August 3–7, 2025
KDD2025
The 33th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

TimeCapsule: Solving the Jigsaw Puzzle of Long-Term Time Series Forecasting with Compressed Predictive Representations

Yihang Lu[1,2], Yangyang Xu[1,2], Qitao Qin[2], *Xianwei Meng[1]

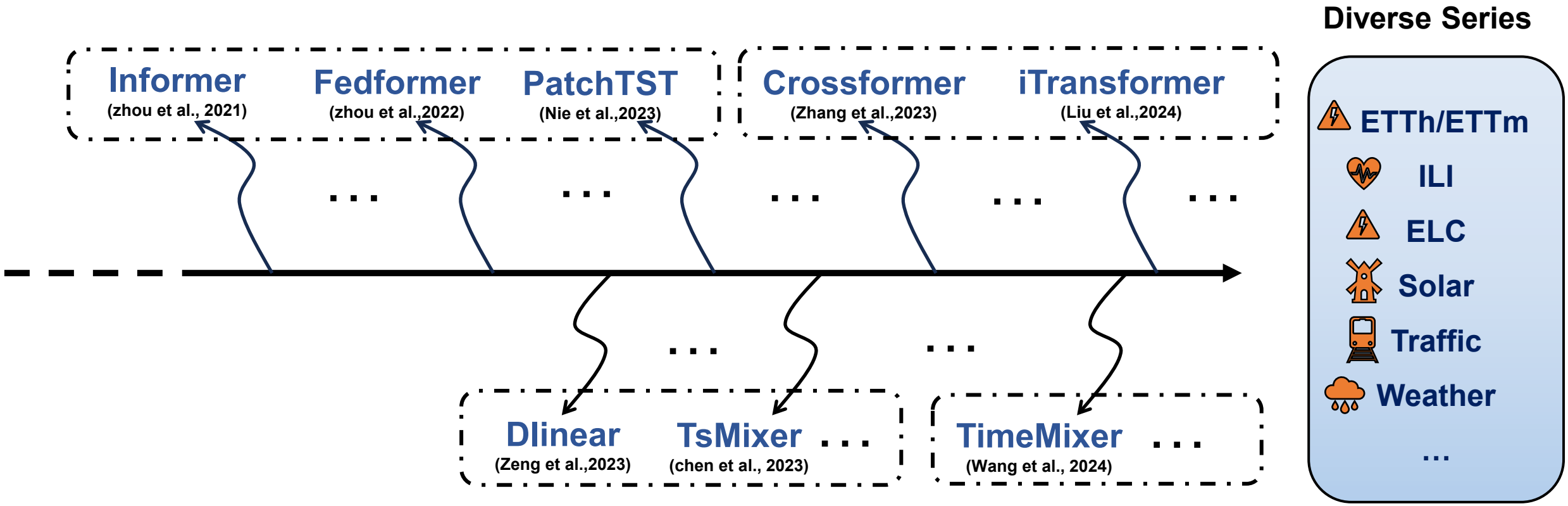
[1] Hefei Institutes of Physical Science, Chinese Academy of Sciences,

[2] University of Science and Technology of China

中国科学技术大学
University of Science and Technology of China

Science Island
中国科学院合肥物质科学研究院
Hefei Institutes of Physical Science, Chinese Academy of Sciences

Review Typical Progress of LTSF



In our view,

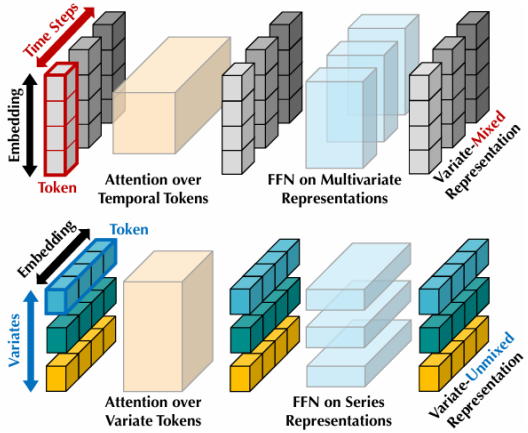
SOTA advances, but the ***underlying focus*** are common in LSTF task.

Each extracts ***one or two critical aspects*** of time series modeling.

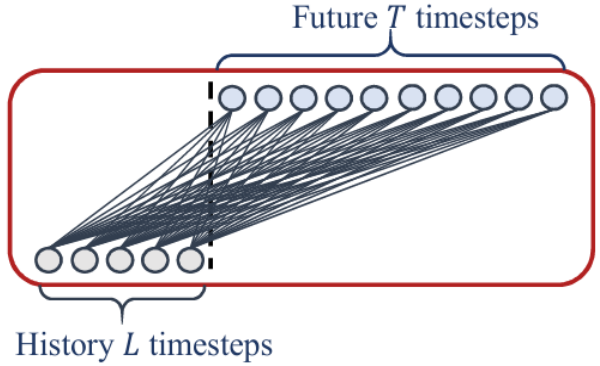
Key Ingredients of "success" LTSF

- Multi-Scale Feature Extraction
- Information Compression / Energy Concentration
  - Informer: condensed attention map
  - Fedformer: compact frequency bases
  - PatchTST: patch-wise aggregation & interaction
- Linear Projection
- Various Series Tokens
- Others (Non-stationary, normalization etc.)

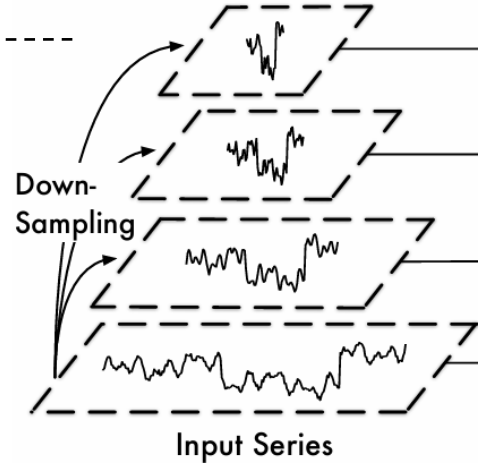
Note:
The follwing figures are taken from the referenced papers.



iTransformer (Liu et al.,)



DLinear (Zeng et al.,)

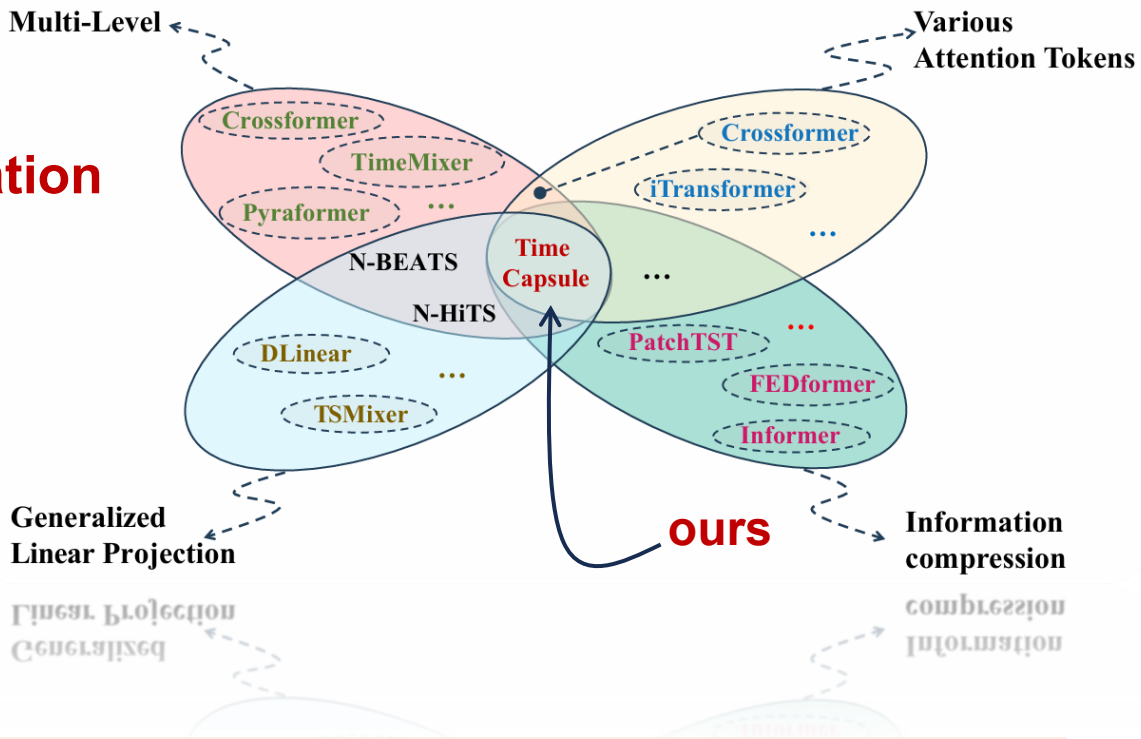


TimeMixer (Wang et al.,)

