Interpretable weather forecasting for worldwide stations with a unified deep model

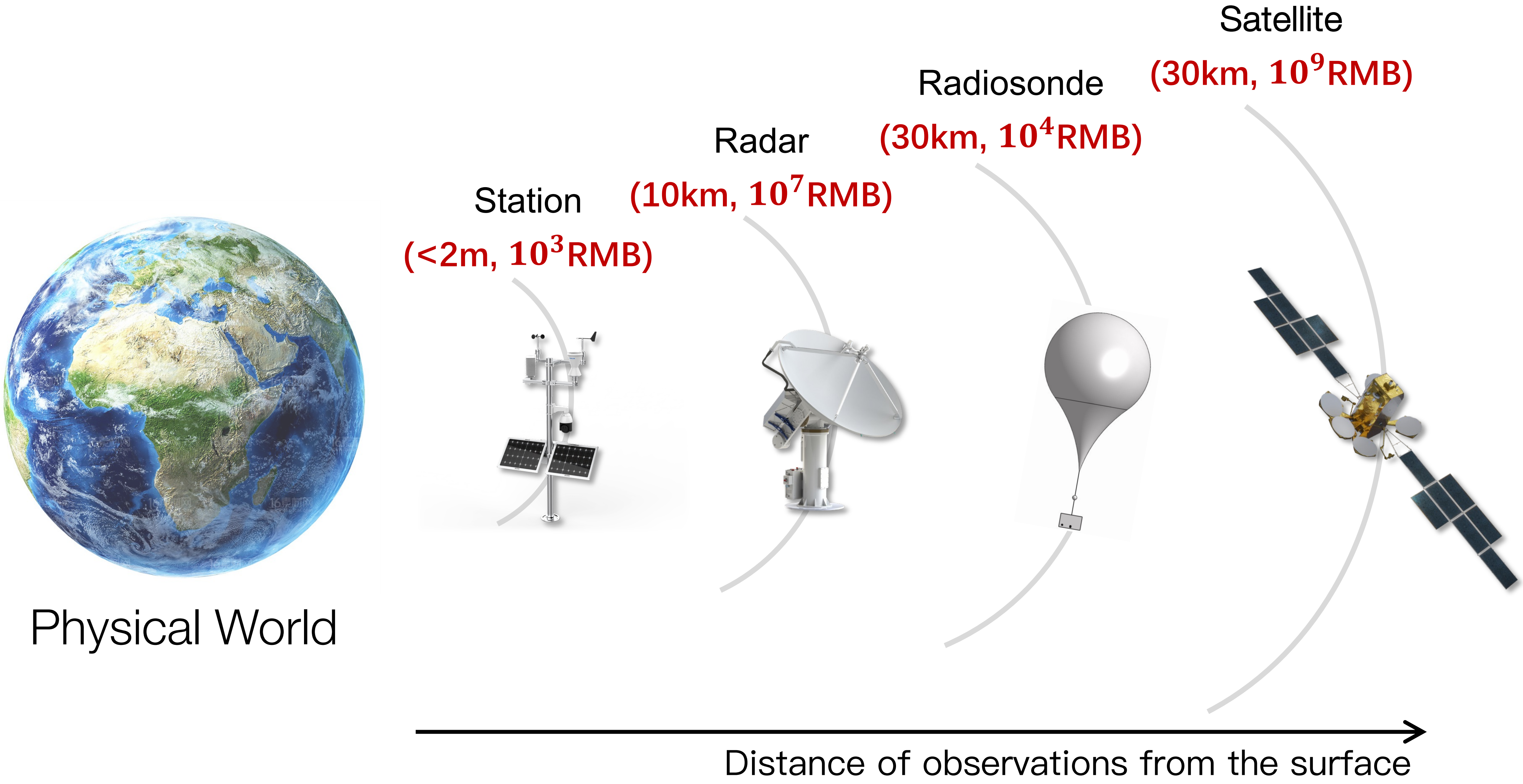
Haixu Wu, Hang Zhou, Mingsheng Long, Jianmin Wang

School of Software, Tsinghua University

Nature Machine Intelligence 2023

https://www.nature.com/articles/s42256-023-00667-9

How to Observe the Weather?

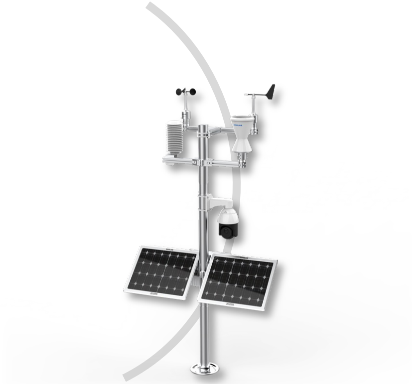


Automatic Weather Station

Physical World

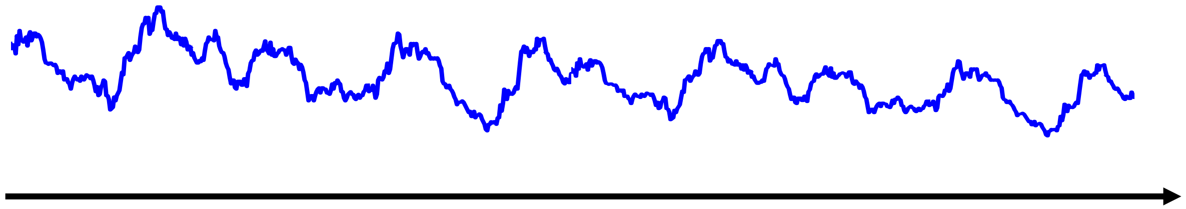
Station

($<$2m, 10^3RMB)



the most accurate and reliable

near-surface weather observation



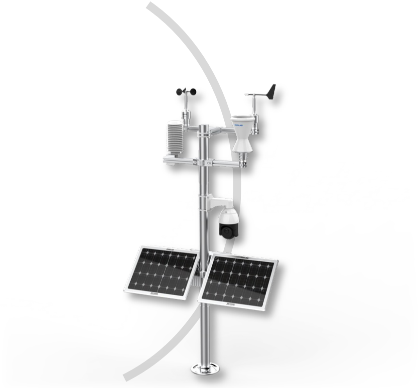
Time

Automatic Weather Station

Physical World

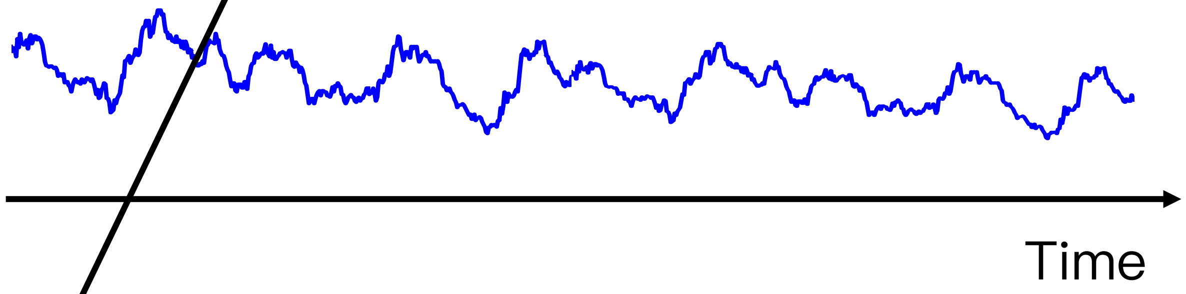
Station

(<2m, 10^3RMB)



