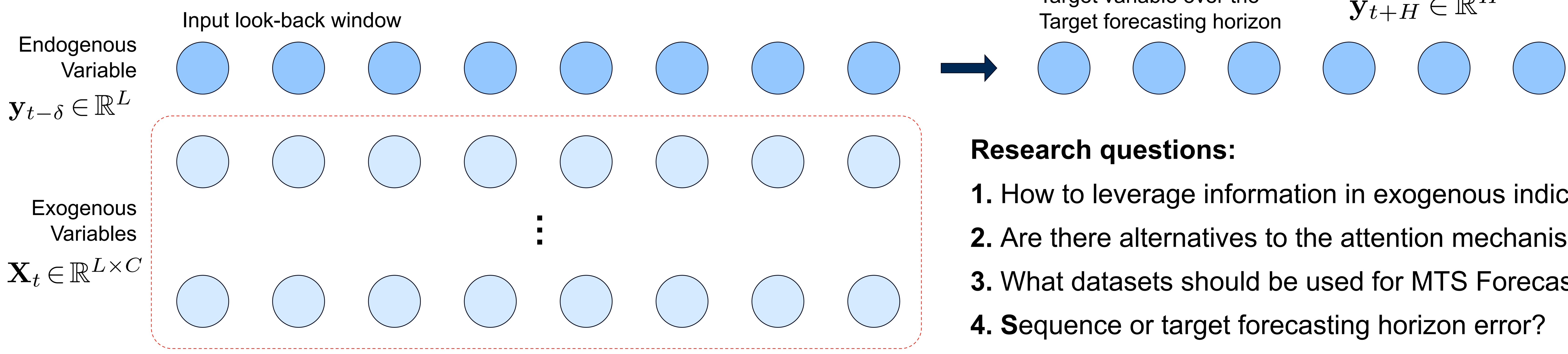
Sonnet: Spectral Operator Neural Network for Multivariable Time Series Forecasting

Yuxuan Shu, Vasileios Lampos

Centre for Artificial Intelligence, Department of Computer Science, University College London, UK

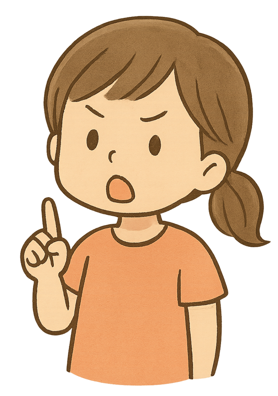
Multivariable Time Series (MTS) Forecasting



Research questions:

1. How to leverage information in exogenous indicators?
2. Are there alternatives to the attention mechanism for MTS?
3. What datasets should be used for MTS Forecasting?
4. Sequence or target forecasting horizon error?

1. Sonnet Model Step-by Step



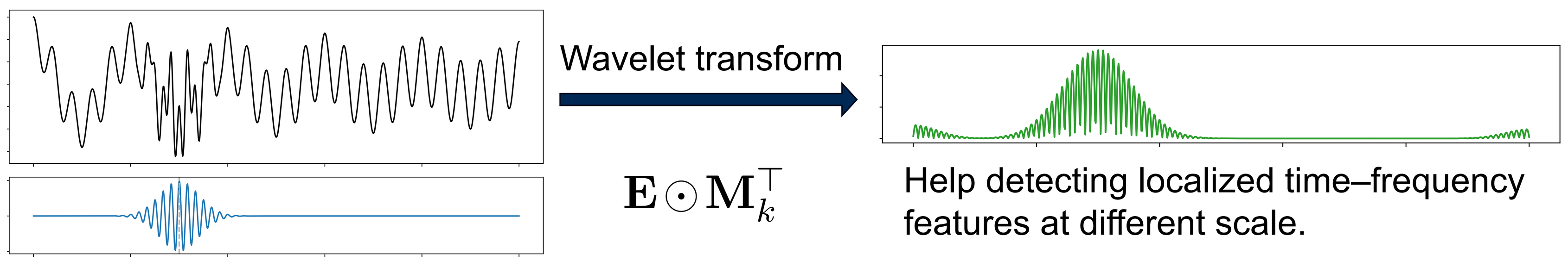
1.1 Embedding covariates

Embedding exo- and endo-genous variables: $\mathbf{E}_x \in \mathbb{R}^{L \times \alpha d}$, $\mathbf{E}_y \in \mathbb{R}^{L \times (1-\alpha) d}$

