NEURAL INFORMATION PROCESSING SYSTEMS

CycleNet

Enhancing Time Series Forecasting through Modeling Periodic Patterns

Shengsheng Lin[1], Weiwei Lin[1,2,*], Xinyi Hu[3]. Wentai Wu[4], Ruichao Mo[1], Haocheng Zhong[1]

[1]South China University of Technology

[2]Pengcheng Laboratory

[3]The Chinese University of Hong Kong

[4]Jinan University

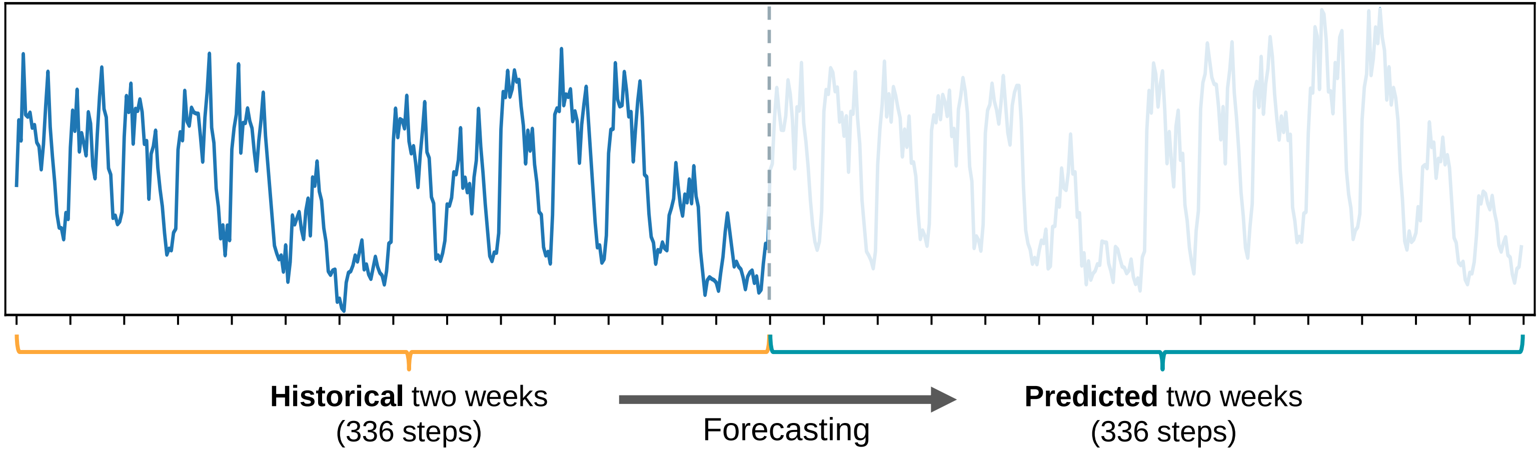
1. Motivations

2. Contributions

3. Method

4. Interpretability

5. Results



- Long-term Time Series Forecasting (**LTSF**) Tasks :
  - **Longer look-back windows** are required for accurate predictions over **extending forecast horizon**
  - **Mainstream methods** rely on ***stacking deep architectures*** to extract long-term dependencies from extended look-back windows, enabling more accurate ***modeling of periodic patterns***

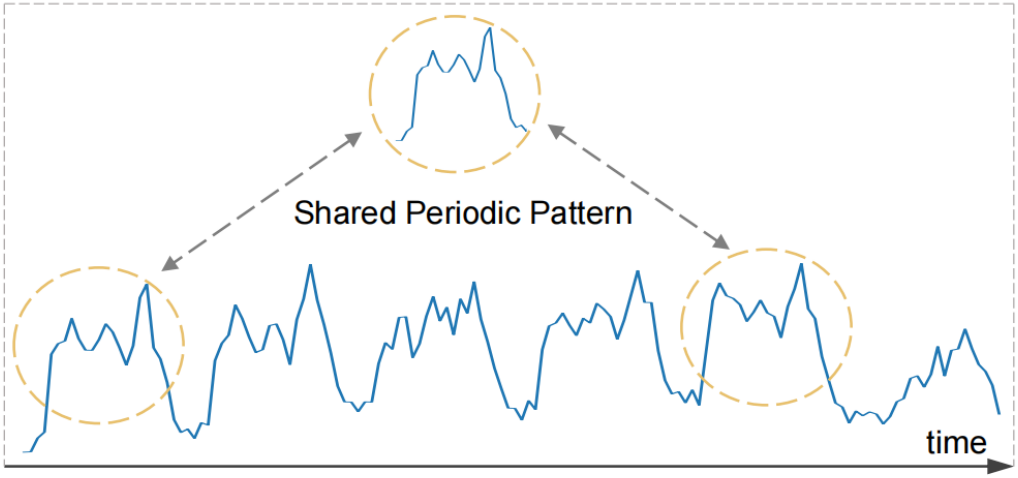
1. Motivations

2. Contributions

3. Method

4. Interpretability

5. Results



- **Intuitions** about Periodic Patterns:
  - Stable **periodic patterns** present in time series data serve as the foundation for **accurate long-horizon forecasts**
  - These **periodic patterns** in time series data can be directly represented through ***globally shared segments***.

- **Explicit modeling** of periodicity:
  - Pioneering **explicit modeling** of ***periodic patterns*** in sequences to enhance time series forecasting tasks.
- **Residual Cycle Forecasting (RCF)** technique:
  - Utilizing ***learnable* recurrent cycles** to explicitly model the inherent ***periodic patterns*** within time series data, followed by predicting the **residual components** of the modeled cycles.
- **CycleNet** model:
  - The proposed **CycleNet** (combined RCF with Linear/MLP) achieves **state-of-the-art** performance with ***significant efficiency advantages***.