the most accurate and reliable

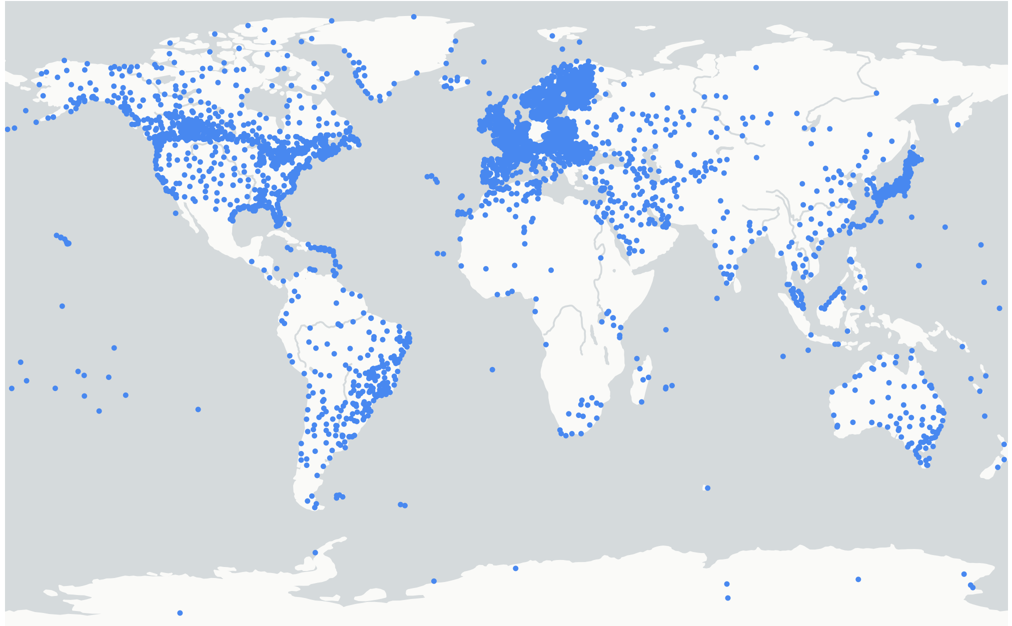
near-surface weather observation

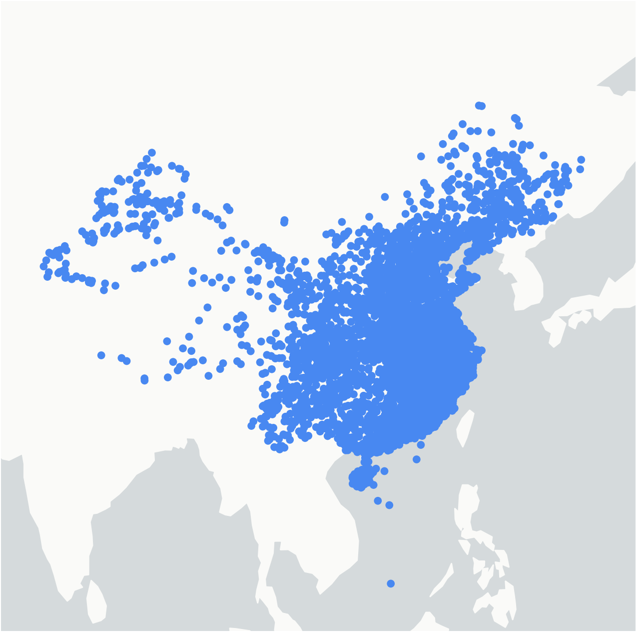


Directly Affect Human Life

Objective

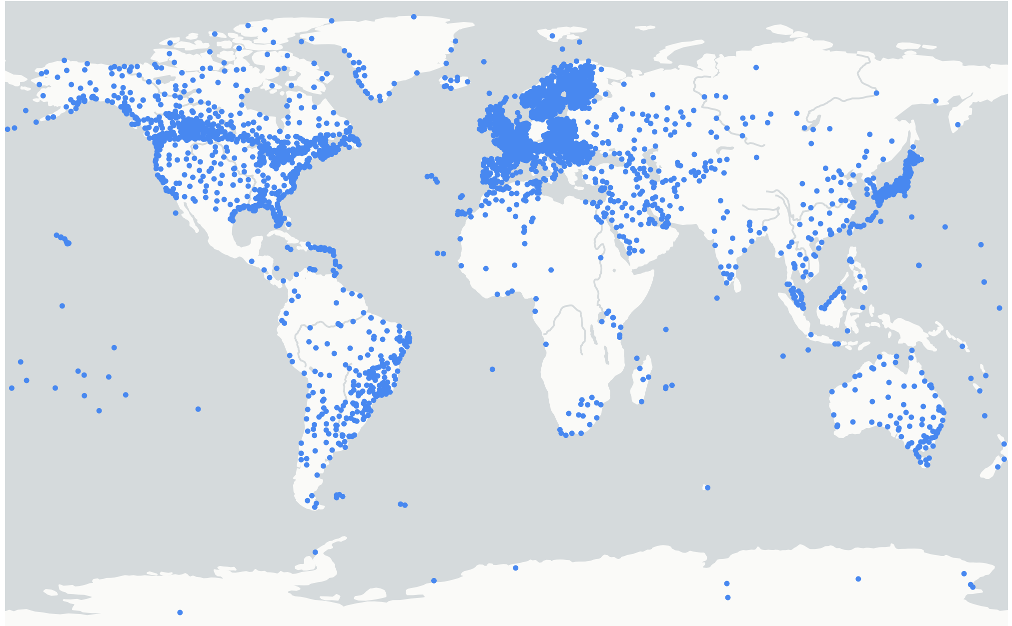
Tens of thousands of weather stations are **scattered** around the world and record the near-surface weather **every minute, even every second**.

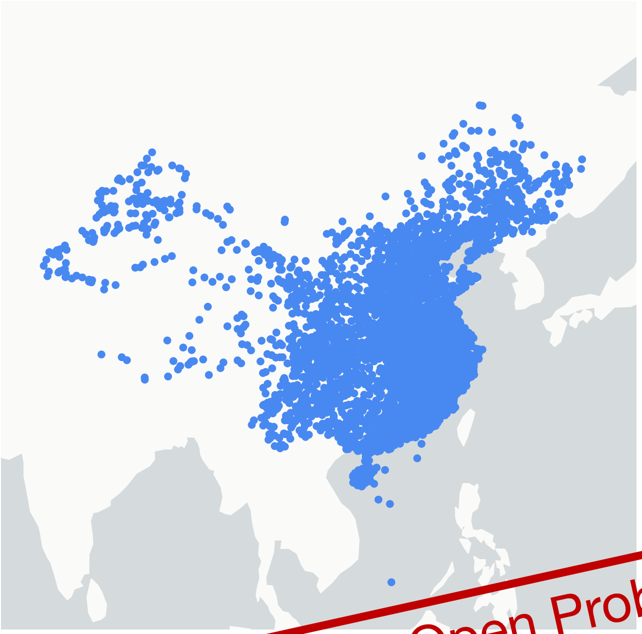




Objective

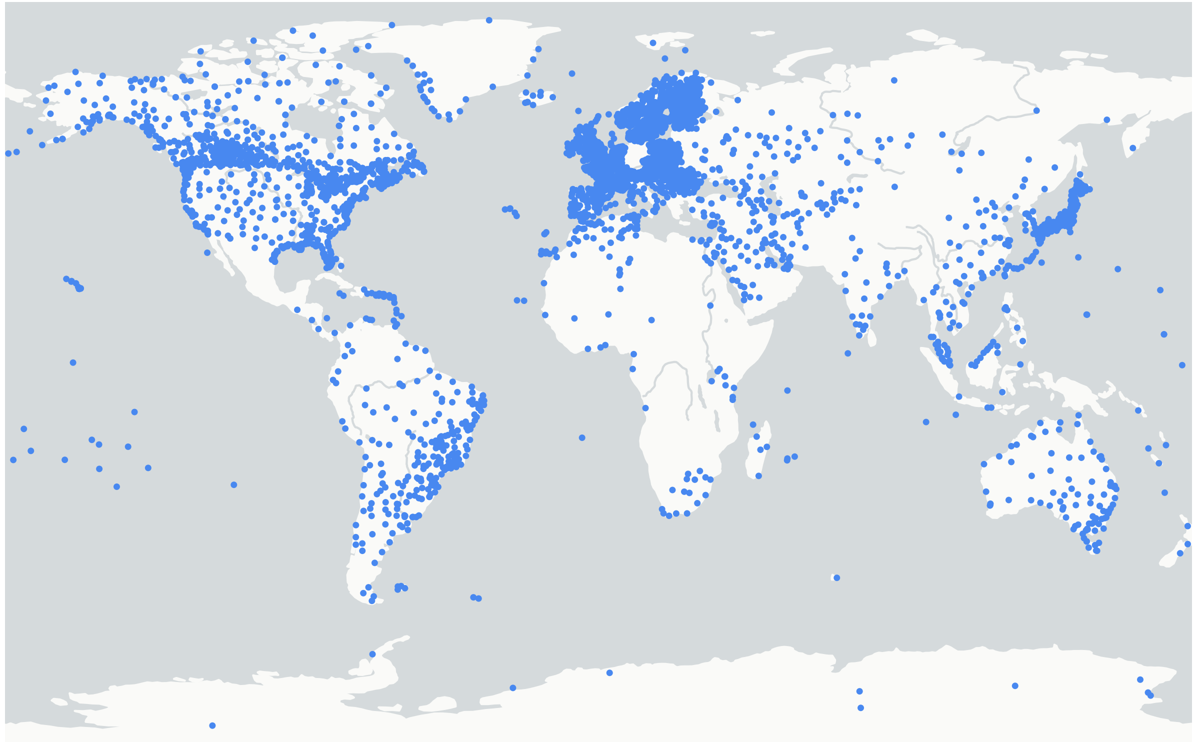
Real-time collaborative forecasts of **worldwide tens of thousands** automatic weather stations (prediction the future **0-24 hours** near-surface weather)



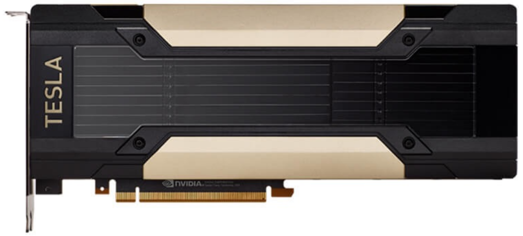


Worldwide Open Problem

Challenge 1: Huge Computation Cost



10,000x computation overhead in both GPU memory and running time



1000 GPU x 400 days

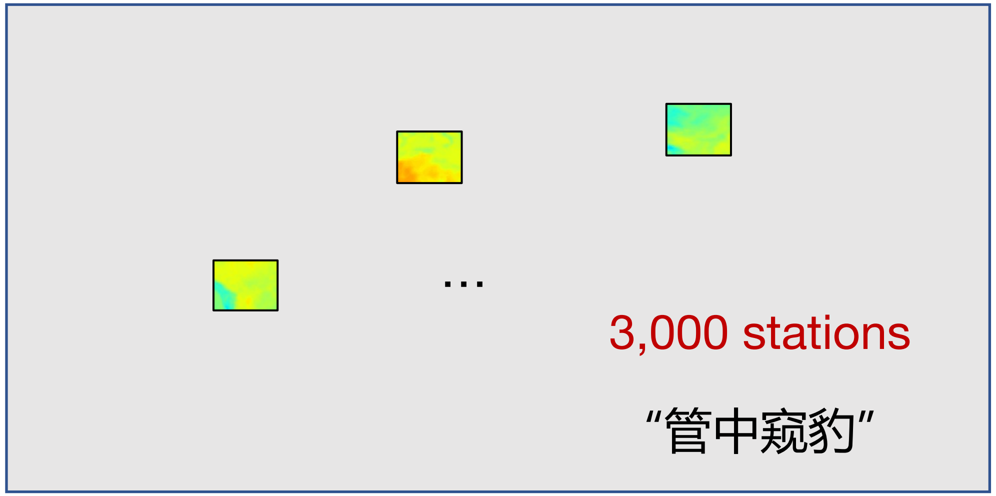
Challenge 2: Partially Observable System

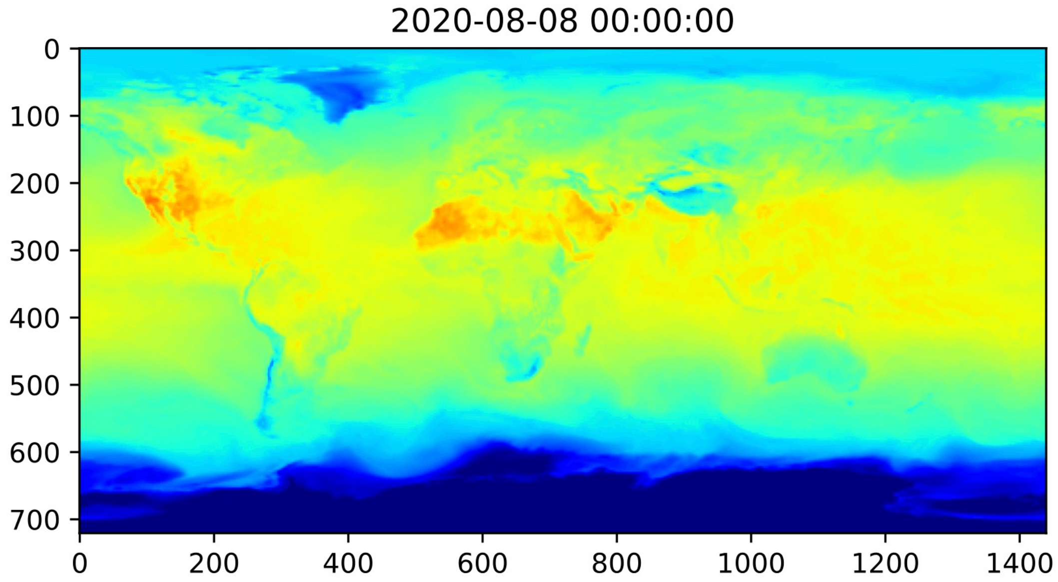
Uncovering the complex system from partial observations

