Spatio-temporal Modeling and Forecasting with Fourier Neural Operators

Pratik Nag

Joint work with Andrew Zammit-Mangion, Sumeetpal Singh, and Noel Cressie

School of Mathematics and Physics
University of Wollongong, Australia

pnag@uow.edu.au

Research supported by Asian Office of Aerospace Research and Development, AFOSR (FA2386-23-1-4100)



NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA



UOW
AUSTRALIA

Dynamical ST Models

Dynamical spatio-temporal (DST) models provide a structured way to describe the temporal evolution of spatial fields through (conditional) dependence across time

Characteristics of classical DST models:

- Evolution modeled via operators; for example, Integro-Difference Equation (IDE) formulations
- Have a strong connection to physical and scientific processes
- Typically assume linear (or simple nonlinear) dynamics
- Has specific model structure such as kernels, Partial Differential Equations (PDEs), etc.
- Widely used in ecology, meteorology, and epidemiology

Proposed Approach to DST Modeling

We introduce a flexible, data-driven, non-linear extension of classical DST models using a Fourier Neural Operator (FNO) (Kovachki et al., 2023), referred as FNO-DST, enabling better forecasting in scenarios where dynamical structure is non-specified

FNO-DST models

- Learn evolution operators directly from data or from simulations
- Have no need for explicit PDE or kernel specifications
- Naturally capture nonlinear and nonstationary dynamics
- Are computationally efficient and scalable to large datasets
- Show strong performance in forecasting complex ST systems

Incorporating History

Suppose γ are the parameters of the evolution process. We incorporate history-dependence by conditioning on the current state and one or more past states:

$$p(\mathbf{Y}_{k+h}|\mathbf{Y}_k, \ldots, \mathbf{Y}_{k-\tau}) = \int p(\mathbf{Y}_{k+h}|\mathbf{Y}_k, \gamma)\, p(\gamma|\mathbf{Y}_k, \ldots, \mathbf{Y}_{k-\tau})\, d\gamma$$

History is incorporated when $\tau \geq 1$.

We can see that

- The posterior of γ is informed by temporal history
- Parameter inference is improved
- Predictive distributions give better forecasts
- The formulation naturally extends DST models beyond first-order dependence

Simple Example of a Nonlinear PDE: Burgers' Equation

To motivate the role of history dependence in DST modeling, we consider a controlled simulation setting based on Burgers' equation which

- Is a simple yet nonlinear PDE capturing advection and diffusion with one parameter γ
- Exhibits nontrivial spatio-temporal dynamics despite its simplicity
- Admits accurate numerical solutions, providing a 'ground truth' that enables direct comparison between model forecasts and true evolution

W e introduce stochasticity by making γ random (i,e., assign a prior to γ)

Simulation Setting

We start with the two-dimensional (1-D space $\times$ time) deterministic nonlinear and non-steady-state Burgers' equation in $(0,1) \times (0,1]$:

$$\partial_t u(s,t) + \partial_s \left(\frac{u^2(s,t)}{2} \right) = \gamma \partial_{ss} u(s,t), \text{ for } (s,t) \in (0,1) \times (0,1],$$

