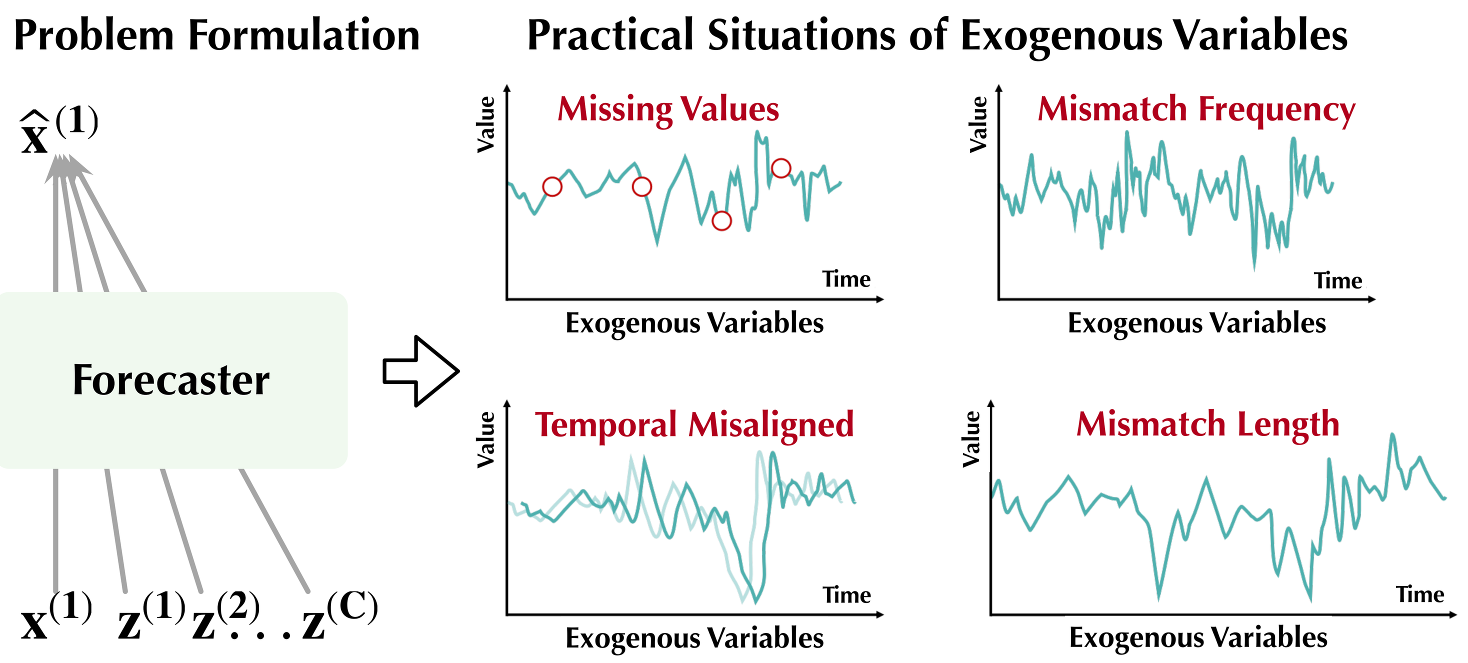
TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables

Yuxuan Wang*, Haixu Wu*, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, Mingsheng Long✉

Introduction

- **Time series forecasting** is of pressing demand in real-world scenarios covering various domains. However, the temporal variations are often influenced by external factors, therefore solely focusing on the target of interest, is insufficient to guarantee accurate prediction.
- **Exogenous variables** are introduced to the time series forecaster for informative purposes and do not need to be predicted.
- **As a practical forecasting scenario**, forecasting with exogenous variables requires the model to reconcile the discrepancy and dependency among *endogenous* and *exogenous* variables.



- **Transformer** has garnered significant interest in time series data due to their ability to capture long-term **temporal dependencies** and complex **multivariate correlations**.