Question: Is it possible to compactly consider them all? – a Jigsaw Puzzle

Streamline, Generalize, and Integrate

- **Multi-Scale Feature Extraction**
- **Information Compression / Energy Concentration**
- **Linear Projection**
- **Various Series Tokens**
- **Others (Non-stationary, normalization etc.)**

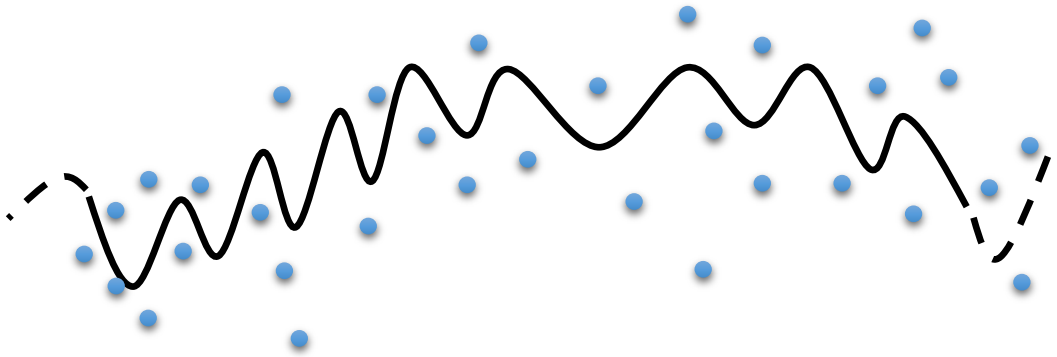


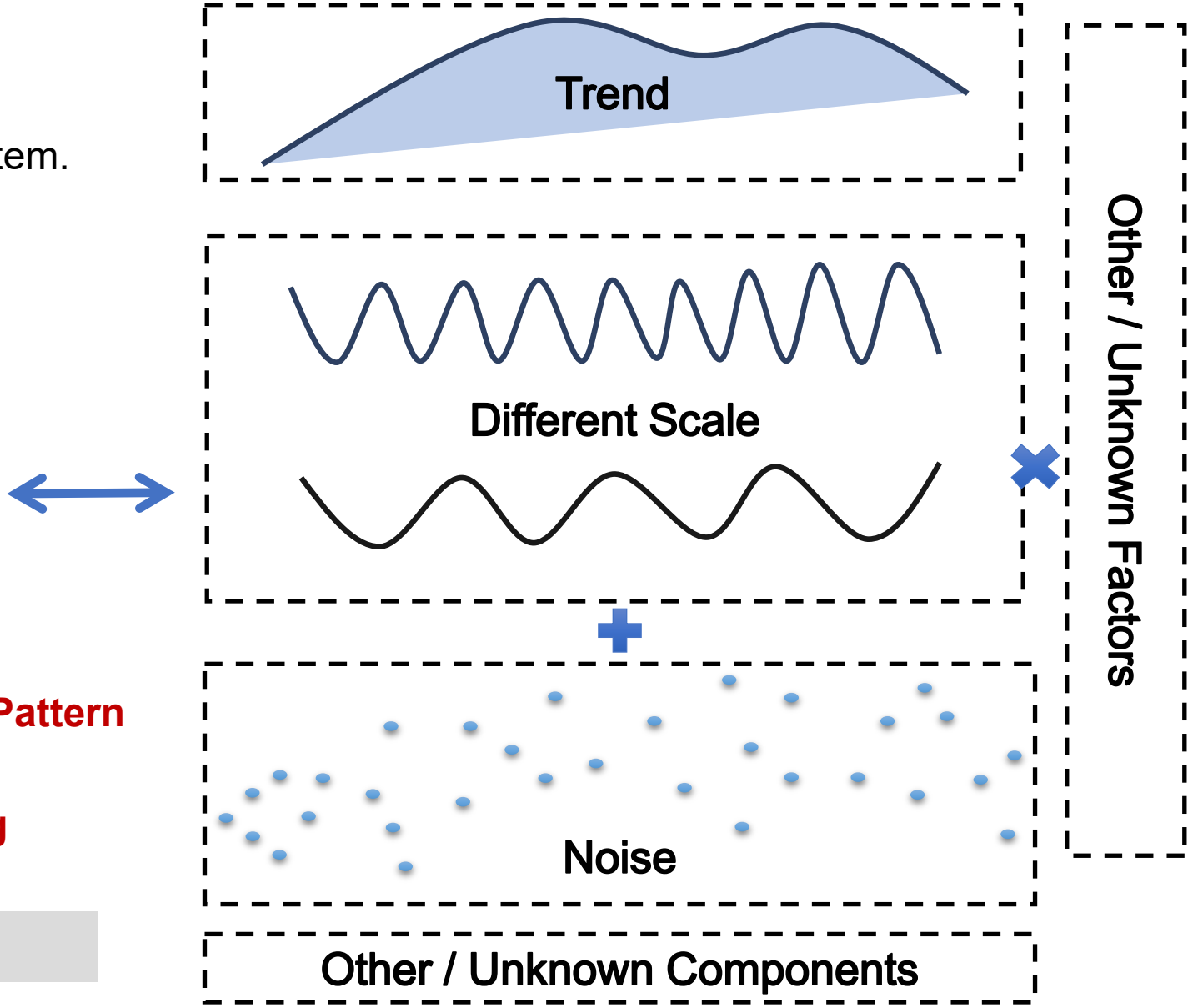
- Component were ***designed and evaluated almost independently.***
- An integral creates a versatile forecaster,

as ***distinct time series often demand different underlying preferences.***

Design Space—Multi-Level

Level typically refers to the expansion centered around a baseline value or steady state of a system.





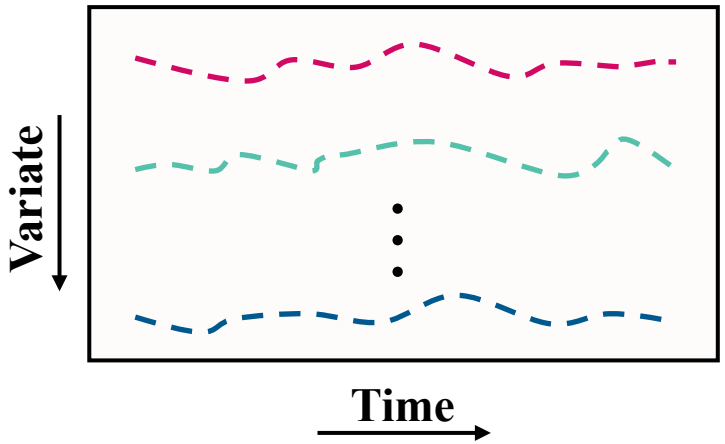
Tradional Decomposition Pattern

&

Multi-Scale Modeling

Manufactually or Automatically?

