Long Sequence Time Series Forecasting Using Spectral-ConvMixer Alongside Weak-stationarizing and Non-stationarity Restoring Blocks

Master's Thesis Defense by:

Ranjai Baidya

Software Major

Department of AI Software

ranjai@gachon.ac.kr

Supervised by:

Sang-Woong Lee

Department of AI Software

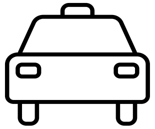
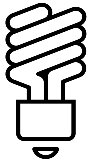
Outline

1. Introduction:
   - Problem of Time Series (Non-stationary problem of time series, Inter and Intra-series Dependencies)
   - Hypothesis
2. Related Works:
   - ARIMA, Autoformer, SCINet, Informer
3. Methodology:
   - Spectral Decomposition, Weak-stationarizing and Non-stationarity Restoring Blocks, ConvMixer, Architecture Overview
4. Contributions
5. Experiments:
   - Datasets, Results, Ablation Study
6. Conclusion and Future Work

1. Introduction

Time Series Forecasting Applications

- Time Series Forecasting holds prominent roles in many real-life applications:

Transportation		Traffic forecasting for traffic management.
Healthcare		Forecasting number of Covid patients to makes beds, isolation centers, oxygen etc. available.
Energy Management		Managing energy production as per forecasted energy usages.

Problem Definition

- Real-life time series data are Non-stationary.
- Two different sequences of same real-life time series have different distribution.

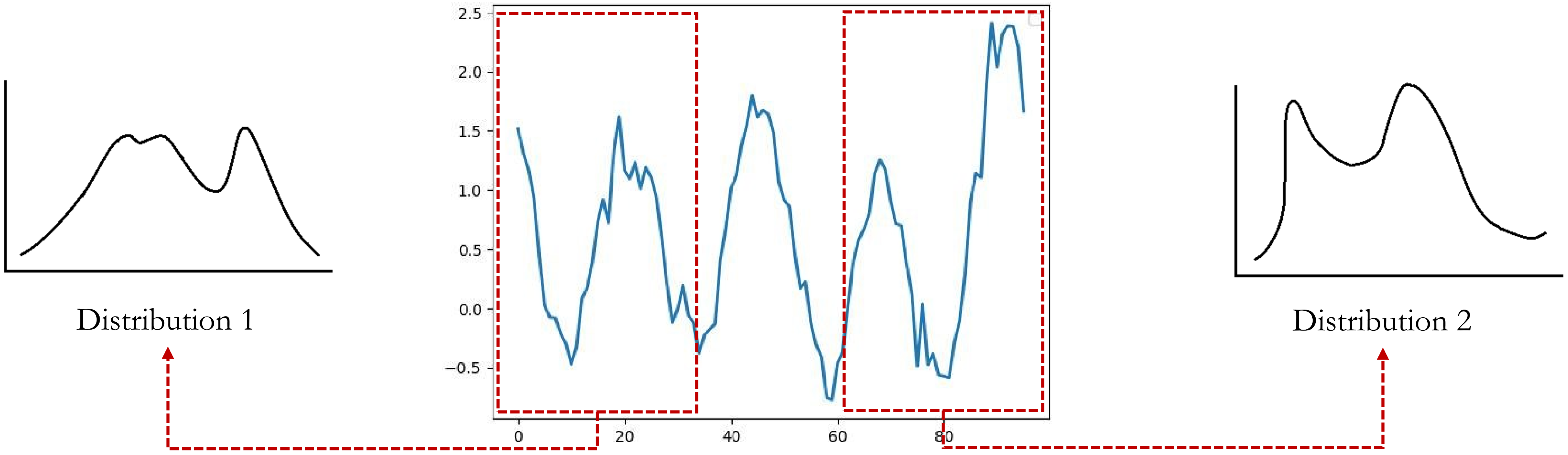


Fig 1: Distribution Shift in Real-life Time Series.

Problem Definition

- Inter and Intra Series dependencies in time series data.