Methods	**TimeXer**	iTran. [23]	PatchTST [28]	Cross. [42]	Auto. [36]	TFT [16]	NBEATSx [29]	TiDE [5]
Univariate	✓	✗	✓	✗	✓	✗	✗	✗
Multivariate	✓	✓	✧	✓	✧	✗	✗	✓
Exogenous	✓	✗	✗	✗	✗	✓	✓	✓

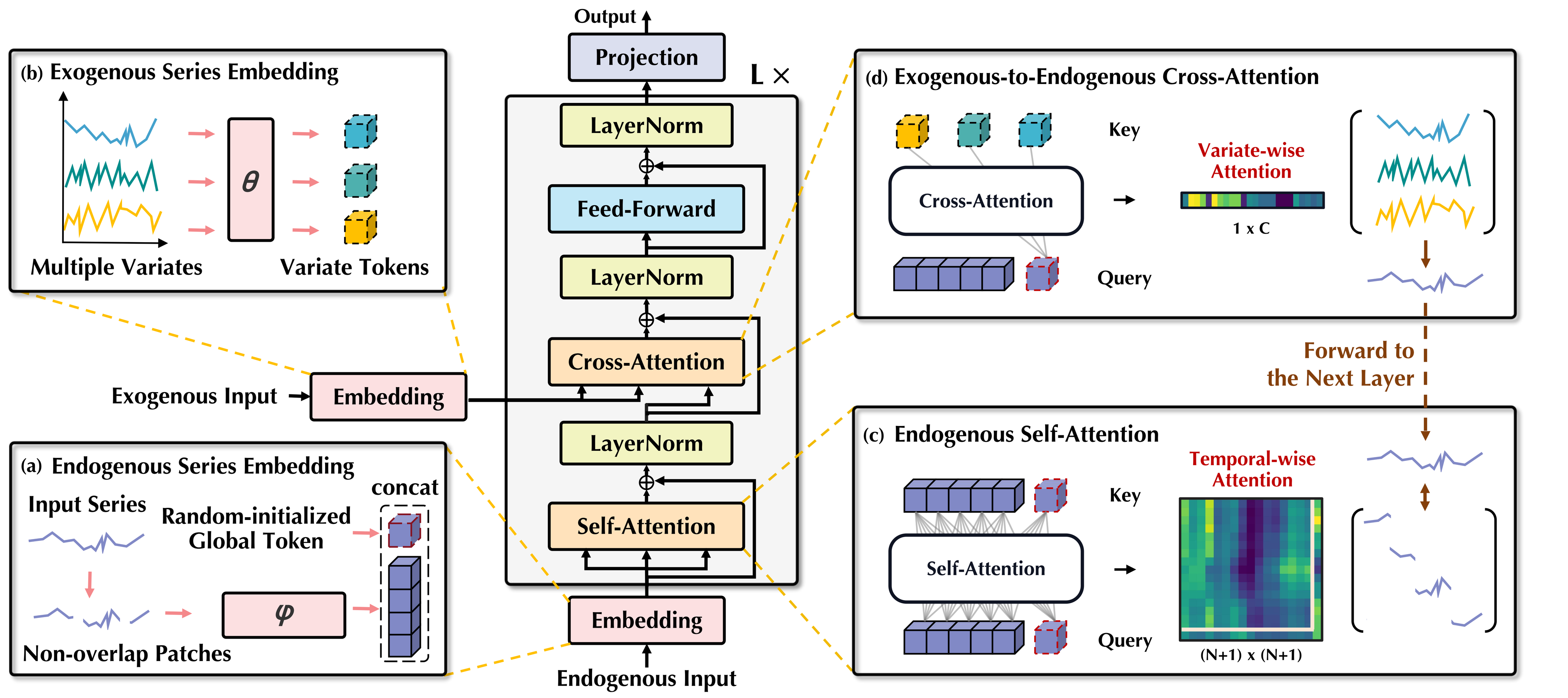
- Existing Transformer-based approaches only focus on multivariate or univariate time series forecasting paradigms and do not conduct special designs for exogenous variables.
- Existing deep models incorporating exogenous variables necessitate the alignment of the endogenous and exogenous series, struggling to handle real-world irregular and heterogeneous data.