Final embedding: $\mathbf{E} = [\mathbf{E}_x, \mathbf{E}_y] \in \mathbb{R}^{L \times d}$

1.2 Learnable Wavelet Transform

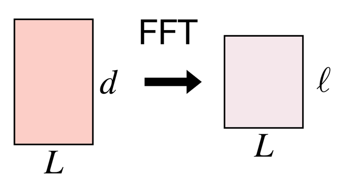
Learnable wavelets: $\mathbf{M}_k = \exp\left(-\mathbf{w}_{\alpha}\mathbf{t}^2\right) \times \cos\left(\mathbf{w}_{\beta}\mathbf{t} + \mathbf{w}_{\gamma}\mathbf{t}^2\right)$



Help detecting localized time–frequency features at different scale.

1.3 Multivariable Coherence Attention (MVCA)

Variable-wise FFT:



$\ell = \lfloor \frac{d}{2} \rfloor + 1$ $\quad \mathbf{Q}_f, \mathbf{K}_f \in \mathbb{C}^{L \times \ell}$

Power-spectral densities: $\mathbf{P}_{qk} = \mathbf{Q}_f \odot \mathbf{K}_f^{*}$, $\mathbf{P}_{qq} = \mathbf{Q}_f \odot \mathbf{Q}_f^{*}$, $\mathbf{P}_{kk} = \mathbf{K}_f \odot \mathbf{K}_f^{*}$

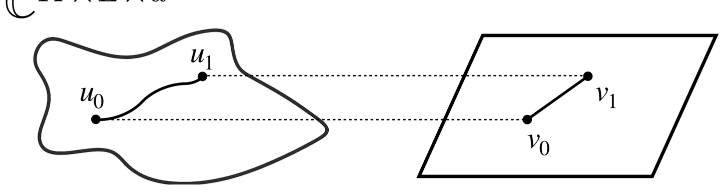
Spectral coherence: $\mathbf{C}_{qk} = |\overline{\mathbf{P}}_{qk}|^2 / (\overline{\mathbf{P}}_{qq} \cdot \overline{\mathbf{P}}_{kk} + \epsilon)$

1.4 Koopman-Guided Spectrum Evolvement

Complex form of the output from MVCA: $\mathbf{O}_c \in \mathbb{C}^{K \times L \times d}$

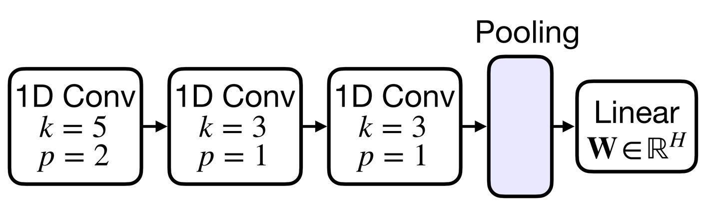
Apply Koopman Operator: $\mathbf{O}_l = \mathbf{K} \times \mathbf{O}_c$

Turns a nonlinear system into a linear one:



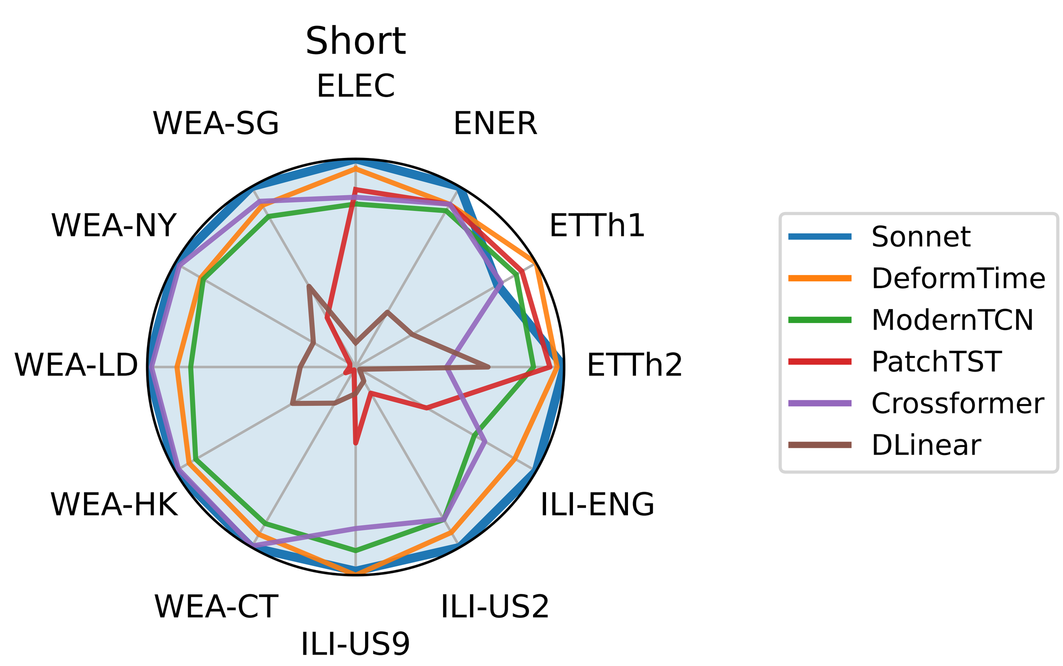
1.5 Sequence reconstruction and decode

- Inverse transform to reconstruct the sequence;
- 3-layer CNN with GELU activation as decoder.