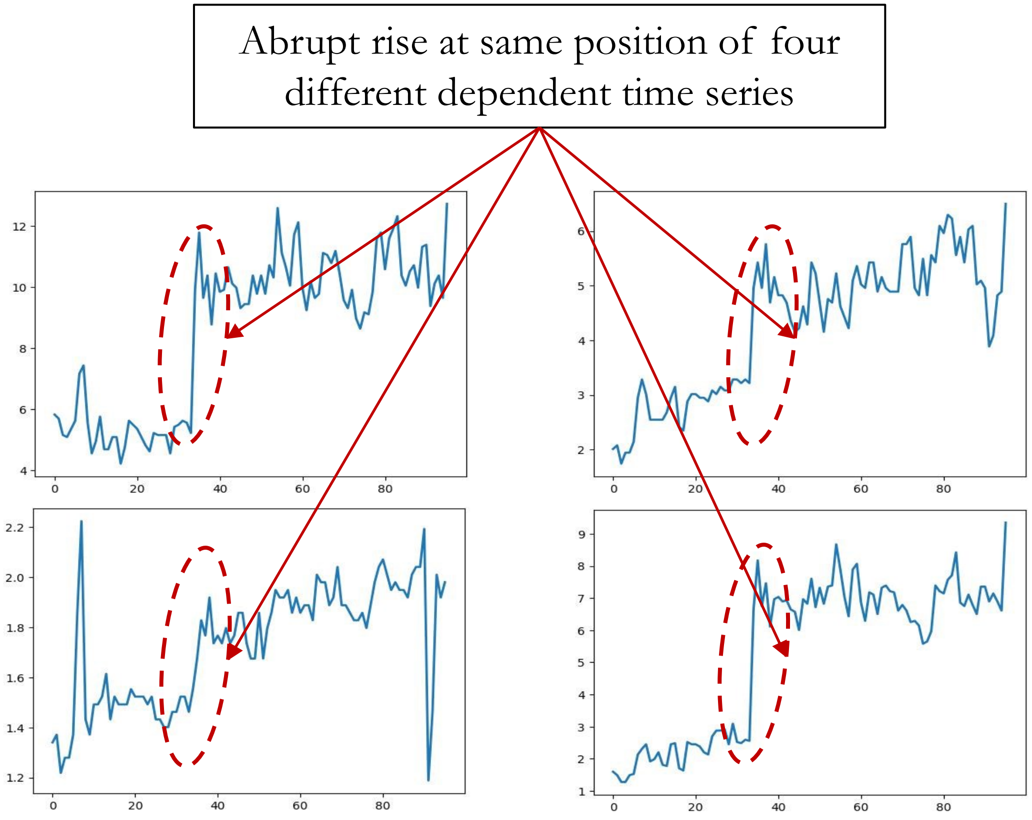


Fig 2: Illustration of Inter-series relationships in Multivariate Scenario.

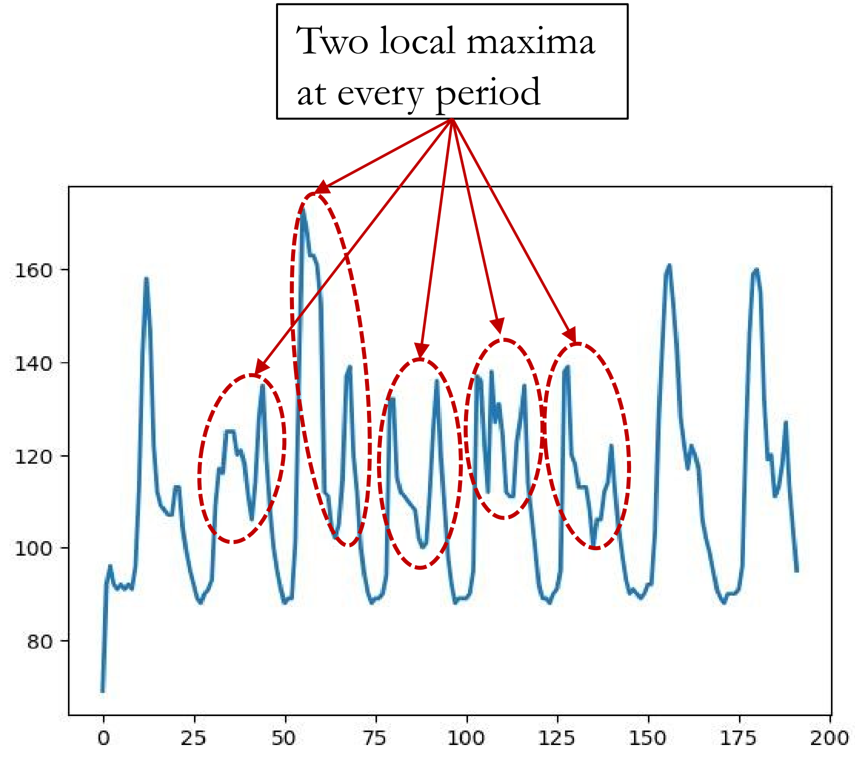


Fig 3: Illustration of Intra-series relationship.

