RMSE
1	1	39.9/**51.5**	50.2/**79.5**
2	2	38.1/53.7	48.8/85.0
3	3	37.1/52.9	47.9/80.4
4	4	38.6/53.2	47.7/80.1
6	6	**35.7**/52.6	**46.1**/80.7
12	12	39.3/53.7	50.3/84.8

Impact of Prompt updating steps.

Method/Dataset		FedPoD	FedPoD-DP	Ave. Variation
AvePRE	Task1	23.7/32.9	24.8/33.9	↓4.33%
	Task2	44.3/65.5	46.1/66.9	↓2.88%
SurTEMP	Task1	35.7/55.0	37.0/56.6	↓2.69%
	Task2	51.4/71.2	52.7/73.0	↓2.53%
SurUPS	Task1	43.9/62.5	45.1/63.7	↓2.33%
	Task2	45.2/63.0	46.4/65.2	↓2.34%

FedPoD can ensure privacy effectively. Differential privacy experiment results.

Step of subgraph S_G	MAE (Task 1/Task 2)	RMSE (Task 1/Task 2)
1	36.9/51.7	47.0/79.4
2	39.1/**51.5**	49.9/**79.0**
4	38.1/51.8	49.7/78.8
6	38.8/54.0	49.1/81.9
8	39.3/54.4	49.7/81.8
10	**35.9**/52.4	**45.6**/79.6

Impact of subgraph step in global aggregation (MAE/RMSE report).

PARAMETERS

Method	Trainable Param.	MAE/RMSE
Train from scratch (FedAvg)	5,284,173	40.3/51.2
Pre-trained FM (FedAvg)	215,089	33.5/44.5
Pre-trained FM & Prompts (FedAvg)	**159,649**	31.1/41.9
FedPoD (Ours)	**159,649**	**27.0/37.6**

Comparsion of training parameters.

(1) Only trains/transmits 3% param.
(2) Nearly 26% of param. can be saved relative to comparable fine-tuning methods.

RESOURCE

Federated Prompt Learning for Weather Foundation Models on Devices

Shengchao Chen, **Guodong Long**, **Tao Shen**, **Jing Jiang** and **Chengqi Zhang**
Australian Artificial Intelligence Institute, FEIT, University of Technology Sydney
shengchao.chen.uts@gmail.com, {guodong.long, tao.shen, jing.jiang, chengqi.zhang}@uts.edu.au

(This Paper)

Foundation Models for Weather and Climate Data Understanding: A Comprehensive Survey

Shengchao Chen, *Member, IEEE,* Guodong Long, Jing Jiang, Dikai Liu, *Senior Member, IEEE,* and Chengqi Zhang, *Senior Member, IEEE*

(Our Survey)

Email: shengchao.chen.uts@gmail.com WeChat: Pavel_Chen | First-year Ph.D. student at University of Technology Sydney