



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✉**



Yuxuan Wang



Haixu Wu



Jiaxiang Dong



Guo Qin



Haoran Zhang



Yong Liu



Yunzhong Qiu



Jianmin Wang



Mingsheng Long

Time Series Forecasting

Wide Applications



Energy Consumption



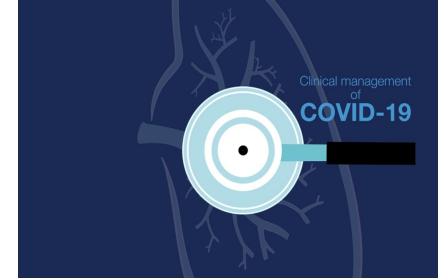
Traffic Flow



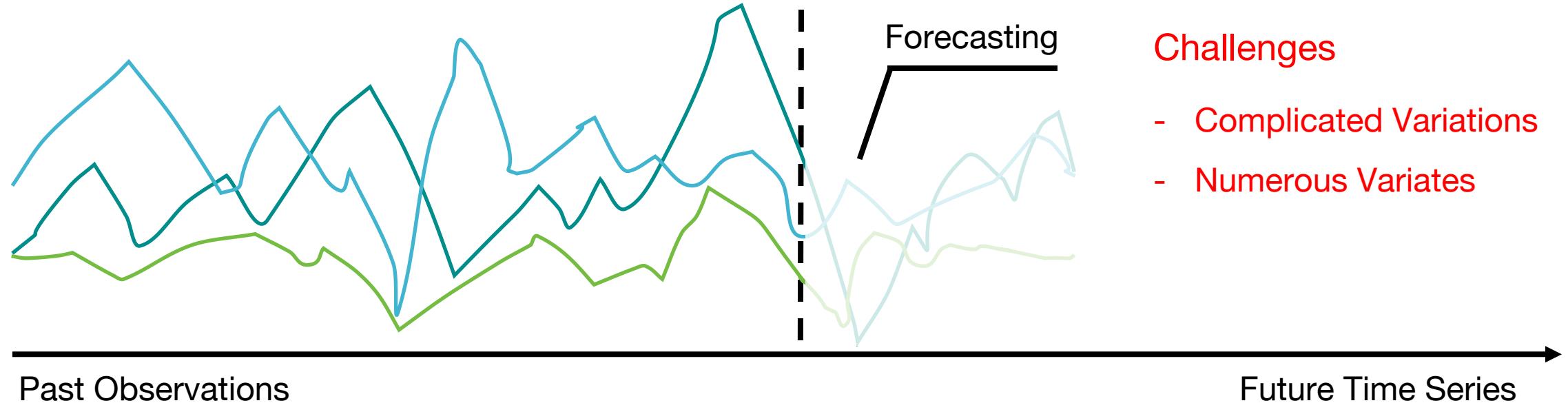
Economic Changes



Weather Variations

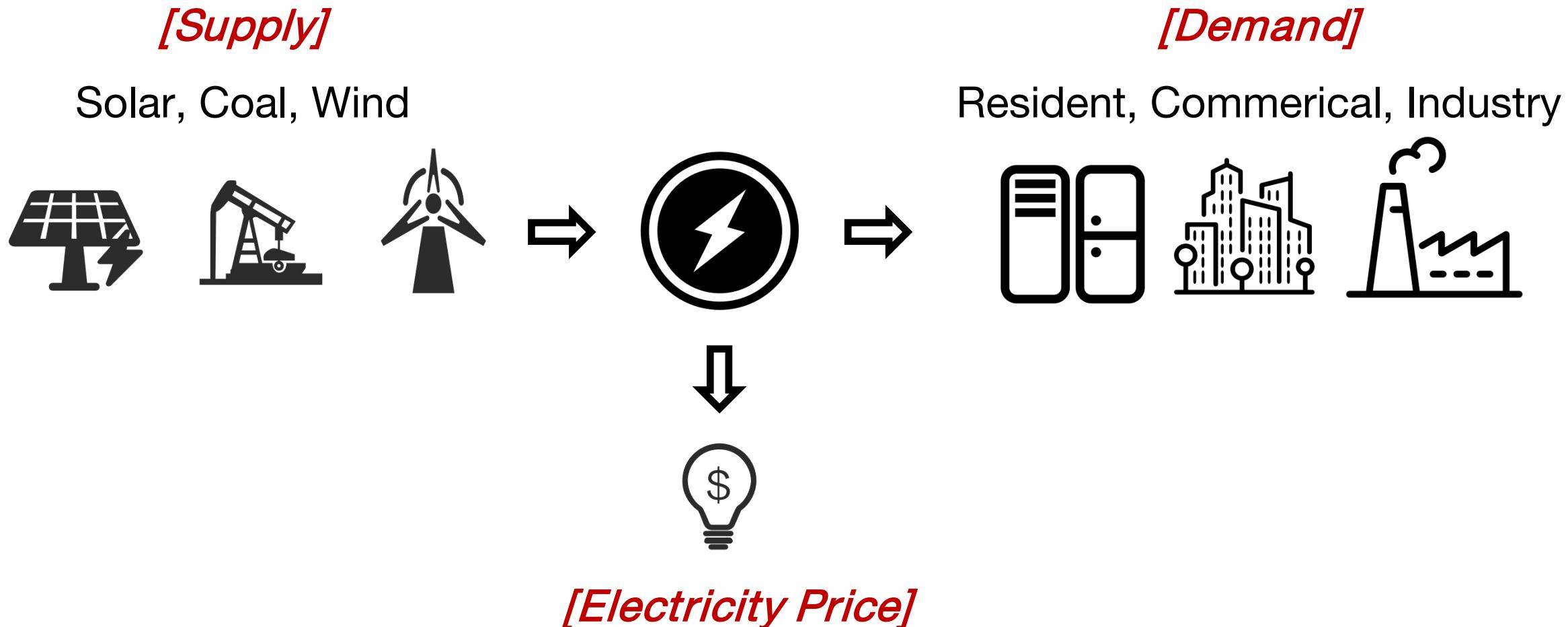


Disease Estimations



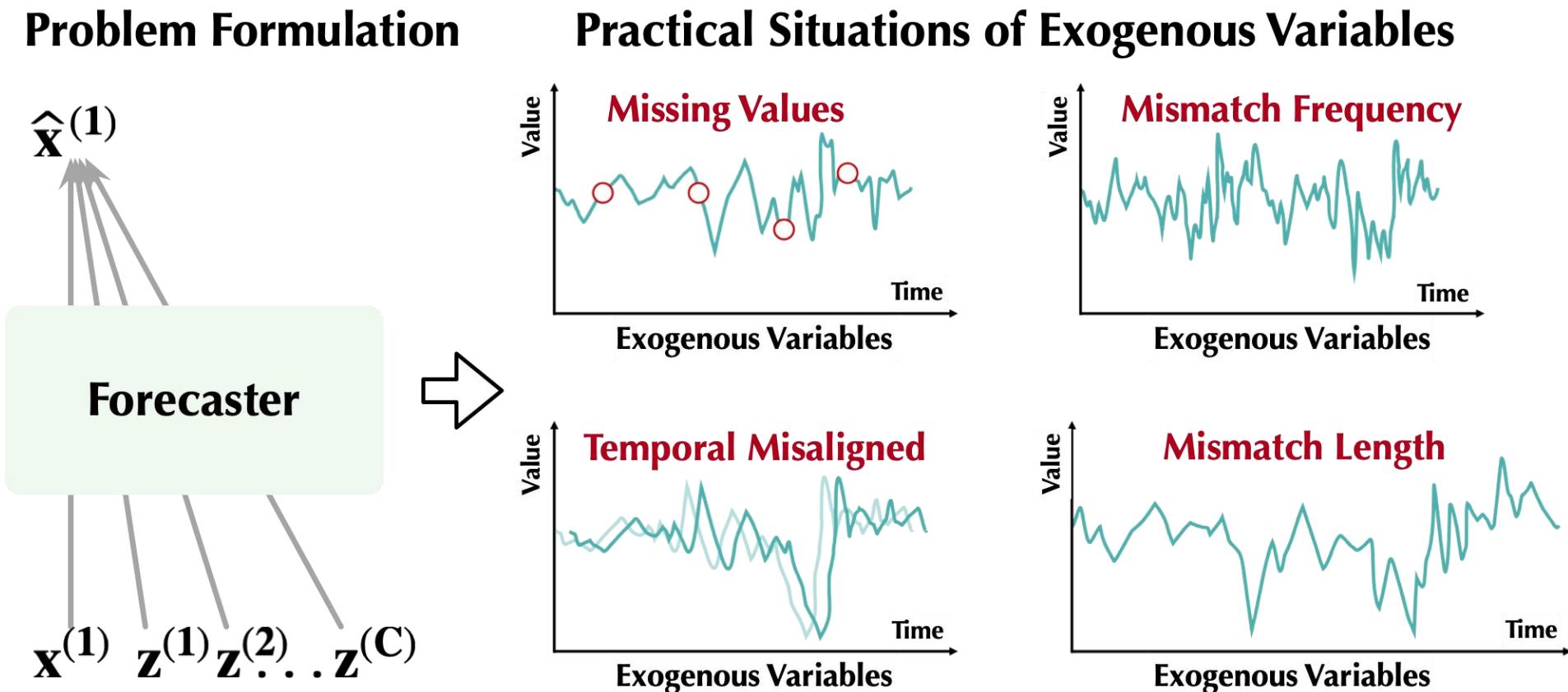
Practical Time Series Forecasting Scenario