Hypothesis

- Differenced datapoints are stationary and easier to forecast than the original datapoints[1].
- Differencing can be employed for obtaining stationarity.
- Simplifying the data into more understandable form can help in forecasts.
- Better forecasts can be obtained if inter-series and intra-series relationship in the time-series data can be modeled.

[1]. Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of statistical software*, *27*, 1-22.

2. Related Works

ARIMA (Autoregressive Integrated Moving Average) [1]

- **A**uto **R**egressive**:** Using lagged values

  $$Y = B_0 + B_1{*}Y_{lag1} + B_2{*}Y_{lag2} + \ldots + B_{\boldsymbol{n}}{*}Y_{lag\boldsymbol{n}}$$

- Integrated: use of differences → **Outputs a stationary time series**

  $$Y_{forward} - Y = B_0 + B_1{*}(Y - Y\_lag_1) + B_2{*}(Y_{lag1} - Y_{lag2}) + \ldots$$

- **M**oving **A**verage: lagged prediction errors

  $$Y = B_0 + B_1{*}E_{lag1} + B_2{*}E_{lag2} + \ldots + B_{\boldsymbol{n}}{*}E_{lag\boldsymbol{n}}$$

[1]. Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of statistical software*, *27*, 1-22.

Informer

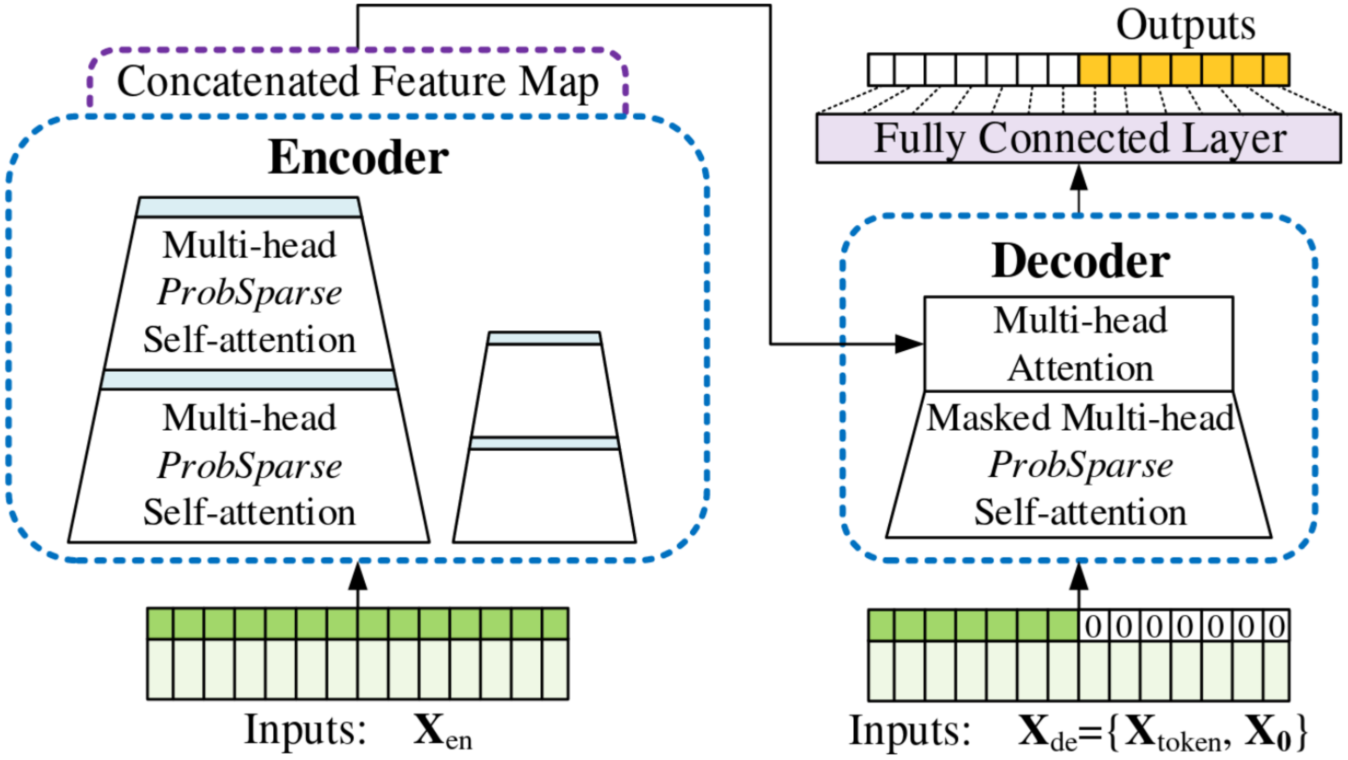


Fig 4: Overall Architecture of Informer [2].