Design Space—Multi-Level

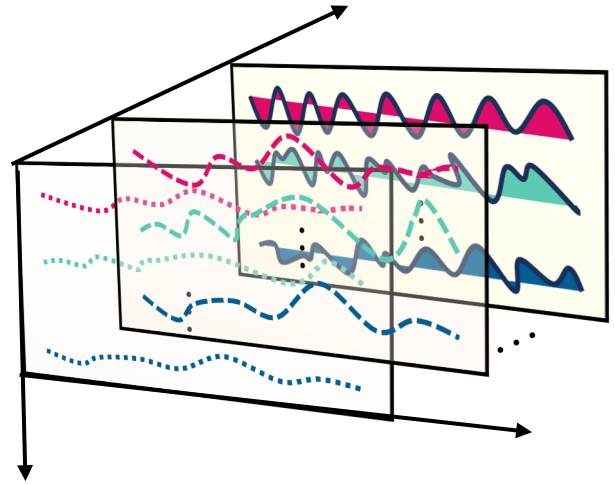


Model Structure

Context

Task

Inductive Bias / Physical Rules



3D modeling

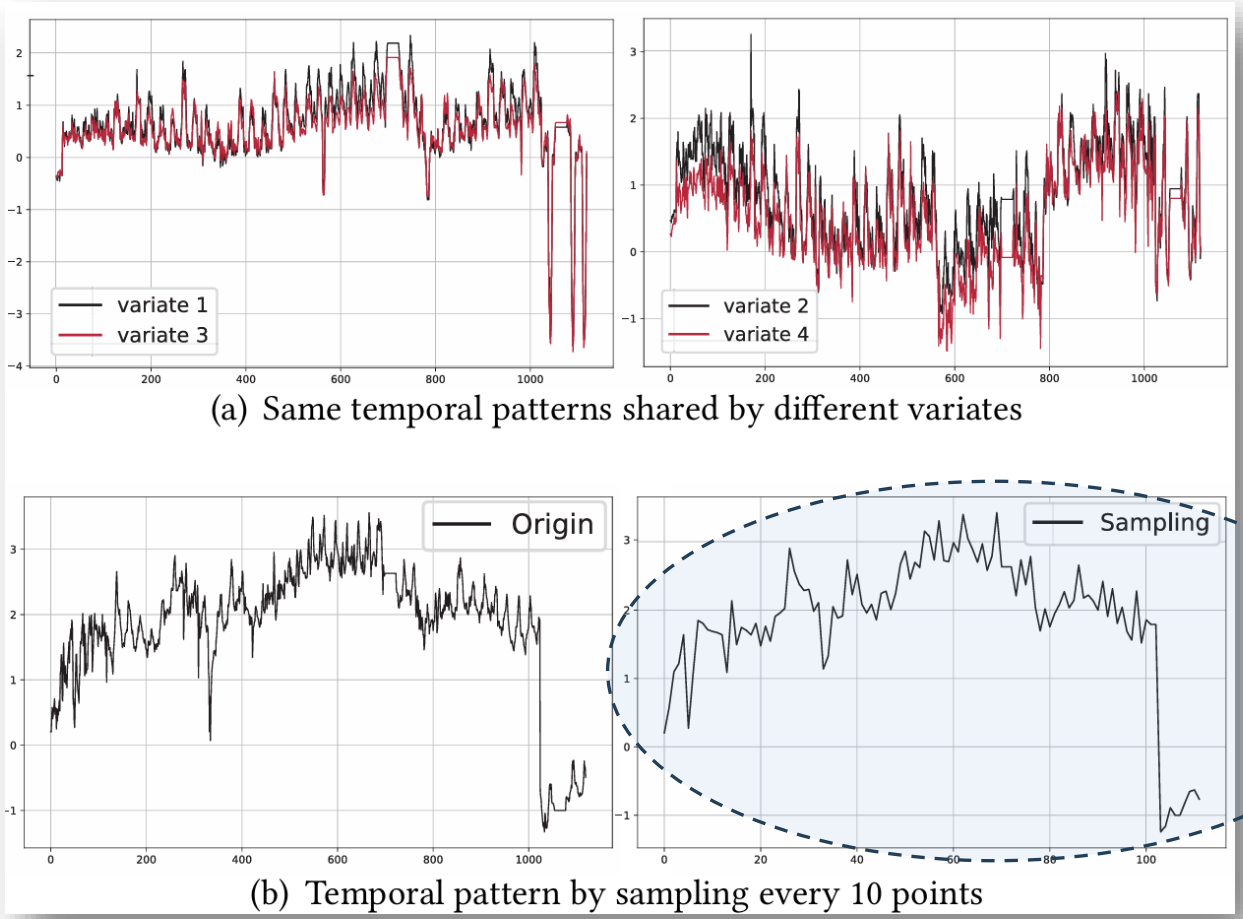
Let time series decide the strategy themselves !

- **Flexible**
- **Explorable**
- **Adaptive**

Design Space—Information Compression

Data Space

Why?



(a) Same temporal patterns shared by different variates

(b) Temporal pattern by sampling every 10 points

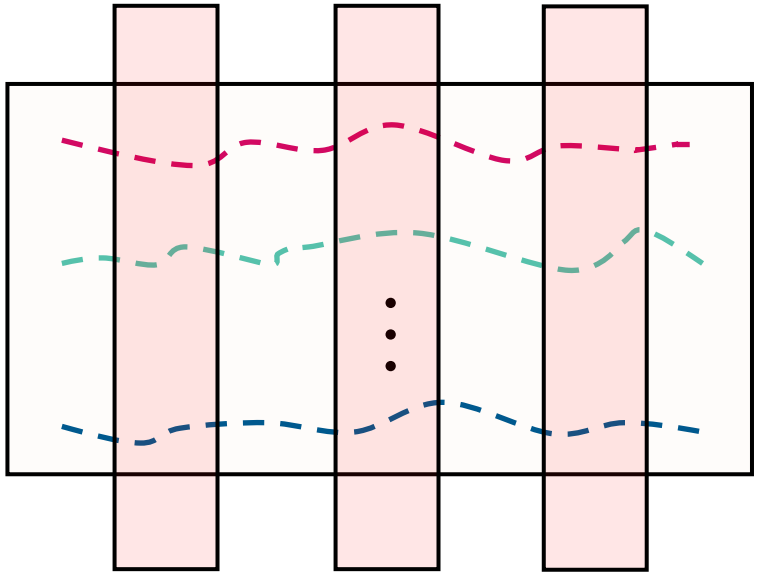
Takeaway:

Fewer bits are required to identify the Long-term trend !