with initial condition,

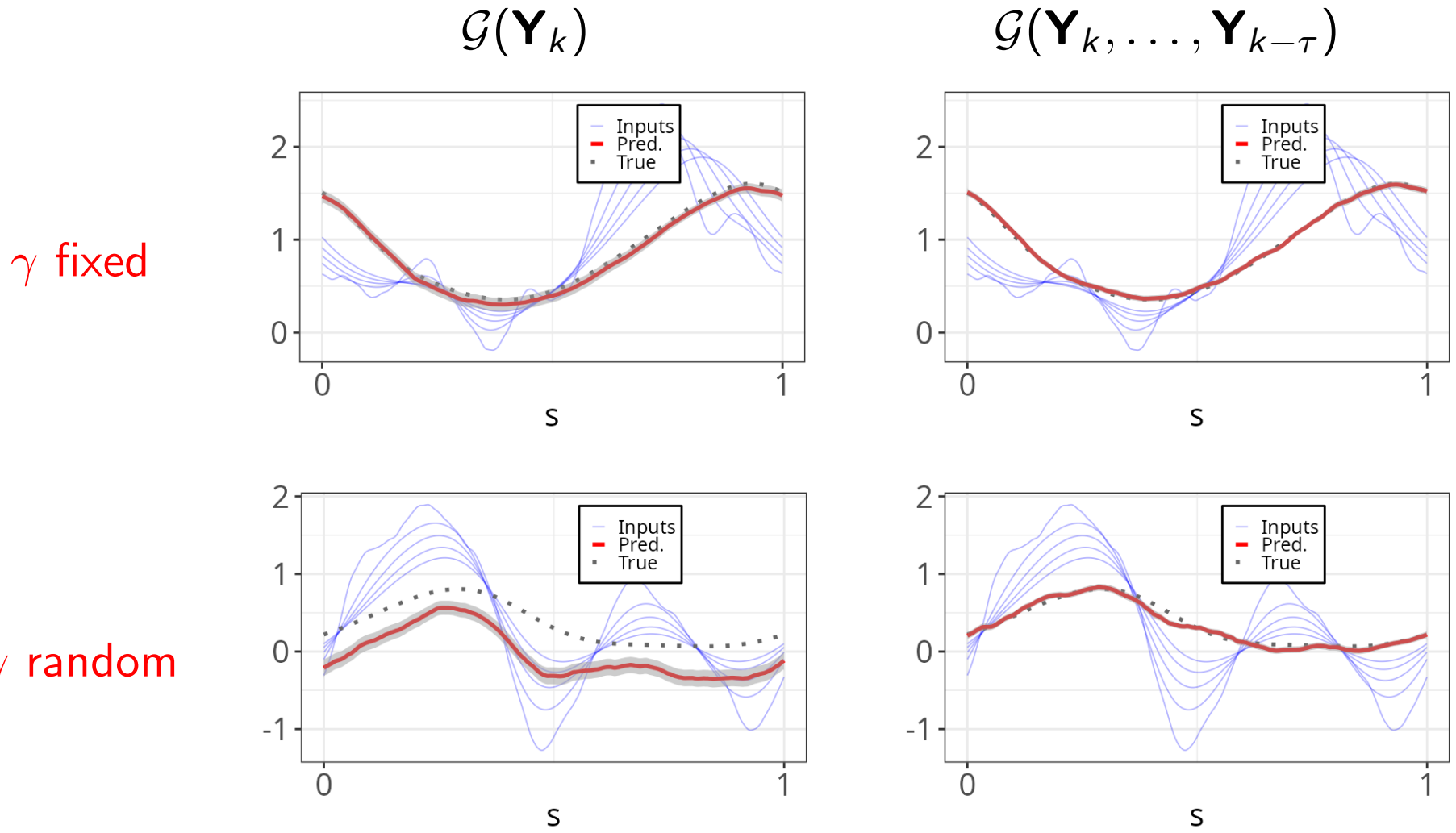
$$u(s,t) = a_0(s), \quad (s,t) \in (0,1) \times \{0\},$$

where γ is the PDE parameter and $a_0(\cdot)$ is given
We assign a prior to γ and aim to learn the underlying evolution operator $\mathcal{G}$ under two settings:

- **Without history:** $\mathbf{Y}_{k+h} = \mathcal{G}(\mathbf{Y}_k)$
- **With history:** $\mathbf{Y}_{k+h} = \mathcal{G}(\mathbf{Y}_k, \ldots, \mathbf{Y}_{k-\tau})$, $\tau \geq 1$

In each of the four cases on the next slide, an FNO-DST model was trained and used to forecast the spatial field at the final time point $t = 1$

Results (γ fixed – degenerate prior; or γ random)