Contribution

- Motivated by the universality and importance of exogenous variables in time series forecasting, we empower the canonical Transformer to simultaneously modeling exogenous and endogenous variables without any architectural modifications.
- We propose a simple and general TimeXer model, which employs patch-level and variate-level representations respectively for endogenous and exogenous variables, with an endogenous global token as a bridge in-between. With this design, TimeXer can capture intra-endogenous temporal dependencies and exogenous-to-endogenous correlations jointly.
- Extensive experiments on twelve datasets show that TimeXer can better utilize exogenous information to facilitate endogenous forecasting, in both univariate and multivariate settings.

TimeXer Architecture

- The goal of forecasting model $\mathcal{F}_\theta$ parameterized by θ is to predict the future S time steps $\widehat{\mathbf{x}} = \{x_{T+1}, \dots, x_{T+S}\}$ based on both historical observations $\mathbf{x}_{1:T}$ and corresponding exogenous series: $\mathbf{z}_{1:T_{ex}}$ $\widehat{\mathbf{x}}_{T+1:T+S} = \mathcal{F}_\theta(\mathbf{x}_{1:T}, \mathbf{z}_{1:T_{ex}})$.
- In TimeXer, the endogenous and exogenous variables are manipulated by different embedding strategies:
  - The Endogenous variable is embedded into a series of patch-wise temporal token $\mathbf{P}_{en}$ and a learnable global token $\mathbf{G}_{en}$.
  - Each exogenous series is embedded in a series-wise variate token $\mathbf{V}_{ex}$.



- **Endogenous Self-Attention** is applied over endogenous temporal tokens and learnable global token to **aggregating** patch-level information across the entire series and (2) **dispatching** the variate-level correlations.

$$\widehat{\mathbf{P}}^l_{en}, \widehat{\mathbf{G}}^l_{en} = \mathrm{LayerNorm}\left(\left[\mathbf{P}^l_{en}, \mathbf{G}^l_{en}\right] + \text{Self-Attention}\left(\left[\mathbf{P}^l_{en}, \mathbf{G}^l_{en}\right]\right)\right). \quad (1)$$

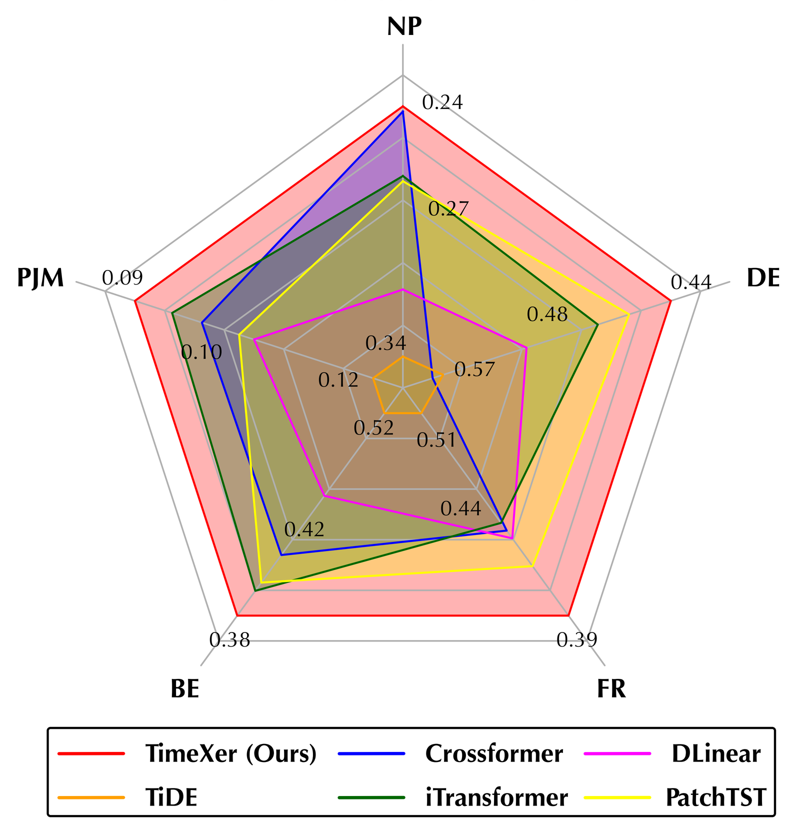
- **Exogenous-to-Endogenous Cross-Attention** is applied between the learned global token of the endogenous and and the exogenous variate tokens to learn correlation between external information and endogenous series.