Information Redundancy

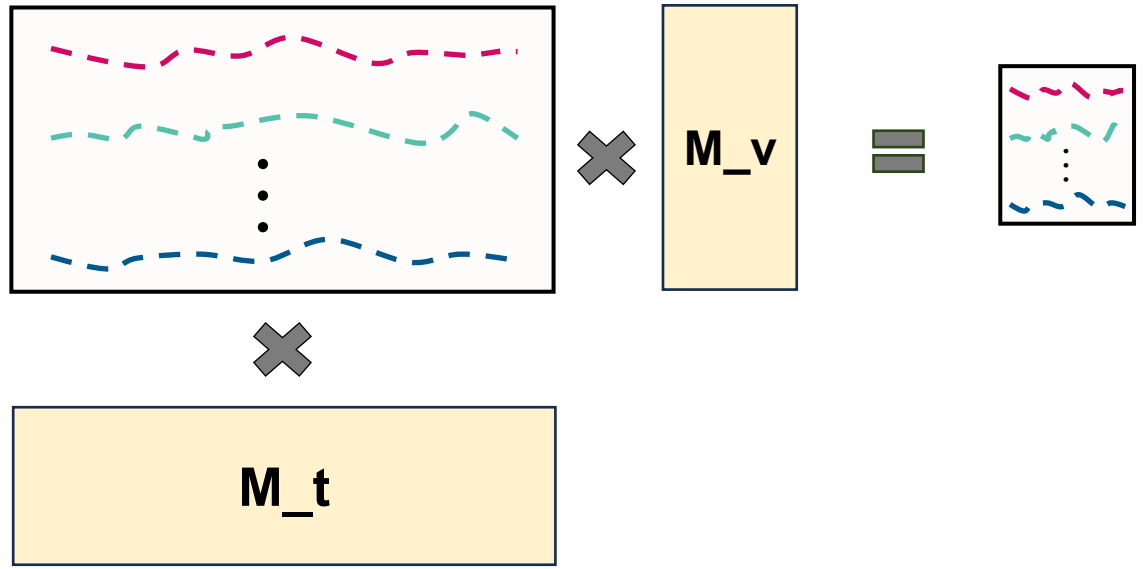
Design Space—Information Compression

Patching



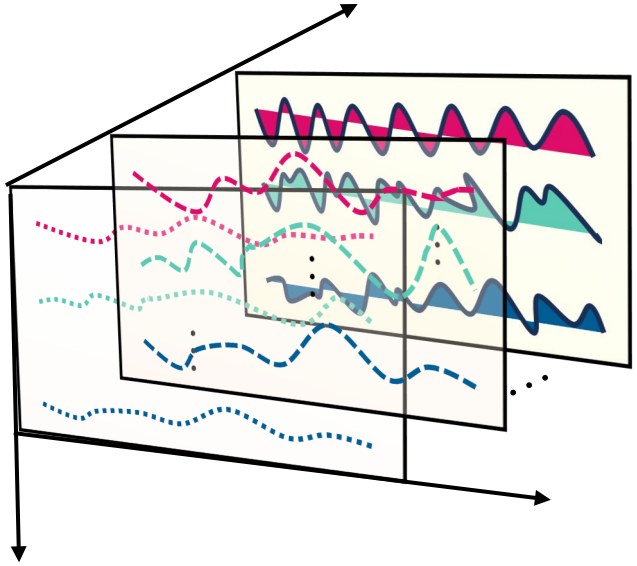
- Inflexible
- Non Differentiable

Low-rank Transform



- Learnable
- Recoverable
- Easy
- Numerically Computable

Design Space—Various Series Tokens



After level expansion

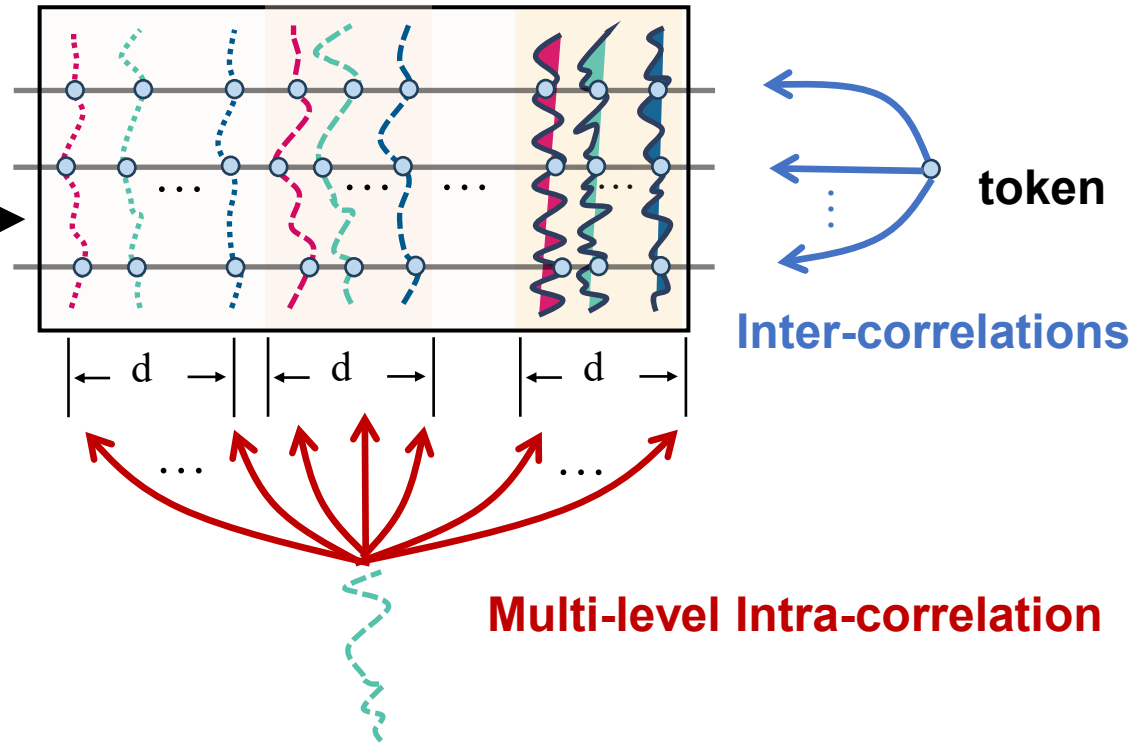
along other dimensions

e.g.,
Unfolding along the time dimension

Multi-mode Learning

Formulated by mode product as:

$$\mathcal{X} \times_3 \boldsymbol{M} = Unfold(\boldsymbol{M}\boldsymbol{X}_{(3)}) \in \mathbb{R}^{n_1 \times n_2 \times m}$$

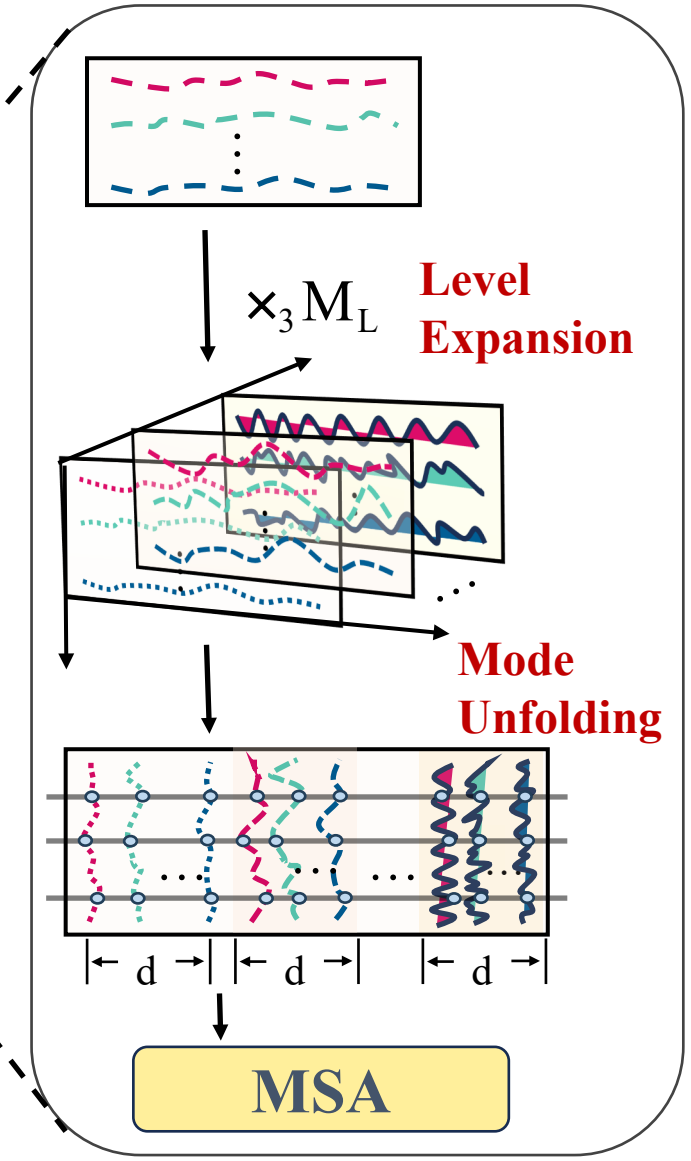


Design Space—Various Series Tokens

Mode Specific Multi-head Self Attention:

- Transform-based Compression
- Level Expansion
- Multi-Mode Learning
- Inter- & Intra- Correlations

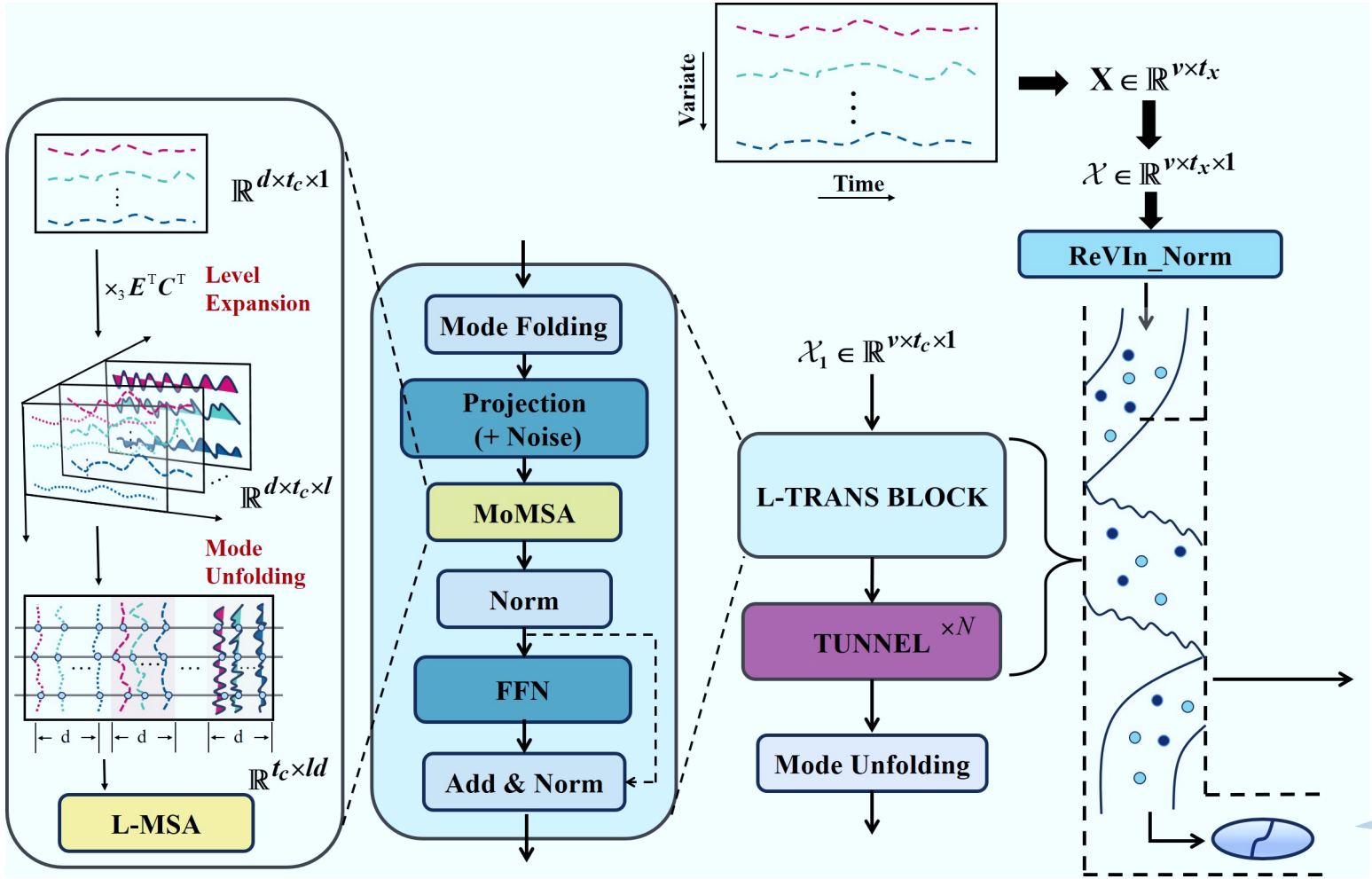
MoMSA



Utilizing Transformer capabilities to develop an encoder for deep, compressed representation learning

Design Space—Encoder

Utilizing Transformer capabilities to develop an encoder for deep, compressed representation learning



What about the decoder?

A Representation *"capsule"*

Design Space—Linear Projection

Decoding with simple MLPs!