Observations from

scattered weather stations

Global weather system

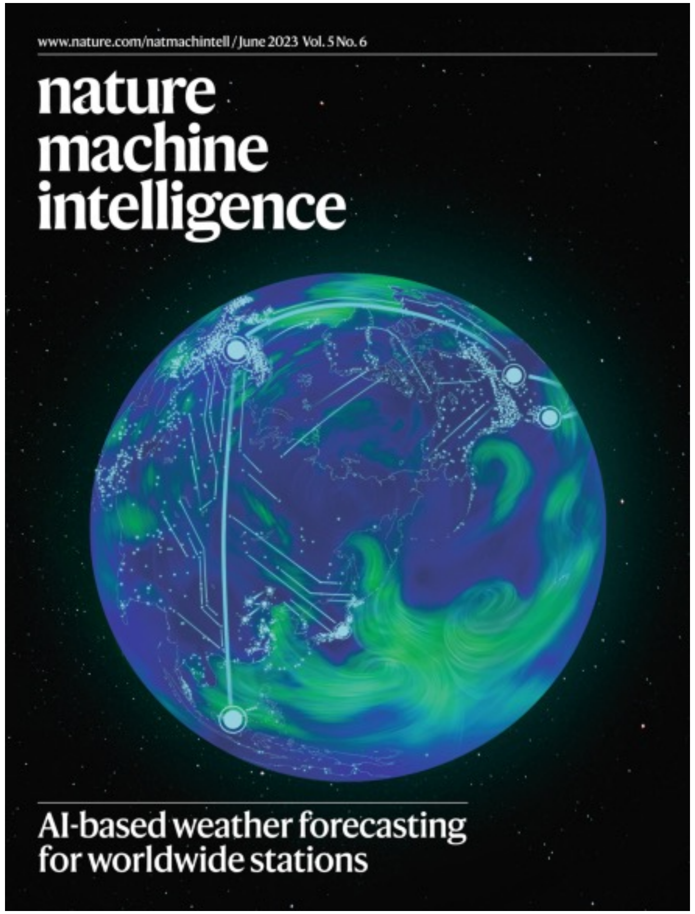




Challenge 3: Near-surface Forecasting

- Extremely complex topography
- Potential chaos effects

Corrformer



Large Meteorology Model for Near-surface Weather

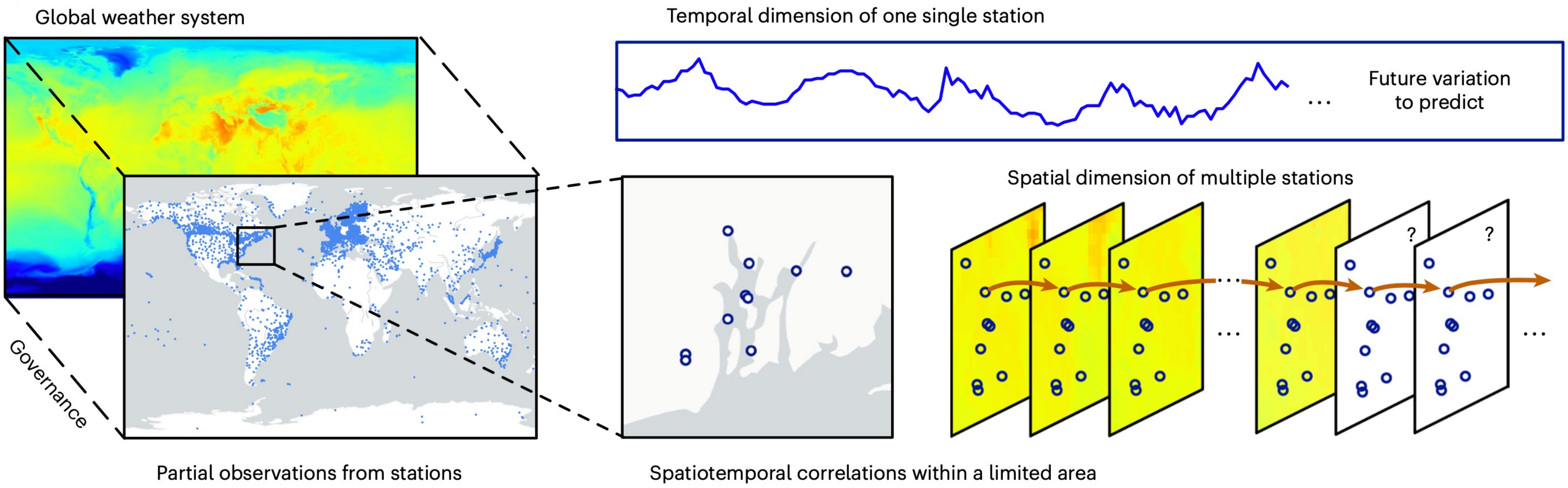
Global collaborative forecasts for the first time

- **Accurate:** Beat EC and GFS in near-surface forecast
- **Efficient:** 1 day for training, 1 second for inference
- **Interpretable:** AI insights for meteorological science

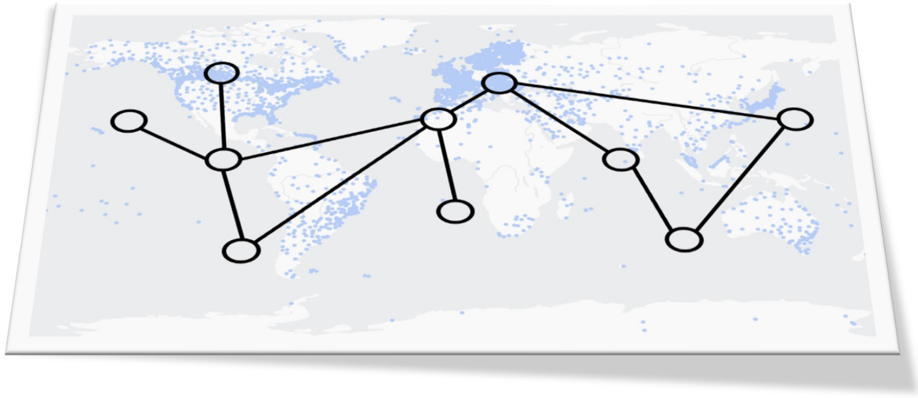
Wu, H. et al. Interpretable weather forecasting for worldwide stations with a unified deep model, ***Nature Machine Intelligence*** 5, 602–611 (2023)

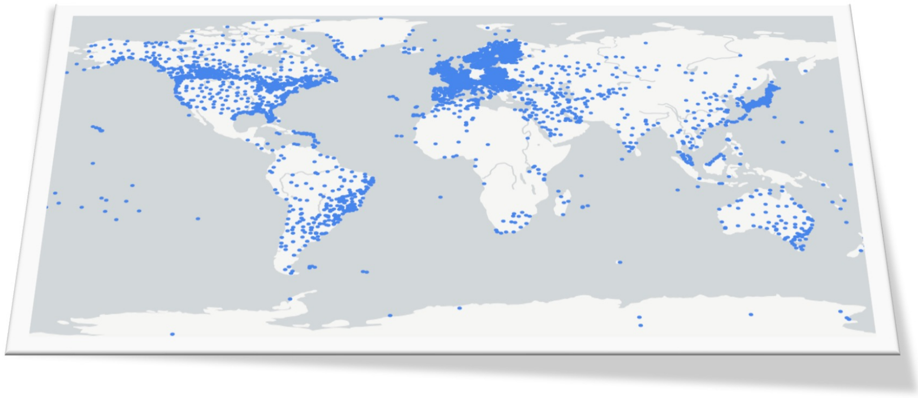
Problem Definition

Capture complex spatiotemporal correlations for worldwide weather stations.



Canonical Spatial Modeling





1. Convolution Neural Networks

- *Inapplicable to scattered stations*

2. Graph Neural Networks

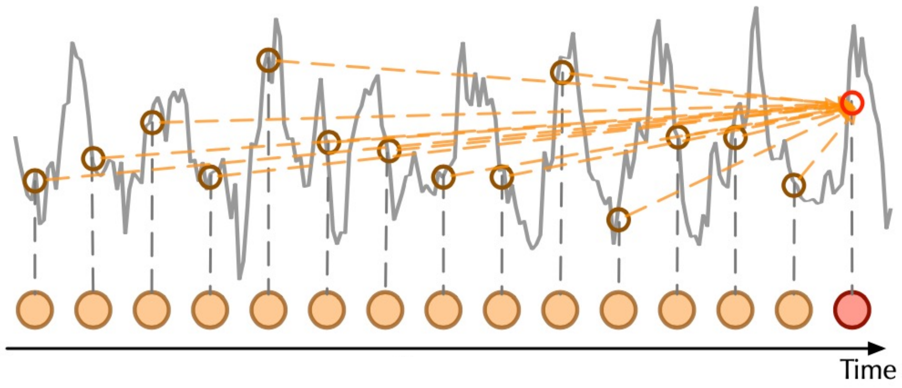
- Massive prior knowledge to construct graph
- Fixed graph vs. ever-changing weather

3. Attention-based models

- Quadratic complexity

How to organize these stations effectively?

Canonical Temporal Modeling



1. Recurrent Neural Networks

- Markov Assumption

2. Attention-based models

- Quadratic complexity
- Scattered steps modeling vs. Continuous weather processes

How to capture dependencies for continuous processes?

Auto-Correlation in Stochastic Process

- Auto-Correlation for wide-sense stationary stochastic process

$$\mathrm{R}_{XX}(\tau) = \mathrm{E}\left[X_{t+\tau}\overline{X}_{t}\right]$$

Time Delay Similarity -> Find Period

- Efficient computation with Fast Fourier Transform (FFT)

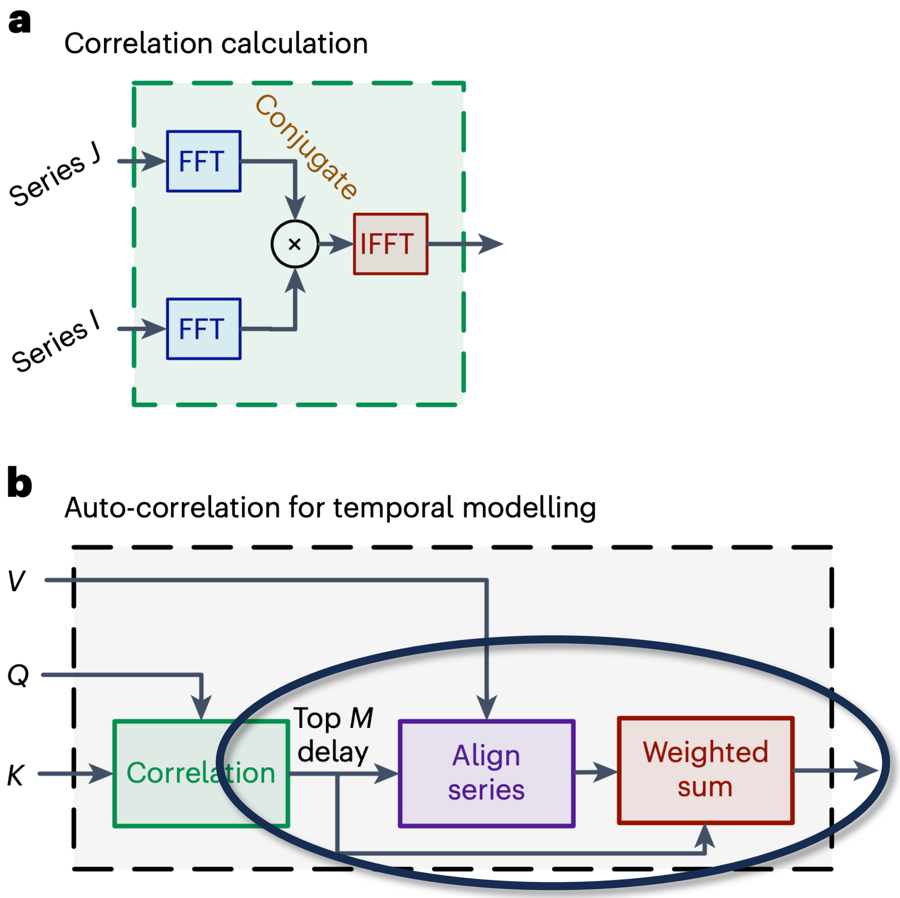
$$\begin{aligned} F_R(f) &= \mathrm{FFT}[X(t)] \\ S(f) &= F_R(f)F_R^*(f) \\ R(\tau) &= \mathrm{IFFT}[S(f)] \end{aligned}$$

