Periodicity Decoupling Framework for Long-term Series Forecasting

REACH OUT!

Code Paper Wechat

Tao Dai[1,2,4], Beiliang Wu[1], Peiyuan Liu[1,3,], Naiqi Li[3,*], Jigang Bao[3], Yong Jiang[3], Shu-tao Xia[3]*

1. College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China

2. National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, China

3. Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

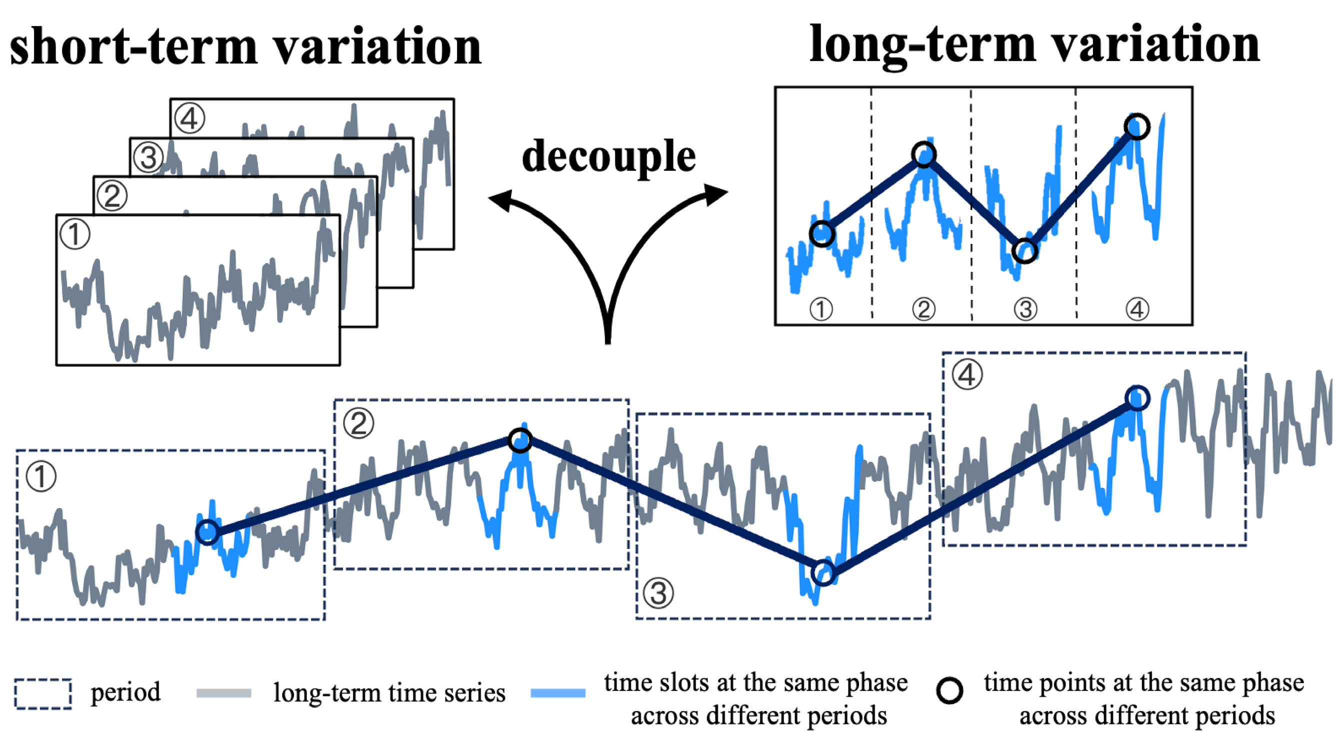
4. WeBank Institute of Financial Technology, Shenzhen University, Shenzhen, China