- Temporal variations are often influenced by external factor (exogenous variables)



Forecasting with Exogenous Variables

- The exogenous variables are introduced to the forecaster for informative purposes without the need for forecasting.

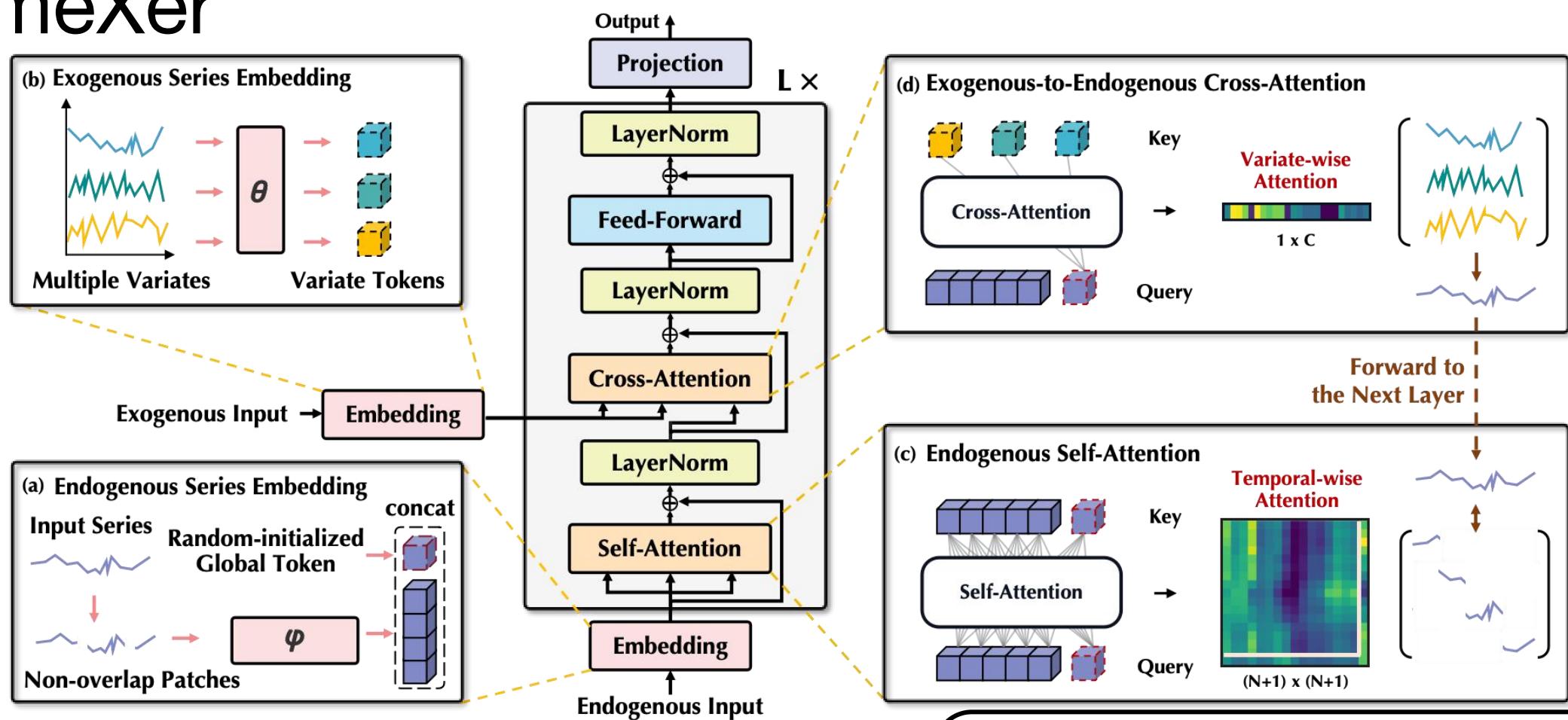


Recent deep time series models

- Most of the existing Transformer-based approaches treat all the variables equally or ignore exogenous information, lacking of special design of exogenous series.
- Previous forecasters designed for exogenous variables overlook the complex situation of exogenous variables.