Main Idea:

- Sparser query and key matrices to calculate attention: Multi-head ProbSparse Attention.

[2]. Liu, M., Zeng, A., Lai, Q., & Xu, Q. (2021). Time Series is a Special Sequence: Forecasting with Sample Convolution and Interaction. arXiv preprint arXiv:2106.09305.

Autoformer

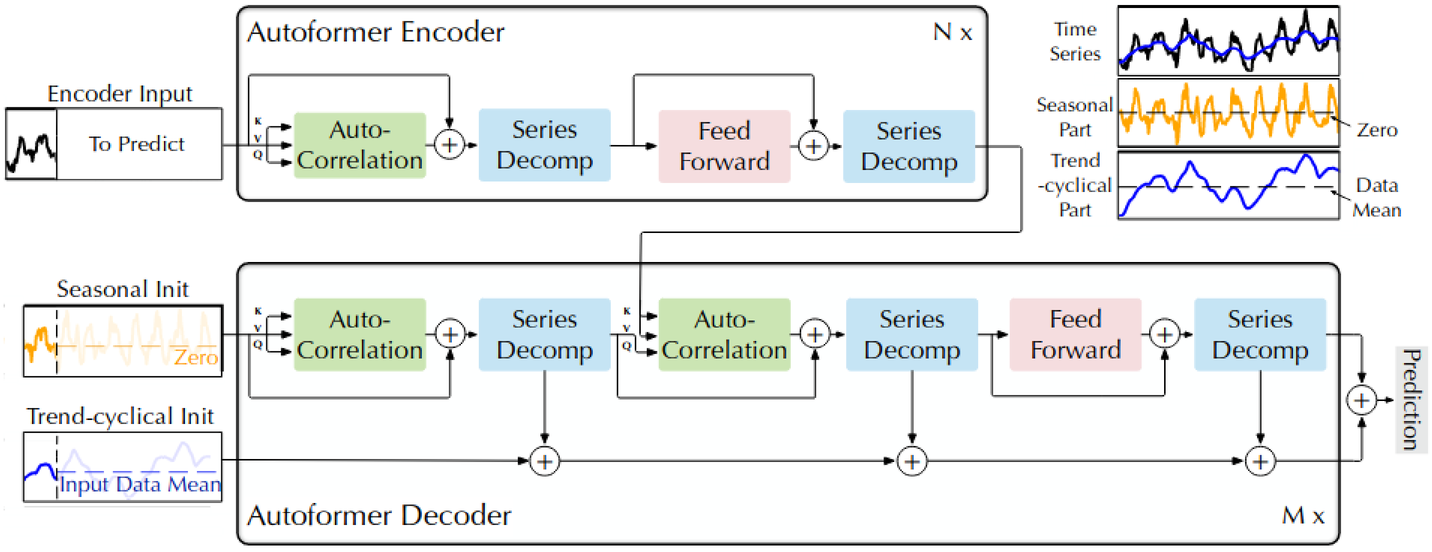


Fig 5: Overall Architecture of Autoformer [3].

Main Ideas:

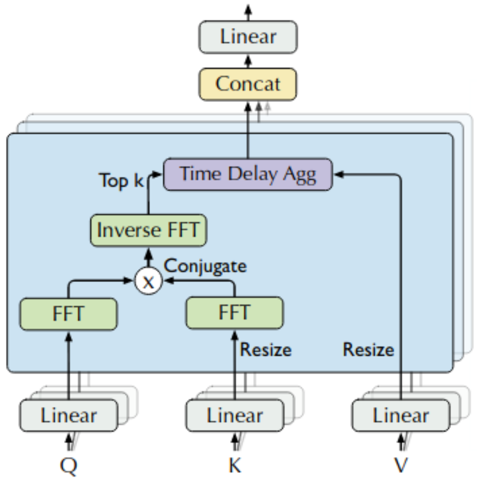
- Series Decomposition
- Autocorrelation