** Correspondence to: {peiyuanliu.edu; linaiqi.thu}@gmail.com*

SHENZHEN UNIVERSITY

Tsinghua University

Motivation

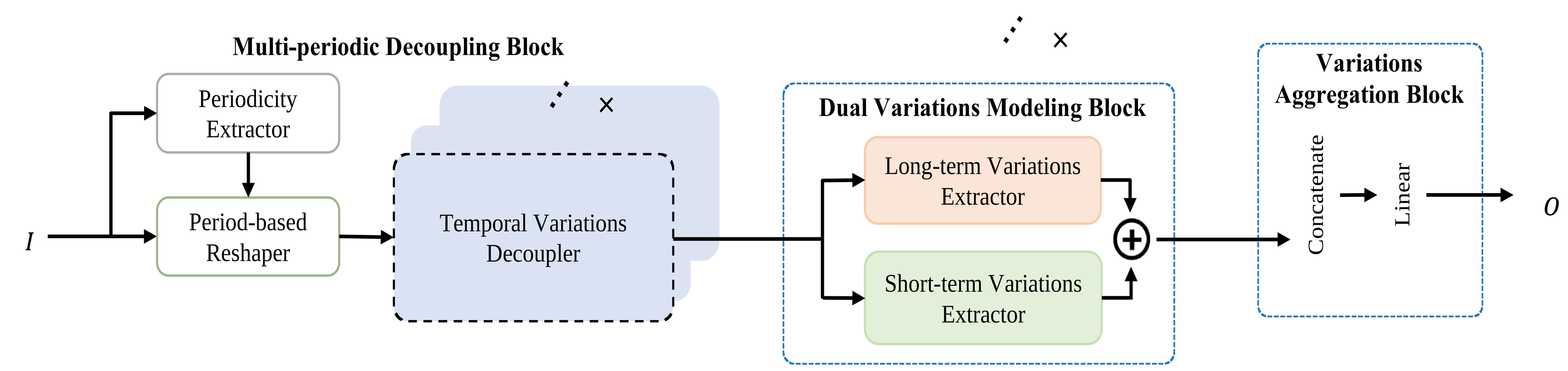


- The ability to accurately predict long-term trends while capturing short-term fluctuations is critical for many application areas such as energy, finance and urban planning.
- Decoupling time series into distinct long-term and short-term variations allows for a more tailored analysis, enhancing prediction accuracy by addressing the unique dynamics of each component separately.
- Transformers excel at globally modeling long-term dependencies, while CNNs are adept at precisely capturing short-term variations with local detail.

Contribution

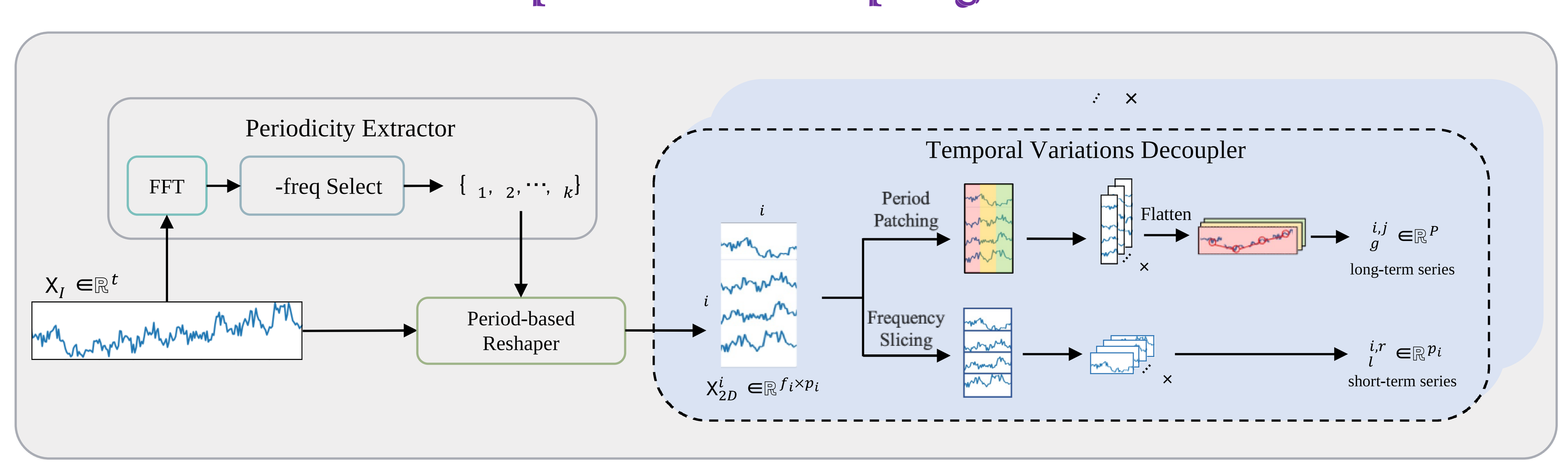
- We introduce the Periodicity Decoupling Framework (PDF) for long-term forecasting, capturing both short-term and long-term variations in 2D form within a parallel architecture.
- Our multi-periodic decoupling block identifies various periods in the frequency domain, converting 1D time series into structured short- and long-term 2D tensors.
- The dual variations modeling block in our framework efficiently extracts and processes both short-term details and long-term dependencies.
- Comprehensive testing across multiple datasets confirms PDF's superior forecasting accuracy and computational efficiency compared to existing methods.