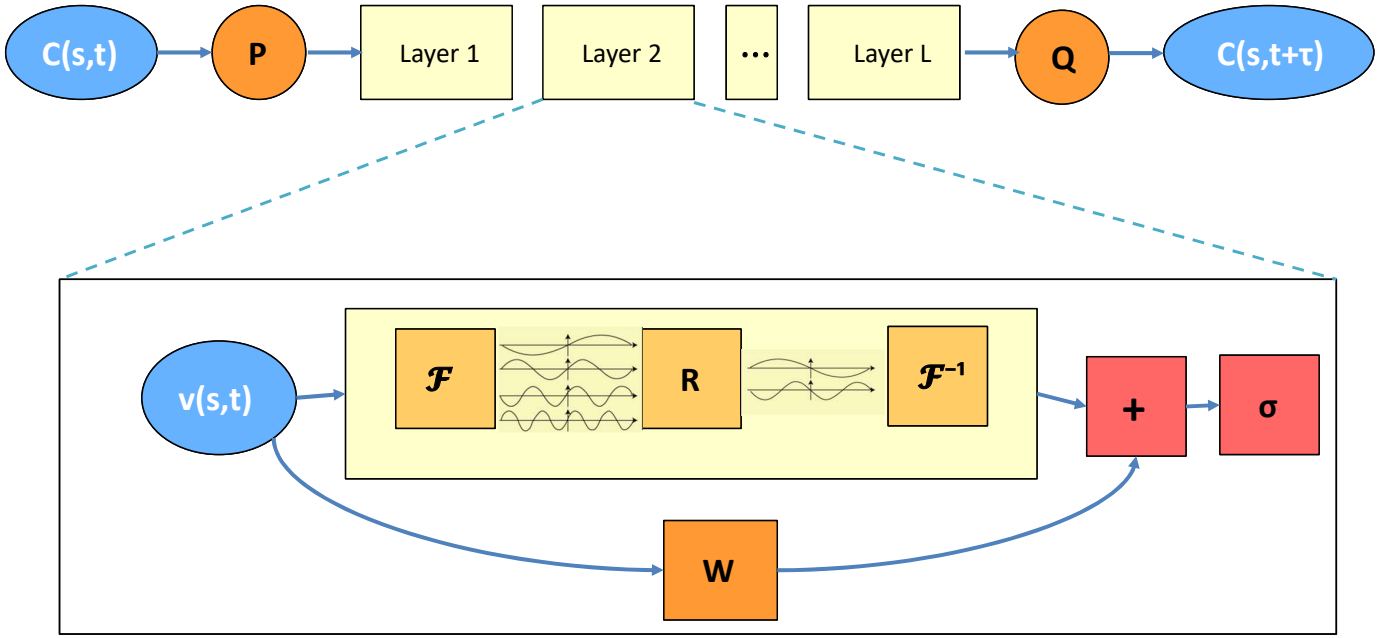


Compare **black dots** (truth) with **red line** (forecast)

Structure of FNO (1/5): Overall Architecture

Fourier Neural Operator (FNO) $\mathcal{G}_{\boldsymbol{\theta}}$ acting on input process $\mathcal{C}$ combines discrete Fourier and inverse Fourier transforms with neural networks (NNs) whose weights/biases are represented as the high-dimensional vector $\boldsymbol{\theta}$

$$\mathcal{G}_{\boldsymbol{\theta}}[\mathcal{C}] = \mathcal{Q} \circ \sigma(\mathcal{W}_L + \mathcal{K}_L) \circ \cdots \circ \sigma(\mathcal{W}_1 + \mathcal{K}_1) \circ \mathcal{P}[\mathcal{C}]$$



Structure of FNO (2/5): Lifting Operator

FNO acting on input process $\mathcal{C}$:

$$\mathcal{G}_{\boldsymbol{\theta}}[\mathcal{C}] = \mathcal{Q} \circ \sigma(\mathcal{W}_L + \mathcal{K}_L) \circ \cdots \circ \sigma(\mathcal{W}_1 + \mathcal{K}_1) \circ \mathcal{P}[\mathcal{C}]$$

Lifting operator $\mathcal{P}$: Maps the input $\mathcal{C}$ from a low-dimensional space to a higher-dimensional feature space

$$\mathcal{P}[\mathcal{C}](\mathbf{x}) = \mathbf{H}_{\mathcal{P}} \begin{pmatrix} \mathbf{x} \\ \\ \mathcal{C}(\mathbf{x}) \end{pmatrix} + \mathbf{b}_{\mathcal{P}},$$

$$\mathbf{x} \in \mathcal{D} \times [t_{k-\tau}, t_k]$$

Structure of FNO (3/5): Local and Kernel Operators

FNO acting on input process $\mathcal{C}$:

$$\mathcal{G}_{\boldsymbol{\theta}}[\mathcal{C}] = \mathcal{Q} \circ \sigma(\mathcal{W}_L + \mathcal{K}_L) \circ \cdots \circ \sigma(\mathcal{W}_1 + \mathcal{K}_1) \circ \mathcal{P}[\mathcal{C}]$$

Local linear operator $\mathcal{W}_l$ and kernel integral operator $\mathcal{K}_l$: These operators capture the dynamics and local dependencies within each layer

$$\mathcal{W}_l[\mathbf{v}_l](\mathbf{x}) = \mathbf{A}_{\mathcal{W},l}\mathbf{v}_l(\mathbf{x}) + \mathbf{b}_{\mathcal{W},l}, \quad l = 1, \ldots, L$$

$$\mathcal{K}_l[\mathbf{v}_{l-1}](\mathbf{x}) = \int_{\mathcal{D}\times[t_{k-\tau}, t_k]} \mathbf{G}_l(\mathbf{x} - \mathbf{x}'; \phi_{\mathcal{K},l})\mathbf{v}_{l-1}(\mathbf{x}')\, d\mathbf{x}', \quad l = 1, \ldots, L$$