Methods	TimeXer	iTran. [23]	PatchTST [28]	Cross. [43]	Auto. [37]	TFT [16]	NBEATSx [29]	TiDE [5]
Univariate	✓	✗	✓	✗	✓	✗	✗	✗
Multivariate	✓	✓	❖	✓	❖	✗	✗	✓
Exogenous	✓	✗	✗	✗	✗	✓	✓	✓

TimeXer



- **Endogenous** time series are embedded into patch-level **Temporal Tokens with a learnable Global Token.**
- **Exogenous** time series are embedded to **Variate Tokens**.

$\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\} = \text{Patchify}(\mathbf{x})$,
 $\mathbf{P}_{\text{en}} = \text{PatchEmbed}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N)$
 $\mathbf{G}_{\text{en}} = \text{Learnable}(\mathbf{x})$.
 $\mathbf{V}_{\text{ex},i} = \text{VariateEmbed}(\mathbf{z}^{(i)}), i \in \{1, \dots, C\}$

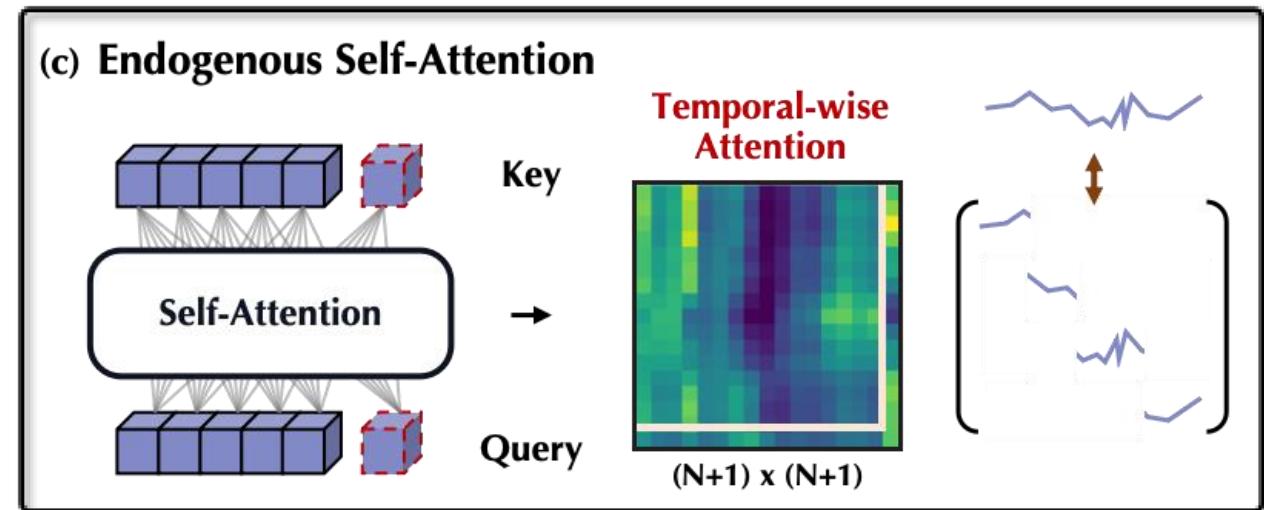
Endogenous Self-Attention

Patch-to-Patch: $\hat{\mathbf{P}}_{\text{en}}^{l,1} = \text{LayerNorm}(\mathbf{P}_{\text{en}}^l + \text{Self-Attention}(\mathbf{P}_{\text{en}}^l)),$

Global-to-Patch: $\hat{\mathbf{P}}_{\text{en}}^{l,2} = \text{LayerNorm}(\mathbf{P}_{\text{en}}^l + \text{Cross-Attention}(\mathbf{P}_{\text{en}}^l, \mathbf{G}_{\text{en}}^l)),$

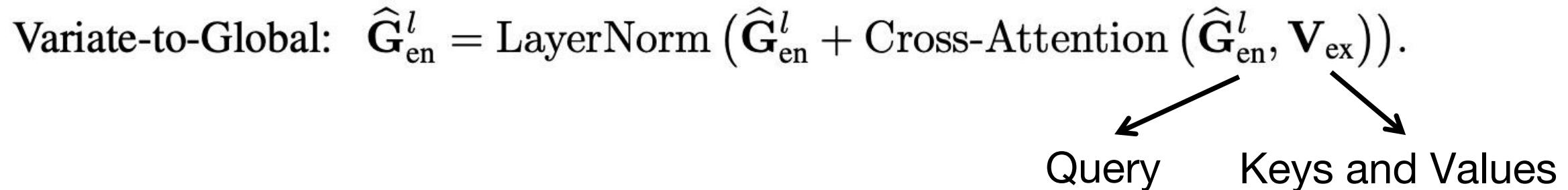
Patch-to-Global: $\hat{\mathbf{G}}_{\text{en}}^l = \text{LayerNorm}(\mathbf{G}_{\text{en}}^l + \text{Cross-Attention}(\mathbf{G}_{\text{en}}^l, \mathbf{P}_{\text{en}}^l)).$

- **Patch-to-Global:** aggregating patch-level information across the entire series.
- **Global-to-Patch:** receiving the variate-level correlations learned by the global token.



Exogenous-to-Endogenous Cross-Attention

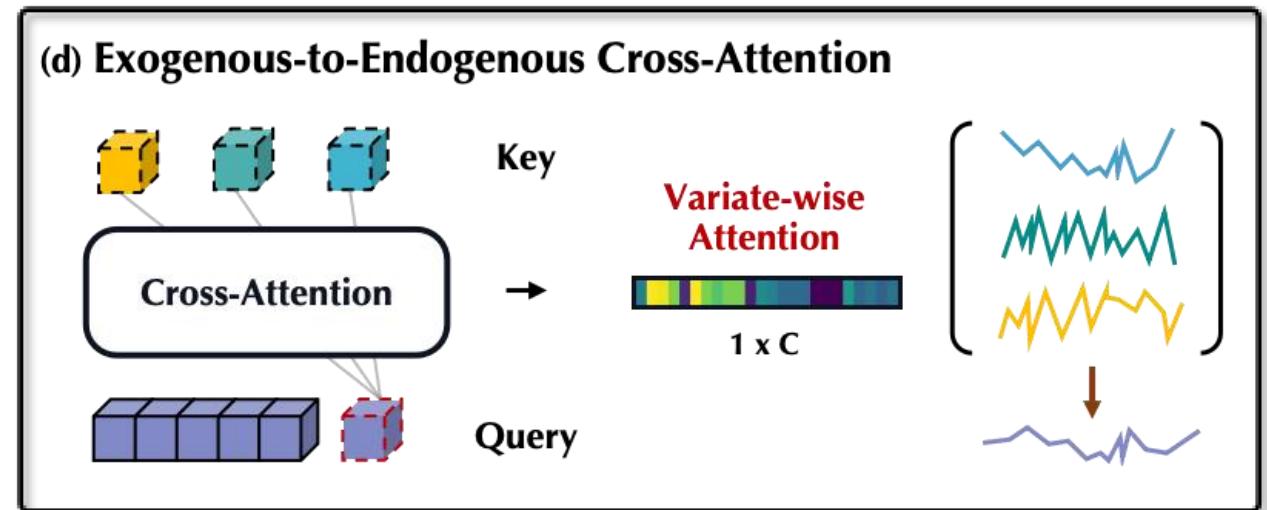
Variate-to-Global: $\hat{\mathbf{G}}_{\text{en}}^l = \text{LayerNorm}(\hat{\mathbf{G}}_{\text{en}}^l + \text{Cross-Attention}(\hat{\mathbf{G}}_{\text{en}}^l, \mathbf{V}_{\text{ex}})).$



