Neural Processes for Short-Term Forecasting of Weather Attributes

Benedetta L. Mussati[1], Helen McKay[2], Stephen Roberts[1,2]

Paper ID: 41

ICLR 2024, Vienna – Workshop: Tackling Climate Change with Machine Learning

[1]University of Oxford [2]Mind Foundry Ltd.

Problem setting: weather forecasting

- Modelling short-term dynamics with limited data

  → Detect unexpected weather conditions and precursors of extreme events

  → Overcome sensors' failure of collecting data

Research questions

Can we use neural processes for short-term forecasting with incomplete context data?

When learning a set of correlated tasks simultaneously, can the model leverage information from their joint distribution?

Neural Processes

Dataset $D = \{C, T\}$

$$C = \{(x, y)_i\}_{i=1}^{I}$$

$$T = \{x_i\}_{i=I+1}^{I+M}$$

Given D, we want to learn $f: X \rightarrow Y$ that generated C to predict the outputs y_* at T

Neural Processes

Assume f is sampled from a distribution over functions P, which defines $p(f(T)|C,T)$

We have $D_P = \{D_f\}_{f=1}^{F} = \{\{C,T\}_f\}_{f=1}^{F}$ for F functions sampled from P

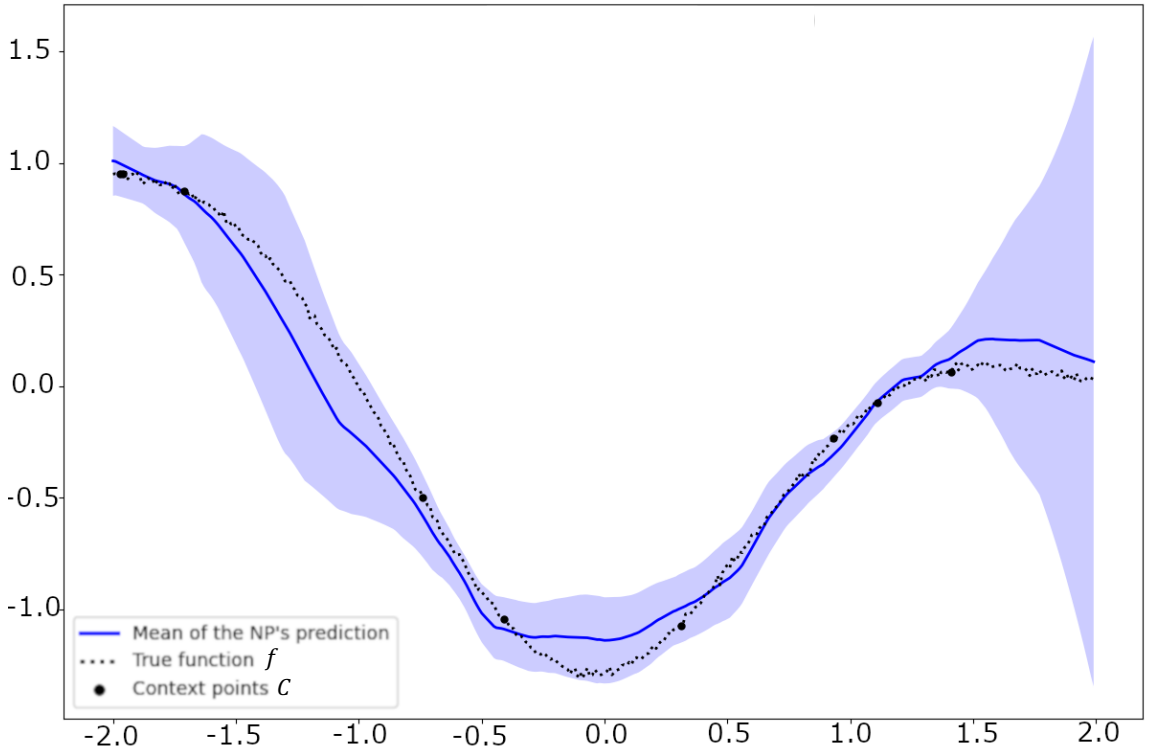
Objective: learn a model whose parameters maximise the likelihood of P

Neural Processes

Assume f is sampled from a distribution over functions P, which defines $p(f(T)|C,T)$