Structure of FNO (4/5): Fast Fourier Transform

FNO acting on input process $\mathcal{C}$:

$$\mathcal{G}_{\boldsymbol{\theta}}[\mathcal{C}] = \mathcal{Q} \circ \sigma(\mathcal{W}_L + \mathcal{K}_L) \circ \cdots \circ \sigma(\mathcal{W}_1 + \mathcal{K}_1) \circ \mathcal{P}[\mathcal{C}]$$

kernel integral operator $\mathcal{K}_l$:

$$\mathcal{K}_l[\mathbf{v}_{l-1}](\mathbf{x}) = \int_{\mathcal{D}\times[t_{k-\tau}, t_k]} \mathbf{G}_l(\mathbf{x} - \mathbf{x}'; \phi_{\mathcal{K},l})\mathbf{v}_{l-1}(\mathbf{x}')\, d\mathbf{x}', \quad l = 1, \ldots, L$$

To speed up computation $\mathcal{K}_l[\cdot]$ is redefined in the Fourier domain where the Fast Fourier Transform (FFT) is used:

$$\mathcal{K}_l[\mathbf{v}_{l-1}](\mathbf{x}) = \mathcal{F}^{-1}\big[\mathcal{F}[\mathbf{G}_l(\cdot; \phi_{\mathcal{K},l})]\, \mathcal{F}[\mathbf{v}_{l-1}]\big](\mathbf{x}),$$

where $\mathcal{F}$, $\mathcal{F}^{-1}$ are the Fourier and inverse Fourier transforms. The spatial domain is discretized and we replace the FFT with the discrete Fourier transform (DFT)

Structure of FNO (5/5): Projection and Output

FNO acting on input process $\mathcal{C}$:

$$\mathcal{G}_{\boldsymbol{\theta}}[\mathcal{C}] = \mathcal{Q} \circ \sigma(\mathcal{W}_L + \mathcal{K}_L) \circ \cdots \circ \sigma(\mathcal{W}_1 + \mathcal{K}_1) \circ \mathcal{P}[\mathcal{C}]$$

Projection operator $\mathcal{Q}$: Projects the final feature representation back to the original solution space

$$\mathcal{Q}[\mathbf{v}_L](\mathbf{x}) = \mathbf{h}_{\mathcal{Q}}^{\top}\mathbf{v_L}(\mathbf{x}) + \mathbf{b}_{\mathcal{Q}}$$

Setup for the FNO-DST Model

FNO-DST model: Leverages FNOs to model complex nonlinear dynamics

The FNO-DST model represents the FNO as a vector autoregressive model

Let $\mathbf{G}(\,\cdot\,;\,\boldsymbol{\theta})$ denote the discretized evaluation of the operator $\mathcal{G}_{\boldsymbol{\theta}}[\,\cdot\,]$. The stochastic FNO-DST represents the evolution from past states more flexibly than the state-dependent-kernel model:

$$\mathbf{Y}_{k+h} = \mathbf{G}(\mathbf{Y}_{k-\tau:k};\,\boldsymbol{\theta}) + \boldsymbol{\eta}_{k+h}, \quad \text{for } k = \tau+1,\,\ldots,\,T-h,$$

where recall $\boldsymbol{\theta}$ are NN parameters

Here, $\mathbf{Y}_{k-\tau:k} \equiv (\mathbf{Y}_{k-\tau}^\top, \ldots, \mathbf{Y}_k^\top)^\top$ is the collection of the current and past spatial fields, recall $\mathbb{E}(\boldsymbol{\eta}_{k+h}) = \mathbf{0}$, and we write

$$\text{var}(\boldsymbol{\eta}_{k+h}) = \boldsymbol{\Sigma}_{k+h}(\mathbf{Y}_k;\boldsymbol{\alpha})$$

FNO-DST Modeling

The stochastic FNO-DST model assumes Gaussian innovations:

$$\mathbf{Y}_{k+h} \mid \mathbf{Y}_{(k-\tau):k}, \boldsymbol{\theta}, \boldsymbol{\alpha} \sim \text{Gau}(\mathbf{G}(\mathbf{Y}_{(k-\tau):k};\, \boldsymbol{\theta}), \boldsymbol{\Sigma}_{k+h}(\mathbf{Y}_k; \boldsymbol{\alpha}))$$