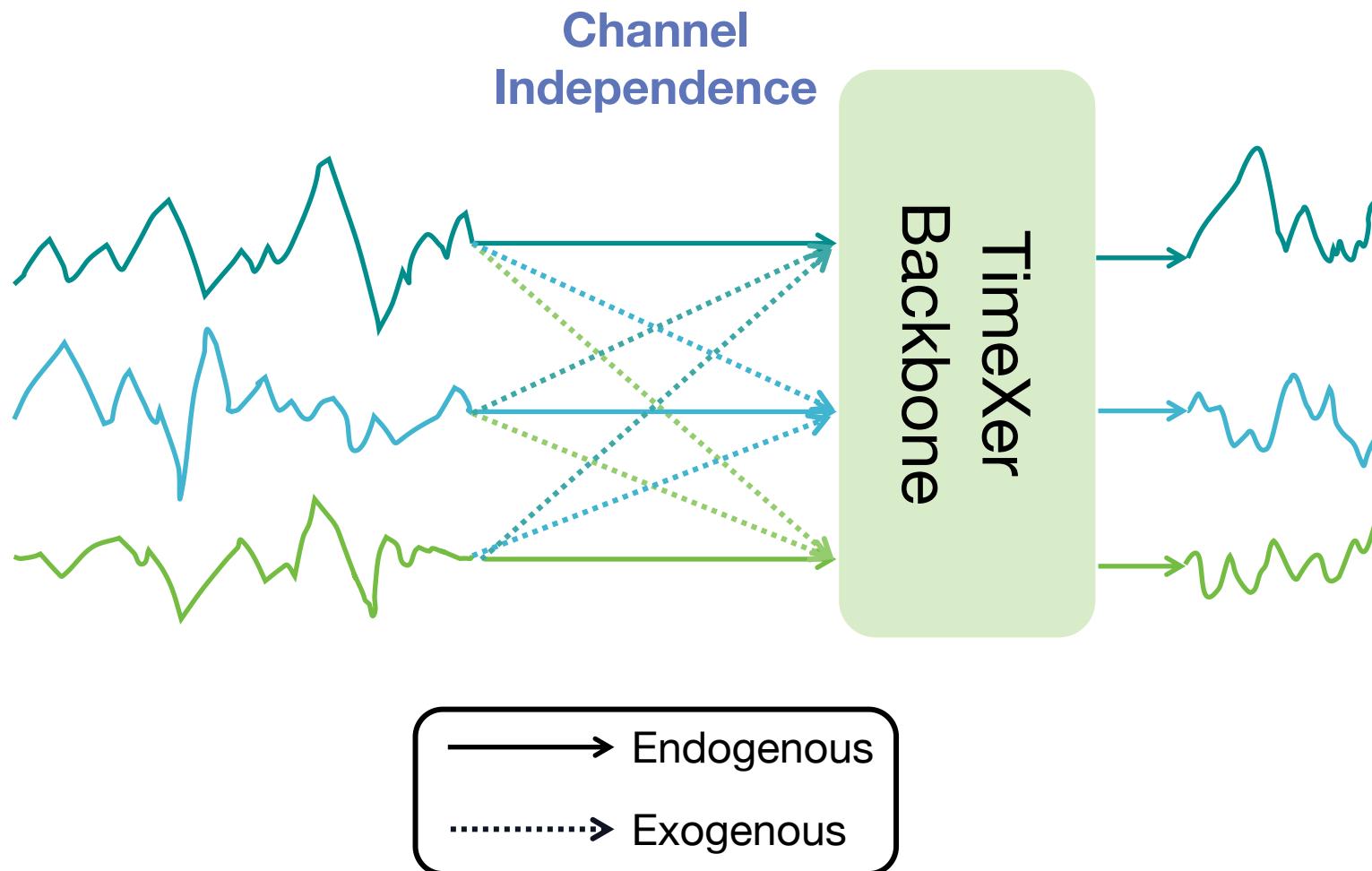
The diagram shows the mathematical expression for the variate-to-global operation. It includes three arrows pointing from the right side of the equation to the words "Query", "Keys and Values", and "Query" again, indicating that the output is composed of these components.

- Omit the interaction among exogenous variate tokens.
- Only the endogenous global token interact with exogenous tokens.

✓ A favourable $O(C)$ efficiency.



Parallel Multivariate Forecasting



- **Extending to multivariate forecasting:** Each variable is treated as the endogenous one which the others are exogenous and use a **shared TimeXer backbone**.