Series Decomposition

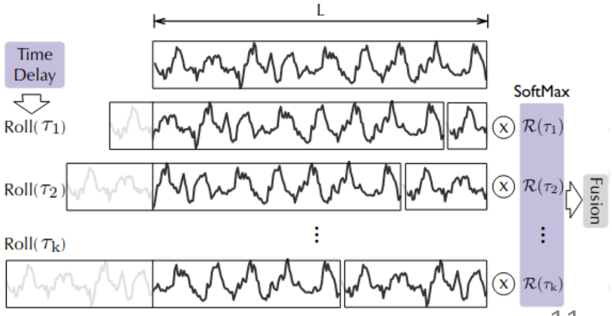
$$\underbrace{X_t = \text{AvgPool}(\text{Padding}(X))}_{\text{Trend}}$$

$$\underbrace{X_s = X - X_t}_{\text{Seasonality}}$$

Auto-correlation



Time Delay Aggregation



[3]. Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems, 34*.

SCINet

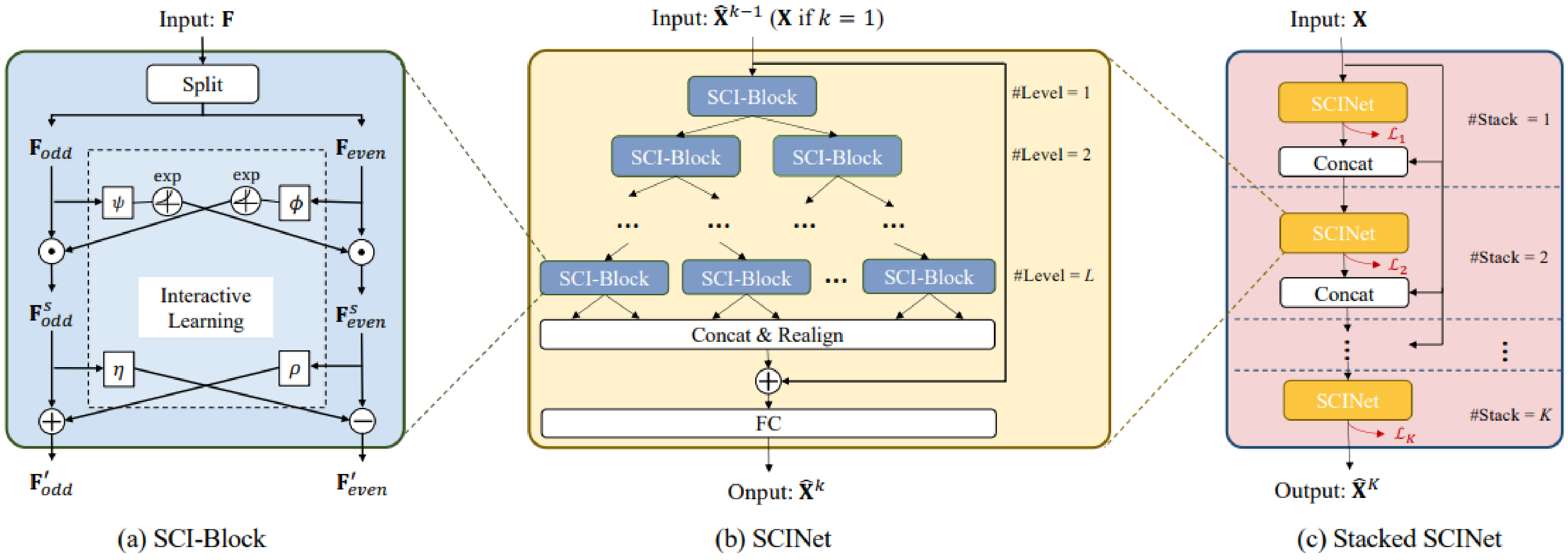


Fig 6: Overall Architecture of SCINet [4].

Main Ideas:

- Capturing multiple temporal dependencies at multiple temporal resolutions.
- Unique Interactive Learning block.

[4]. Liu, M., Zeng, A., Lai, Q., & Xu, Q. (2021). Time Series is a Special Sequence: Forecasting with Sample Convolution and Interaction. arXiv preprint arXiv:2106.09305.