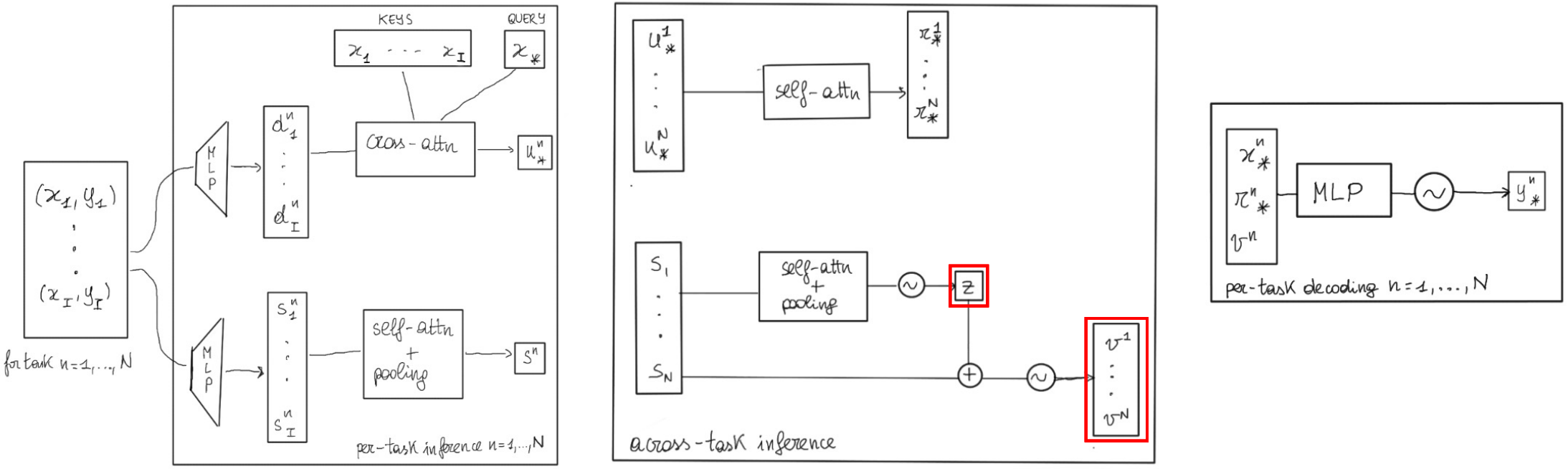
We have $D_P = \{D_f\}_{f=1}^{F} = \{\{C,T\}_f\}_{f=1}^{F}$ for F functions sampled from P

Objective: learn a model whose parameters maximise the likelihood of P



Output of an NP: $p(y_*|C,x_*)$
A realisation at target location x_* conditioned on C can be sampled from this predictive distribution

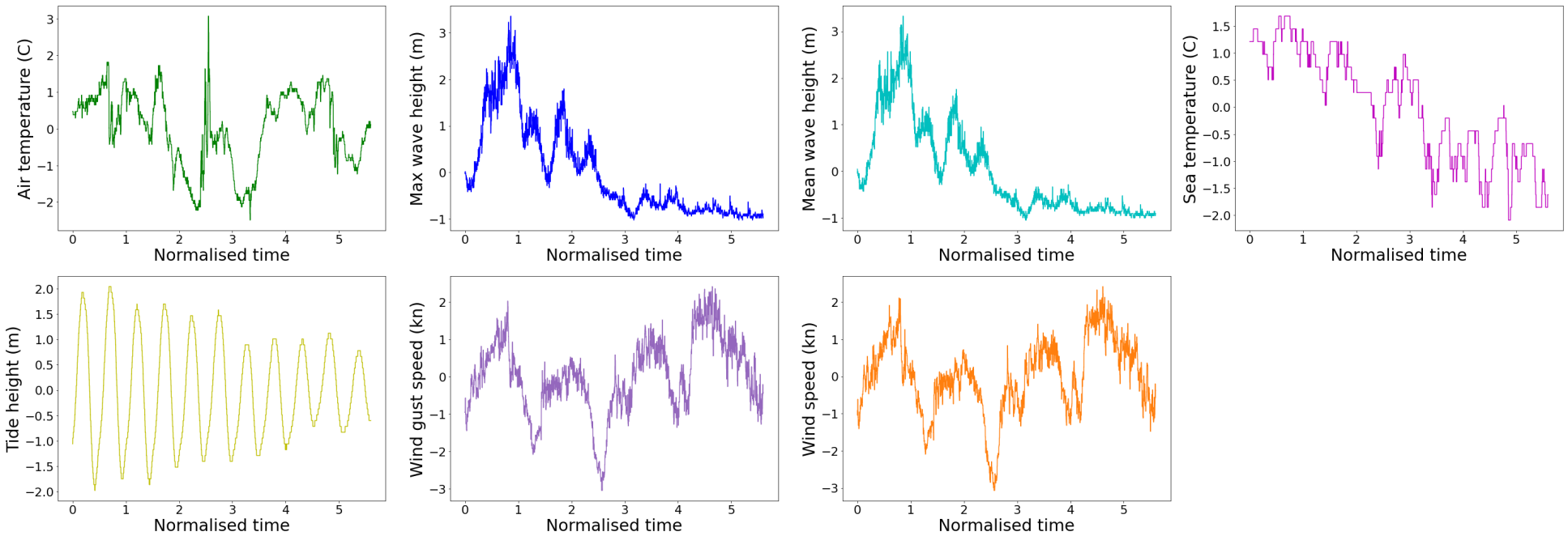
Multi-task neural processes (MTNP)