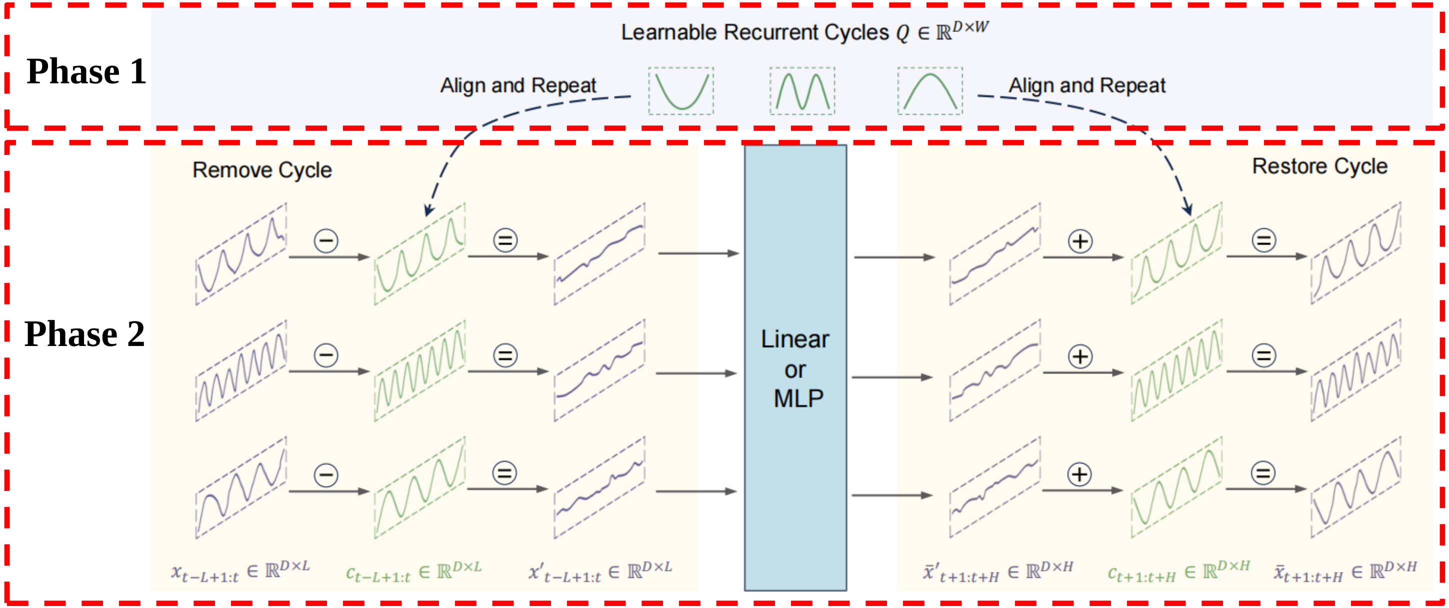
1. Motivations

2. Contributions

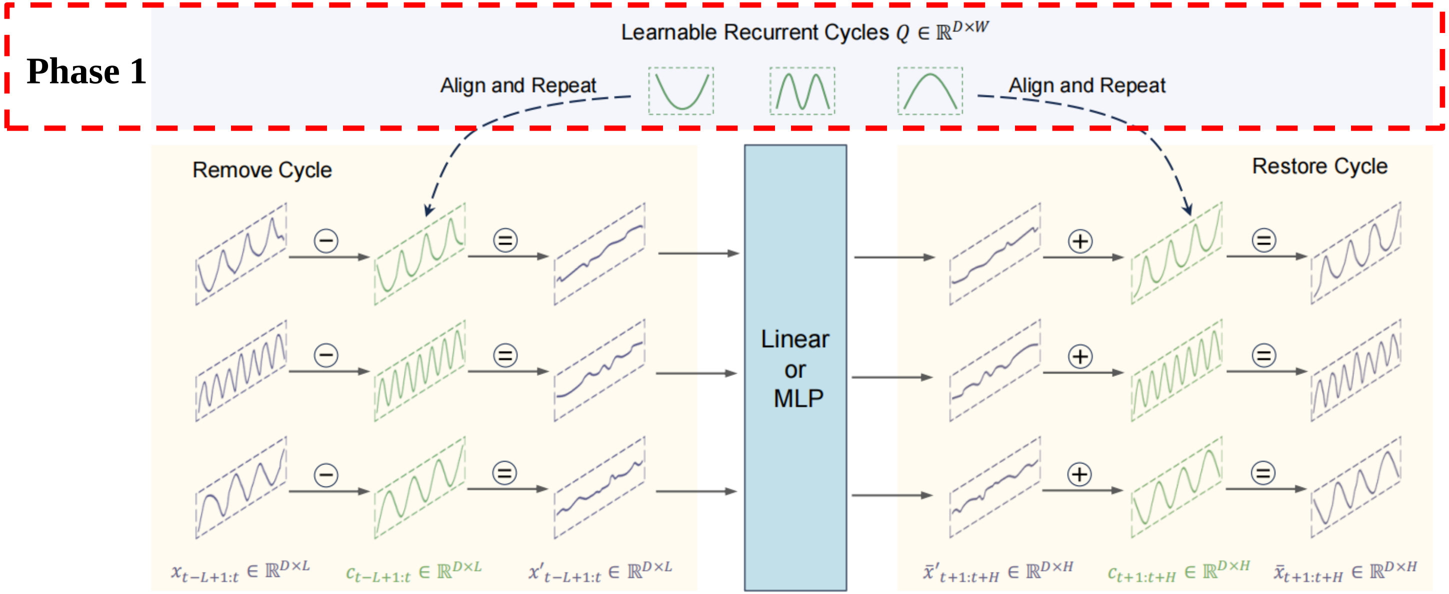
3. Method

4. Interpretability

5. Results



- **The core idea of *RCF* technique:**
  - **Phase 1 (Periodic patterns modeling):** Utilizing ***learnable* recurrent cycles** to explicitly model the inherent ***periodic patterns*** within time series data
  - **Phase 2 (Residual forecasting):** Predicting the **residual components** of the modeled cycles.



- **Phase 1 (Periodic patterns modeling)**:
  - Generate ***learnable* recurrent cycles** $Q \in \mathbb{R}^{W \times D}$ and all initialized to **zeros**, where W is the length of periodicity and D is the number of channels.
  - The ***learnable* recurrent cycles** Q will undergo **gradient backpropagation** training along with the backbone module for prediction, **yielding learned representations**.