Pipeline



- Periodicity Decoupling Framework (PDF) initially decouples time series into short- and long-term variations using Multi-periodic Decoupling Block that identifies periodic patterns in the frequency domain and reshapes them into 2D tensors.
- These tensors are then processed through several Dual Variations Modeling Blocks to simultaneously capture short- and long-term variations in parallel.
- Finally, a Variations Aggregation Block combines these outputs to form the predicted future series.

Multi-periodic Decoupling Block



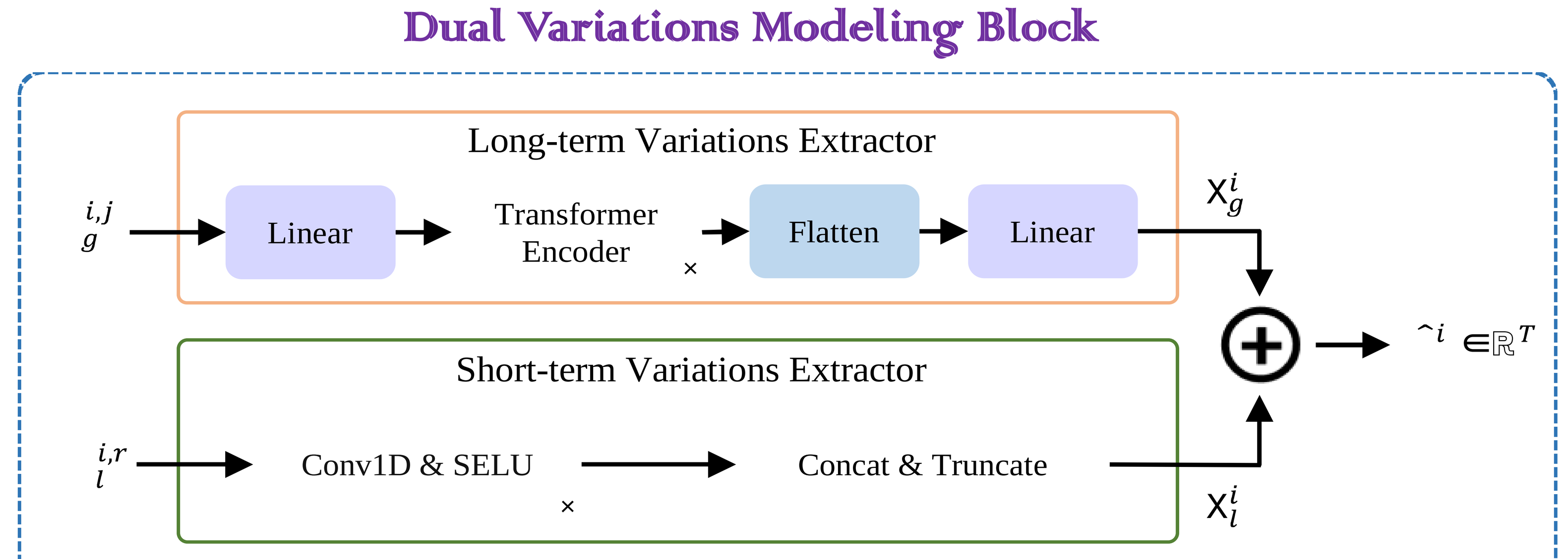
$$\mathbf{A} = \mathrm{Avg}(\mathrm{Amp}(\mathrm{FFT}(\mathbf{X}_I)))$$