Experiments

- 5 Short-term Forecasting with Exogenous Benchmark

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

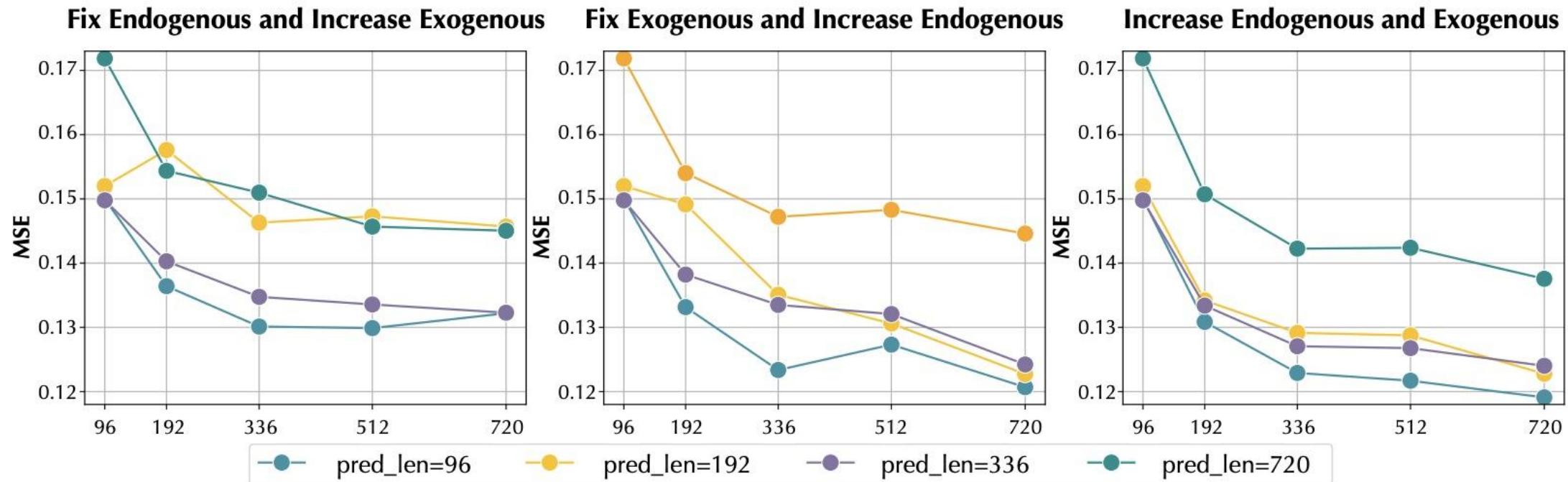
Experiments

- 7 Long-term Multivariate Forecasting Benchmark

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

TimeXer Generality

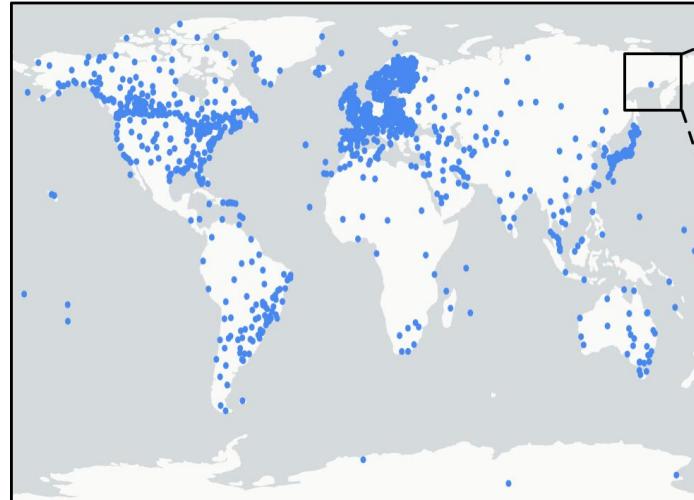
- Increasing the look back length of endogenous or / and exogenous series.
- TimeXer can be adapted to situations where the look-back are mismatched
- The forecasting performance benefits from enlarged look-back lengths of both endogenous and exogenous series.



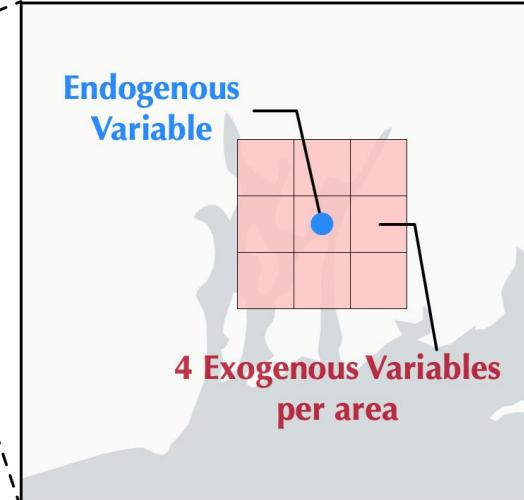
TimeXer Scalability

- The large-scale meteorology datasets
 - Endogenous Variable: temperature collected from weather stations from NCEI
 - Exogenous Variables: meteorological indicators of corresponding adjacent areas derived from ERA5

Partial Observations From Stations



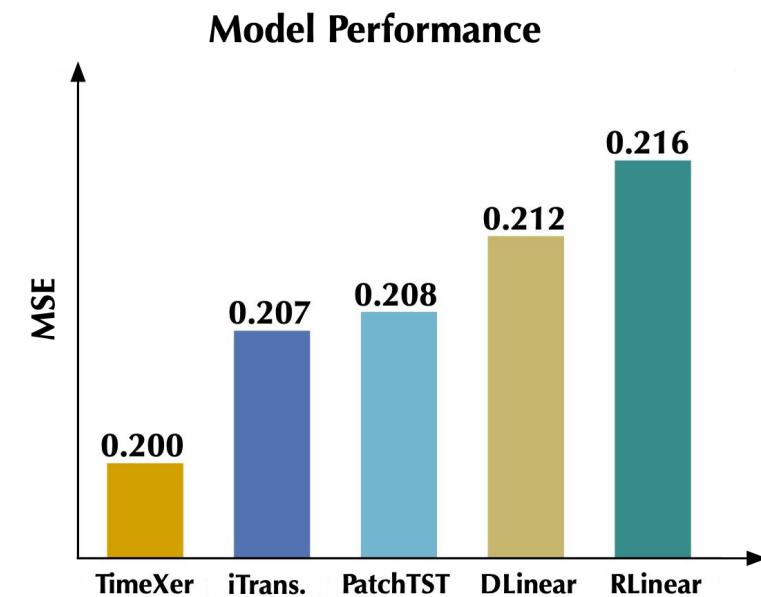
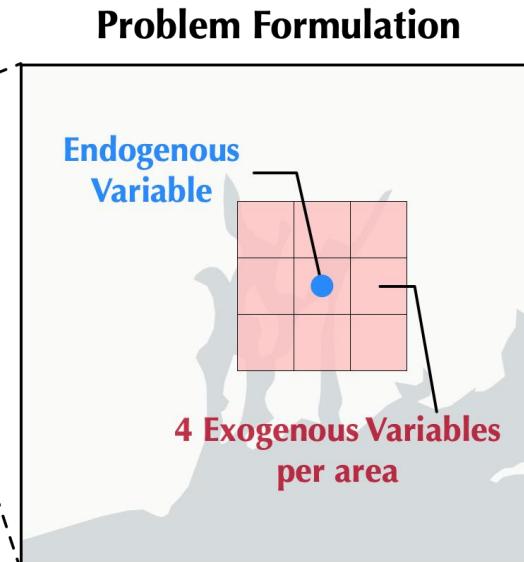
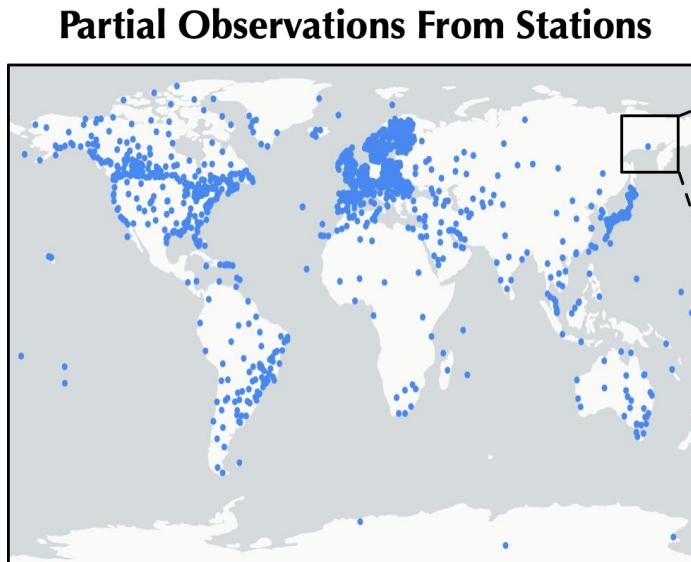
Problem Formulation



- Date Period: January 1, 2019 to December 31, 2020.
- Sample Frequency: 1 hour for endogenous variable and 3 hours for exogenous variables.
- Task: Past 7 days predict future 3 days.