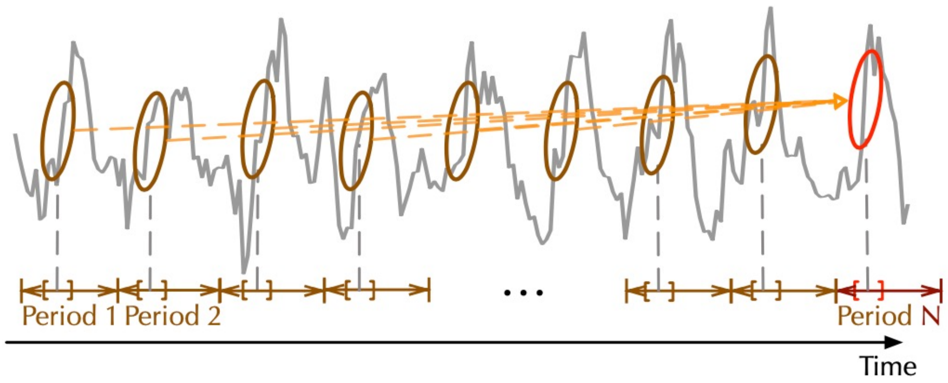
Wiener–Khinchin Theorem

Wu, H. et al. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting, NeurIPS 2021

PaperDigest Most Influential Paper, Citation 600+

Auto-Correlation for Series-wise Dependencies Modeling





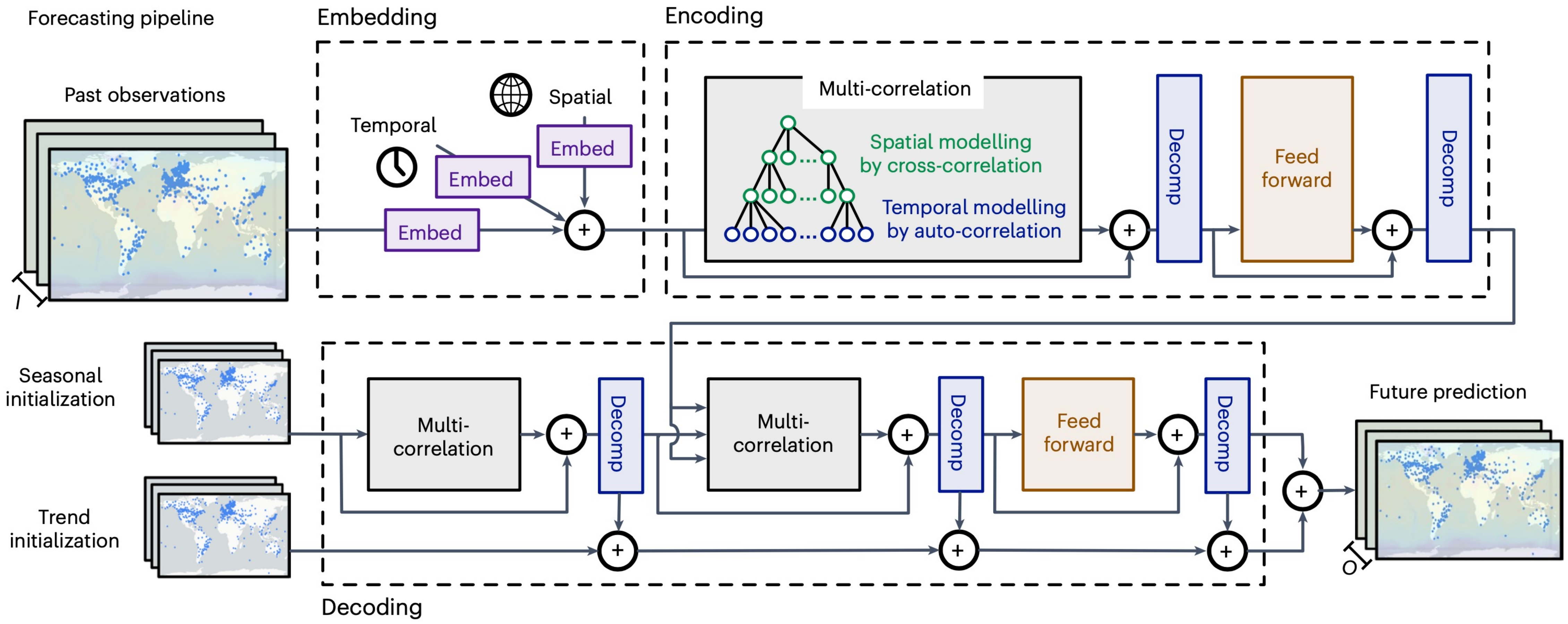
Auto-Correlation Mechanism

- Series-wise dependencies modeling
- Efficiency computation with complexity $O(L \log L)$

Wu, H. et al. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting, NeurIPS 2021

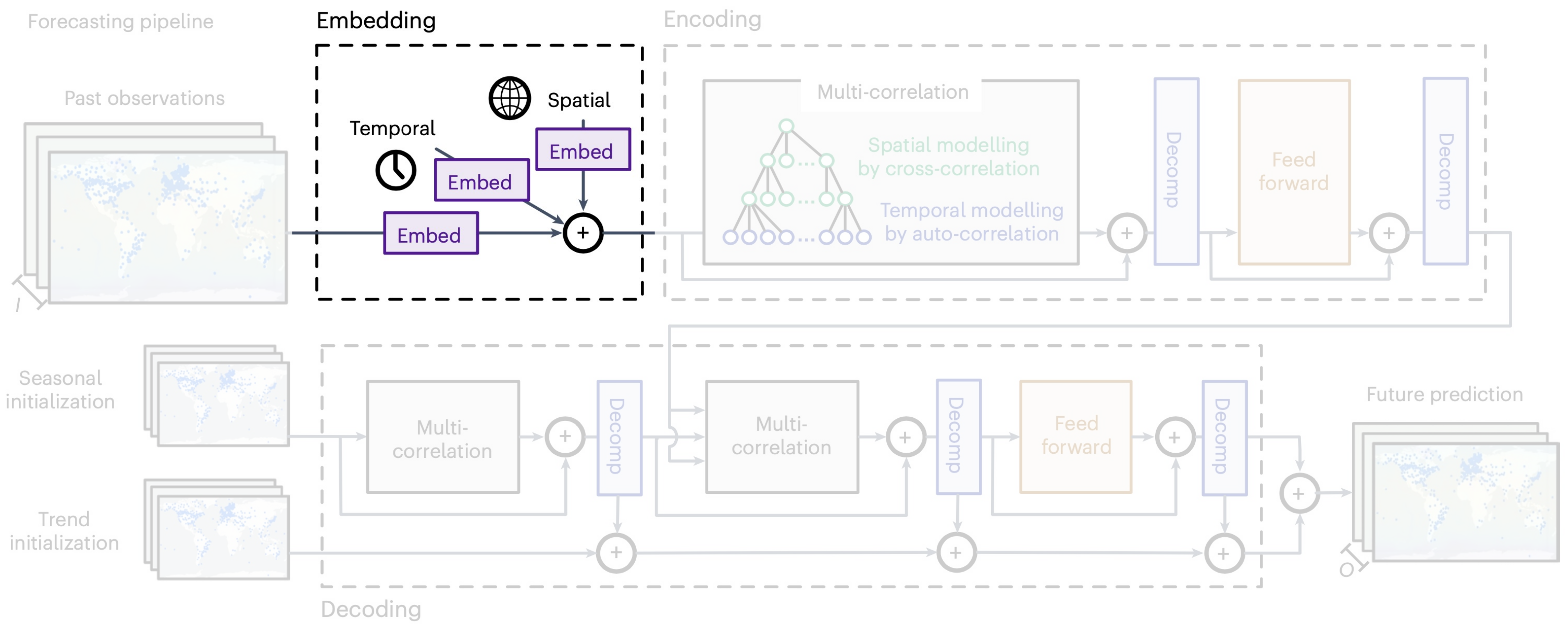
PaperDigest Most Influential Paper, Citation 600+

Corrformer



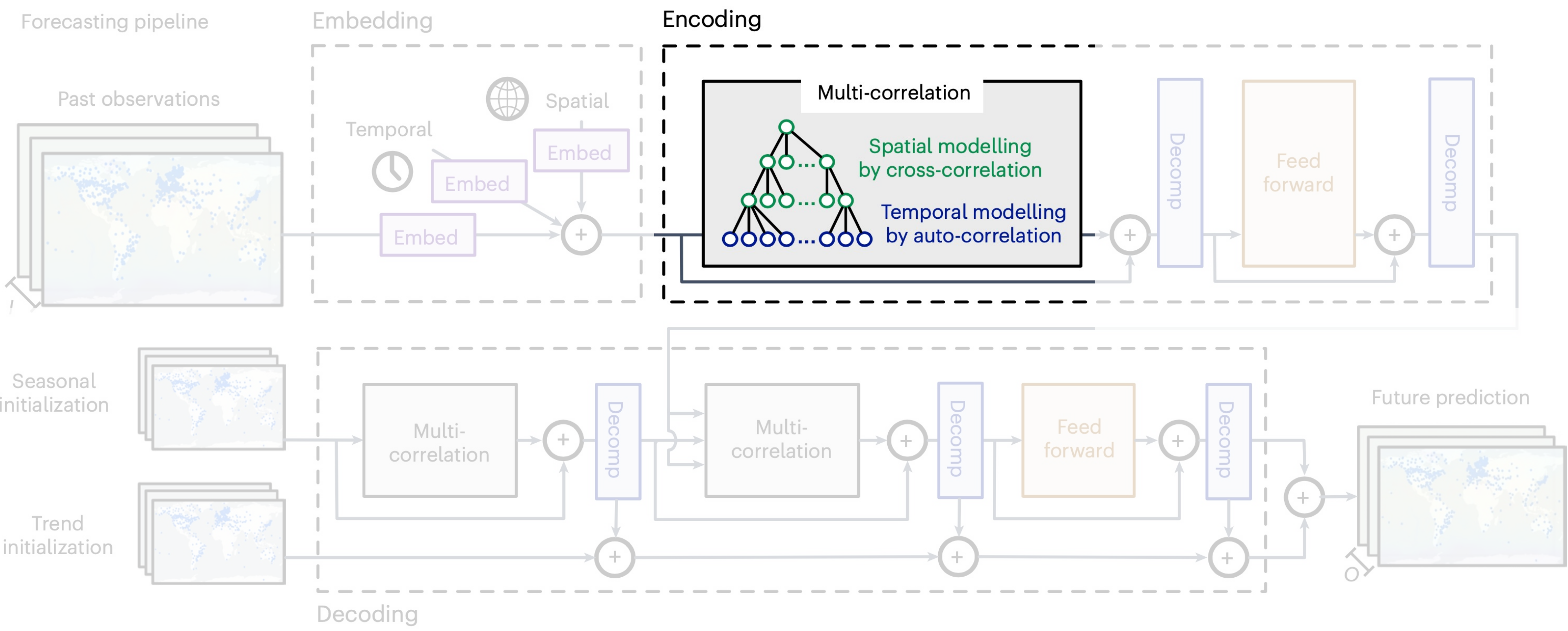
Forecasting pipeline
Embedding
Encoding
Past observations
Spatial
Temporal
Embed
Embed
Embed
Multi-correlation
Spatial modelling
by cross-correlation
Temporal modelling
by auto-correlation
Decomp
Feed
forward
Decomp
I
Seasonal
initialization
Trend
initialization
Multi-
correlation
Decomp
Multi-
correlation
Decomp
Feed
forward
Decomp
Future prediction
O
Decoding