3. Methodology

Spectral Decomposition

$$X_{t:t+T} = \sum_{f} A(f)\cos\big(2\pi f(t:t+T)\big) + B(f)\sin(2\pi f(t:t+T))$$

Where, A(f) and B(f) are the amplitudes of the sine and cosine components at frequency f.

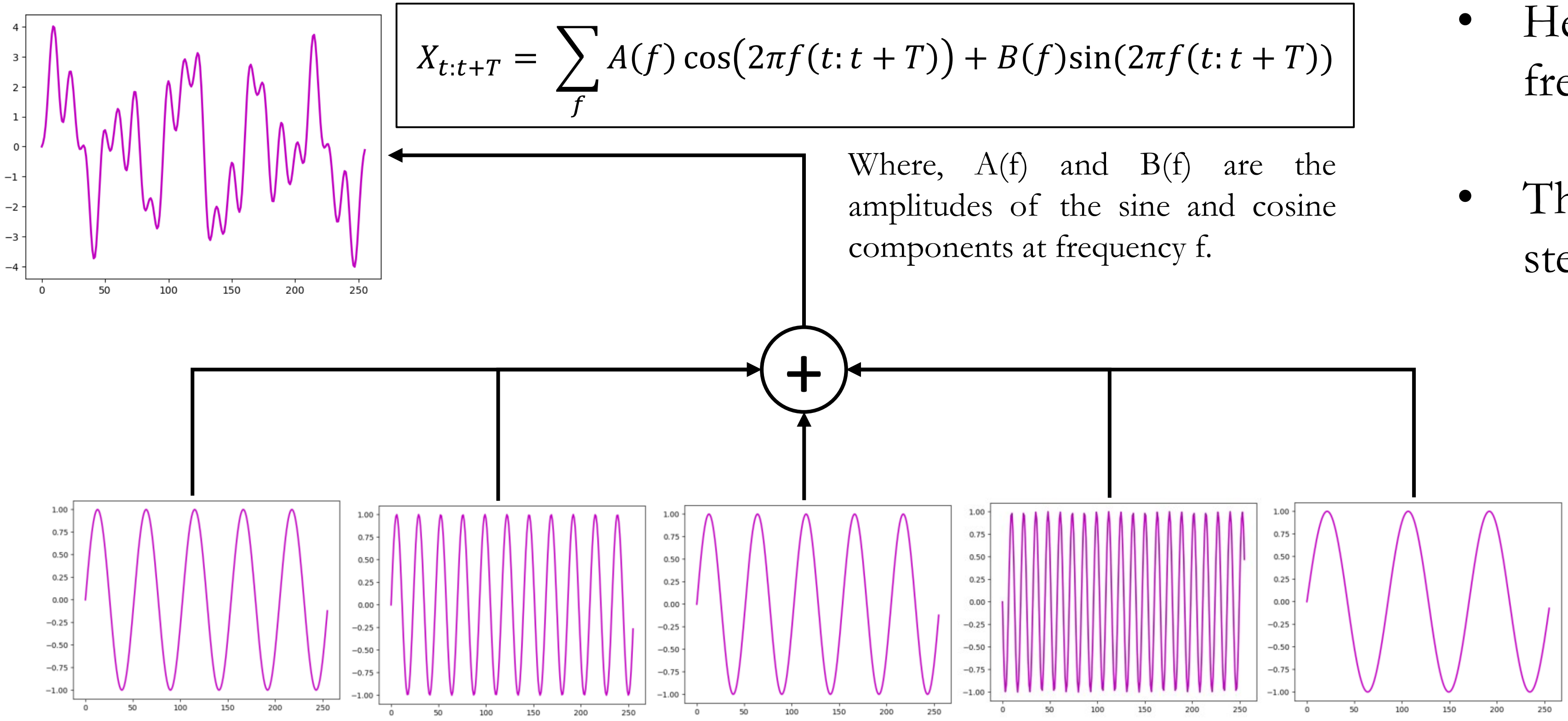


Fig 7: Visualization of Spectral Decomposition.

- Helps in realizing the comprised frequency components.
- This is used as a simplification step.

Equation Source:

[5]. Koopmans, L. H. (1995). The spectral analysis of time series. Elsevier.

Weak-stationarizing Block

- Use of differencing inspired from the ARIMA [1] model.