Parameter Estimation:
Parameters $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are estimated together by maximizing the Pseudo Likelihood:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\theta}, \boldsymbol{\alpha}} \prod_{k=\tau+1}^{T-h} p(\mathbf{Y}_{k+h} \mid \mathbf{Y}_{(k-\tau):k}, \boldsymbol{\theta}, \boldsymbol{\alpha})$$

Forecasting (with uncertainty quantified):

$$\hat{\mathbf{Y}}_{k+h} = \mathbb{E}(\mathbf{Y}_{k+h} \mid \mathbf{Y}_{(k-\tau):k}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) = \mathbf{G}(\mathbf{Y}_{(k-\tau):k}; \hat{\boldsymbol{\theta}})$$
$$\text{var}(\mathbf{Y}_{k+h} \mid \mathbf{Y}_{(k-\tau):k}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}_{k+h}(\mathbf{Y}_k; \hat{\boldsymbol{\alpha}}),$$

which are the posterior mean and posterior covariance matrix

Simulation Experiments: Return to Burgers' Equation

A numerical solution to Burgers' equation with $\gamma \sim U(0.05, 0.7)$ provides the 'ground-truth': In the experiment, we do not know that the data were generated using Burgers' equation

We compare forecasts from the proposed history-dependent FNO-DST model with those from a range of state-of-the-art statistical models

- Competing models are generally history-dependent, relying on the past τ time steps (e.g., $\tau = 4$)
- Competing models include NN and classical approaches used in spatio-temporal forecasting
- **Evaluation focus:**
  - Forecast accuracy for h-step-ahead forecasts (e.g., $h = 5$)
  - Uncertainty quantification

Key objective: Assess the forecast accuracy and validity achieved by the FNO-DST model in comparison to other competing models

Benchmark Models for Comparison

We compare the proposed FNO-DST model with several established approaches:

- **ConvLSTM** Model (Shi et al., 2015):
  - Incorporates convolutional structure within LSTM units
  - Captures spatio-temporal dependencies via recurrent dynamics
  - Is implemented by replacing $\mathbf{G}(\cdot;\boldsymbol{\theta})$ with ConvLSTM blocks
- **Space-Time DeepKriging (STDK)** Model (Nag et al., 2023):
  - Learns mapping from embedded locations $(\mathbf{s}^\top, t)^\top$ to the value $Y(\mathbf{s}, t)$
  - Has demonstrated strong predictive performance compared to Gaussian-process based methods
- **Persistence** Model:
  - Uses the current observation as the forecast; it serves as a simple benchmark

Results

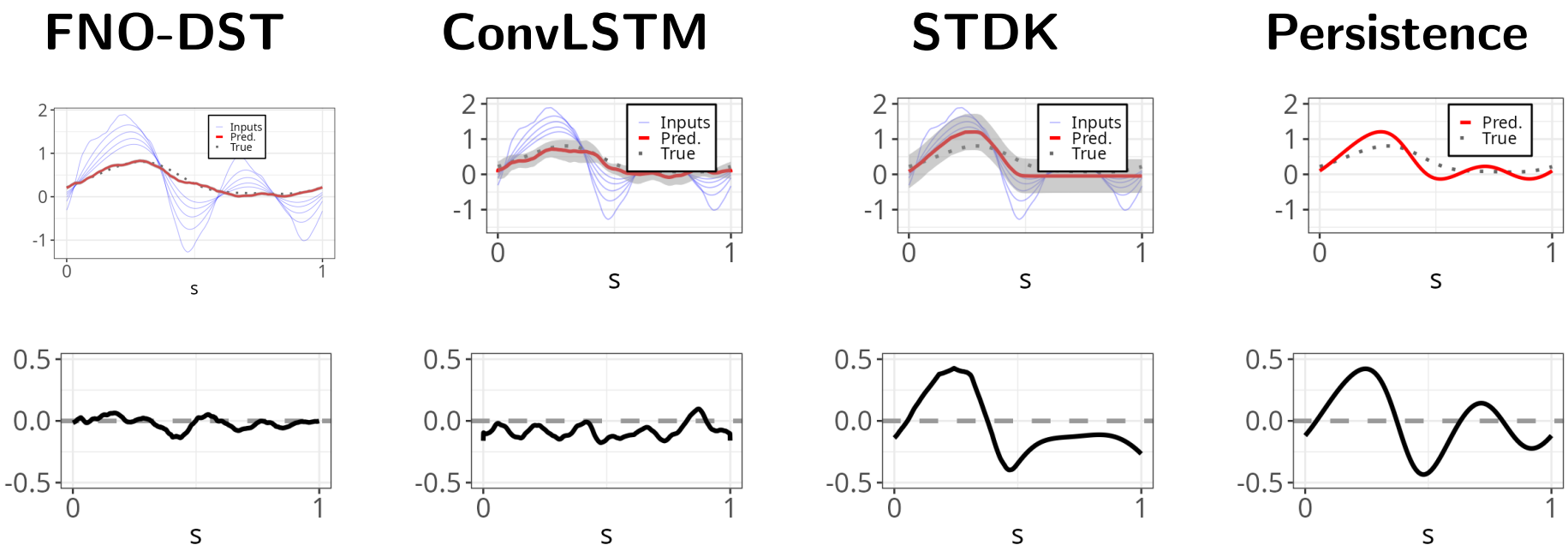
Discretize $[0, 1]$ with $\delta = 0.1$. Then $\{t_1, \ldots, t_T\} = \{0.1, 0.2, \ldots, 0.9, 1.0\}$ and $T = 10$