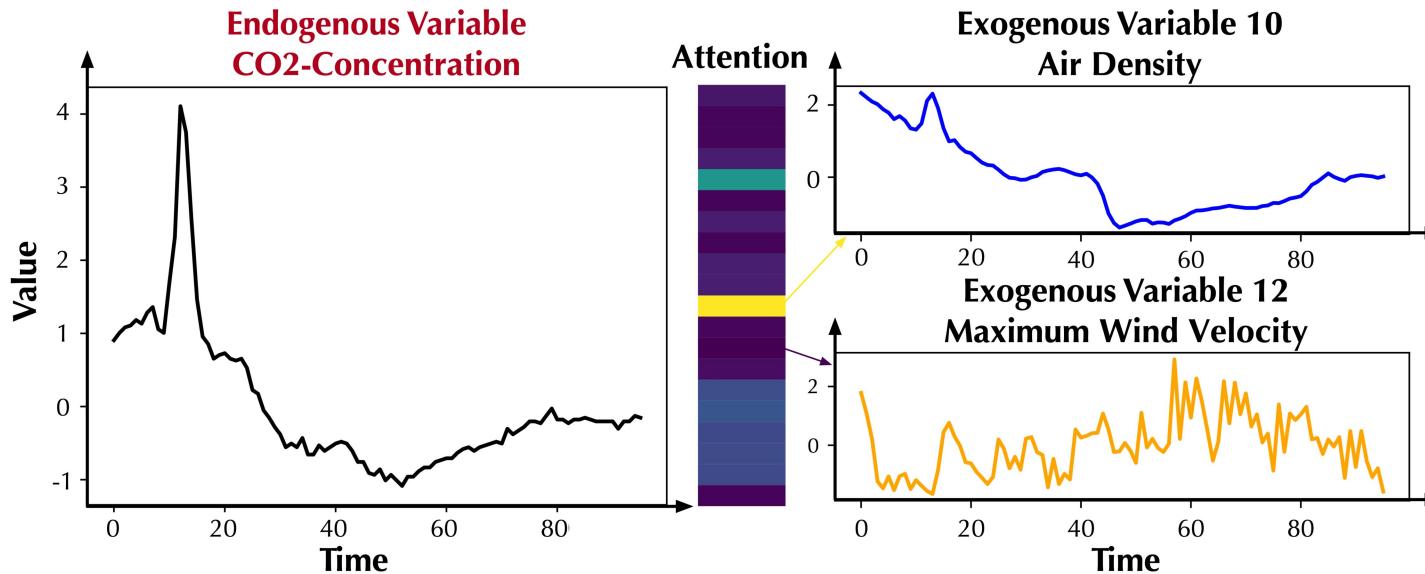
TimeXer Scalability

- The large-scale meteorology datasets
 - Endogenous Variable: temperature collected from weather stations from NCEI
 - Exogenous Variables: meteorological indicators of corresponding adjacent areas derived from ERA5



TimeXer outperforms other advanced forecasters

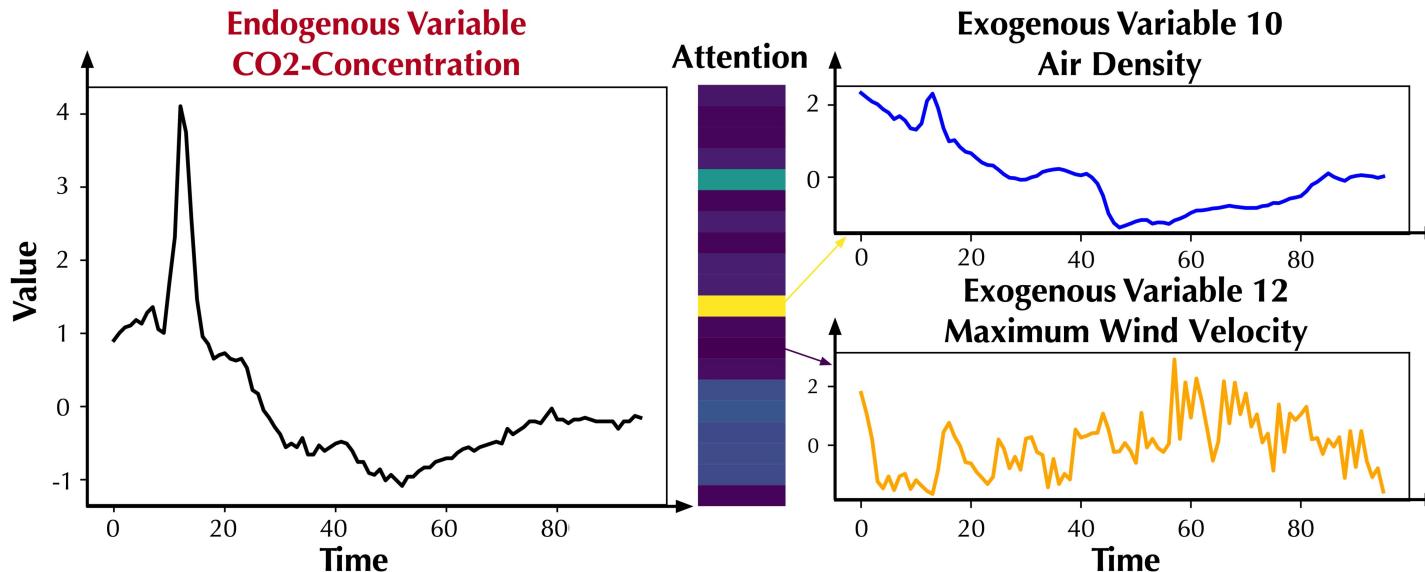
Model Analysis



Attention map can reflect the correlation between endogenous and exogenous variables

- ✓ A notable distinction can be observed in the attention maps of endogenous variables with different exogenous variables.
- ✓ TimeXer has the ability to distinguish between exogenous variables, resulting in a more focused and interpretable attention map.