Corrformer



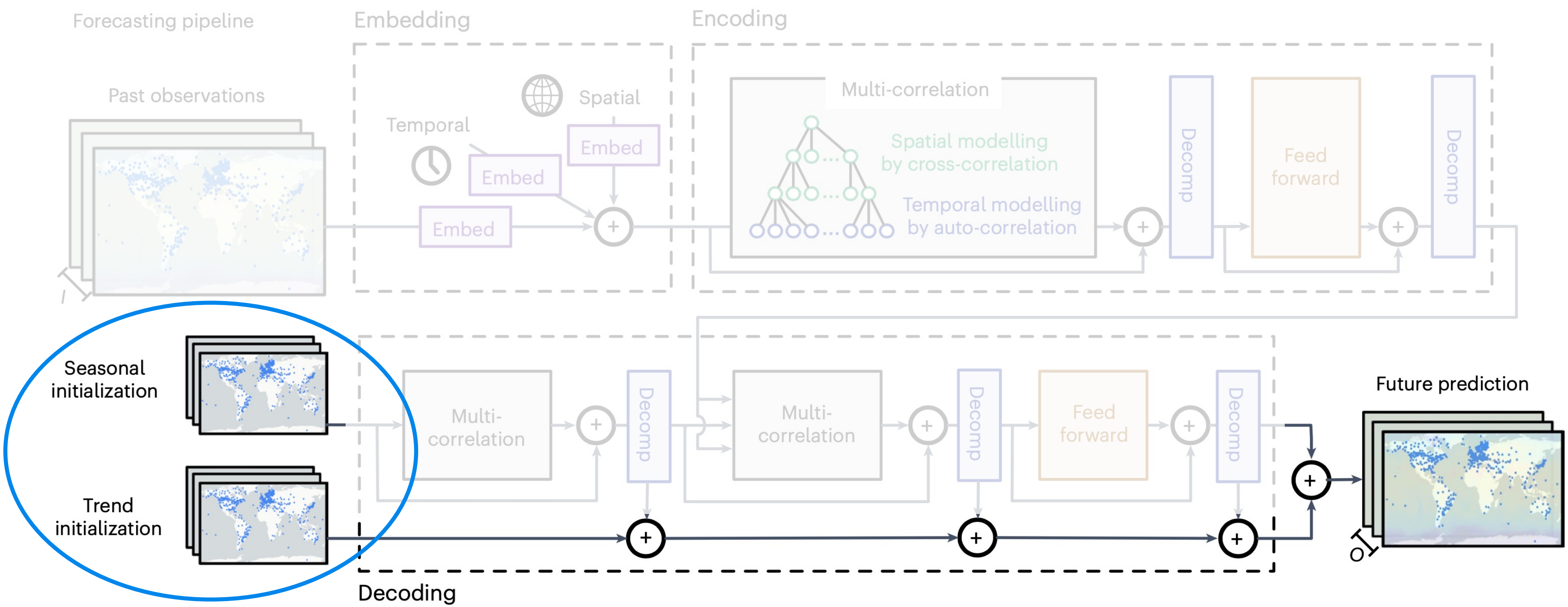
1. Embedding to incorporate **temporal, geographic and topography** information

Corrformer



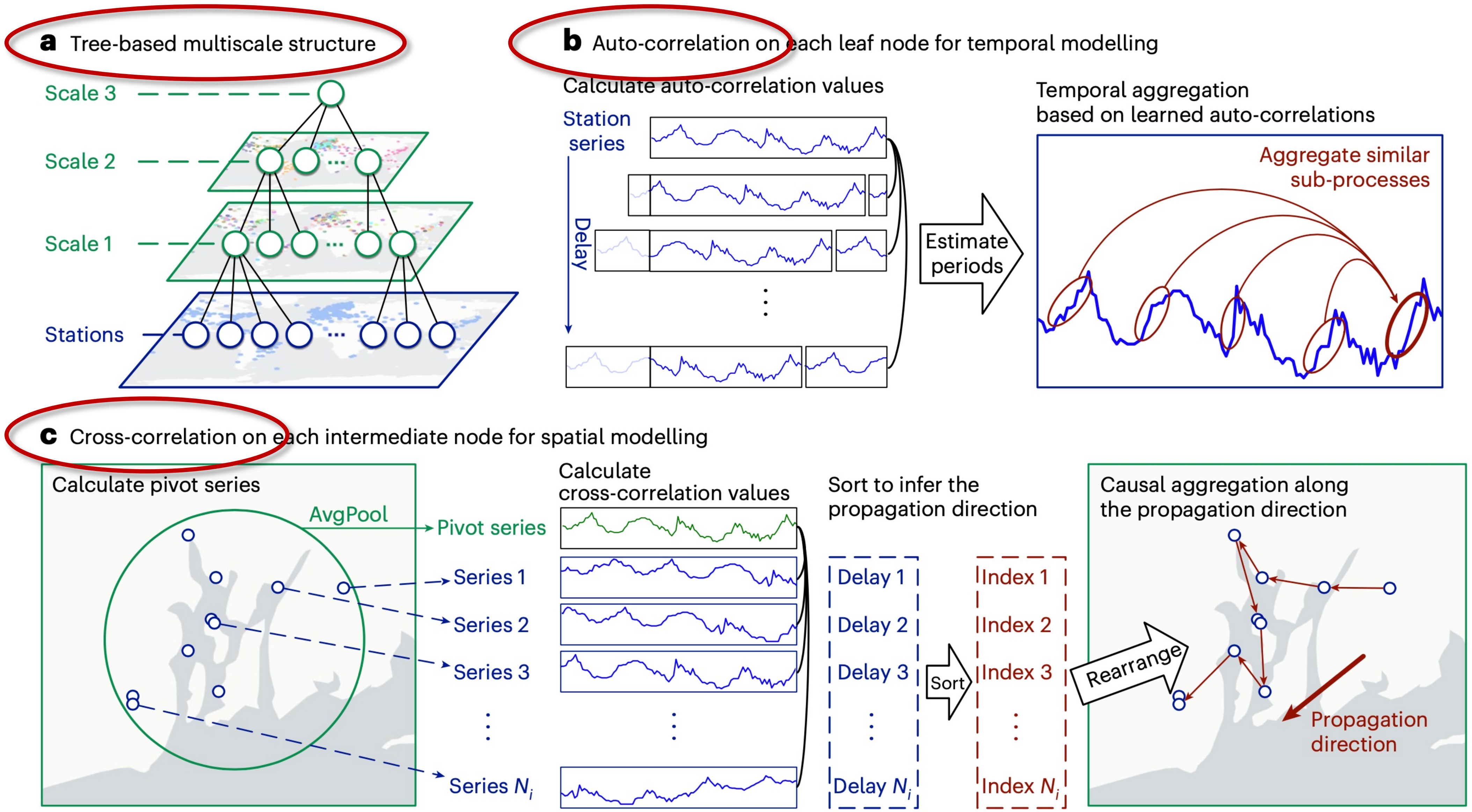
2. **Multi-correlation** to capture complex spatiotemporal correlations

Corrformer



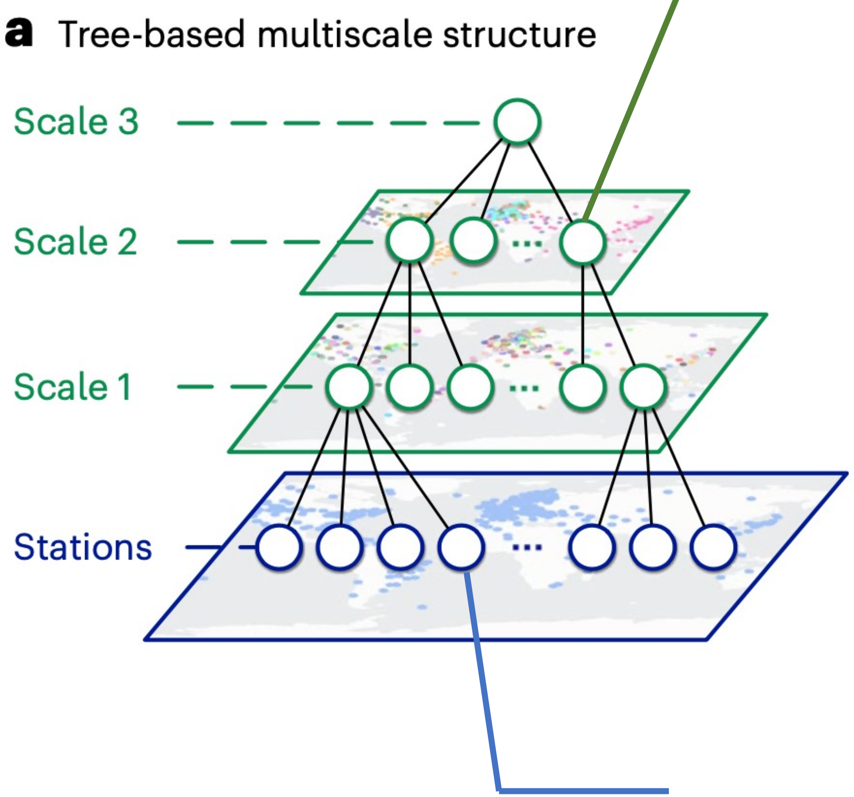
3. Deep decomposition architecture from Autoformer to **enhance stationarity**

Multi-correlation



Multi-correlation

Intermediate node: summed from child nodes for comprehensive meteorological characteristics of a region