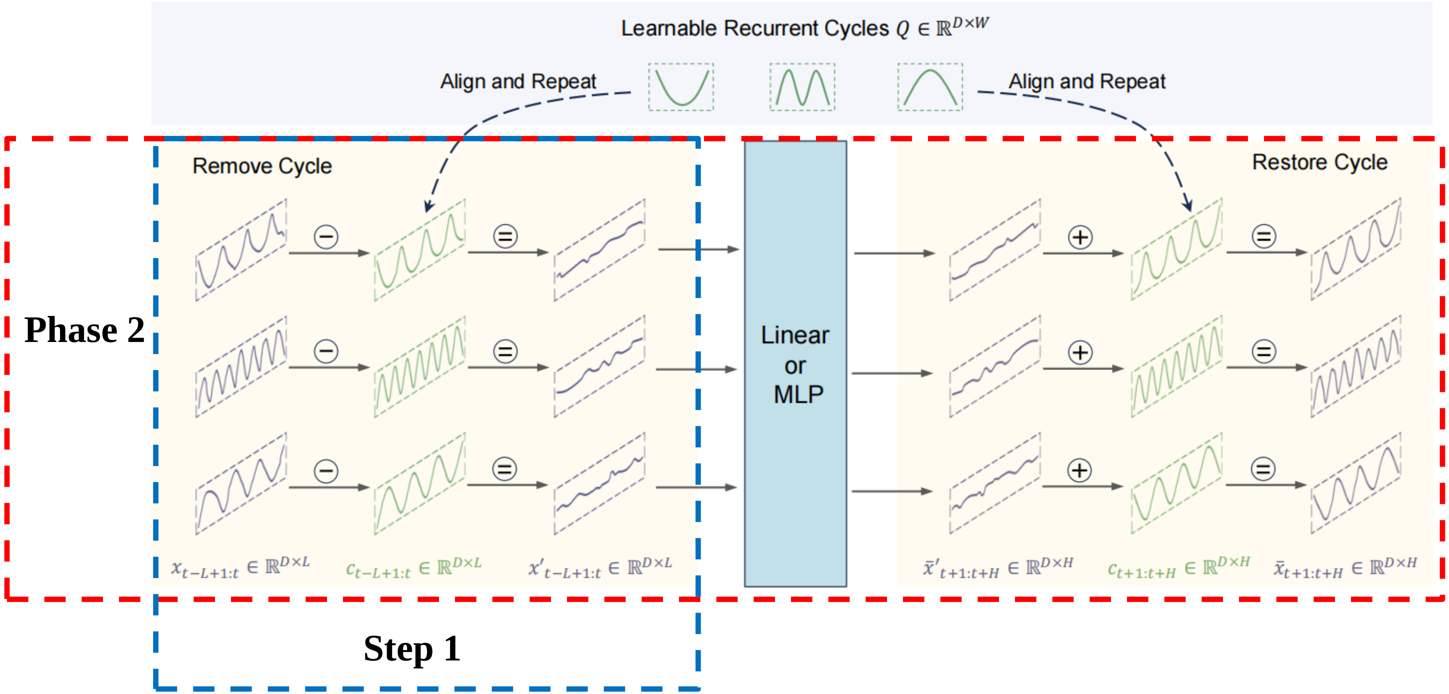
1. Motivations

2. Contributions

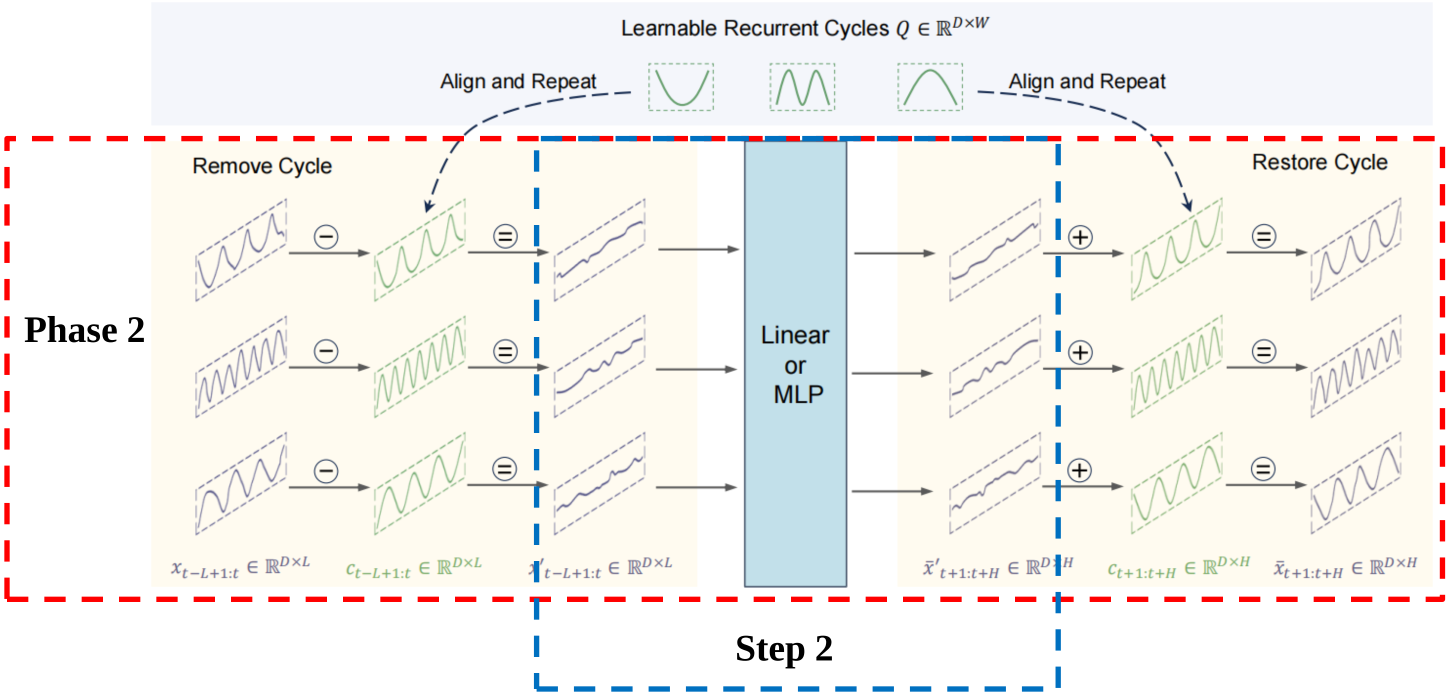
3. Method

4. Interpretability

5. Results



- **Phase 2 (Residual forecasting)**:
  - **(Step 1**) Remove the **cyclic components** $c_{t-L+1:t}$ from the original input $x_{t-L+1:t}$ to obtain **residual components** $x'_{t-L+1:t}$



- **Phase 2 (Residual forecasting)**:
  - **(Step 2)** Pass $x'_{t-L+1:t}$ through the ***backbone*** to obtain **predictions** for the **residual components**, $\bar{x}'_{t+1:t+H}$