Model	MSPE	PICP	MPIW
FNO-DST	**0.001**	**0.97**	**0.088**
ConvLSTM	0.054	0.98	0.574
STDK	0.075	0.99	1.013
Persistence	0.089	-	-

Evaluation metrics:

- **MSPE:** Mean Squared Prediction Error (lower is better)
- **PICP:** Prediction Interval Coverage Probability (closer to nominal level is better; e.g., 95%)
- **MPIW:** Mean Prediction Interval Width (lower is better – indicates lower uncertainty)

Forecasts (red line) Minus Truth (black dots)



Top: Forecasts of the spatial field on $\mathcal{D}^G \subset (0,1)$ at $T = 10$ (i.e, end point $t = 1$). Bottom: Corresponding residual fields. Comparison across models for one randomly chosen test instance

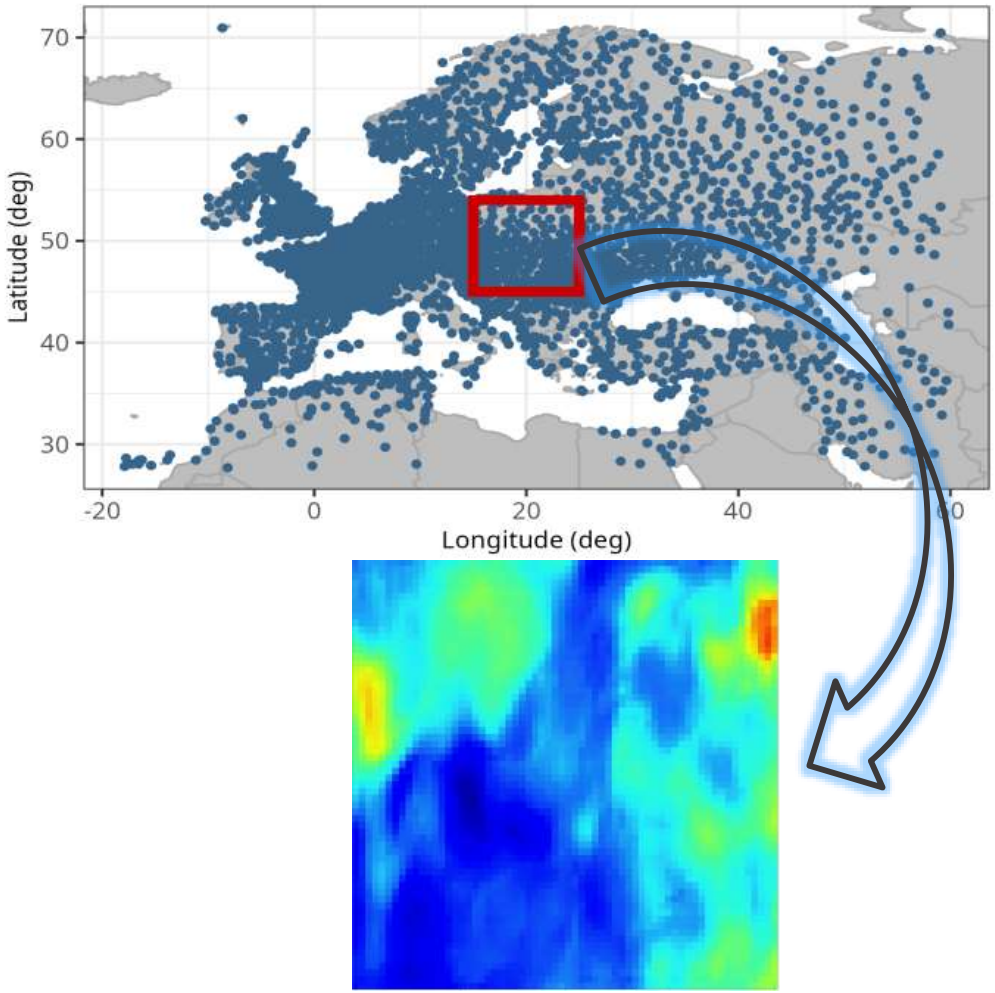
Spatio-Temporal Data Application

We evaluate the proposed methodology on a real-world spatio-temporal dataset exhibiting complex dynamics:

Precipitation: Central Europe, including a region around Germany. Precipitation data exhibit highly variable and intermittent behaviour. The stochastic FNO-DST model captures unknown dynamics, probably nonlinear and nonstationary

Data Description - Precipitation Across Europe

- The full dataset consists of daily precipitation measurements from weather stations across Europe, from 01 Jan 14 to 31 Dec 24
- A 10-day rolling average is applied to smooth out short-term fluctuations, the data are standardized for uniform scaling, and missing values are filled in using the Space-Time DeepKriging model
- Top map shows weather station locations and forecast region over central Europe. Bottom map shows precipitation (mm) on 18 Nov 2023