Model Analysis

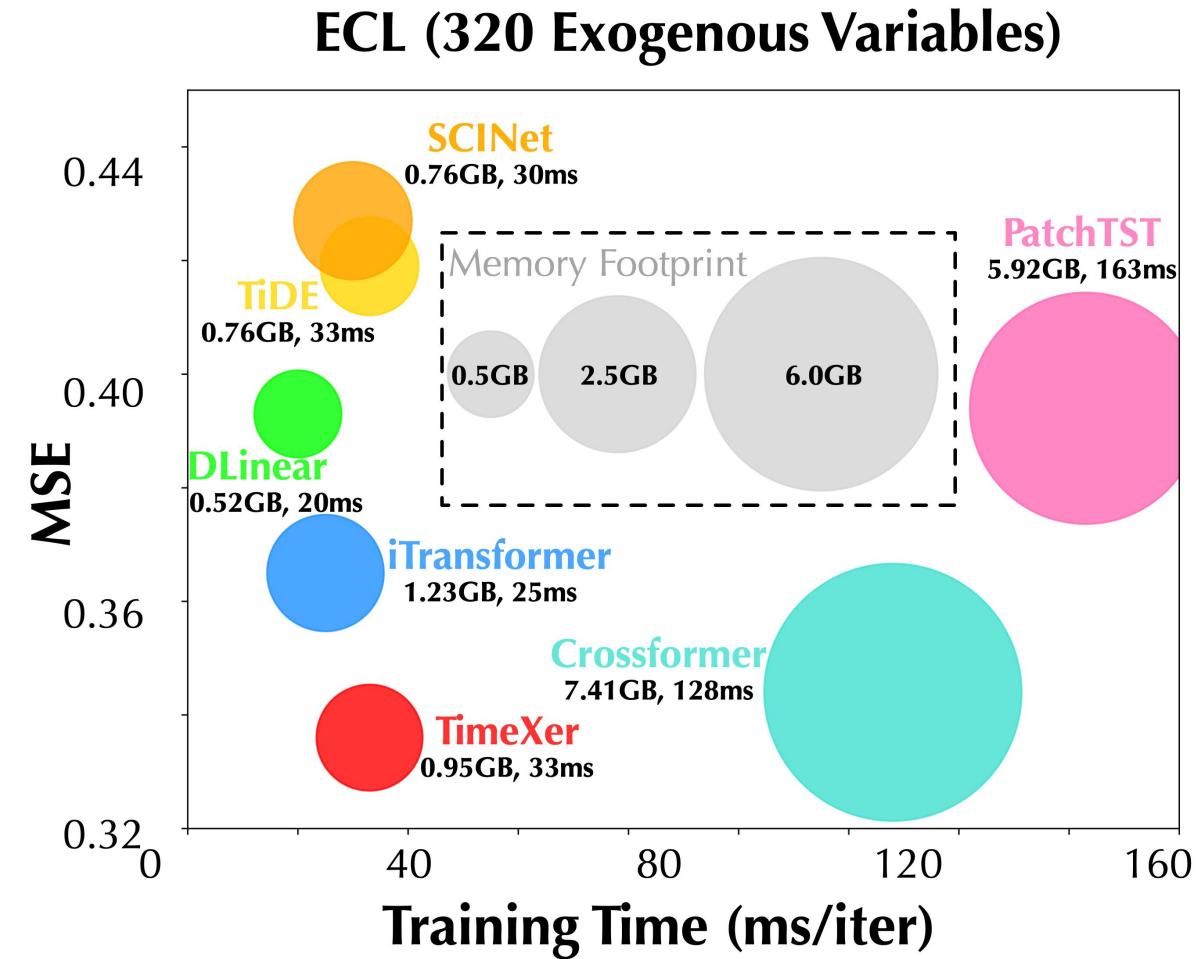


Attention map can reflect the correlation between endogenous and exogenous variables

- ✓ A notable distinction can be observed in the attention maps of endogenous variables with different exogenous variables.
- ✓ **TimeXer has the ability to distinguish between exogenous variables, resulting in a more focused and interpretable attention map.**

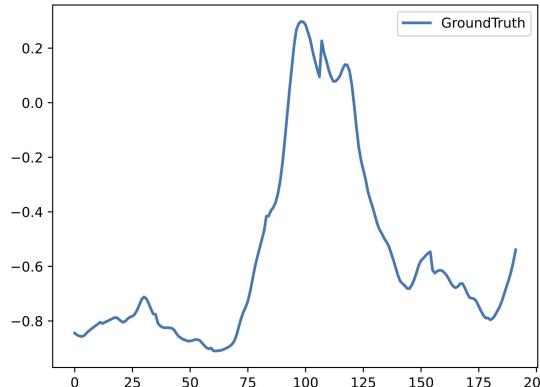
Model Efficiency

- ✓ All models use identical hidden dimension and batch size
- ✓ When faced with numerous variables TimeXer exhibits its advantage since it omits the interaction among learned exogenous variate tokens, resulting in favorable efficiency

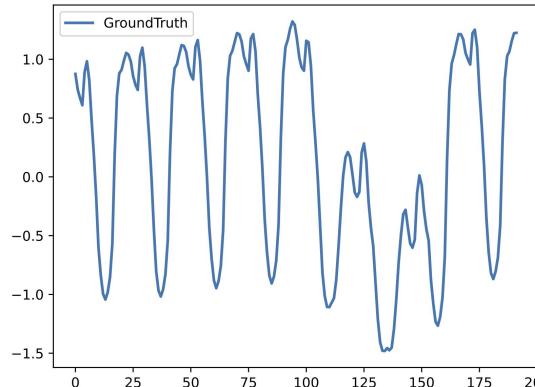


Showcase

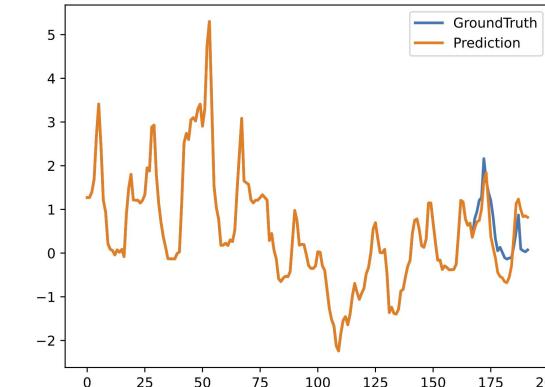
(a) Wind power (Exogenous)