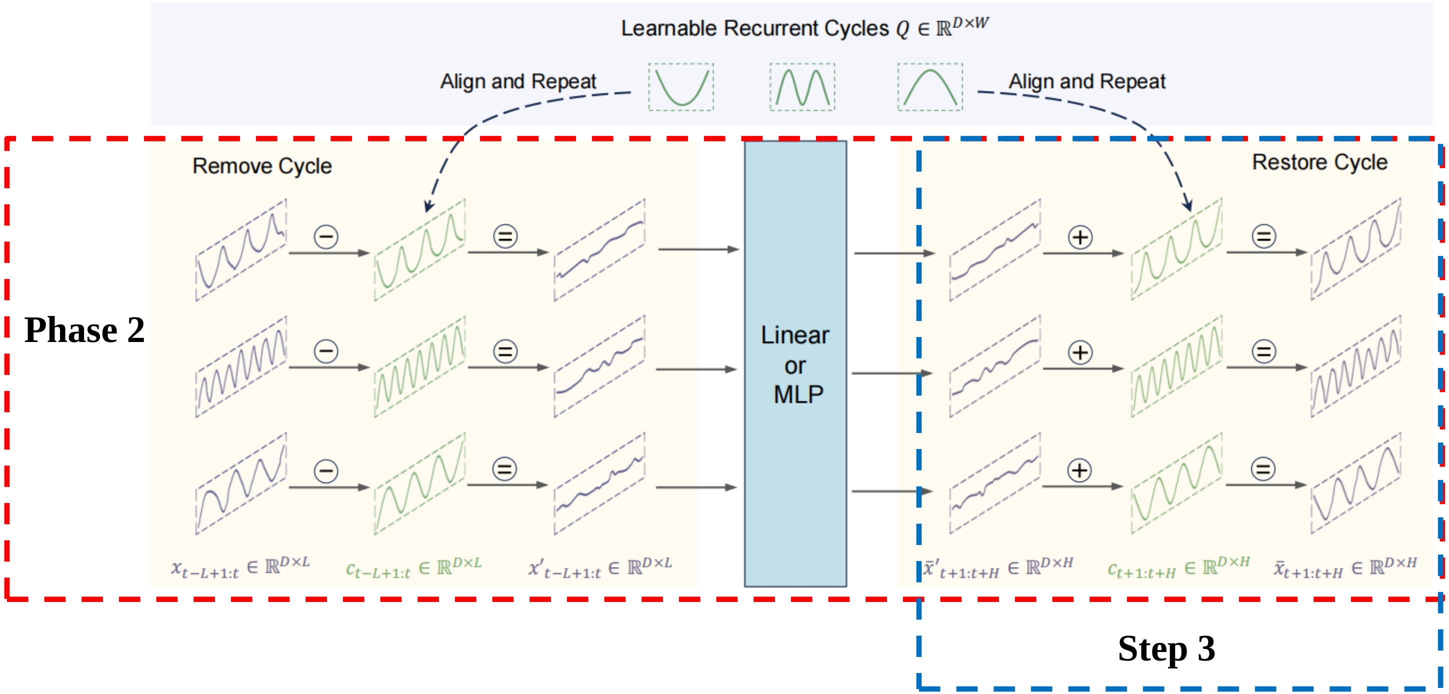
1. Motivations

2. Contributions

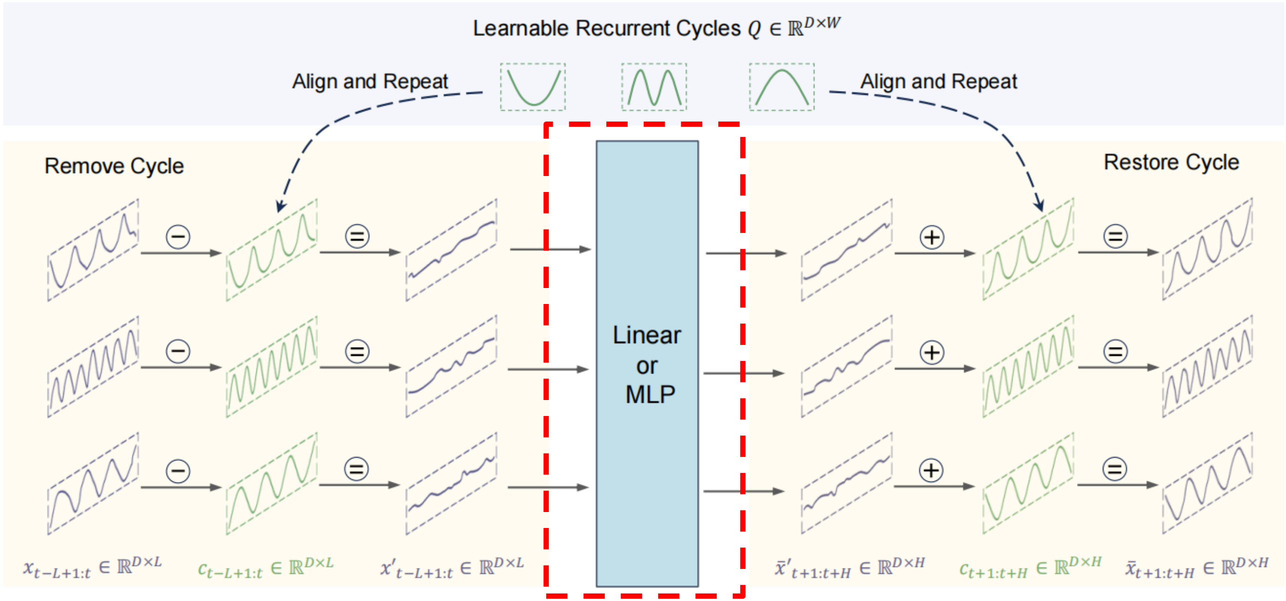
3. Method

4. Interpretability

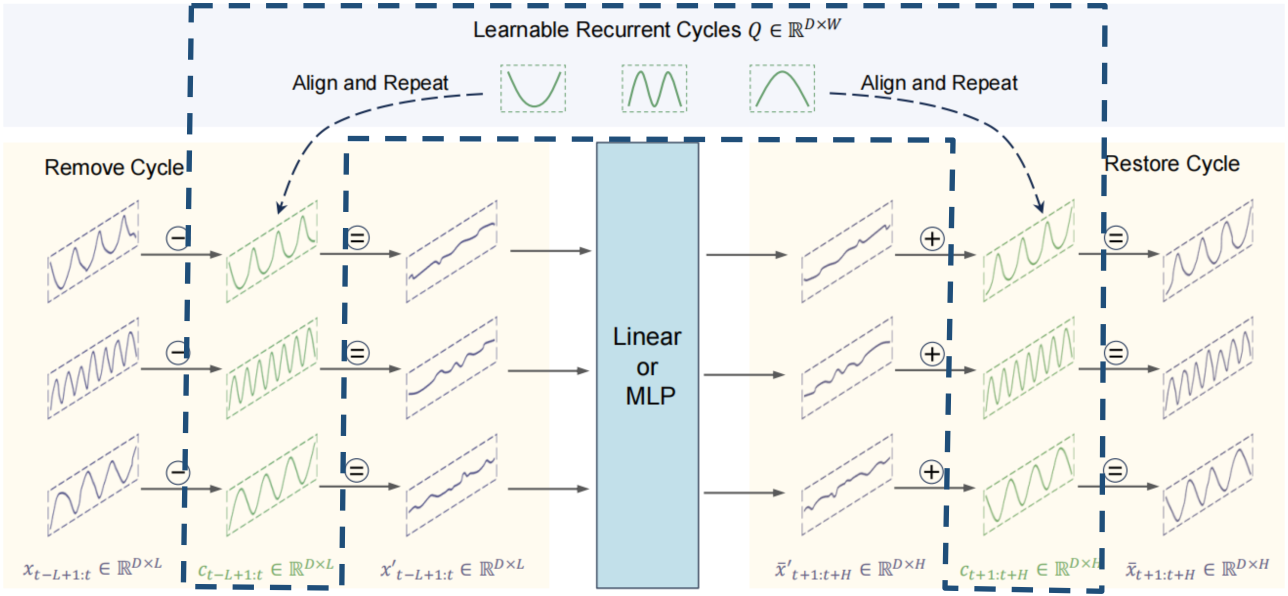
5. Results



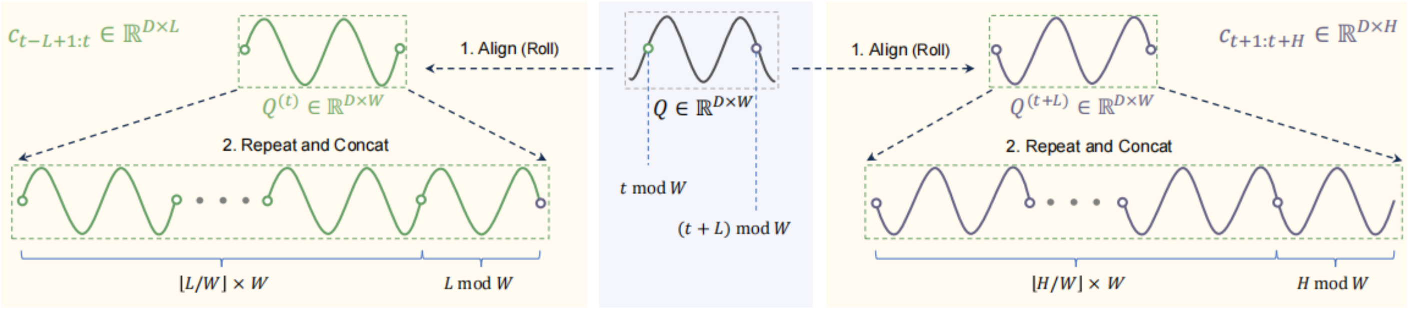
- **Phase 2 (Residual forecasting)**:
  - **(Step 3)** Add the **predicted residual components** $\bar{x}'_{t+1:t+H}$ to the **cyclic components** $c_{t+1:t+H}$ to obtain final prediction $\bar{x}_{t+1:t+H}$