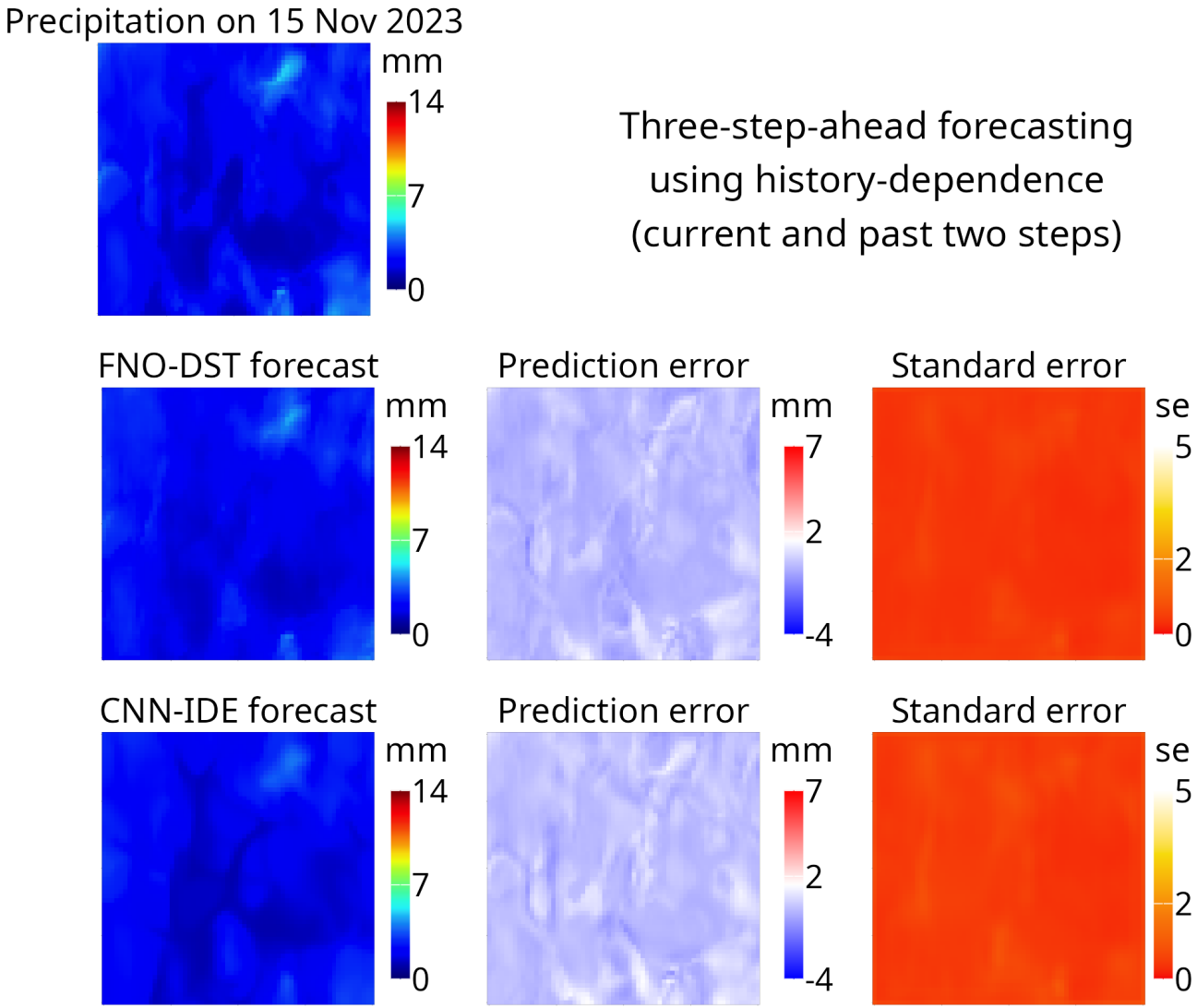


Benchmark Models for Comparison

We compare the proposed **FNO-DST** model with previous models plus an additional model

- **CNN-IDE** Model (Zammit-Mangion and Wikle, 2020):
  - Is an extension of integro-difference equation models
  - Serves as a gold standard for the SST application
  - Uses known advection–diffusion-type dynamics incorporated into the model
- **ConvLSTM** Model (Shi et al., 2015):
  - Incorporates convolutional structure within LSTM units
  - Captures spatio-temporal dependencies via recurrent dynamics
  - Is implemented by replacing $\mathbf{G}(\cdot;\boldsymbol{\theta})$ with ConvLSTM blocks
- **Space-Time DeepKriging (STDK)** Model (Nag et al., 2023):
  - Learns mapping from embedded locations $(\mathbf{s}^\top, t)^\top$ to the value $Y(\mathbf{s}, t)$
  - Has demonstrated strong predictive performance compared to GP-based methods
- **Persistence** Model:
  - Uses the current observation as the forecast and serves as a simple benchmark

Forecast-Precipitation



Model Comparison on Precipitation Data

We compare **FNO-DST** with competing models: **CNN-IDE**, **ConvLSTM**, **STDK**, and **Persistence** using MSPE, coverage probability, prediction interval width, and computing time (in sec)

Model name	MSPE	PICP	MPIW	Comp.time
FNO-DST	**0.32**	**0.96**	**1.95**	1368.3
CNN-IDE	0.46	0.96	2.43	7720.7
ConvLSTM	0.38	0.94	2.23	472.1
STDK	0.82	0.92	1.97	136.3
Persistence	0.84	-	-	**1**

Conclusions and Future Work

- The FNO-DST model effectively captures complex spatio-temporal dependencies after it is trained
- It can infer the underlying dynamics without knowing the true operator