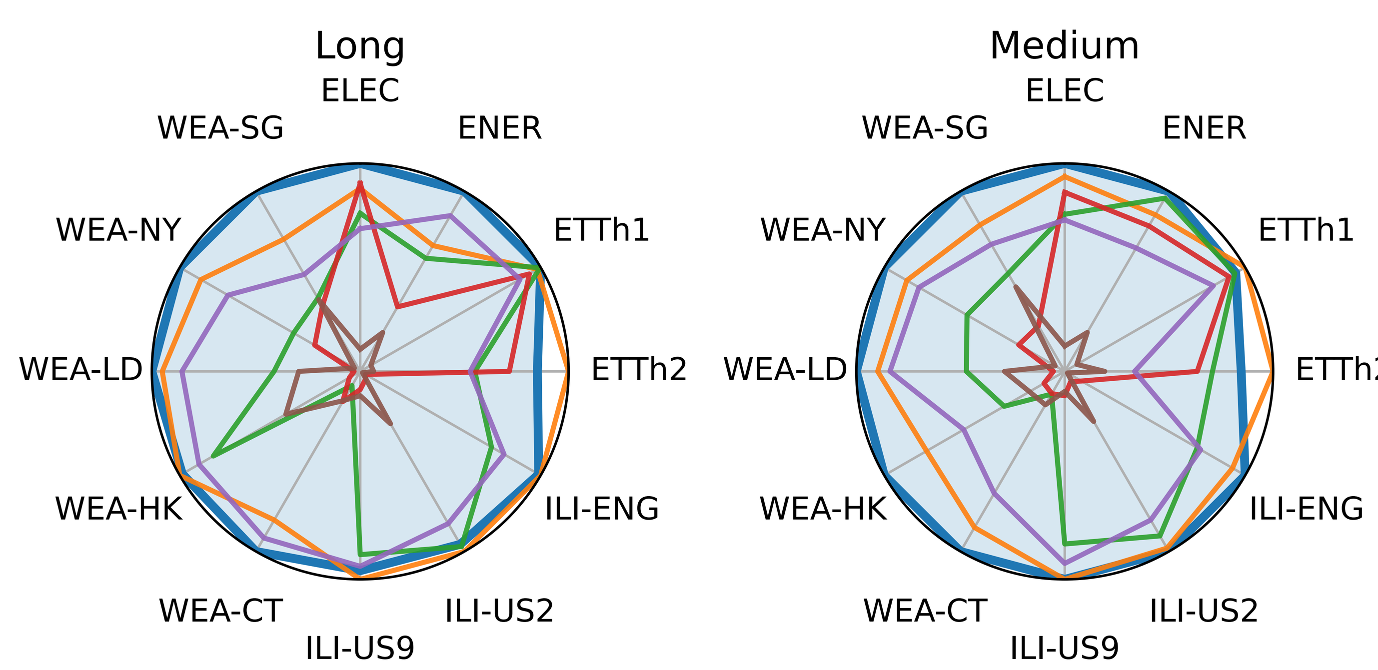
2. Results Overview – Sonnet

Radar plots w.r.t. different forecasting horizons:



Performance Ranking

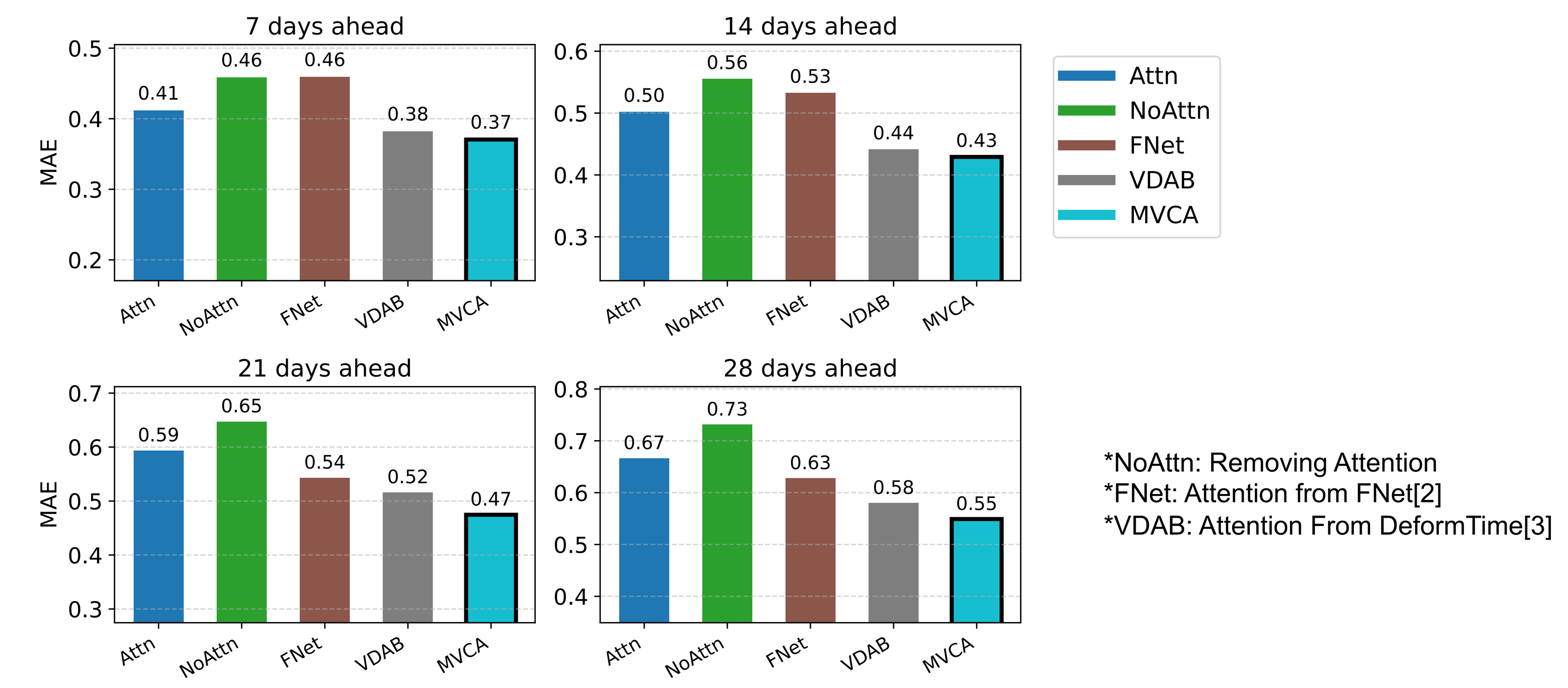
Model	1st	2nd
Sonnet	**34**	10
DeformTime	12	**23**
TimeXer	1	0
Crossformer	0	7
ModernTCN	0	3
Samformer	0	2
PatchTST	0	2
iTransformer	0	0
DLinear	0	0