Adapted from Kim2022

Data

Chimet dataset → week-long dataset of weather attributes timeseries.
Collected from Chichester Harbour in August 2007



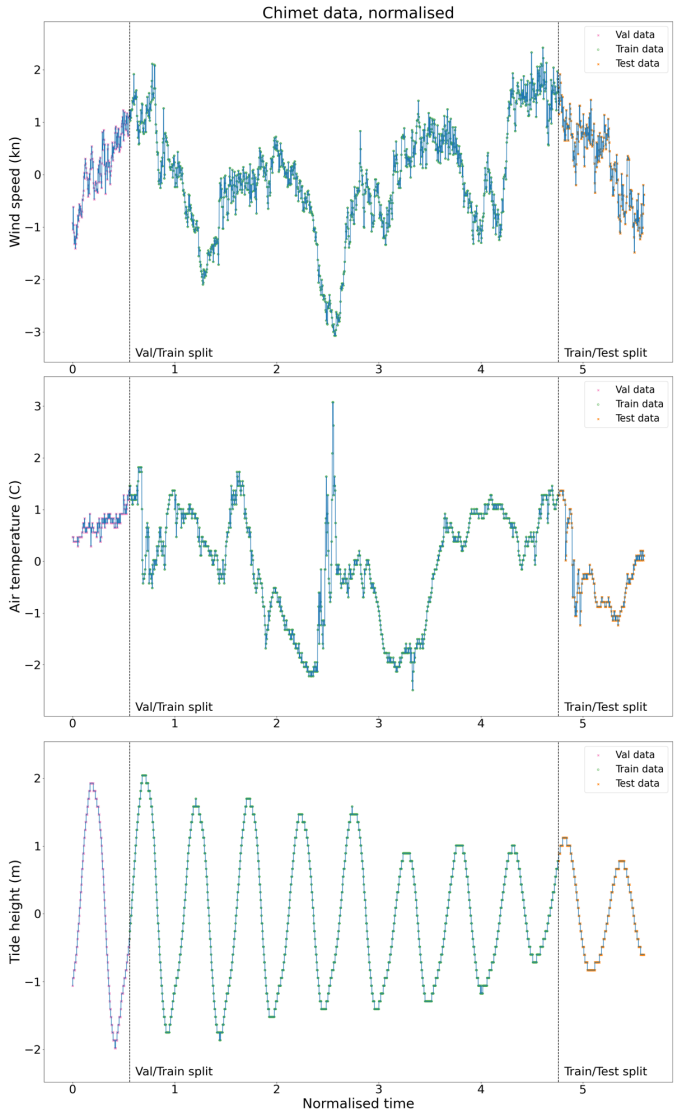
Engineering and Physical Sciences Research Council

EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines & Systems

mind FOUNDRY

UNIVERSITY OF OXFORD

Datasets



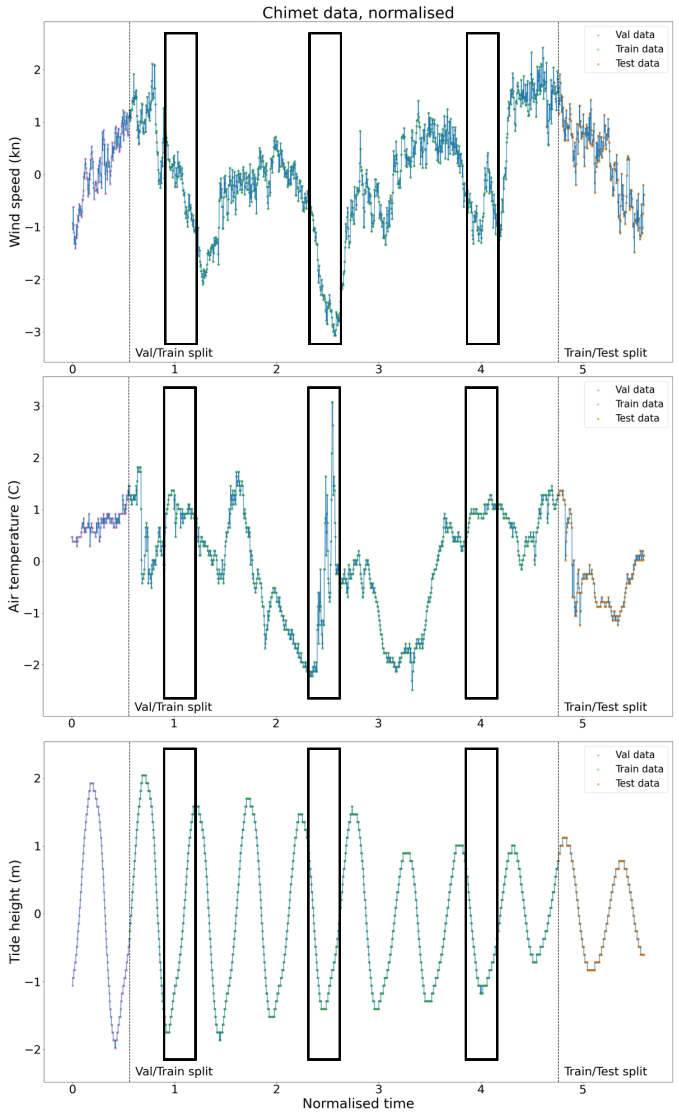
Engineering and Physical Sciences Research Council

AIMS EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines & Systems

mind FOUNDRY

UNIVERSITY OF OXFORD

Datasets



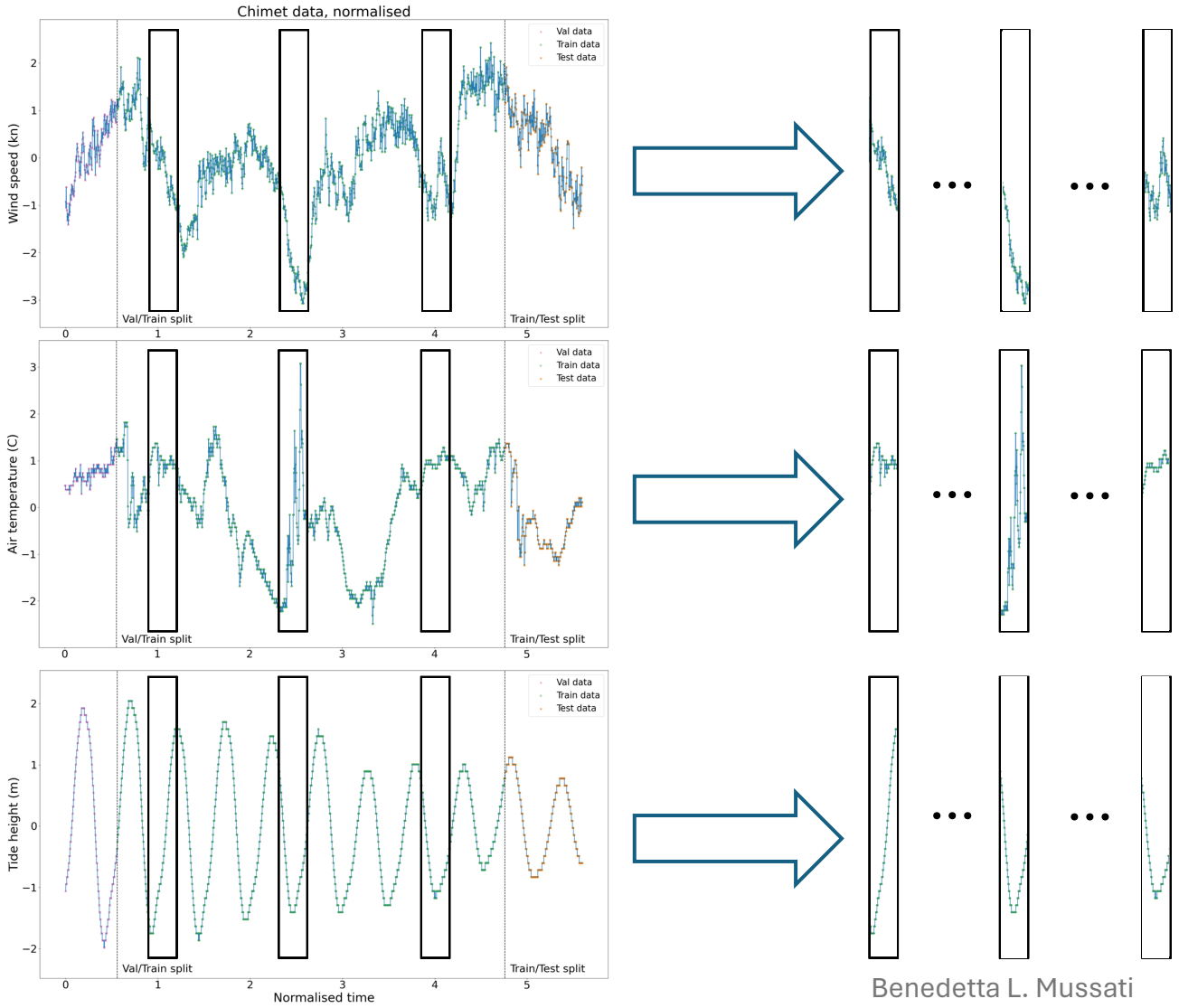
Engineering and Physical Sciences Research Council