$$\widehat{\mathbf{G}}^l_{en} = \mathrm{LayerNorm}\left(\widehat{\mathbf{G}}^l_{en} + \text{Cross-Attention}\left(\widehat{\mathbf{G}}^l_{en}, \mathbf{V}_{ex}\right)\right). \quad (2)$$

- **Parallel Multivariate Forecast** can be achieved by employing the channel independence mechanism, for each variable of the multivariate, it is treated as the endogenous one.

Forecasting Results

Performance of TimeXer



5 Short-term Forecasting with Exogenous Variables Benchmarks

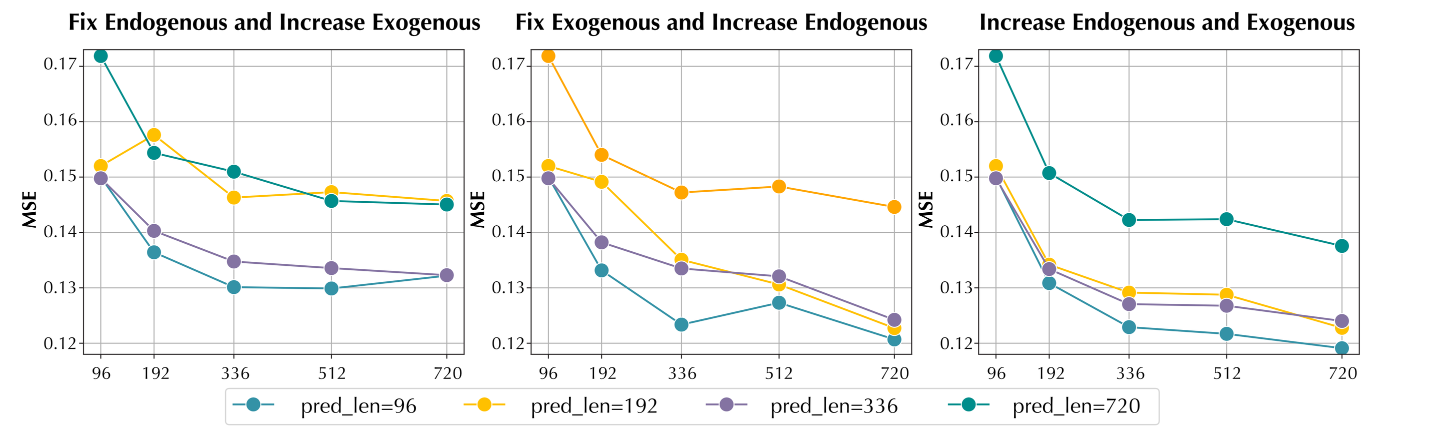
Model	TimeXer		iTransformer		RLinear		PatchTST		Crossformer		TiDE		TimesNet		DLinear		SCINet		Autoformer	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	**0.236**	**0.268**	0.265	0.300	0.335	0.340	0.267	0.284	0.240	0.285	0.335	0.340	0.250	0.289	0.309	0.321	0.373	0.368	0.402	0.398
PJM	**0.093**	**0.192**	0.097	0.197	0.124	0.229	0.106	0.209	0.101	0.199	0.124	0.228	0.097	0.195	0.108	0.215	0.143	0.259	0.168	0.267
BE	**0.379**	**0.243**	0.394	0.270	0.520	0.337	0.400	0.262	0.420	0.290	0.523	0.336	0.419	0.288	0.463	0.313	0.731	0.412	0.500	0.333
FR	**0.385**	**0.208**	0.439	0.233	0.507	0.290	0.411	0.220	0.434	0.208	0.510	0.290	0.431	0.234	0.429	0.260	0.855	0.384	0.519	0.295
DE	**0.440**	**0.415**	0.479	0.443	0.574	0.498	0.461	0.432	0.574	0.430	0.568	0.496	0.502	0.446	0.520	0.463	0.565	0.497	0.674	0.544
AVG	**0.307**	**0.265**	0.335	0.289	0.412	0.339	0.330	0.282	0.354	0.284	0.412	0.338	0.340	0.290	0.366	0.314	0.533	0.384	0.453	0.368