3. Results Overview – MVCA

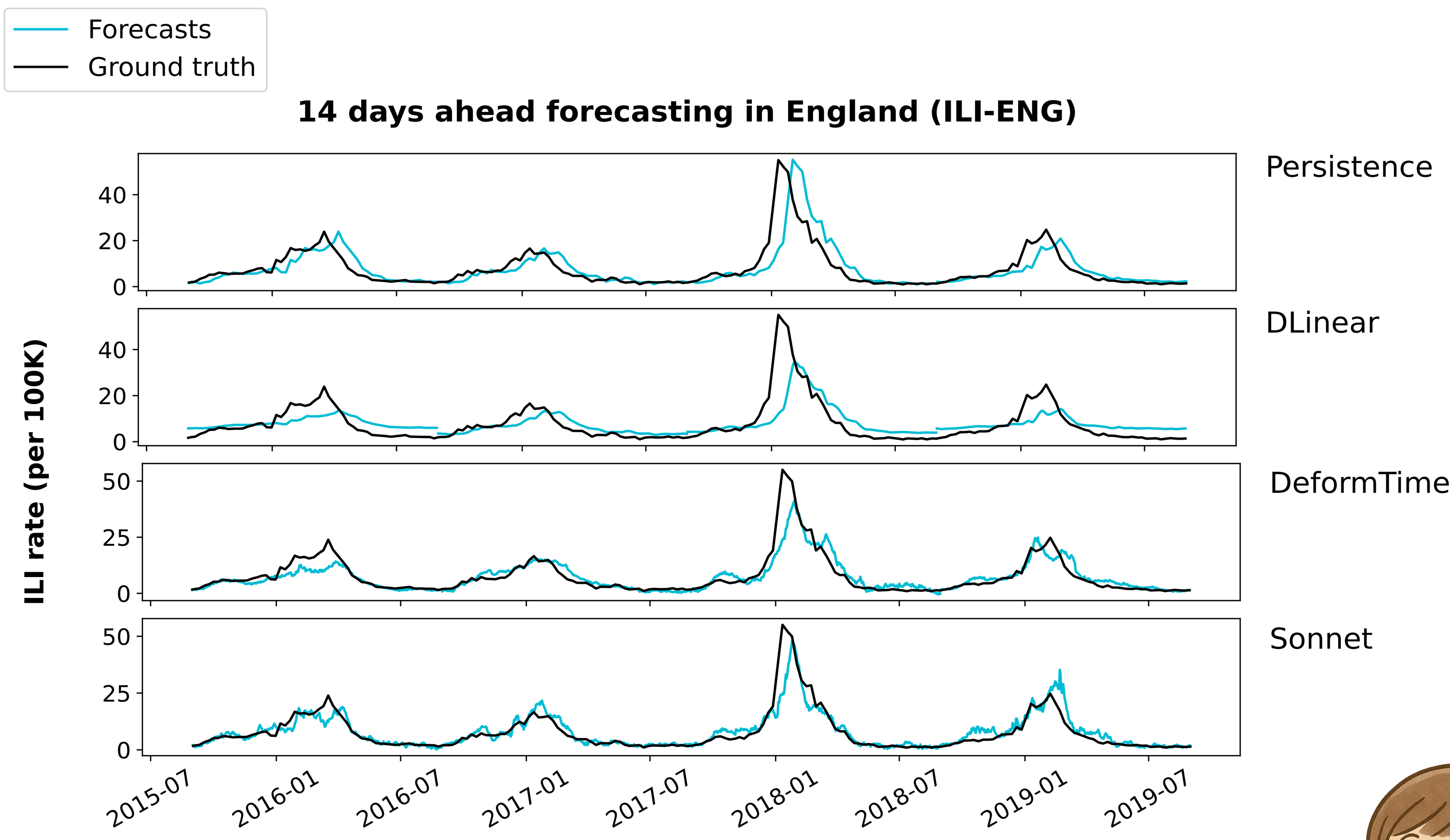
Replacing the Attention module with different modifications to naïve attention