AIMS EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines & Systems

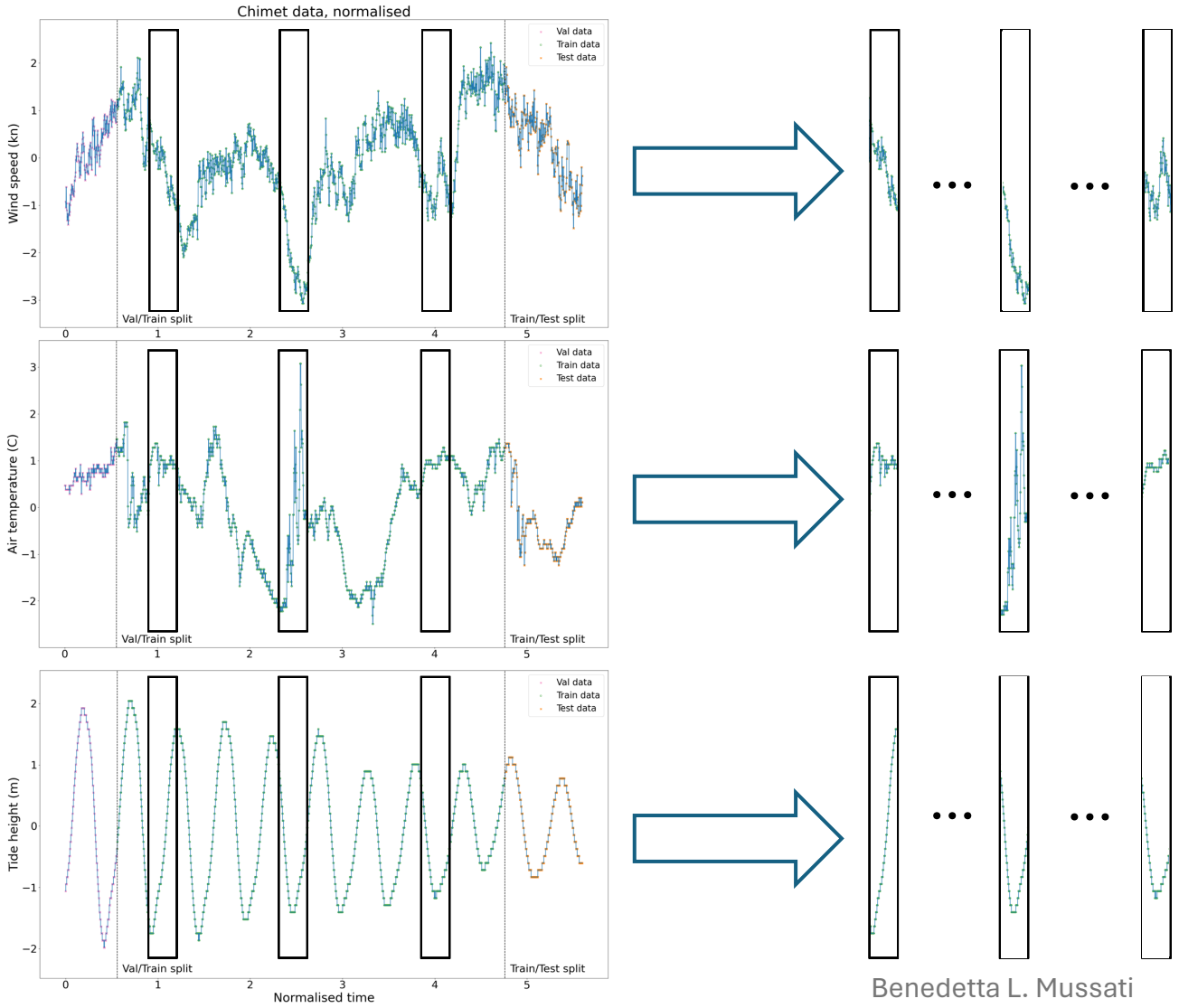
mind FOUNDRY

UNIVERSITY OF OXFORD

Datasets



Datasets

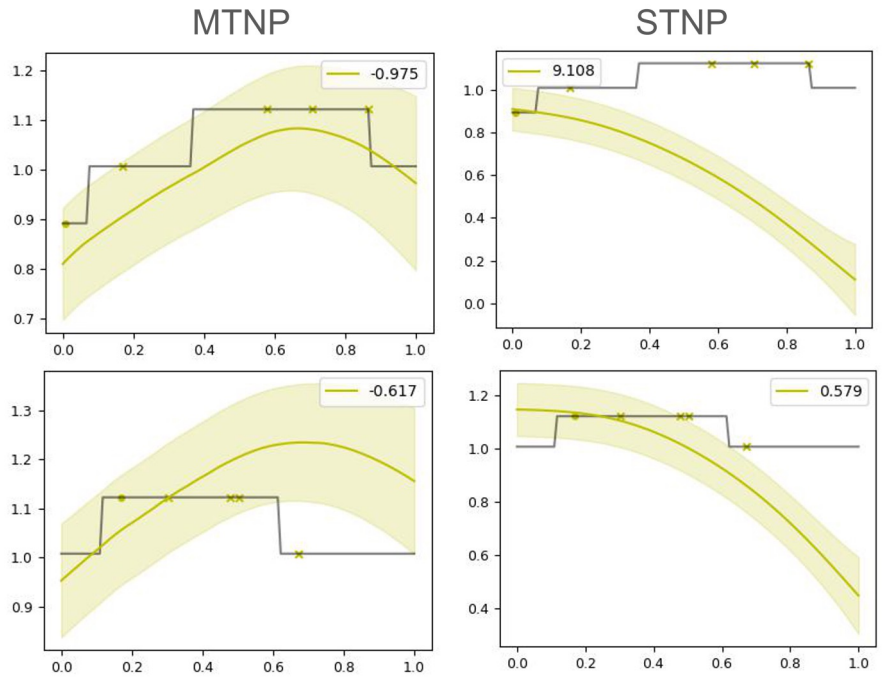


We experiment with datasets of 2-hour long slices with 5 minute intervals and with datasets of 6-hour long slices with 30 minute intervals.

Results: 2-hour forecast horizon

cs_{eval}	Model	NLL	MSE
5	STNP	0.9839	**0.1455**
	MTNP	**0.5502**	0.1515
10	STNP	0.1910	**0.0724**
	MTNP	**0.1089**	0.0870
20	STNP	**−0.3140**	**0.0505**
	MTNP	−0.1668	0.0694

Table 1: Comparison of cumulative NLL and MSE, with varying cs_{eval} and fixed $\gamma_{train}, \gamma_{eval} = 0.8$



Comparison of MTNP's and STNP's predictions on two evaluation functions of *Tide height (m)*, with $cs_{eval} = 5$ and $\gamma_{eval} = 0.8$

Terminology

- $\gamma \in [0, 1]$: missing rate with whom context points are randomly dropped
- C^n, T^n : context set and target locations for task n
- cs_{eval} size of the context set at evaluation