7 Long-term Multivariate Forecasting Benchmarks

Model	TimeXer		iTransformer		RLinear		PatchTST		Crossformer		TiDE		TimesNet		DLinear		SCINet		Autoformer	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ECL	**0.171**	**0.270**	0.178	0.270	0.219	0.298	0.205	0.290	0.244	0.334	0.251	0.244	0.192	0.295	0.212	0.300	0.268	0.365	0.227	0.338
Weather	**0.241**	**0.271**	0.258	0.278	0.272	0.291	0.259	0.281	0.259	0.315	0.271	0.320	0.259	0.287	0.265	0.317	0.292	0.363	0.338	0.382
ETTh1	**0.437**	**0.437**	0.454	0.447	0.446	0.434	0.469	0.454	0.529	0.522	0.541	0.507	0.458	0.450	0.456	0.452	0.747	0.647	0.496	0.487
ETTh2	**0.367**	**0.396**	0.383	0.407	0.374	0.398	0.387	0.407	0.942	0.684	0.611	0.550	0.414	0.427	0.559	0.515	0.954	0.723	0.450	0.459
ETTm1	**0.382**	**0.397**	0.407	0.410	0.414	0.407	0.387	0.400	0.512	0.496	0.419	0.419	0.400	0.406	0.403	0.407	0.485	0.481	0.588	0.517
ETTm2	**0.274**	**0.322**	0.288	0.332	0.286	0.327	0.281	0.326	0.757	0.610	0.358	0.404	0.291	0.333	0.350	0.401	0.571	0.537	0.327	0.371
Traffic	0.466	0.287	**0.428**	**0.282**	0.626	0.378	0.481	0.304	0.550	0.304	0.760	0.473	0.620	0.336	0.625	0.383	0.804	0.509	0.628	0.379

Generality of TimeXer

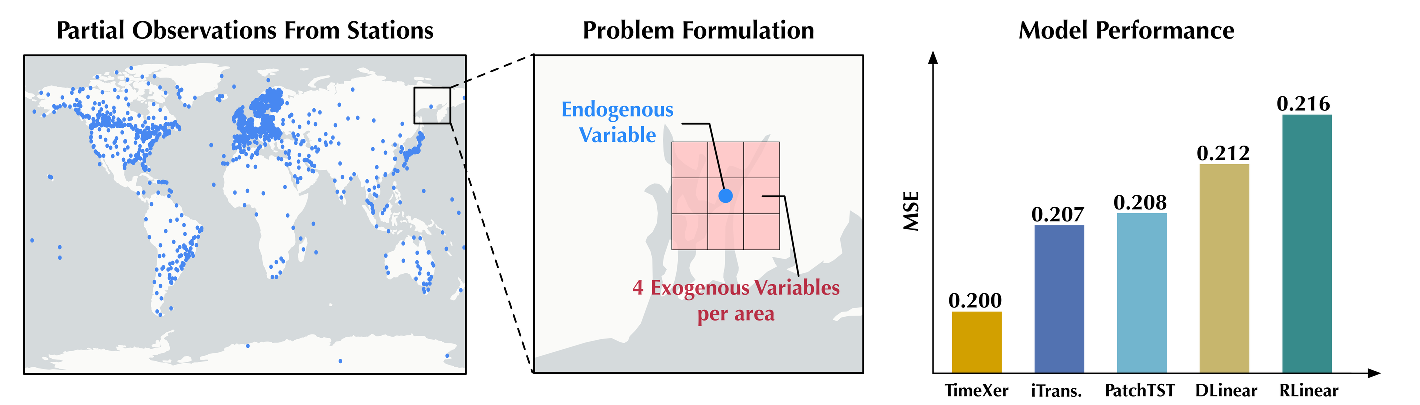
- **Increasing Look-back Length:** TimeXer can be adapted to situations where the look-back of endogenous and exogenous are mismatched. The forecasting performance benefits from enlarged look-back lengths of both endogenous and exogenous series.