$$\mathbf{F}_u = \underset{f_* \in \{1,\cdots,[\frac{t}{2}]\}}{\arg\mathrm{top}\text{-}m}(\mathbf{A}), \quad \mathbf{F}_{k_1} = \underset{f_* \in \{1,\cdots,[\frac{t}{2}]\}}{\arg\mathrm{top}\text{-}k_1}(\mathbf{A})$$

$$\{f_1, \cdots, f_k\} = \mathbf{F}_{k_1} \cup \mathrm{top}\text{-}k_2(\mathbf{F}_u \setminus \mathbf{F}_{k_1})$$

$$X^i_{2D} = \underset{i \in \{1,\cdots,k\}}{\mathrm{Reshape}}(\mathrm{Padding}(X_I))$$

Dual Variations Modeling Block



Long-term Prediction

$$\hat{x}^{i,j}_g = \mathrm{BatchNorm}(x^{i,j}_g + \mathrm{MSA}(x^{i,j}_g, x^{i,j}_g, x^{i,j}_g))$$
$$\hat{x}^{i,j}_g = \mathrm{BatchNorm}(\hat{x}^{i,j}_g + \mathrm{MLP}(\hat{x}^{i,j}_g))$$
$$\boxed{X^i_g} = \mathrm{Linear}(\mathrm{Flatten}(\hat{x}^{i,j}_g)) \in \mathbb{R}^t$$

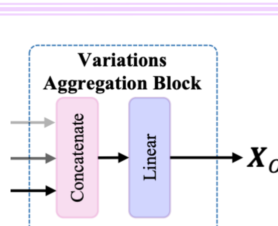
Summation

$$\boxed{\hat{X}^i = X^i_g + X^i_l}$$

Short-term Prediction

$$\hat{x}^{i,r}_l = \mathrm{SELU}(\mathrm{Conv1d}(x^{i,r}_l))$$
$$\boxed{X^i_l} = \mathrm{Truncate}(\mathrm{Concat}(\hat{x}^{i,r}_l))$$

Variations Aggregation Block



The Variations Aggregation Block consolidates the results from k DVMBs. Specifically, we concatenate these k results and then map them through a parameter-shared linear layer to produce univariate prediction $X_O \in \mathbb{R}^T$:

$$X_O = \mathrm{Linear}(\mathrm{Concat}(\hat{X}^i))$$

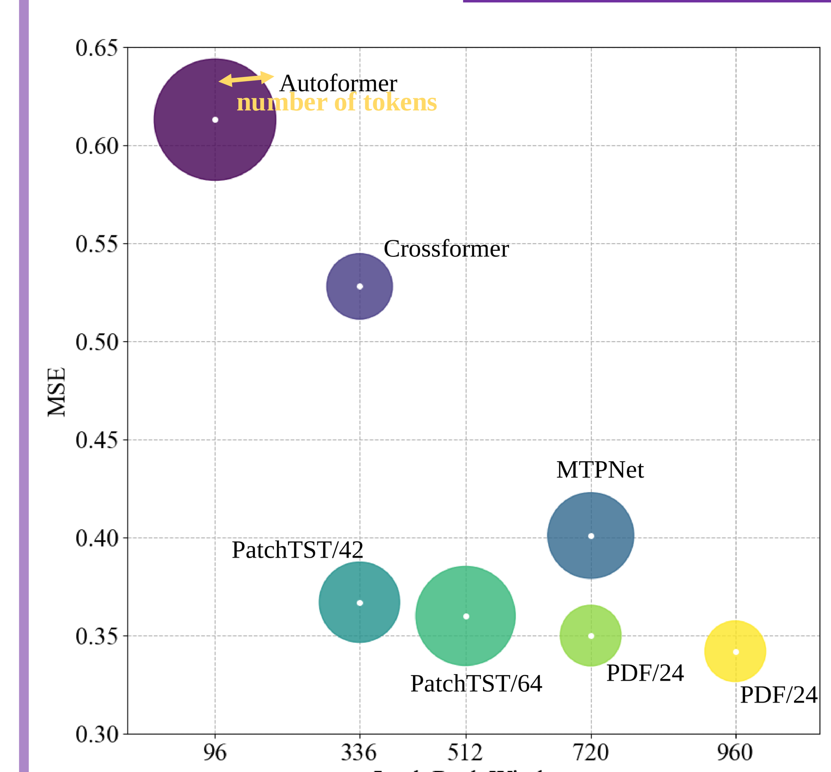
The final multivariate prediction $\mathbf{X}_O \in \mathbb{R}^{T \times d}$ is obtained by stacking d univariate prediction X_O.