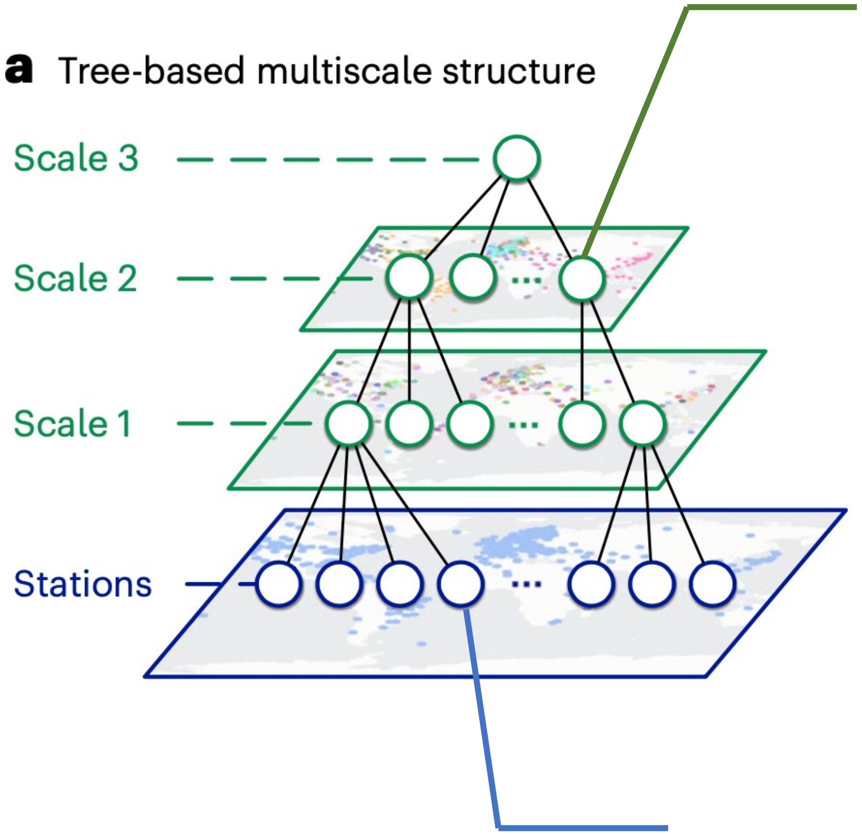


Leaf node: record the real observations

Multi-correlation

Intermediate node: summed from child nodes for comprehensive meteorological characteristics of a region

a Tree-based multiscale structure

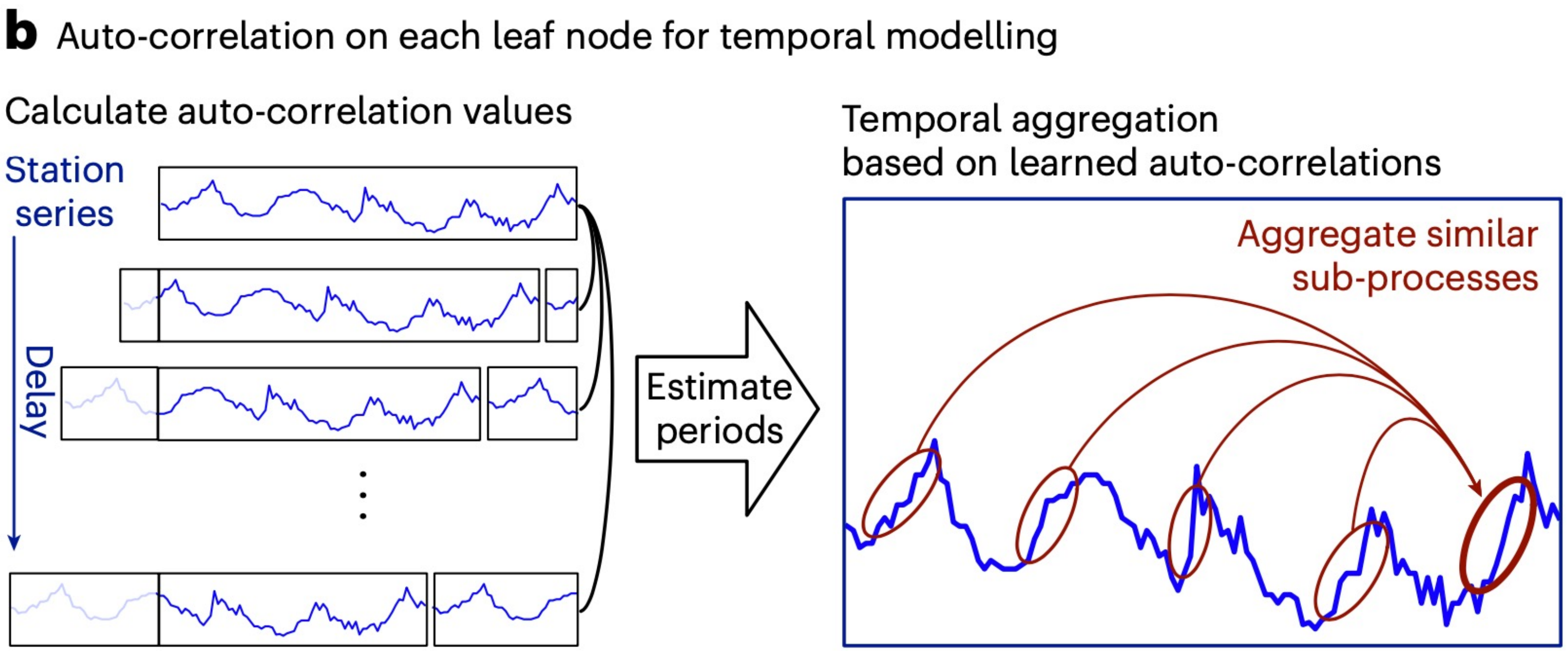


Benefits of tree structure

- More flexible for scattered stations
- Inherently model the multi-scale property of weather
- **Provide an elegant framework for spatiotemporal correlation modeling**
- **Significantly reduce computing overhead**

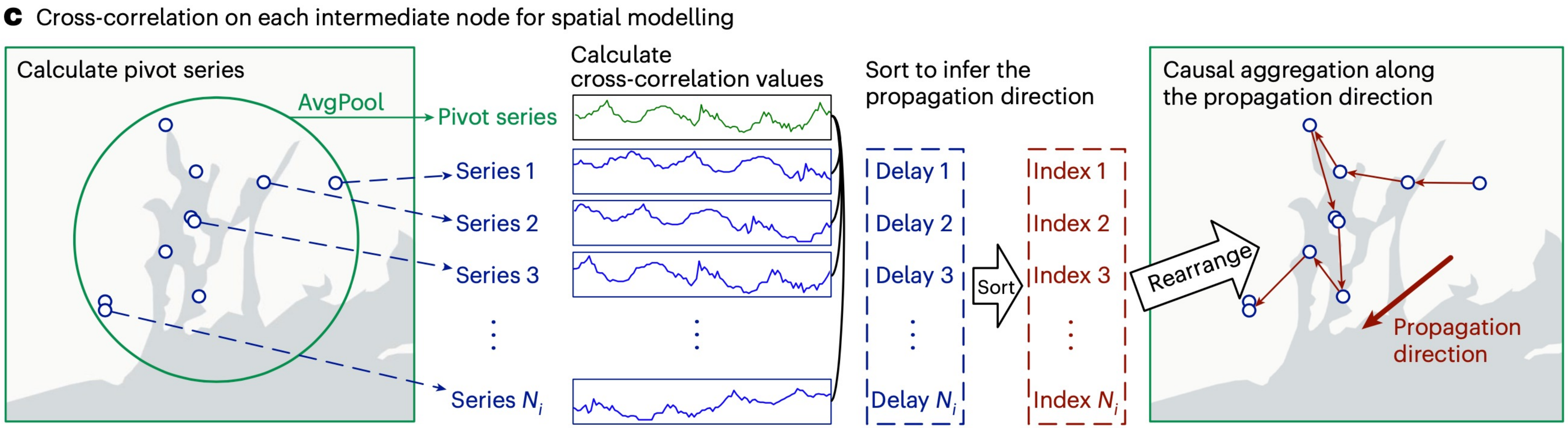
Leaf node: record the real observations

Multi-correlation



For the leaf node, calculate Auto-correlation

Multi-correlation



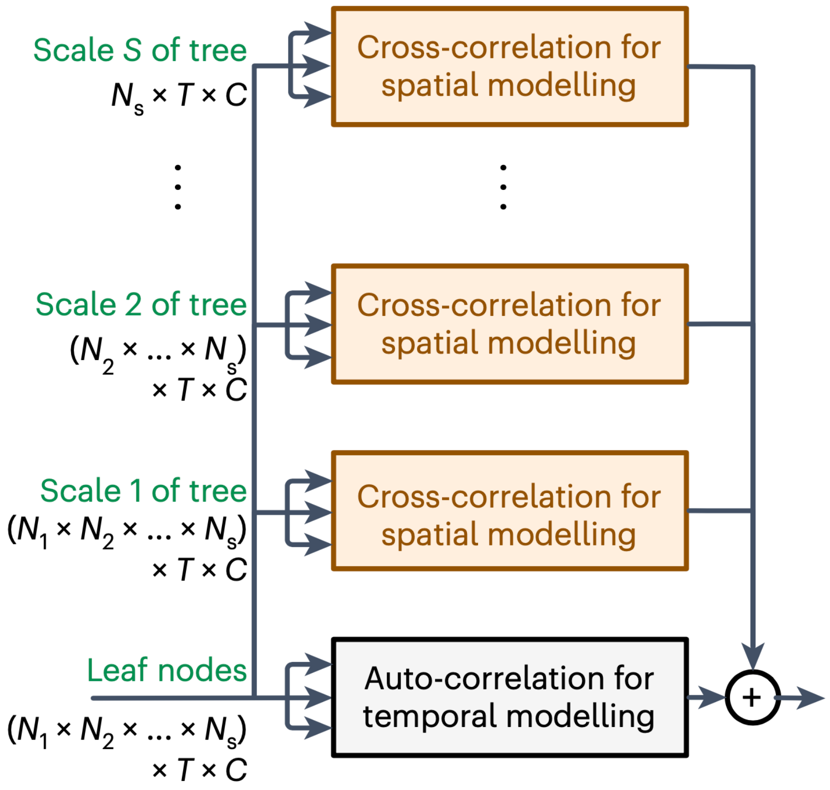
For each intermediate node, calculate Cross-correlation

1. Calculate the temporal delay **between intermediate node and its child nodes**
2. **Sort relative delay values** to infer the propagation direction.
3. **Causal aggregation** along the propagation direction.