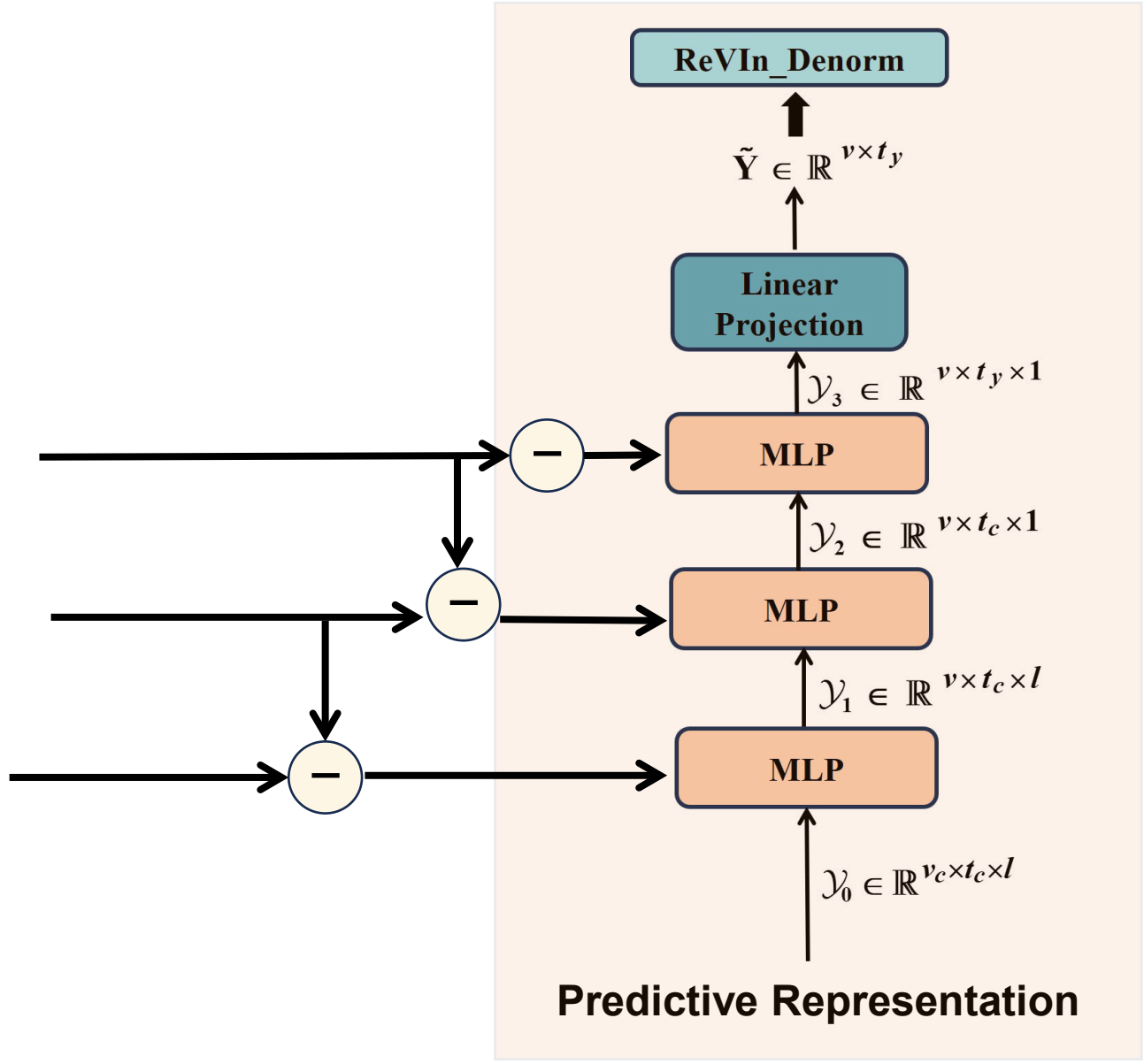
$$\mathcal{B}_1 = (\mathcal{X}_0 - \mathcal{X}_1 \times_2 \mathrm{C_T E_T}) \in \mathbb{R}^{v \times t_x \times 1}$$

$$\mathcal{B}_2 = (\mathcal{X}_1 - \mathcal{X}_2 \times_3 \mathrm{C_L E_L}) \in \mathbb{R}^{v \times t_c \times 1}$$

$$\mathcal{B}_3 = (\mathcal{X}_2 - \mathcal{X}_3 \times_1 \mathrm{C_V E_V}) \in \mathbb{R}^{v \times t_c \times l}$$

Information from the encoder:
Residuals,
transform factors,
To ensure meaningful recovery.

Is it enough?



Design Space—JEPA

Does the system know they are doing the prediction ?

Injecting Predictive Bias:

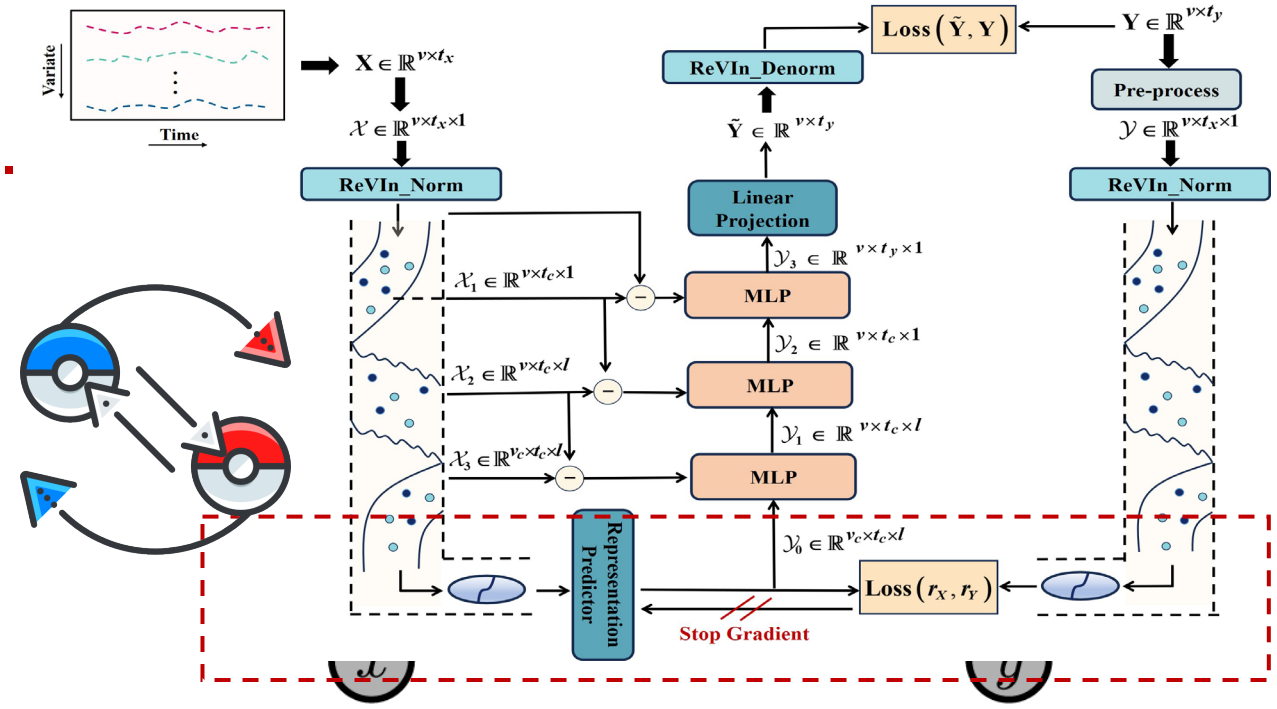
- **What?**

  Joint Embedding Predictive Architecture [1].

- **Where?**

  Between the encoder (history) and decoder (future), as a bridge.

- **Other benefits**
  - **Supervision out of normal L^p space**
  - **Monitoring of the representation learning**
  - **Predict non-stationary factors. (we expect)**