- **Backbone for residual forecasting:**
  - Can be any **existing** time series forecasting model.
  - Combining **Linear** or dual-layer **MLP** forms the proposed simple yet powerful methods, **CycleNet**.



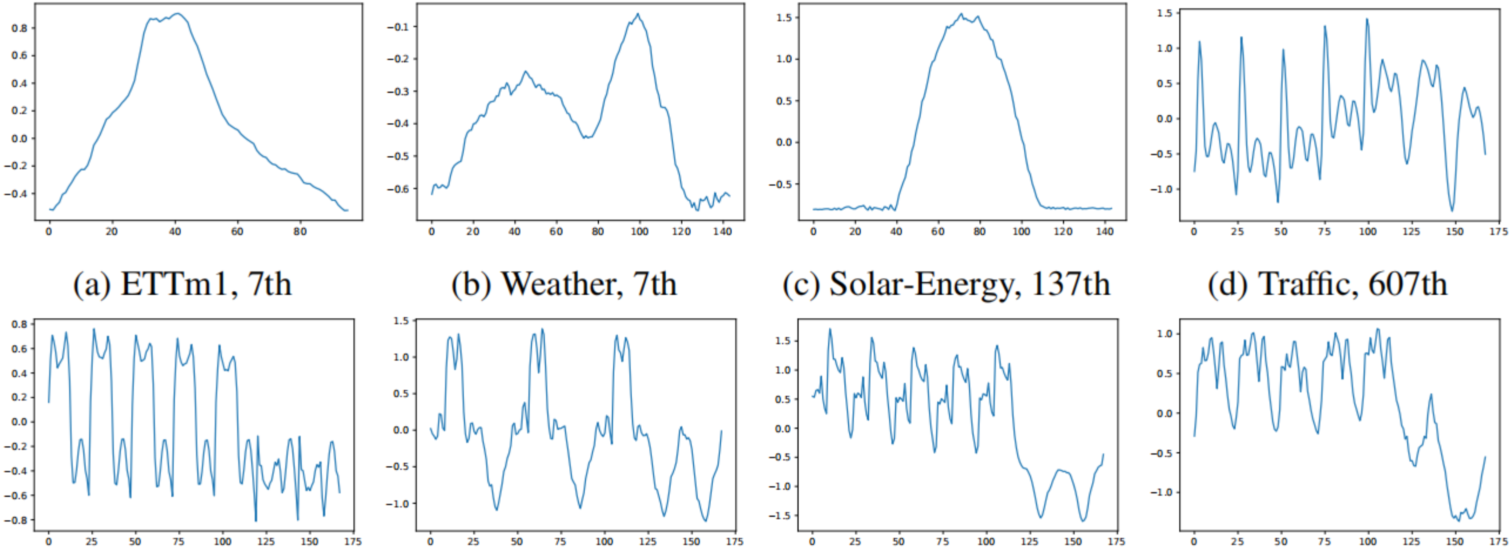
- **Generation of the cyclic components**:
  - Appropriate **alignments and repetitions (**according to **relative positional index** t mod W**)** of the **recurrent cycles** Q are needed to obtain equivalent cyclic components.