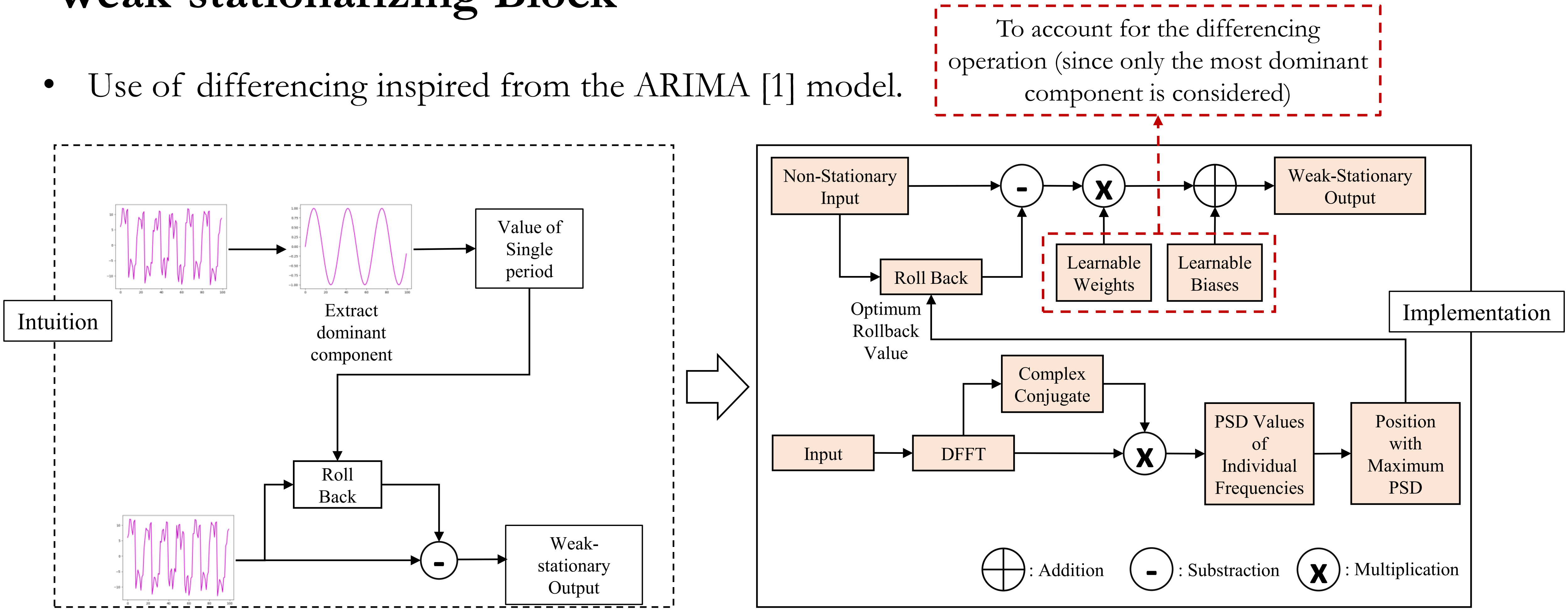


Fig 8: Concept and Structure of Weak-Stationarizing Block.

Non-stationarity Restoring Block

- Non-stationary information removed by Weak-stationarizing block important for forecasting.
- Restores the non-stationary information before the forecast is made.