The MTNP leverages inter-task correlation to maximise $p(f(T^n)|C^n, T^n)$ for task n when C^n is small

Results: 2-hour forecast horizon

Terminology

- $\gamma \in [0, 1]$: missing rate with whom context points are randomly dropped
- C^n, T^n : context set and target locations for task n
- cs_{eval} size of the context set at evaluation

$\gamma_{train}, \gamma_{eval}$	Model	NLL	MSE
0.2	STNP	**−0.5913**	**0.0343**
	MTNP	−0.4438	0.0390
0.5	STNP	**−0.4467**	**0.0443**
	MTNP	−0.2967	0.0484
0.8	STNP	0.1910	**0.0724**
	MTNP	**0.1089**	0.0870

Table 2: Comparison of cumulative NLL and MSE, with varying $\gamma_{train}, \gamma_{eval}$ and fixed $cs_{eval} = 10$

However, the STNP performs better than the MTNP when provided with sufficient context information.

Results: 6-hour forecast horizon

Fix $\gamma_{train}, \gamma_{eval} = 0.5$ and compare the MTNP and STNP to an ensemble of GPs.

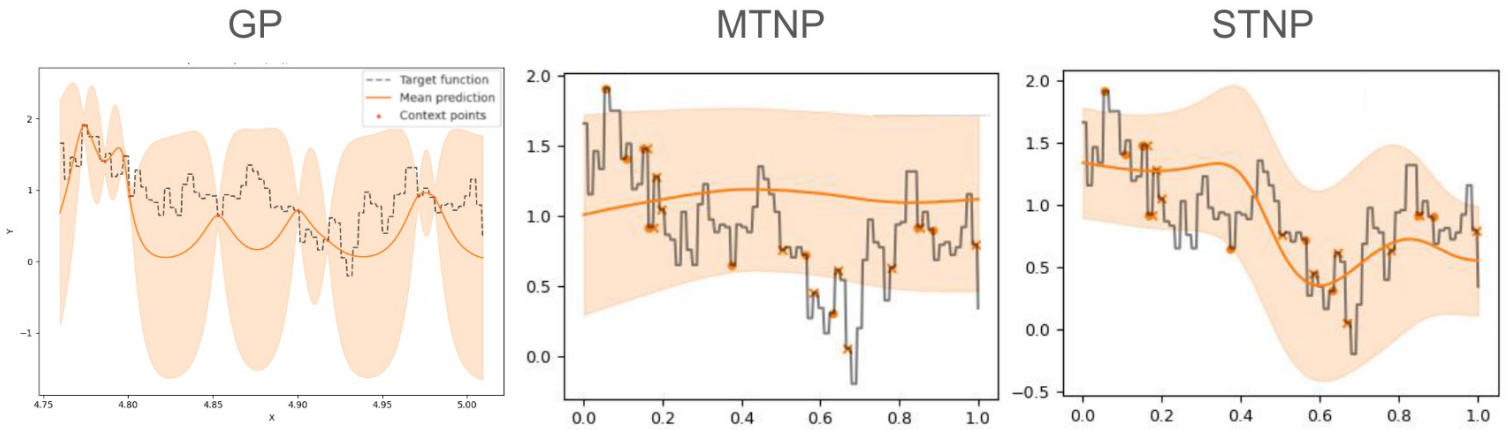
cs_{eval}	Model	NLL	MSE
10	MTNP	0.5542	0.1740
	STNP	**0.1449**	**0.0933**
	GP	0.4252	0.2955
20	MTNP	0.3058	0.1139
	STNP	**-0.0911**	**0.0725**
	GP	0.0348	0.1667
50	MTNP	0.2561	0.1086
	STNP	-0.2281	0.0632
	GP	**-0.5705**	**0.0475**
100	MTNP	0.2493	0.1069
	STNP	-0.2649	0.0616
	GP	**-0.9895**	**0.0203**
200	MTNP	0.2188	0.1044
	STNP	-0.2660	0.0613
	GP	**-1.3624**	**0.0080**

Table 3: Comparison of cumulative NLL and MSE, with varying cs_{eval}

Terminology

- $\gamma \in [0, 1]$: missing rate with whom context points are randomly dropped
- C^n, T^n : context set and target locations for task n
- cs_{eval} size of the context set at evaluation

The STNP outperforms the MTNP and GP when cs_{eval} is small



Comparison of GP's, MTNP's and STNP's predictions on the same evaluation function of the *Wind speed (kn)* weather attribute with $cs_{eval} = 20$.