Experiment

Catagories		Ours				Transformers				CNNs				Linears			
Models		PDF(720)		PDF(336)		PatchTST(512)		FEDformer(96)		TimesNet(96)		MICN(96)		TiDE(720)		DLinear(336)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.356**	0.391	0.357	**0.388**	0.370	0.400	0.376	0.419	0.384	0.402	0.421	0.431	0.375	0.398	0.375	0.399
	192	**0.390**	0.413	0.397	**0.412**	0.413	0.429	0.420	0.448	0.436	0.429	0.474	0.487	0.412	0.422	0.405	0.416
	336	**0.402**	**0.421**	0.409	0.422	0.422	0.440	0.459	0.465	0.491	0.469	0.569	0.551	0.435	0.433	0.439	0.443
	720	0.462	0.477	**0.432**	**0.455**	0.447	0.468	0.506	0.507	0.521	0.500	0.770	0.672	0.454	0.465	0.472	0.490
ETTh2	96	**0.270**	**0.332**	0.272	0.333	0.274	0.337	0.358	0.397	0.340	0.374	0.299	0.364	**0.270**	0.336	0.289	0.353
	192	0.334	**0.375**	0.335	**0.375**	0.341	0.382	0.429	0.439	0.402	0.414	0.441	0.454	**0.332**	0.380	0.383	0.418
	336	**0.324**	0.379	0.325	**0.377**	0.329	0.384	0.496	0.487	0.452	0.452	0.654	0.567	0.360	0.407	0.448	0.465
	720	0.378	0.422	**0.375**	**0.417**	0.379	0.422	0.463	0.474	0.462	0.468	0.956	0.716	0.419	0.451	0.605	0.551
ETTm1	96	**0.277**	0.337	0.280	**0.335**	0.293	0.346	0.379	0.419	0.338	0.375	0.316	0.362	0.306	0.349	0.299	0.343
	192	**0.316**	0.364	0.317	**0.359**	0.333	0.370	0.426	0.441	0.374	0.387	0.363	0.390	0.335	0.366	0.335	0.365
	336	**0.346**	**0.381**	0.354	0.382	0.369	0.392	0.445	0.459	0.410	0.411	0.408	0.426	0.364	0.384	0.369	0.386
	720	**0.402**	**0.409**	0.405	0.413	0.416	0.420	0.543	0.490	0.478	0.450	0.481	0.476	0.413	0.413	0.425	0.421
ETTm2	96	**0.159**	**0.251**	0.162	0.253	0.166	0.256	0.203	0.287	0.187	0.267	0.179	0.275	0.161	**0.251**	0.167	0.260
	192	0.217	0.292	0.219	0.291	0.223	0.296	0.269	0.328	0.249	0.309	0.307	0.376	**0.215**	**0.289**	0.224	0.303
	336	**0.266**	**0.325**	0.270	0.326	0.274	0.329	0.325	0.366	0.321	0.351	0.325	0.388	0.267	0.326	0.281	0.342
	720	**0.345**	**0.375**	0.358	0.380	0.362	0.385	0.421	0.415	0.408	0.403	0.502	0.490	0.352	0.383	0.397	0.421
Weather	96	**0.143**	**0.193**	0.147	0.194	0.149	0.198	0.217	0.296	0.172	0.220	0.161	0.229	0.166	0.222	0.176	0.237
	192	**0.188**	**0.236**	0.192	0.239	0.194	0.241	0.276	0.336	0.219	0.261	0.220	0.281	0.209	0.263	0.220	0.282
	336	**0.240**	**0.279**	0.244	**0.279**	0.245	0.282	0.339	0.380	0.280	0.306	0.278	0.331	0.254	0.301	0.265	0.319
	720	**0.308**	**0.328**	0.318	0.330	0.314	0.334	0.403	0.428	0.365	0.359	0.311	0.356	0.313	0.340	0.323	0.362
Electricity	96	**0.126**	0.220	0.127	**0.219**	0.129	0.222	0.193	0.308	0.168	0.272	0.164	0.269	0.132	0.229	0.140	0.237
	192	**0.145**	0.238	**0.145**	**0.237**	0.147	0.240	0.201	0.315	0.184	0.289	0.177	0.285	0.147	0.243	0.153	0.249
	336	**0.159**	**0.255**	0.162	**0.255**	0.163	0.259	0.214	0.329	0.198	0.300	0.193	0.304	0.161	0.261	0.169	0.267
	720	**0.194**	**0.287**	0.200	0.290	0.197	0.290	0.246	0.355	0.220	0.320	0.212	0.321	0.196	0.294	0.203	0.301
Traffic	96	0.350	0.239	0.351	**0.238**	0.360	0.249	0.587	0.366	0.593	0.321	0.519	0.309	**0.336**	0.253	0.410	0.282
	192	0.363	**0.247**	0.374	0.248	0.379	0.256	0.604	0.373	0.617	0.336	0.537	0.315	**0.346**	0.257	0.423	0.287
	336	0.376	0.258	0.386	**0.253**	0.392	0.264	0.621	0.383	0.629	0.336	0.534	0.313	**0.355**	0.260	0.436	0.296
	720	0.419	0.279	0.421	0.278	0.432	0.286	0.626	0.382	0.640	0.350	0.577	0.325	**0.386**	**0.273**	0.466	0.315
Count		**52**		44		6		0		0		1		20		0	