Interpretable Evidence

Multi-correlation

Multi-correlation mechanism



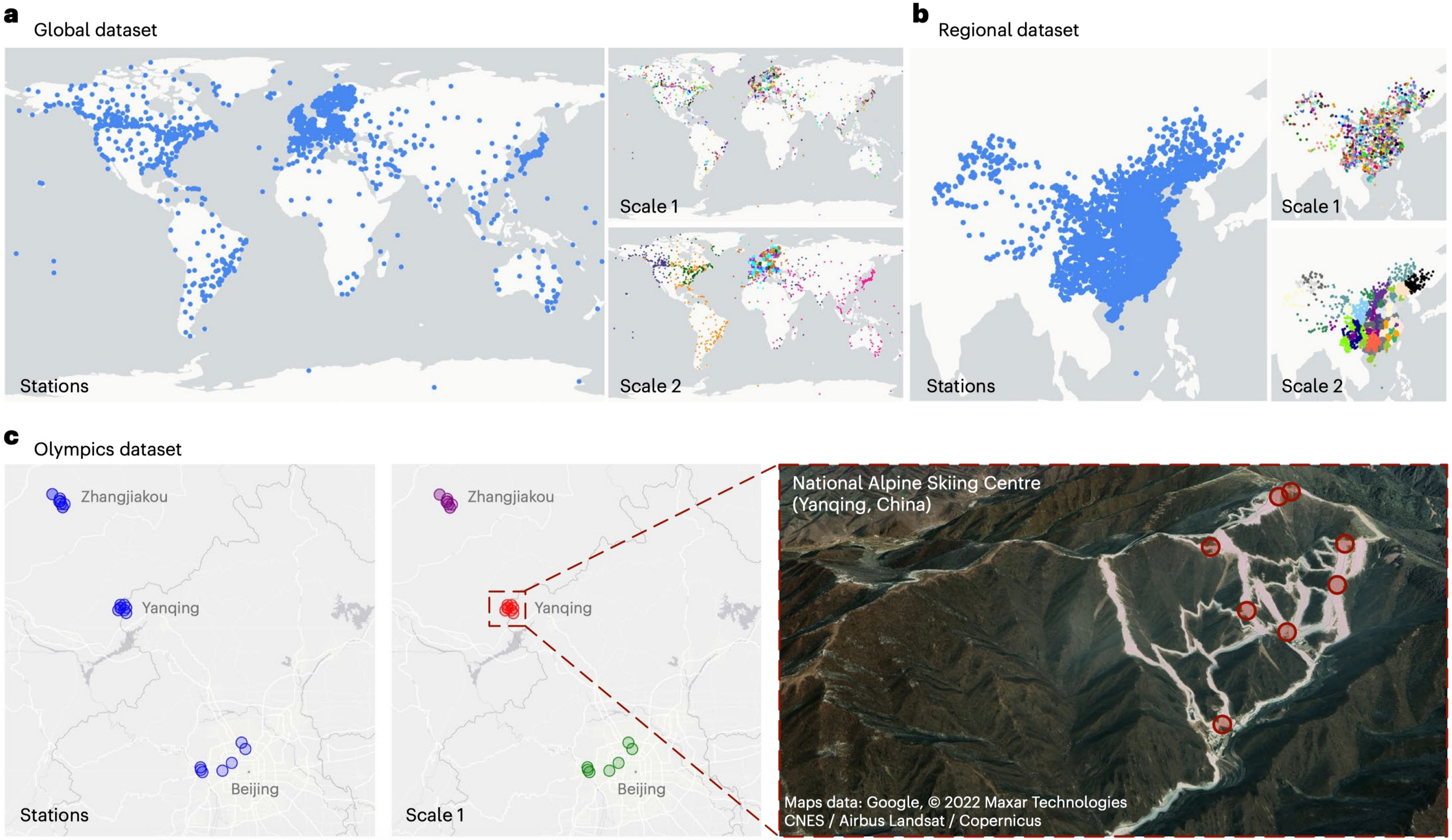
Benefits of Multi-correlation

- **Multiscale modeling** for scattered stations
- **Interpretable evidence** for forecasting
- **Efficient computation** (*almost optimal*)

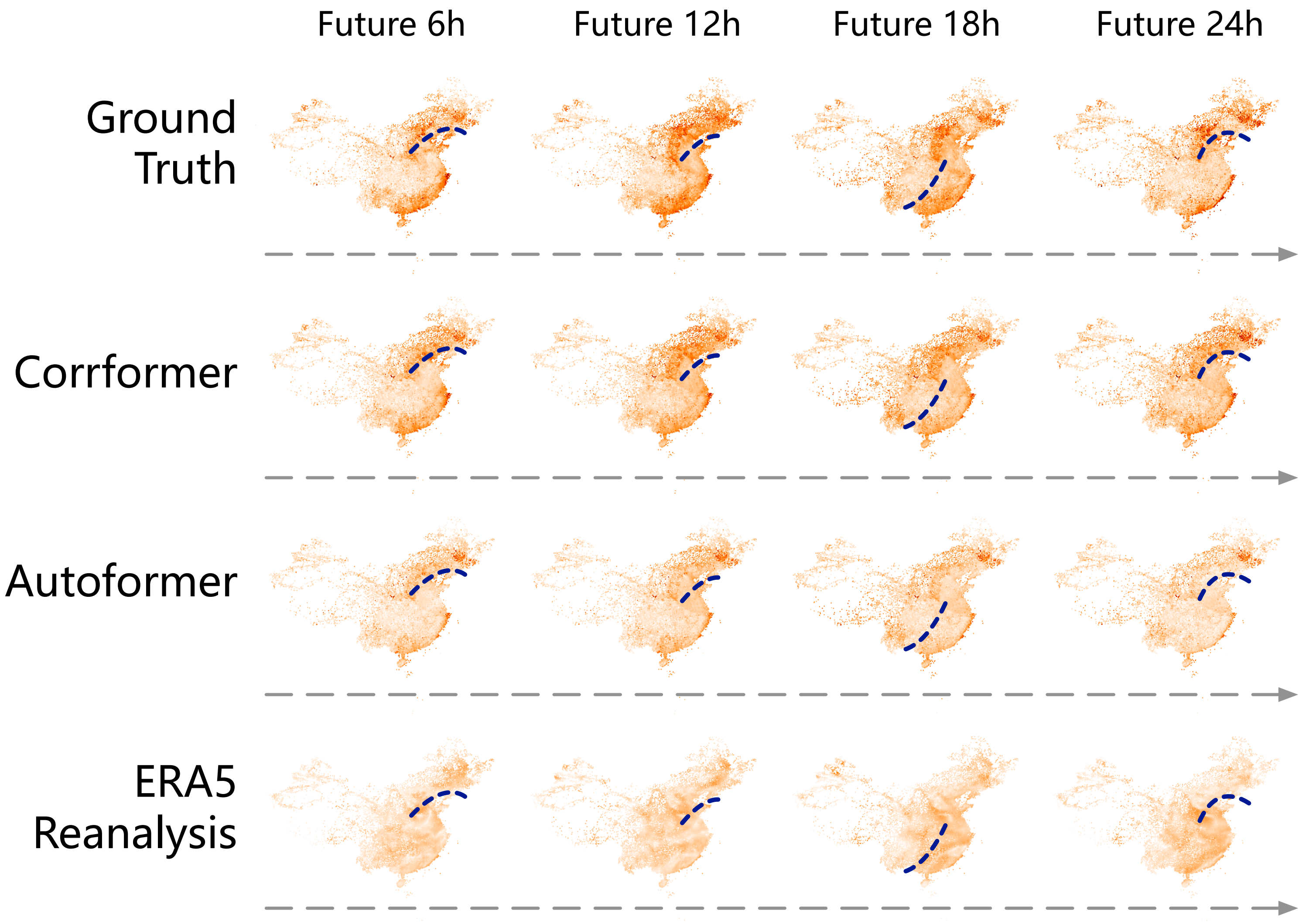
$$O(N^2L^2) \rightarrow O(NL\log L)$$

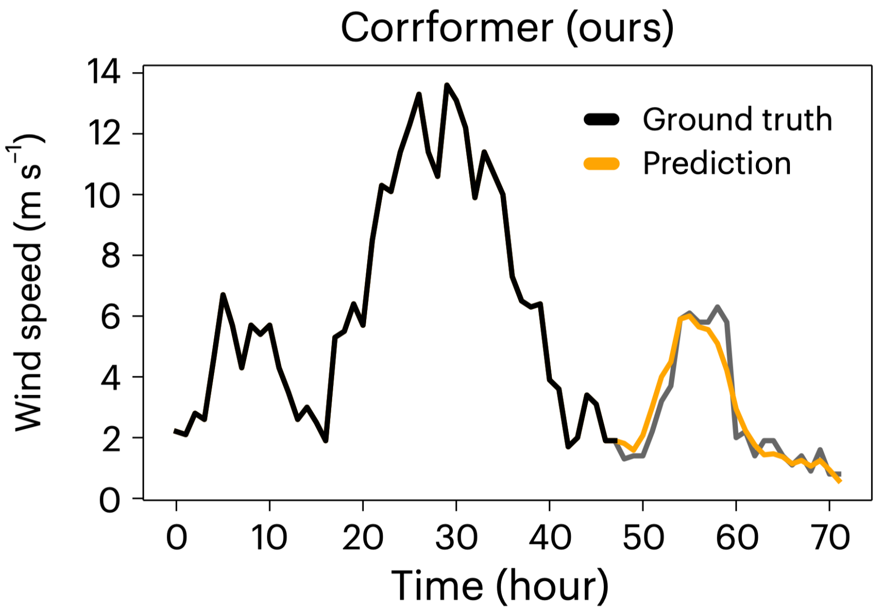
$$\mathcal{O}\left(\sum_{i=1}^{S}\left(\underbrace{N_{i+1}\times\cdots\times N_S}_{\text{intermediate nodes at scale } i}\right)\times\left(\underbrace{N_iL\log L}_{\text{cross-correlation}}+\underbrace{N_i\log N_i}_{\text{sort}}+\underbrace{N_iL}_{\text{aggregation}}\right)\right)=\mathcal{O}(NL\log L),$$