- **Generation of the cyclic components**:
  - **Step1: Left-shift** Q by t mod W positions to obtain $Q^{(t)}$. Here, t mod W can be viewed as the **relative positional index** of the current sequence sample within Q.
  - **Step2: Repeat** $Q^{(t)}$ $\lfloor L/W \rfloor$ times and **concatenate** $Q^{(t)}_{0:L \bmod W}$.

$$c_{t-L+1:t} = [\underbrace{Q^{(t)}, \cdots, Q^{(t)}}_{\lfloor L/W \rfloor}, Q^{(t)}_{0:L \bmod W}],$$

$$c_{t+1:t+H} = [\underbrace{Q^{(t+L)}, \cdots, Q^{(t+L)}}_{\lfloor H/W \rfloor}, Q^{(t+L)}_{0:H \bmod W}].$$



(a) ETTm1, 7th (b) Weather, 7th (c) Solar-Energy, 137th (d) Traffic, 607th

(e) Electricity, 311st (f) Electricity, 318th (g) Electricity, 320th (h) Electricity, 321st

The *learnable* recurrent cycles Q in RCF technique can effectively learn the inherent periodic patterns!

1. Motivations
2. Contributions
3. Method
4. Interpretability
5. Results

● Main Results (Multivariate long-term time series forecasting)

Dataset	ETTh1		ETTh2		ETTm1		ETTm2		Electricity		Solar-Energy		Traffic		Weather	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Autoformer [2021]	0.496	0.487	0.450	0.459	0.588	0.517	0.327	0.371	0.227	0.338	0.885	0.711	0.628	0.379	0.338	0.382
FEDformer [2022]	0.440	0.460	0.437	0.449	0.448	0.452	0.305	0.349	0.214	0.327	0.291	0.381	0.610	0.376	0.309	0.360
SCINet [2022]	0.747	0.647	0.954	0.723	0.485	0.481	0.571	0.537	0.268	0.365	0.282	0.375	0.804	0.509	0.292	0.363
DLinear [2023]	0.456	0.452	0.559	0.515	0.403	0.407	0.350	0.401	0.212	0.300	0.330	0.401	0.625	0.383	0.265	0.317
TimesNet [2023]	0.458	0.450	0.414	0.427	0.400	0.406	0.291	0.333	0.192	0.295	0.301	0.319	0.620	0.336	0.259	0.287
TiDE [2023]	0.541	0.507	0.611	0.550	0.419	0.419	0.358	0.404	0.251	0.344	0.347	0.417	0.760	0.473	0.271	0.320
Crossformer [2023]	0.529	0.522	0.942	0.684	0.513	0.496	0.757	0.610	0.244	0.334	0.641	0.639	0.550	0.304	0.259	0.315
PatchTST [2023]	0.469	0.454	0.387	0.407	0.387	0.400	0.281	0.326	0.205	0.290	0.270	0.307	0.481	0.304	0.259	0.281
TimeMixer [2024]	0.447	0.440	**0.364**	**0.395**	0.381	0.395	0.275	0.323	0.182	0.272	0.216	0.280	0.484	0.297	**0.240**	0.271
iTransformer [2024]	0.454	0.447	0.383	0.407	0.407	0.410	0.288	0.332	0.178	0.270	0.233	0.262	**0.428**	**0.282**	0.258	0.278
CycleNet/Linear	**0.432**	**0.427**	0.383	0.404	0.386	**0.395**	0.272	0.315	0.170	0.260	0.235	0.270	0.485	0.313	0.254	0.279
CycleNet/MLP	0.457	0.441	0.388	0.409	**0.379**	0.396	**0.266**	**0.314**	**0.168**	**0.259**	**0.210**	**0.261**	0.472	0.301	0.243	**0.271**

Achieving State-of-the-Art Performance with Minimal Computational Resources

Model	Parameters	MACs	Training Time(s)
Informer [2021]	12.53M	3.97G	70.1
Autoformer [2021]	12.22M	4.41G	107.7
FEDformer [2022]	17.98M	4.41G	238.7
DLinear [2023]	139.6K	44.91M	18.1
PatchTST [2023]	10.74M	25.87G	129.5
iTransformer [2024]	5.15M	1.65G	35.1
CycleNet/MLP	472.9K	134.84M	30.8
CycleNet/Linear	123.7K	22.42M	29.6
RCF part	53.9K	0	12.8

1. Motivations

2. Contributions

3. Method

4. Interpretability

5. Results

- **Ablation study** (Datasets with significant periodicity)

Dataset	Electricity								Traffic							
Horizon	96		192		336		720		96		192		336		720	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Linear	0.197	0.274	0.197	0.277	0.212	0.292	0.253	0.324	0.645	0.383	0.598	0.361	0.605	0.362	0.643	0.381
+ RCF	0.141	0.234	0.155	0.247	0.172	0.264	0.210	0.296	0.480	0.314	0.482	0.313	0.476	0.303	0.503	0.320
Improve	**28.6%**	**14.6%**	**21.4%**	**10.8%**	**18.8%**	**9.5%**	**17.1%**	**8.7%**	**25.6%**	**18.0%**	**19.5%**	**13.2%**	**21.3%**	**16.2%**	**21.8%**	**16.1%**
MLP	0.175	0.259	0.181	0.265	0.197	0.282	0.240	0.317	0.500	0.325	0.496	0.321	0.509	0.325	0.542	0.342
+ RCF	0.136	0.229	0.152	0.244	0.170	0.264	0.212	0.299	0.458	0.296	0.457	0.294	0.470	0.299	0.502	0.314
Improve	**22.2%**	**11.6%**	**15.9%**	**8.0%**	**13.6%**	**6.3%**	**11.6%**	**5.7%**	**8.5%**	**8.9%**	**7.9%**	**8.3%**	**7.7%**	**8.0%**	**7.3%**	**8.1%**
DLinear	0.195	0.278	0.194	0.281	0.207	0.297	0.243	0.331	0.649	0.398	0.599	0.372	0.606	0.375	0.646	0.396
+ RCF	0.143	0.240	0.156	0.253	0.171	0.270	0.204	0.302	0.506	0.317	0.499	0.317	0.512	0.325	0.545	0.343
Improve	**26.6%**	**13.6%**	**19.7%**	**10.0%**	**17.4%**	**8.9%**	**16.3%**	**8.8%**	**22.1%**	**20.4%**	**16.6%**	**14.6%**	**15.4%**	**13.3%**	**15.6%**	**13.5%**
PatchTST	0.168	0.260	0.176	0.266	0.193	0.282	0.233	0.317	0.436	0.281	0.449	0.285	0.464	0.293	0.499	0.310
+ RCF	0.136	0.231	0.153	0.246	0.170	0.264	0.211	0.299	0.438	0.264	0.457	0.270	0.469	0.275	0.509	0.292
Improve	**19.0%**	**11.0%**	**13.0%**	**7.6%**	**11.7%**	**6.6%**	**9.4%**	**5.7%**	-0.5%	**6.1%**	-1.8%	**5.5%**	-1.0%	**6.3%**	-2.0%	**6.1%**
iTransformer	0.148	0.240	0.162	0.253	0.178	0.269	0.225	0.317	0.395	0.268	0.417	0.276	0.433	0.283	0.467	0.302
+ RCF	0.136	0.231	0.153	0.247	0.168	0.263	0.194	0.287	0.415	0.263	0.440	0.271	0.456	0.278	0.491	0.294
Improve	**8.1%**	**3.7%**	**5.6%**	**2.4%**	**5.8%**	**2.2%**	**13.8%**	**9.5%**	-5.1%	**1.9%**	-5.5%	**1.8%**	-5.3%	**1.8%**	-5.1%	**2.6%**