- Based on various simulations and real-data studies, it performs best or nearly best in accuracy and uncertainty quantification. And its computational efficiency is acceptable
- Future work:
  - We are currently developing a hierarchical FNO-DST framework based on Ensemble Kalman Filtering to handle noisy and incomplete data
  - We are looking at the sensitivity of forecast and uncertainty quantification of the FNO-DST model when discretization and NN models are varied

References

Nag, P., Zammit-Mangion, A., Singh, S., and Cressie, N. (2026). Spatio-temporal modeling and forecasting with Fourier neural operators. arXiv:2601.01813.

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97.

Nag, P., Sun, Y., and Reich, B. J. (2023). Spatio-temporal DeepKriging for interpolation and probabilistic forecasting. *Spatial Statistics*, 57:100773.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., New York.

Zammit-Mangion, A. and Wikle, C. K. (2020). Deep integro-difference equation models for spatio-temporal forecasting. *Spatial Statistics*, 37:100408.

FNO-DST for Noisy and Incomplete Data

The FNO-DST framework presented so far assumes observations are **noiseless** and available on a **regular spatial grid**

In practice, environmental observations are **noisy** and at **irregular spatial locations**

To address this, we introduce a **data model**, a conditional distribution linking the noiseless spatially regular FNO-DST process to the observed data

This yields a hierarchical model that accounts for irregular, noisy observations while retaining the FNO-DST process model

The hierarchical FNO-DST model is naturally integrated with an Ensemble Kalman Filter (EnKF) for sequential data assimilation and forecasting