Experiments



Regional Wind Forecasting

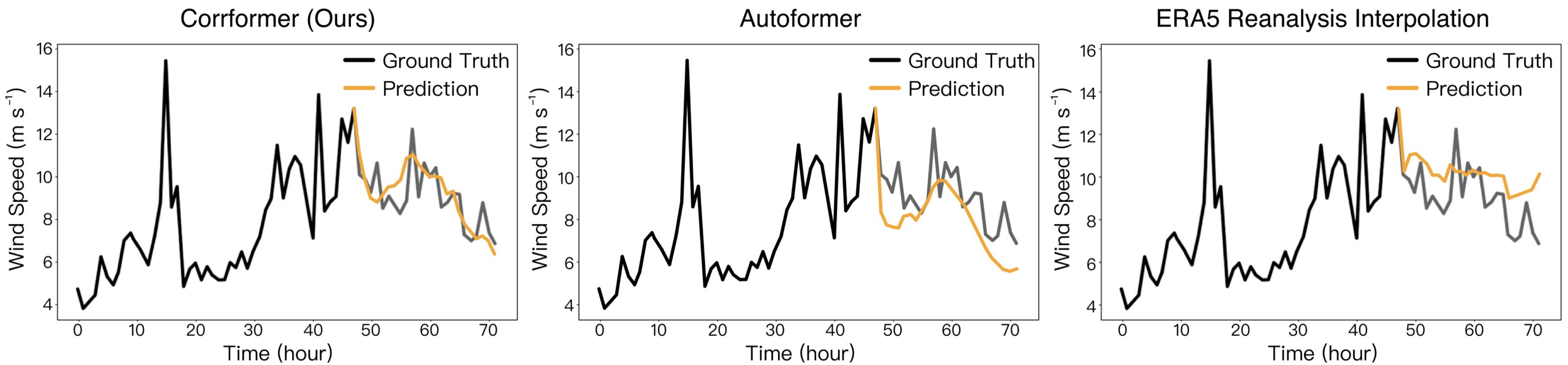




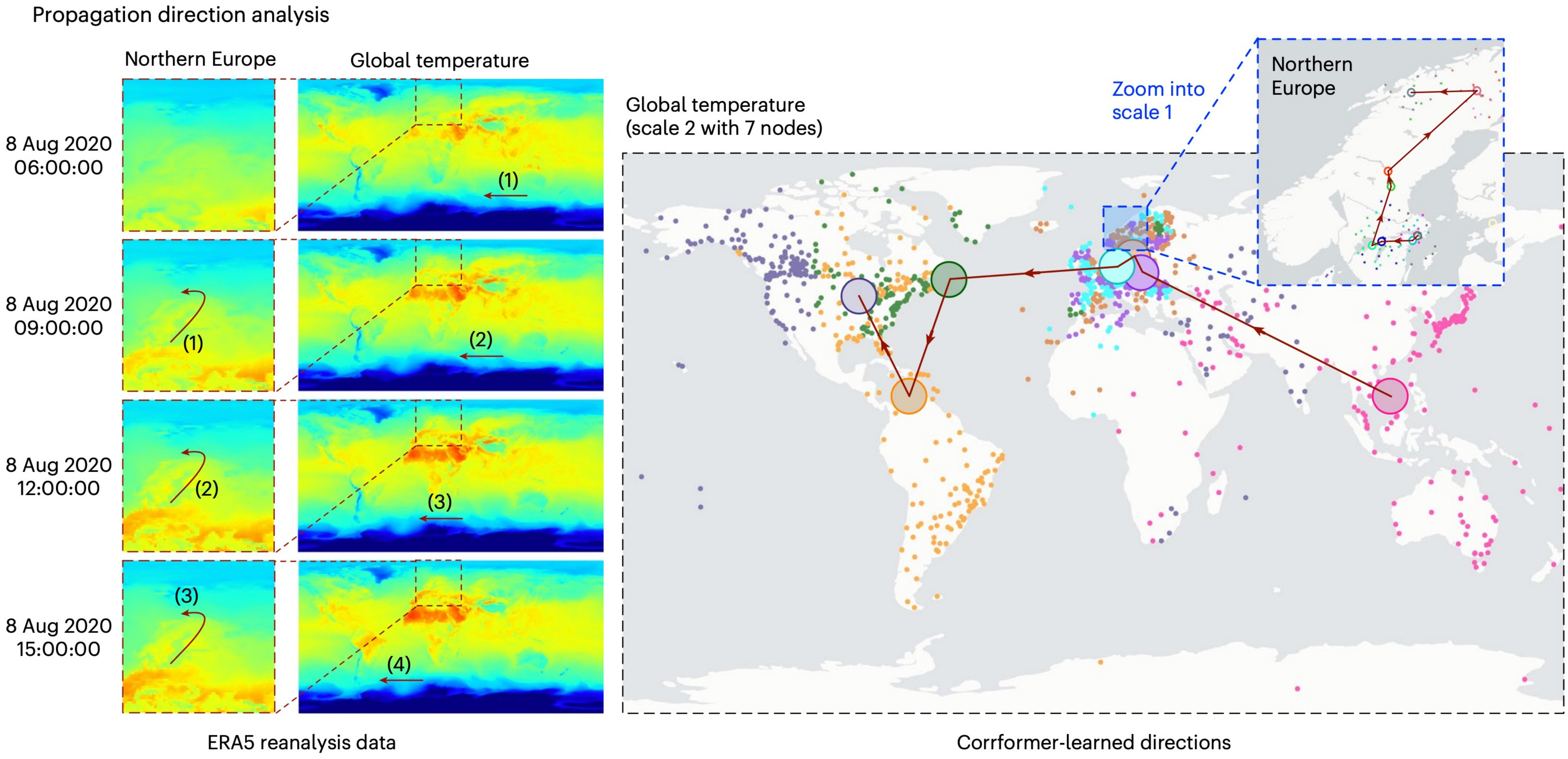
- ✓ Overall 34,040 Stations
- ✓ Training one day on single GPU, real-time forecast in 1 second
- ✓ Achieve the state-of-the-art

Olympics Wind Forecasting

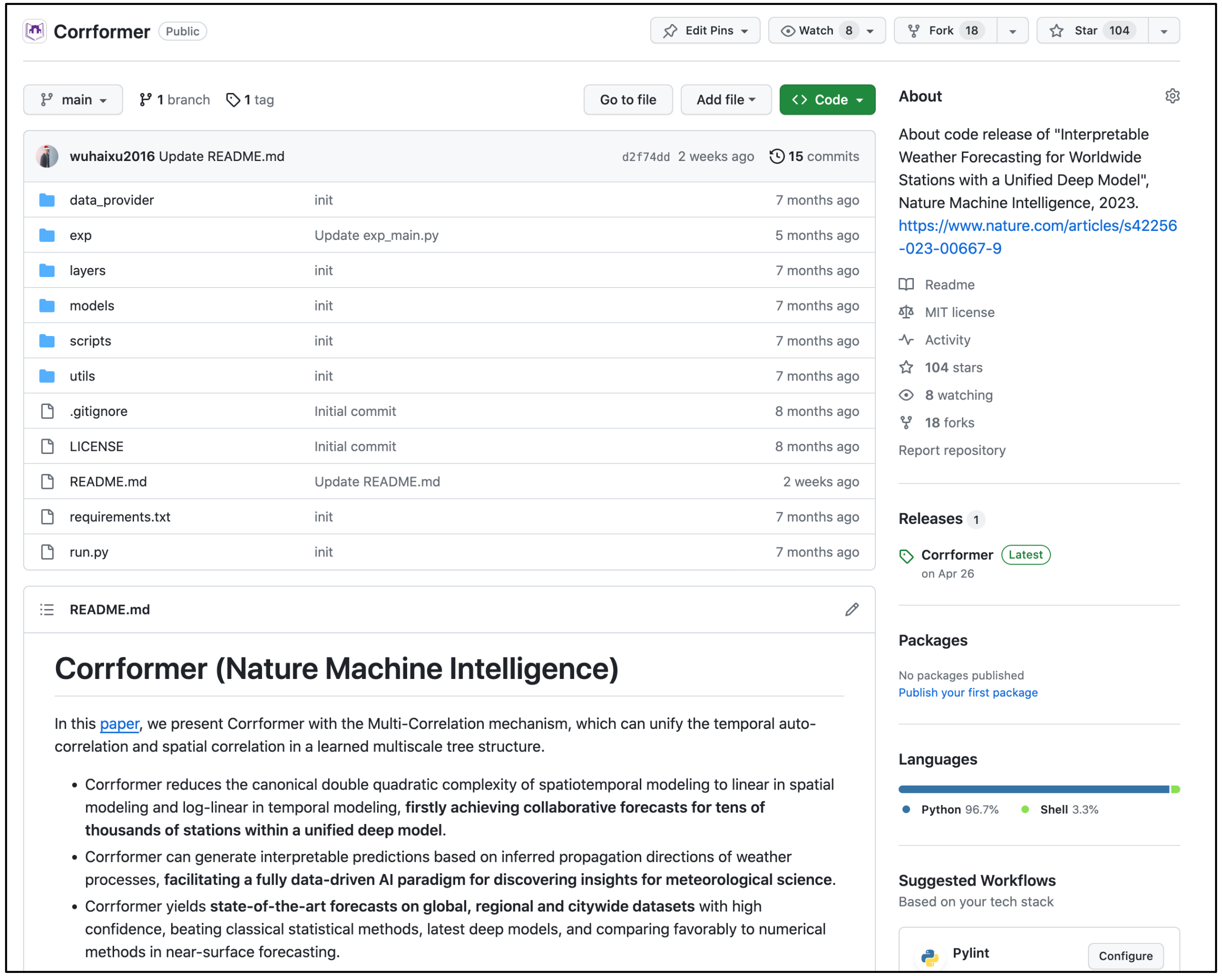
- Alpine weather is seriously affected by chaos effects
- Provide skillful weather forecasting for extremely complex topography



Interpretable Forecasting for Global Weather



Open Source

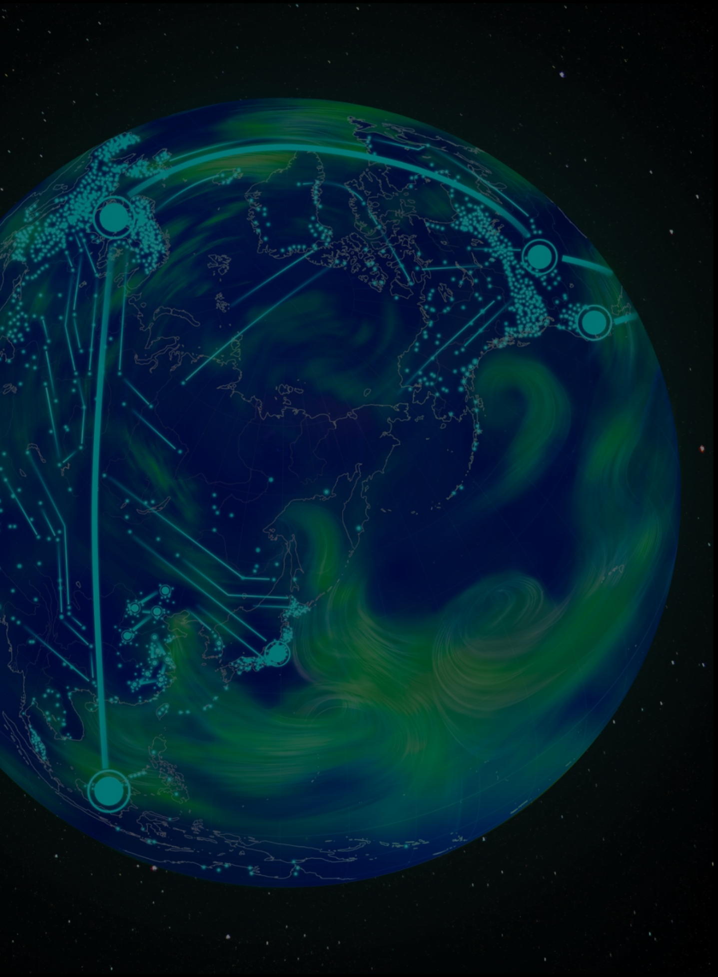


GitHub

CODE OCEAN

https://github.com/thuml/Corrformer

https://codeocean.com/capsule/0341365/tree/v1



Thank You

wuhx23@mails.tsinghua.edu.cn