- The RCF technique significantly **enhances** the predictive performance of **basic models** like *Linear* and *MLP*.

- For more **advanced models**, such as *PatchTST*, the RCF technique can also achieve further improvements..

Ablation study (Comparison of different STD techniques)

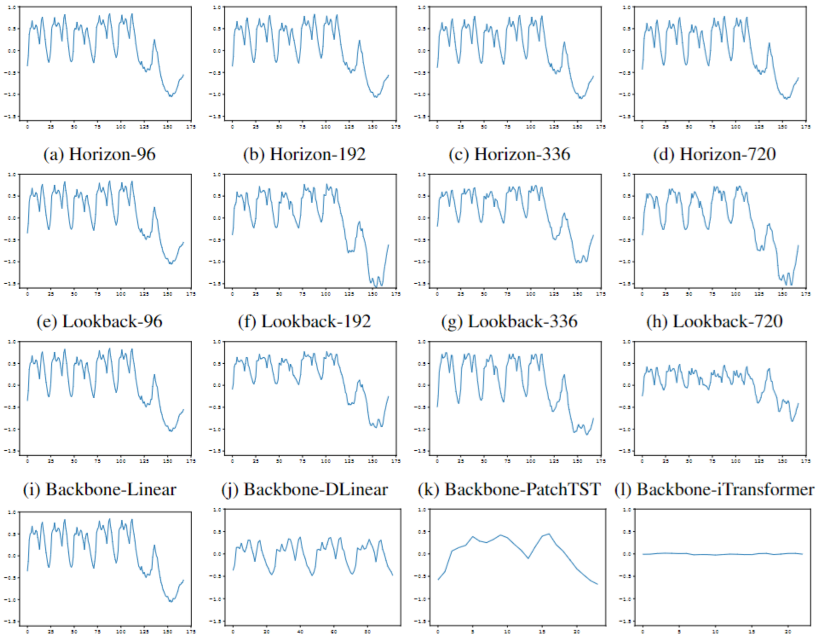
Model		CLinear (RCF+Linear)		LDLinear (LD+Linear)		DLinear (MOV+Linear)		SLinear (Sparse+Linear)		Linear	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.370	0.395	0.372	0.394	0.372	0.394	**0.366**	**0.388**	0.374	0.395
	192	**0.404**	0.417	0.410	0.420	0.408	0.417	0.406	**0.414**	0.409	0.418
	336	**0.434**	**0.440**	0.449	0.452	0.441	0.442	0.440	0.442	0.442	0.444
	720	**0.465**	**0.486**	0.476	0.492	0.480	0.494	0.483	0.501	0.484	0.498
	Avg	**0.418**	**0.434**	0.427	0.439	0.425	0.437	0.424	0.436	0.427	0.439
ETTh2	96	0.308	0.369	**0.292**	**0.357**	0.297	0.362	0.340	0.389	0.305	0.368
	192	0.382	0.416	**0.372**	**0.409**	0.398	0.426	0.379	0.413	0.385	0.419
	336	0.454	0.465	0.479	0.480	0.496	0.489	**0.404**	**0.437**	0.458	0.470
	720	**0.661**	**0.575**	0.675	0.582	0.694	0.592	0.720	0.600	0.691	0.592
	Avg	**0.451**	**0.456**	0.455	0.457	0.471	0.467	0.460	0.460	0.460	0.462
ETTm1	96	**0.298**	0.350	0.305	0.350	0.309	0.356	0.306	**0.349**	0.305	0.349
	192	**0.330**	0.370	0.335	**0.366**	0.346	0.380	0.339	0.370	0.338	0.369
	336	**0.359**	**0.388**	0.372	0.390	0.373	0.391	0.372	0.389	0.371	0.389
	720	**0.410**	**0.421**	0.445	0.443	0.439	0.435	0.430	0.426	0.433	0.428
	Avg	**0.349**	**0.382**	0.365	0.387	0.367	0.390	0.362	0.383	0.362	0.384
ETTm2	96	**0.164**	0.260	0.165	**0.257**	0.165	0.257	0.177	0.272	0.166	0.259
	192	**0.225**	**0.304**	0.240	0.318	0.232	0.310	0.246	0.325	0.228	0.305
	336	**0.271**	**0.332**	0.290	0.349	0.295	0.356	0.309	0.370	0.275	0.334
	720	0.406	0.423	**0.396**	**0.419**	0.427	0.442	0.427	0.440	0.407	0.425
	Avg	**0.266**	**0.330**	0.273	0.336	0.280	0.341	0.290	0.352	0.269	0.331
Electricity	96	**0.131**	**0.228**	0.140	0.237	0.140	0.237	0.148	0.243	0.140	0.238
	192	**0.145**	**0.242**	0.154	0.250	0.154	0.250	0.159	0.254	0.154	0.251
	336	**0.160**	**0.260**	0.170	0.268	0.169	0.268	0.173	0.271	0.170	0.269
	720	**0.193**	**0.292**	0.204	0.300	0.204	0.301	0.207	0.303	0.204	0.301
	Avg	**0.157**	**0.255**	0.167	0.264	0.167	0.264	0.172	0.268	0.167	0.265
Solar-Energy	96	**0.192**	**0.251**	0.222	0.294	0.222	0.298	0.226	0.296	0.224	0.302
	192	**0.218**	**0.258**	0.249	0.315	0.250	0.312	0.252	0.312	0.250	0.310
	336	**0.231**	**0.262**	0.268	0.326	0.270	0.335	0.270	0.326	0.269	0.325
	720	**0.239**	**0.265**	0.271	0.327	0.272	0.327	0.271	0.327	0.270	0.333
	Avg	**0.220**	**0.259**	0.253	0.316	0.254	0.318	0.255	0.315	0.253	0.318
Traffic	96	**0.397**	**0.275**	0.411	0.285	0.411	0.284	0.414	0.281	0.411	0.283
	192	**0.412**	**0.282**	0.423	0.288	0.423	0.289	0.425	0.285	0.423	0.289
	336	**0.426**	**0.290**	0.436	0.296	0.436	0.296	0.436	0.293	0.437	0.297
	720	**0.456**	**0.308**	0.466	0.315	0.466	0.316	0.464	0.310	0.466	0.316
	Avg	**0.423**	**0.289**	0.434	0.296	0.434	0.296	0.435	0.292	0.434	0.296
Weather	96	0.174	0.240	**0.174**	**0.235**	0.175	0.237	0.176	0.235	0.175	0.235
	192	0.218	0.279	**0.215**	**0.271**	0.215	0.273	0.218	0.277	0.218	0.276
	336	0.262	0.314	0.263	0.315	**0.261**	**0.311**	0.265	0.316	0.262	0.312
	720	0.328	0.367	0.325	0.365	**0.324**	**0.363**	0.325	0.363	0.327	0.366
	Avg	0.245	0.300	0.244	0.297	**0.244**	**0.296**	0.246	0.298	0.245	0.297