The GP's predictive mean does not approximate the true function as well as the STNP in areas with no context points.

Results: 6-hour forecast horizon

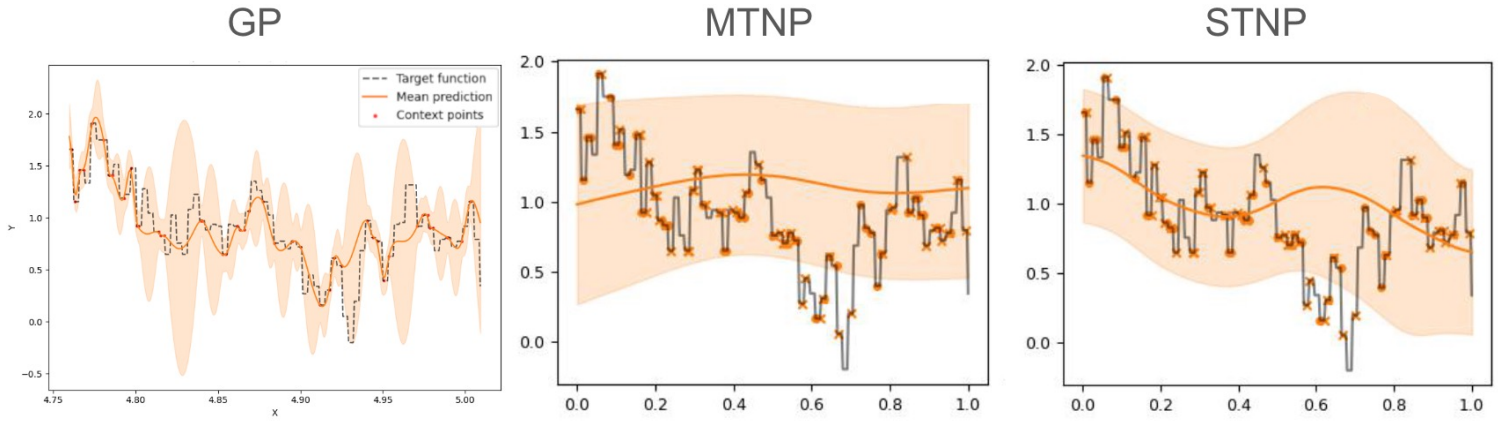
Fix $\gamma_{train}, \gamma_{eval} = 0.5$ and compare the MTNP and STNP to an ensemble of GPs.

cs_{eval}	Model	NLL	MSE
10	MTNP	0.5542	0.1740
	STNP	**0.1449**	**0.0933**
	GP	0.4252	0.2955
20	MTNP	0.3058	0.1139
	STNP	**-0.0911**	**0.0725**
	GP	0.0348	0.1667
50	MTNP	0.2561	0.1086
	STNP	-0.2281	0.0632
	GP	**-0.5705**	**0.0475**
100	MTNP	0.2493	0.1069
	STNP	-0.2649	0.0616
	GP	**-0.9895**	**0.0203**
200	MTNP	0.2188	0.1044
	STNP	-0.2660	0.0613
	GP	**-1.3624**	**0.0080**

Table 3: Comparison of cumulative NLL and MSE, with varying cs_{eval}

Terminology

- $\gamma \in [0, 1]$: missing rate with whom context points are randomly dropped
- C^n, T^n : context set and target locations for task n
- cs_{eval} size of the context set at evaluation



Comparison of GP's, MTNP's and STNP's predictions on the same evaluation function of the *Wind speed (kn)* weather attribute with $cs_{eval} = 100$.

When cs_{eval} is large, we find the NPs are outperformed by the GP.

Conclusions

- Preliminary results on how NPs can be used as a data-driven approach to learn short-term dynamics for forecasting purposes.
- Limited context information → the MTNP leverages inter-task knowledge and outperforms the STNP.
- Sufficient but not too exceeding context information → the STNP outperforms both the MTNP and the GPs ensemble.
- Excessive context information → it may hinder the STNP and the MTNP.

Thank you!

Neural Processes for Short-Term Forecasting of Weather Attributes

Paper ID:41

Benedetta L. Mussati

Bibliography:
Kim2022: MULTI-TASK NEURAL PROCESSES – Donggyun Kim et al. ICLR 2022