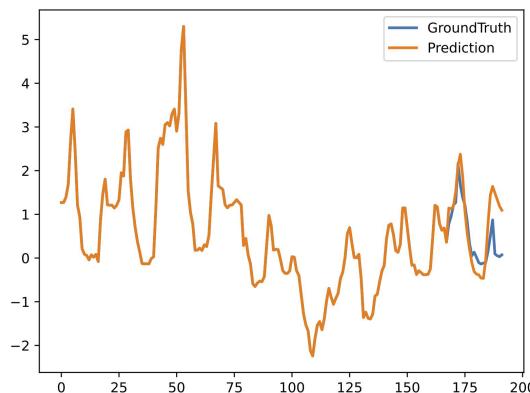
(b) Ampirion zonal load (Exogenous)



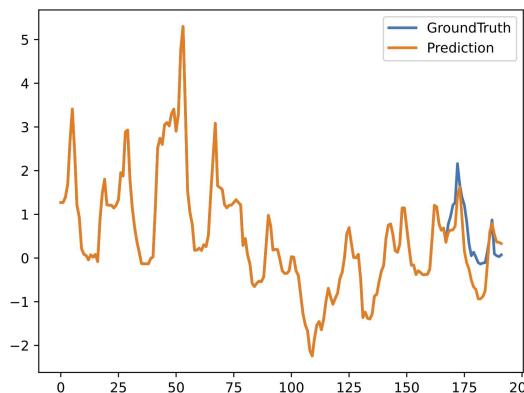
(c) TimeXer (PastEx, PastEn)



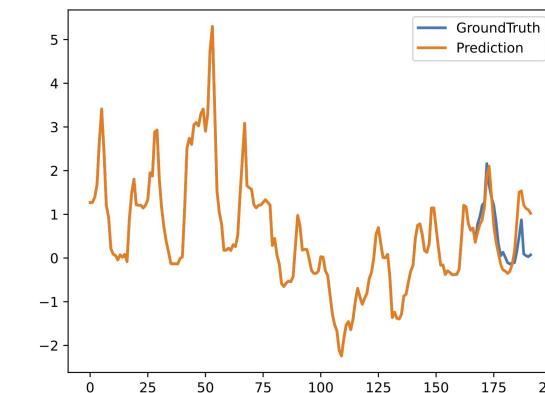
(d) TimeXer (PastEx, None)



(e) TimeXer (None, PastEn)

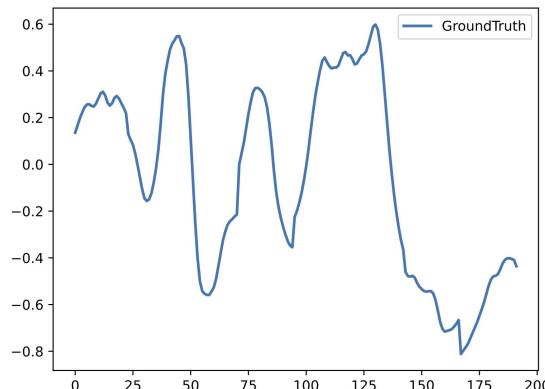


(f) TimeXer (Past+FutureEx, PastEN)

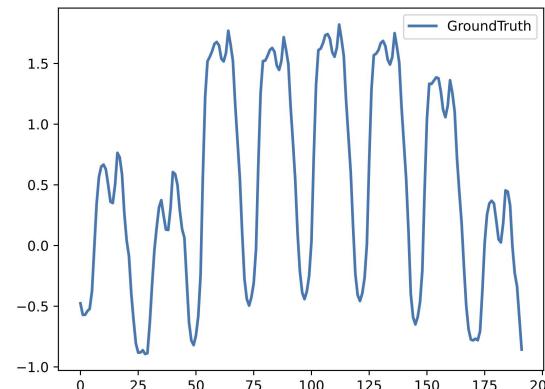


Showcase

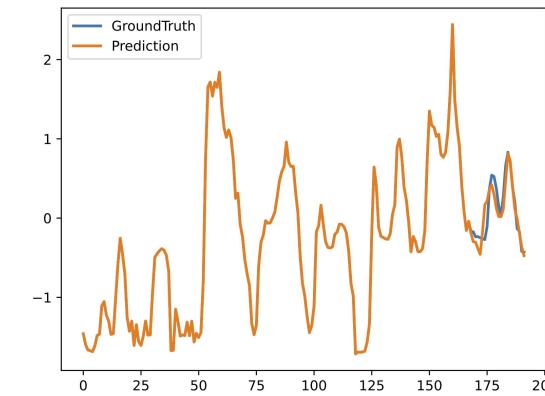
(a) Wind power (Exogenous)



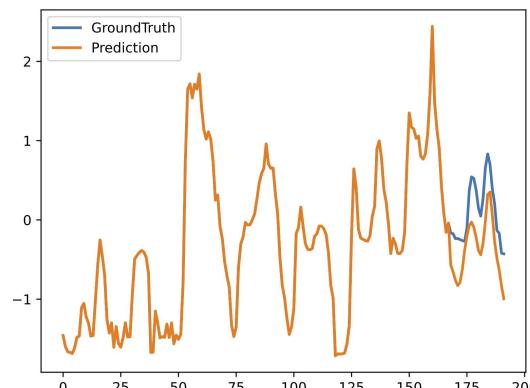
(b) Ampirion zonal load (Exogenous)



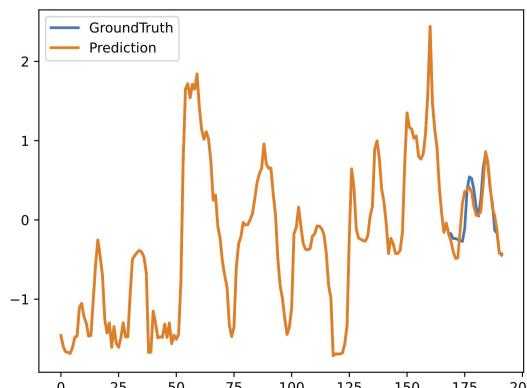
(c) TimeXer (PastEx, PastEn)



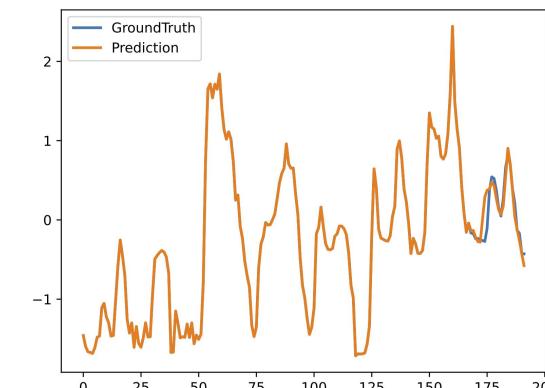
(d) TimeXer (PastEx, None)



(e) TimeXer (None, PastEn)



(f) TimeXer (Past+FutureEx, PastEN)



Thank You!

wangyuxu22@mails.tsinghua.edu.cn



长按关注，获取最新资讯

Code and datasets are available at <https://github.com/thuml/TimeXer>