- Datasets: ETTh1, ETTh2, ETTm1, ETTm2, Weather, Electricity, Traffic
- Metric: Mean Square Error (MSE) and Mean Absolute Error (MAE)
- PDF(720) achieves significant improvements over existing models, reducing MSE by 14.59%, 24.61%, and 7.05% and MAE by 10.77%, 19.91%, and 5.51% compared to Transformer, CNN, and Linear-based models respectively.

Complexity Analysis



Method	Encoder layer	Decoder layer
Trans. (Vaswani et al., 2017)	$O(t^2)$	$O(T(t+T))$
In. (Zhou et al., 2021)	$O(t \log t)$	$O(T(T+\log t))$
Auto. (Wu et al., 2021)	$O(t \log t)$	$O((\frac{t}{2}+T)\log(\frac{t}{2}+T))$
Pyra. (Liu et al., 2021)	$O(t)$	$O(t(t+T))$
FED. (Zhou et al., 2022)	$O(t)$	$O(\frac{t}{2}+T)$
ETS. (Woo et al., 2022)	$O(t \log t)$	$O(T \log T)$
Cross. (Zhang & Yan, 2023)	$O(\frac{d}{p^2}t^2)$	$O(\frac{d}{p^2}T(t+T))$
MTP. (Zhang et al., 2023)	$O((\frac{t}{p})^2)$	$O((\frac{t+T}{p})^2)$
PET. (Lin et al., 2023)	$O((\frac{t}{p})^2)$	-
Patch. (Nie et al., 2023)	$O((\frac{t}{p})^2)$	-
PDF (Ours)	$O((\frac{\max(p_i)}{p})^2)$	-

- Left: Comparison of MSE and token number of our PDF over other Transformer-based methods to predict future 96 time steps on Traffic dataset. Other Transformer-based methods obtain worse MSE results with more patch numbers. By contrast, our PDF achieves the lowest MSE with only 24 tokens on the look-back window of 960 length.
- Right: Computational complexity per layer in Transformer-based models. t and T denote the length of the look-back window and prediction window. d denotes the number of variates. p denotes the length of each patch. The complexity of PDF is solely dependent on the length of the longest period selected.