- **Missing Values:** Introducing information-rich exogenous variables is the key to TimeXer's forecasting performance improvement.

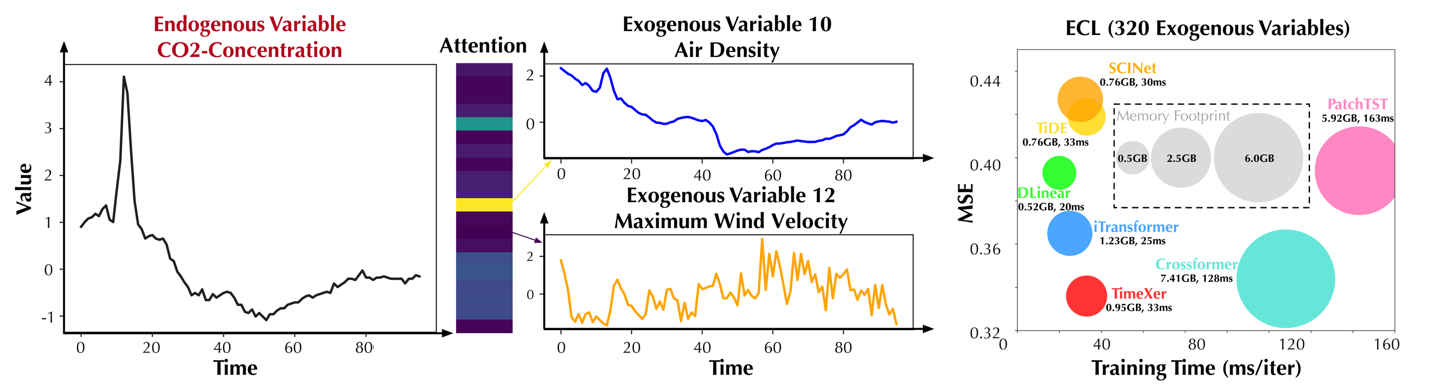
Variate	Strategies	NP		PJM		BE		FR		DE		AVG	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Endogenous	Zeros	2.954	1.396	0.188	0.288	0.930	0.664	0.781	0.534	0.774	0.559	1.125	0.688
	Random	3.140	1.450	0.233	0.325	0.926	0.667	0.761	0.527	0.692	0.533	1.150	0.701
Exogenous	Zeros	0.257	0.278	0.108	0.210	0.400	0.254	0.416	0.214	0.471	0.430	0.330	0.277
	Random	0.258	0.280	0.110	0.212	0.399	0.253	0.424	0.221	0.475	0.432	0.333	0.280
TimeXer		**0.236**	**0.268**	**0.093**	**0.192**	**0.379**	**0.243**	**0.385**	**0.208**	**0.440**	**0.415**	**0.307**	**0.265**

- **Scalability:** TimeXer achieves superior performance on large-scale meteorology datasets, where the endogenous series is the hourly temperature from the weather station and the exogenous information is corresponding meteorological indicators from adjacent areas.



Model Analysis

- **Variate-wise Correlations:** The learned cross-attention map reveals that TimeXer has the ability to distinguish between exogenous variables, enhancing the models' interpretability.



- **Efficiency:** Benefiting from the cross-attention design between endogenous and exogenous series, TimeXer omits the interaction among various exogenous variate tokens, resulting in favorable model efficiency with optimal forecasting performance.