Here we show results using **PatchTST**[1] on the **ILI** forecasting task in **US Region 9**

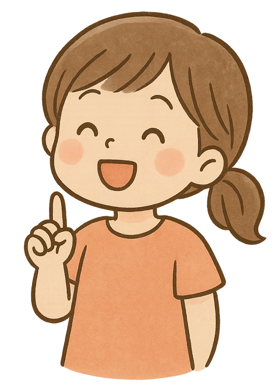


*NoAttn: Removing Attention
*FNet: Attention from FNet[2]
*VDAB: Attention From DeformTime[3]

4. Visualisation over the entire test set



5. Back to our questions …

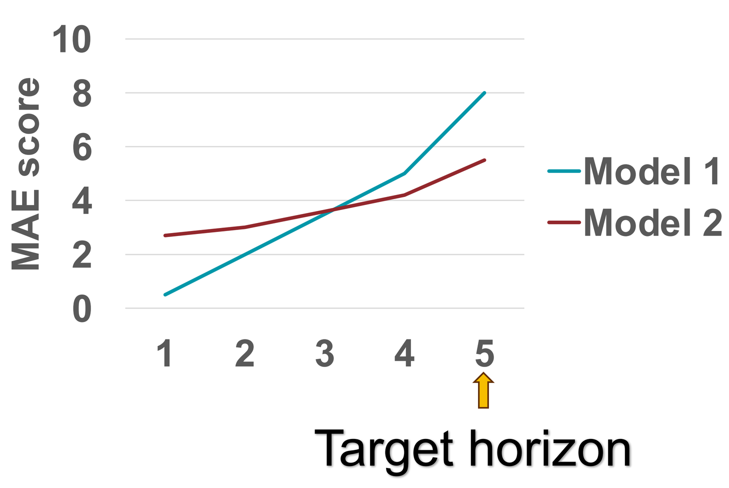


About the datasets

1. Exogenous variables should be **predictive of the target**
   - Check for → Lead lag, Mutual Information, etc.
2. Dataset should **cover a long enough time span**
   - To capture seasonality & ensure generalisablility.

We formed a weather dataset (1979-2018) from **WeatherBench** for multivariable forecasting → **Check it out!**

About evaluation



1. Forecasting becomes more difficult as horizon increases
2. Averaging error across all time steps mixes short- and long-term errors

Errors should be evaluated both over the sequence and at the **target horizon**

code paper dataset

[1] Nie, Y, et al. (2022). A Time Series is Worth 64Words: Long-term Forecasting with Transformers.
[2] Lee-Thorp, J, et al. (2022). Fnet: Mixing tokens with fourier transforms.
[3] Shu, Y., & Lampos, V. (2025). DeformTime: capturing variable dependencies with deformable attention for time series forecasting.

AAAI-26 / IAAI-26 / EAAI-26
JANUARY 20–27, 2026 | SINGAPORE