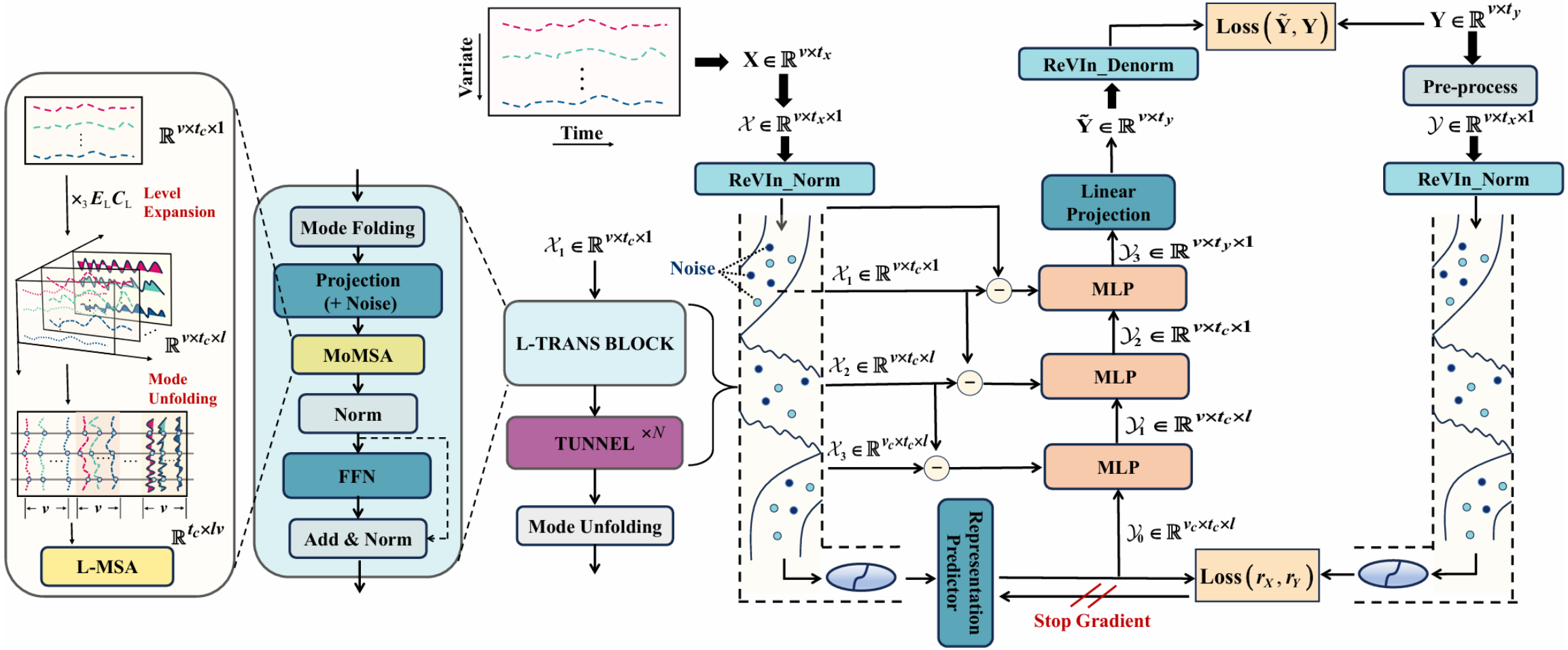


Inner prediction in representation space

[1] Figure in (Yann LeCun, 2022)

[1] LeCun, Yann. "A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27

Design Space—The Whole Model



Augmented representation learning system that explicitly aware of the prediction.

For more details, please see our paper.

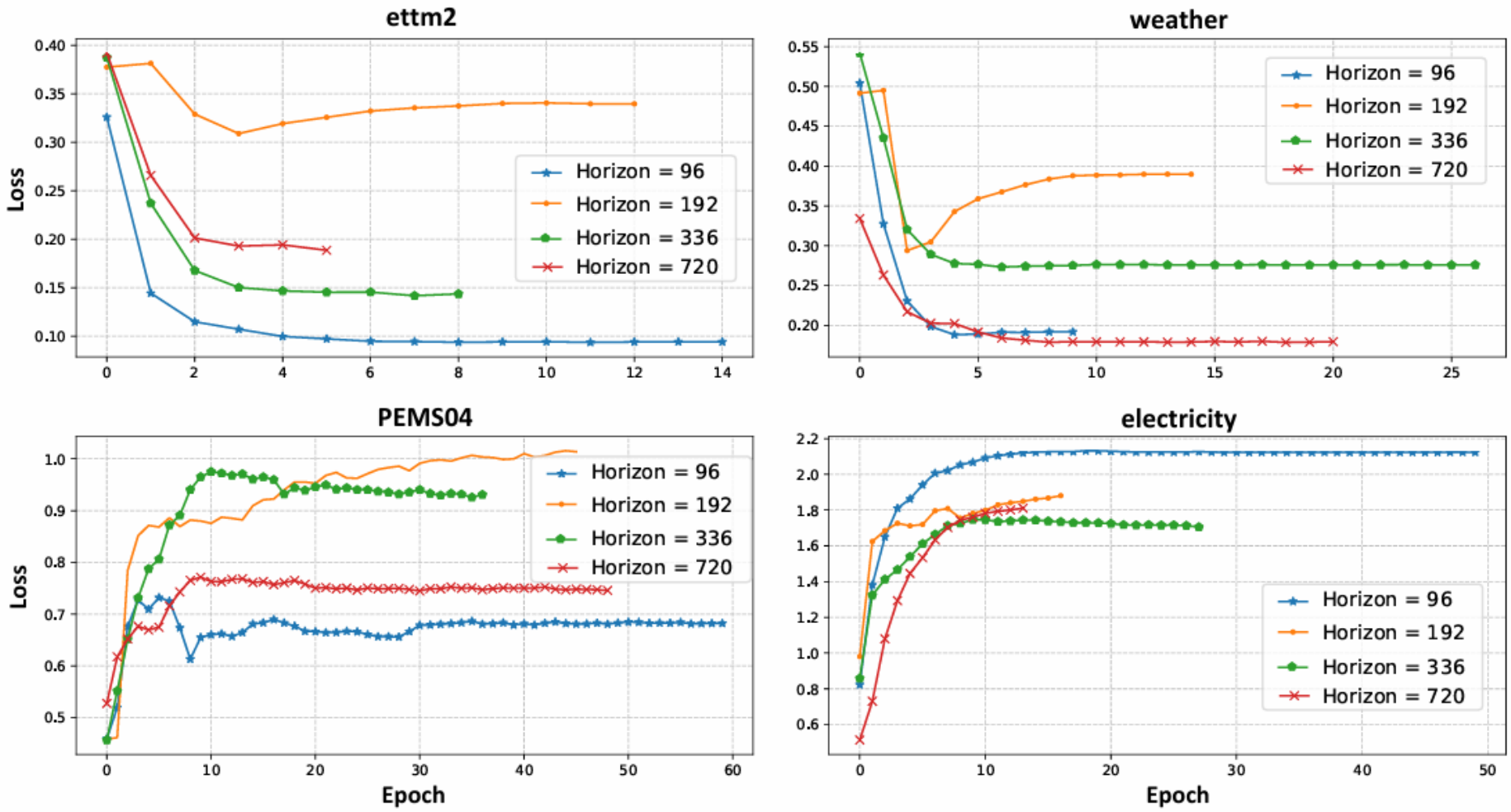
How about the Results

Table 1: Full results of multivariate forecasting. For TimeCapsule, we use superscript † to denote the employment of JEPA training. The lookback length $T = 96$ and prediction lengths $S \in \{24, 36, 48, 60\}$ for ILI, $S \in \{96, 192, 336, 720\}$ and fixed lookback length $T = 512$ for others; For other models, lookback lengths are searched for the best performance.