- The RCF technique can essentially be considered a type of **Seasonal-Trend Decomposition (STD)** method.
- RCF **significantly outperforms** other existing STD methods, particularly on datasets with strong periodicity.
- RCF enables models to overcome the limitations of finite-length look-back windows, as periodic components are **globally estimated** from the entire training set.

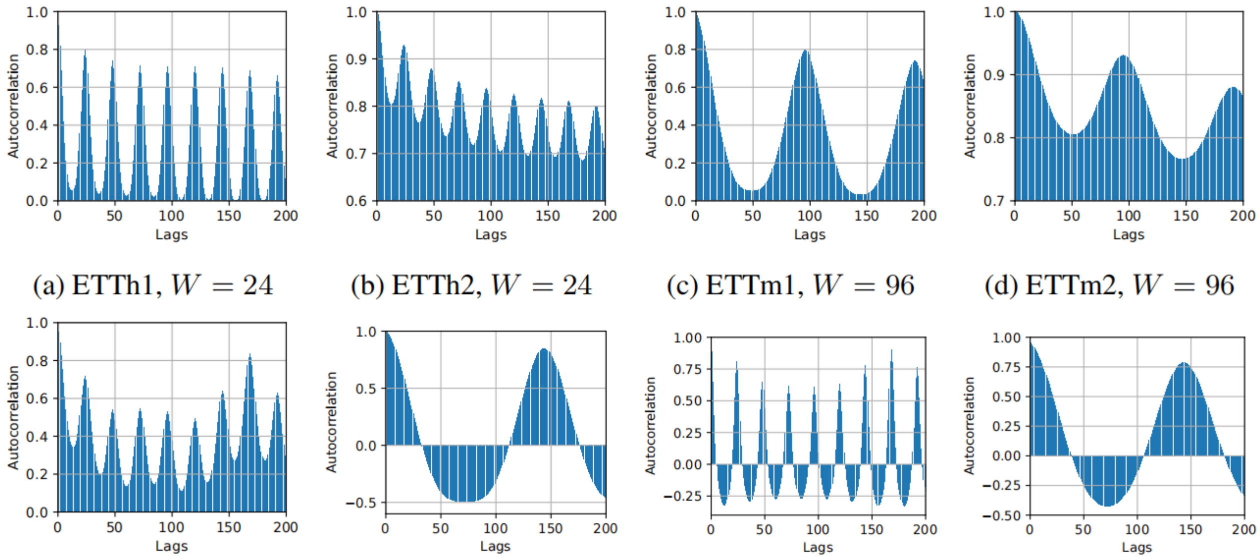
1. Motivations
2. Contributions
3. Method
4. Interpretability
5. Results

● **Further analysis** （The Impact of hyperparameters W）

Setup	RCF/W=168		RCF/W=144		RCF/W=96		RCF/W=24		W/o. RCF	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	**0.142**	**0.234**	0.196	0.275	0.196	0.274	0.195	0.274	0.197	0.274
Trafiic	**0.480**	**0.314**	0.617	0.386	0.617	0.385	0.618	0.385	0.645	0.383
Solar-Energy	0.289	0.376	**0.208**	**0.256**	0.276	0.365	0.287	0.375	0.286	0.375
ETTm1	0.350	0.369	0.340	0.366	**0.325**	**0.363**	0.348	0.367	0.351	0.372
ETTh1	0.395	0.402	0.384	0.395	0.383	0.393	**0.377**	**0.391**	0.384	0.392


(a) Horizon-96 (b) Horizon-192 (c) Horizon-336 (d) Horizon-720
(e) Lookback-96 (f) Lookback-192 (g) Lookback-336 (h) Lookback-720
(i) Backbone-Linear (j) Backbone-DLinear (k) Backbone-PatchTST (l) Backbone-iTransformer
(m) W-168 (n) W-96 (o) W-24 (p) W-23

- ➢ When the length of the learnable recurrent cycles (W) is correctly set to **match the inherent periodicity** of the data,
- ➢ RCF can effectively **learn the correct periodic patterns**, leading to significant performance improvements.
- ➢ RCF is **robust** to other hyperparameters.



(a) ETTh1, $W = 24$ (b) ETTh2, $W = 24$ (c) ETTm1, $W = 96$ (d) ETTm2, $W = 96$

(e) Electricity, $W = 168$ (f) Solar-Energy, $W = 144$ (g) Traffic, $W = 168$ (h) Weather, $W = 144$

- **Determine** the cycle length W through **Autocorrelation Function** (**ACF**):

$$ACF = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}$$

- The hyperparameter W should be set to the ***lag*** corresponding to the **observed maximum peak**.

Thank You!

Paper

Code

Our Team

https://github.com/ACAT-SCUT