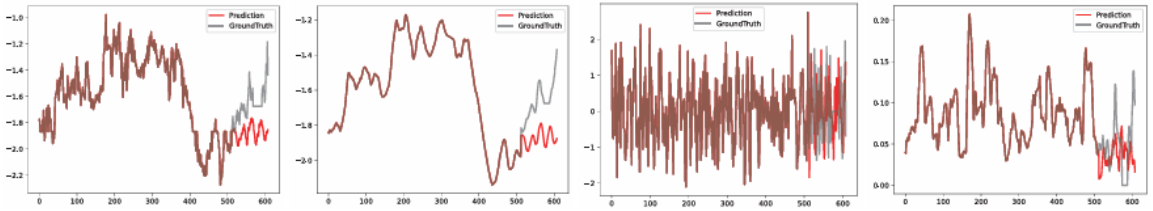
Models		TimeCapsule†		TimeCapsule		iTransformer (2024b)		TimeMixer (2024)		PatchTST (2023)		Crossformer (2023)		DLinear (2023)		TimesNet (2022)		FEDformer (2022b)		Informer (2021)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS04	96	**0.099**	**0.202**	0.110	0.211	0.164	0.280	0.122	0.229	0.161	0.280	0.112	0.224	0.196	0.296	0.159	0.266	0.573	0.565	0.189	0.304
	192	**0.117**	**0.222**	0.127	0.224	0.216	0.316	0.141	0.239	0.178	0.290	0.134	0.236	0.213	0.310	0.179	0.282	0.655	0.624	0.229	0.335
	336	**0.126**	**0.229**	0.133	0.230	0.189	0.288	0.153	0.254	0.193	0.302	0.190	0.286	0.235	0.327	0.169	0.269	1.365	0.920	0.217	0.323
	720	**0.137**	**0.239**	0.187	0.285	0.251	0.351	0.174	0.276	0.233	0.338	0.235	0.331	0.327	0.395	0.187	0.286	0.873	0.728	0.310	0.391
Weather	96	**0.141**	**0.186**	0.142	0.188	0.159	0.208	0.147	0.198	0.149	0.196	0.146	0.212	0.170	0.230	0.170	0.219	0.223	0.292	0.218	0.255
	192	**0.187**	**0.232**	0.188	0.235	0.200	0.248	0.192	0.243	0.193	0.240	0.195	0.261	0.212	0.267	0.222	0.264	0.252	0.322	0.269	0.306
	336	**0.239**	**0.272**	0.241	0.274	0.253	0.289	0.247	0.284	0.244	0.281	0.268	0.325	0.257	0.305	0.293	0.310	0.327	0.371	0.320	0.340
	720	**0.309**	**0.323**	0.311	0.324	0.321	0.338	0.318	0.330	0.314	0.332	0.330	0.380	0.318	0.356	0.360	0.355	0.424	0.419	0.392	0.390
Traffic	96	0.361	0.246	**0.355**	**0.244**	0.363	0.265	0.369	0.257	0.370	0.262	0.514	0.282	0.410	0.282	0.600	0.313	0.593	0.365	0.664	0.371
	192	0.383	0.257	**0.378**	**0.256**	0.385	0.273	0.400	0.272	0.386	0.269	0.501	0.273	0.423	0.288	0.619	0.328	0.614	0.375	0.724	0.396
	336	0.393	0.262	**0.390**	**0.262**	0.396	0.277	0.407	0.272	0.396	0.275	0.507	0.278	0.436	0.296	0.627	0.330	0.609	0.373	0.796	0.435
	720	0.430	**0.282**	**0.429**	0.282	0.445	0.312	0.461	0.316	0.435	0.295	0.571	0.301	0.466	0.315	0.659	0.342	0.646	0.394	0.823	0.453
Electricity	96	**0.125**	**0.218**	0.126	0.219	0.138	0.237	0.131	0.224	0.133	0.233	0.135	0.237	0.140	0.237	0.164	0.267	0.186	0.302	0.214	0.321
	192	**0.146**	**0.238**	0.149	0.242	0.157	0.256	0.151	0.242	0.150	0.248	0.160	0.262	0.154	0.250	0.180	0.280	0.201	0.315	0.245	0.350
	336	**0.158**	**0.255**	0.171	0.269	0.167	0.264	0.169	0.260	0.168	0.267	0.182	0.282	0.169	0.268	0.190	0.292	0.218	0.330	0.294	0.393
	720	**0.187**	**0.280**	0.194	0.287	0.194	0.286	0.227	0.312	0.202	0.295	0.246	0.337	0.204	0.301	0.209	0.307	0.241	0.350	0.306	0.393
ILI	24	**1.675**	**0.793**	3.115	1.110	1.783	0.846	1.807	0.820	1.840	0.835	2.981	1.096	2.208	1.031	2.009	0.926	2.400	1.020	2.738	1.151
	36	**1.619**	**0.796**	1.740	0.841	1.746	0.860	1.896	0.927	1.724	0.845	3.295	1.162	2.032	0.981	2.552	0.997	2.410	1.005	2.890	1.145
	48	**1.653**	**0.835**	1.682	0.856	1.716	0.898	1.753	0.866	1.762	0.863	3.586	1.230	2.209	1.063	1.956	0.919	2.592	1.033	2.742	1.136
	60	1.653	0.830	**1.627**	**0.827**	1.960	0.977	1.828	0.930	1.752	0.894	3.693	1.256	2.292	1.086	2.178	0.962	2.539	1.070	2.825	1.139
Solar	96	0.173	0.229	**0.170**	**0.225**	0.188	0.242	0.178	0.231	0.170	0.234	0.183	0.230	0.216	0.287	0.285	0.330	0.509	0.530	0.338	0.373
	192	**0.188**	0.242	0.189	0.245	0.201	0.259	0.209	0.273	0.204	0.302	0.208	**0.226**	0.244	0.305	0.309	0.342	0.474	0.500	0.375	0.391
	336	0.194	**0.248**	0.195	0.248	0.195	0.259	**0.190**	0.256	0.212	0.293	0.203	0.260	0.263	0.319	0.335	0.365	0.338	0.439	0.417	0.416
	720	0.204	**0.254**	**0.203**	0.255	0.223	0.281	0.203	0.261	0.217	0.307	0.215	0.256	0.264	0.324	0.346	0.355	0.365	0.459	0.390	0.407

Like PatchTST, TimeCapsule excels by leveraging long-range information.

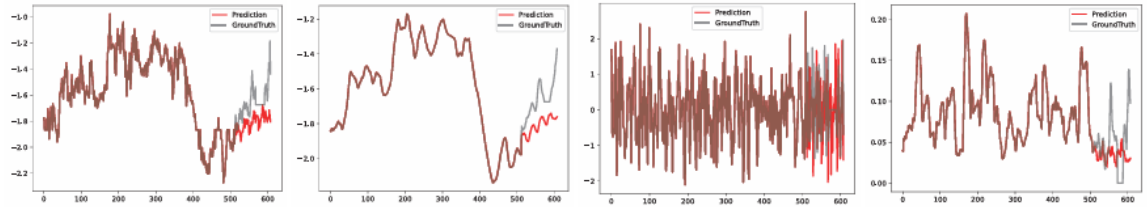
Monitor the Representation Learning



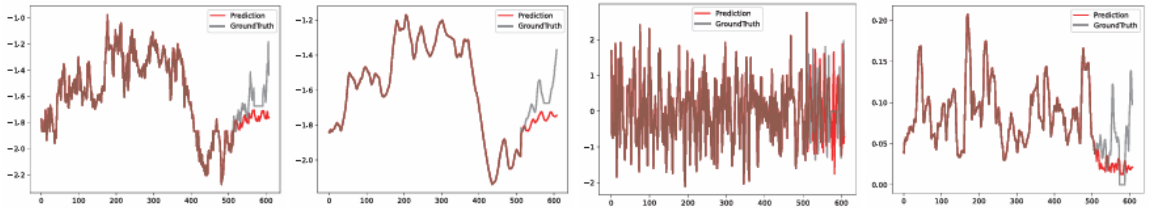
Robusteness to non-stationary



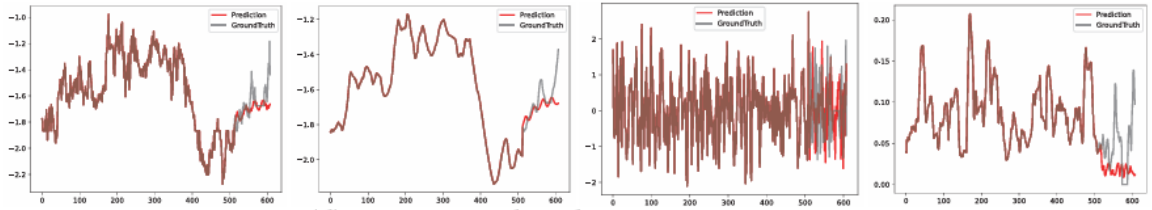
(a) MLP (SAN) achieves MSE = 0.378, this model explicitly predicts non-stationary statistics using SAN [22].



(b) iTransformer (version of non-stationary transformer) achieves MSE = 0.391, this model explicitly models the non-stationary statistics within the non-stationary transformer [21].



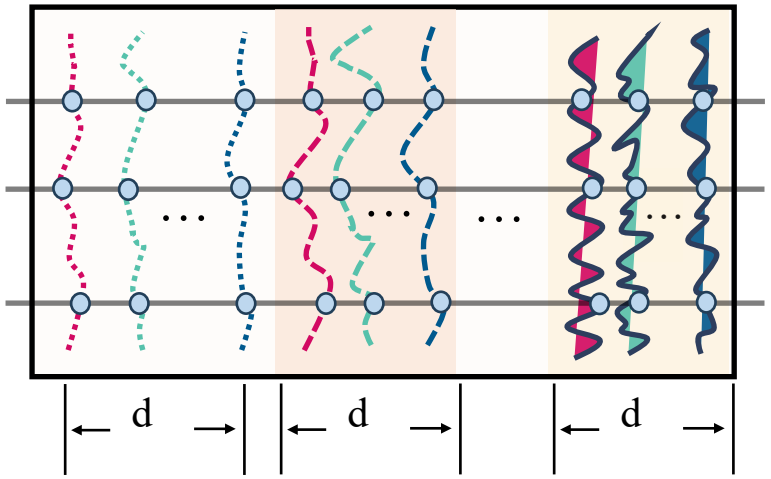
(c) MLP (RevIn) achieves MSE = 0.366, this model is selected as a control group for the use of RevIn [15].



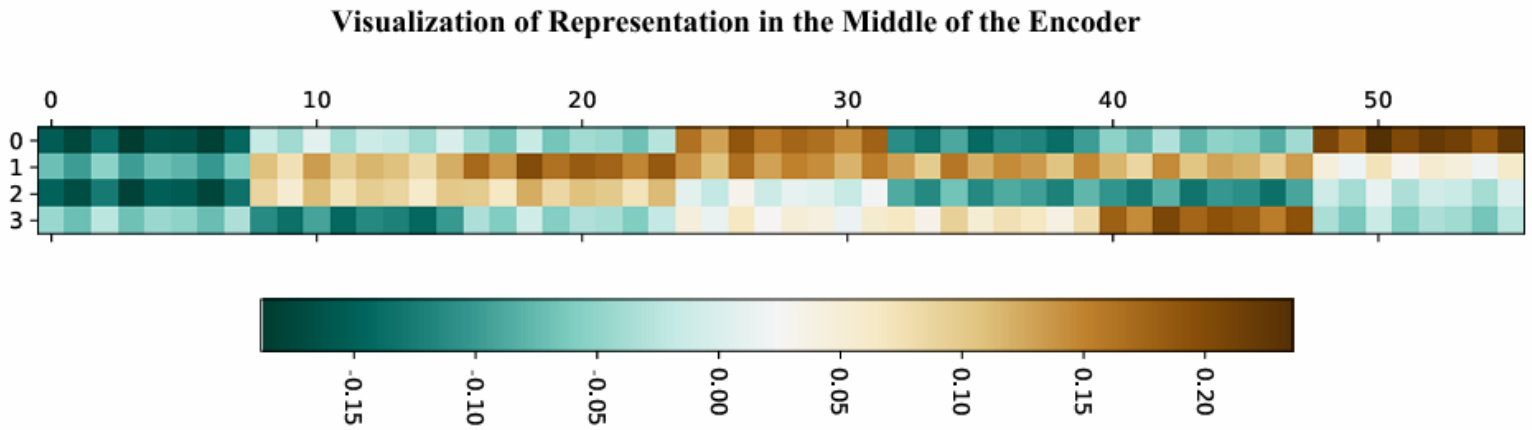
(d) TimeCapsule achieves MSE = 0.362

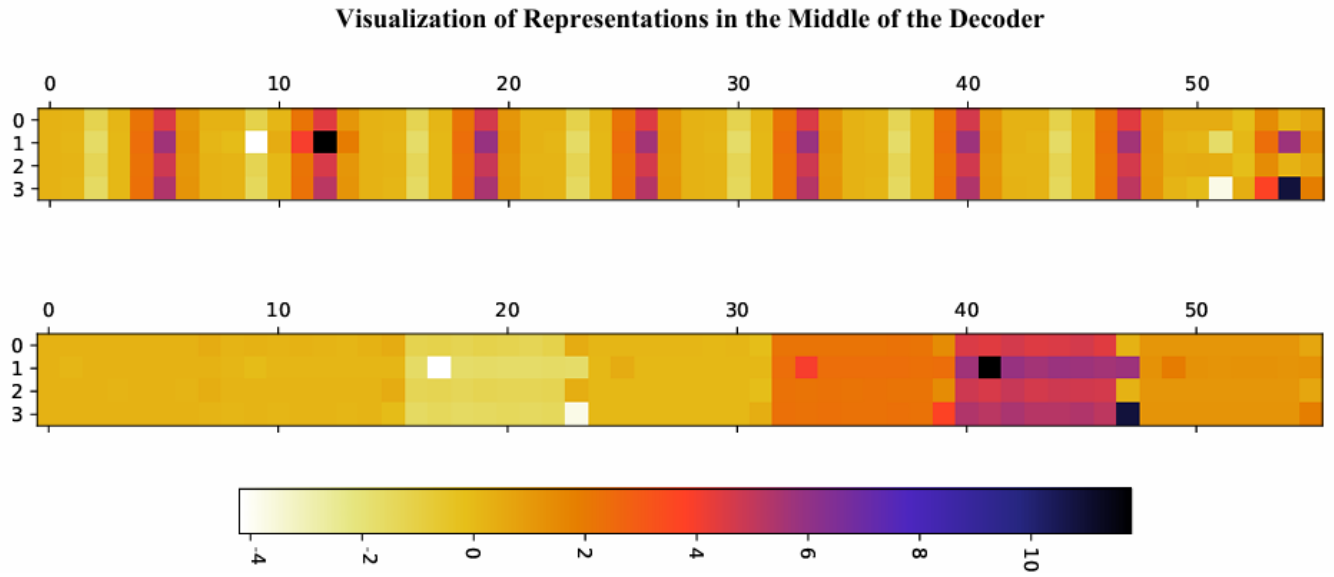
- **NNs primarily predicts the trend**
- **TimeCapsule can potentially refine the non-stationry trend prediction**

TimeCapsule learns what?

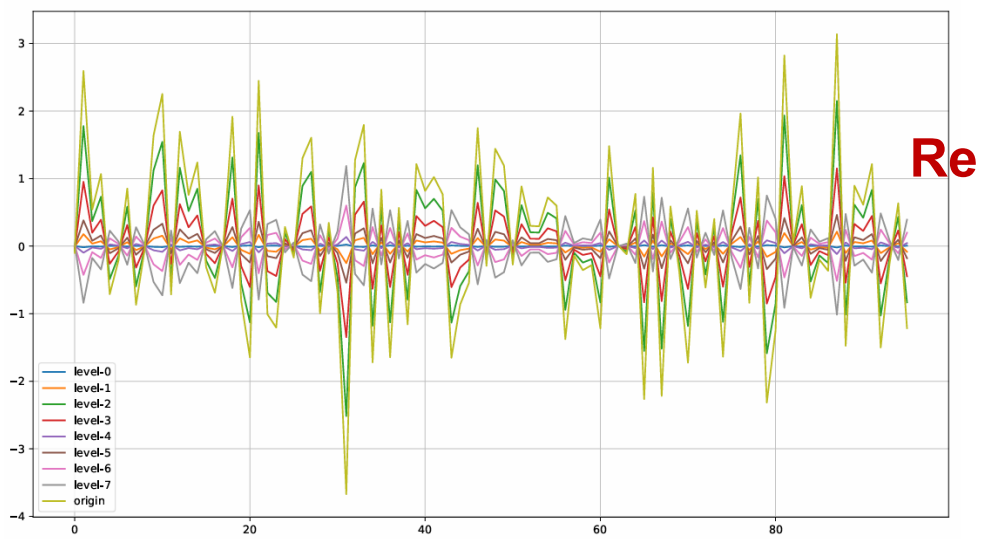


Multi-mode Learning

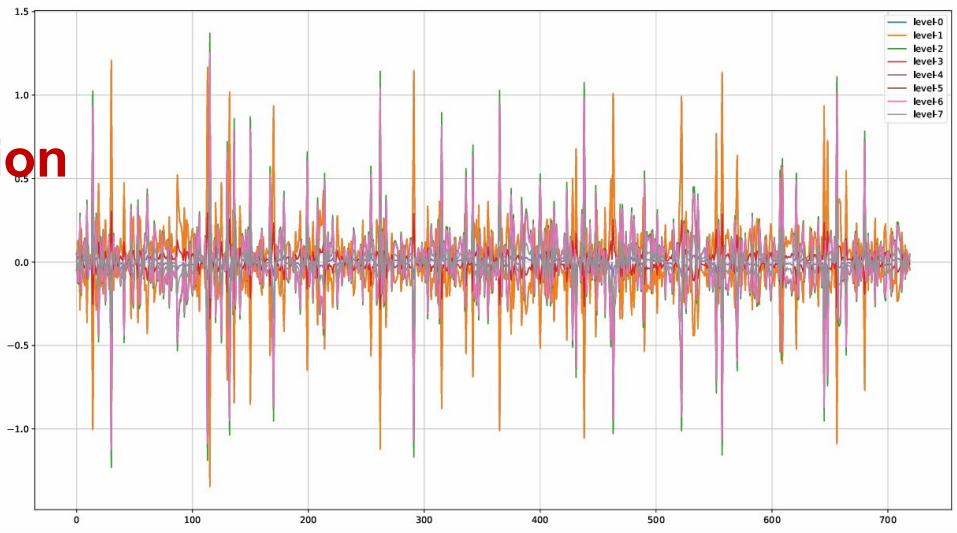




TimeCapsule learns what?

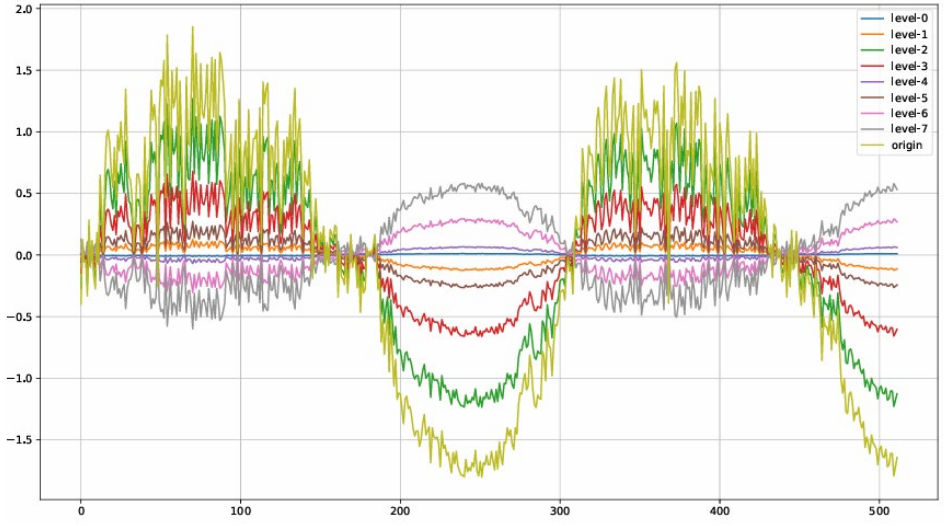


Representation space



Multi-Level

Data Space



What next ?

- The initial designs are coarse, necessitating further refinement.
- To avoid complex non-convex optimization landscapes, the model's structure needs further simplification.
- The learning system can adapt to diverse time series applications.
- Concrete theoretical analyses are lacking. (Lossy compression, Tensor recovery).

It is hard to assess the true improvements for LSTF.

We hope our work will inspire further research in this area.

Thanks for listening to what we had to say about LSTF!