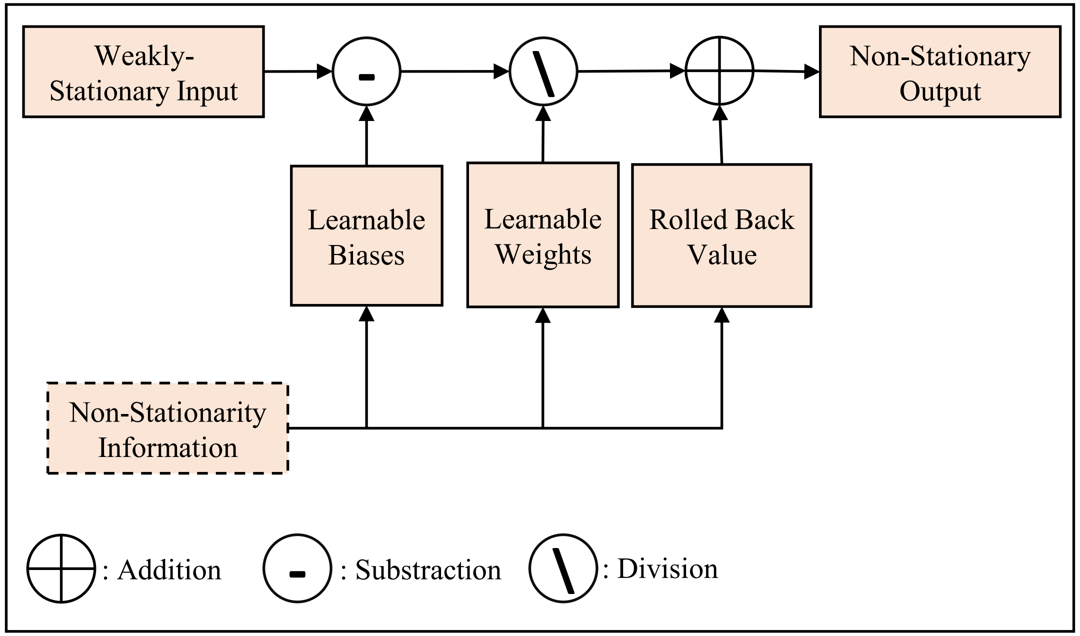


Fig 9: Structure of Non-stationarity Restoring Block.

ConvMixer

- Mixer architectures shuffle data spatially and channel-wise [6].
- This shuffling is equivalent to mixing the data in terms of the dependent time series and in terms of temporal location.

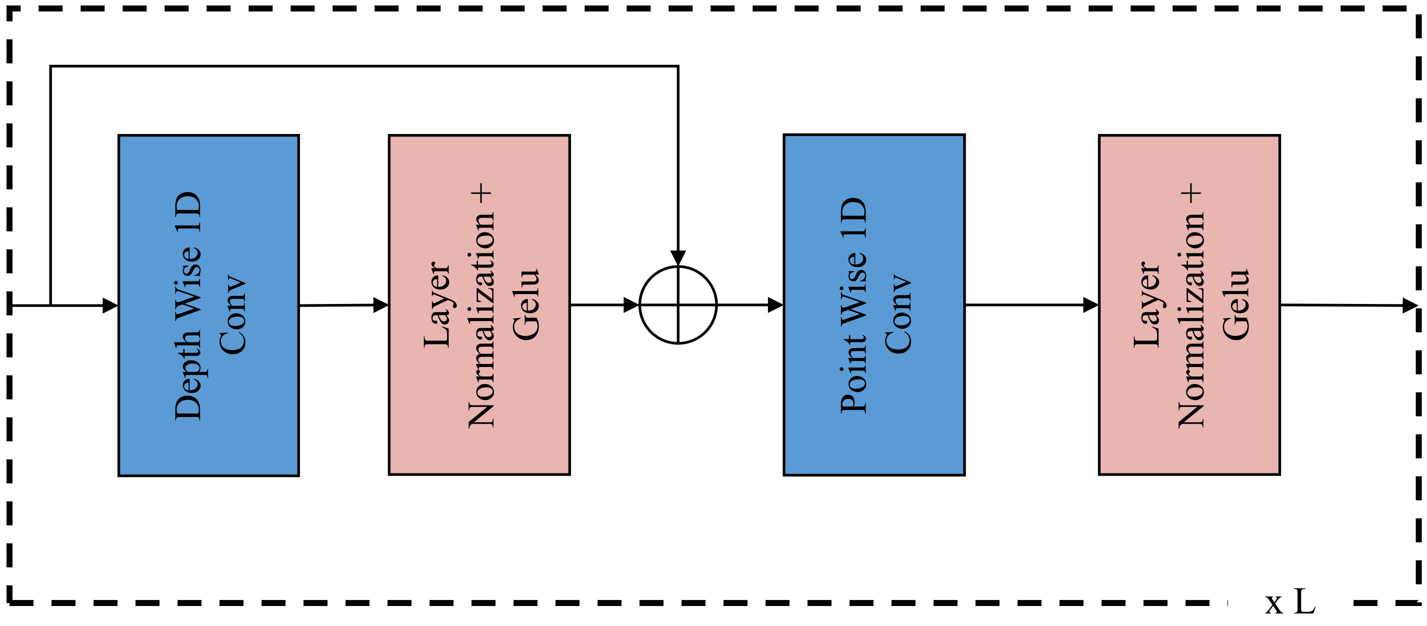


Fig 10: Structure of Mixer Layer.

[6]. Trockman, A., & Kolter, J. Z. (2022). Patches Are All You Need?. *arXiv preprint arXiv:2201.09792*.

Overall Architecture

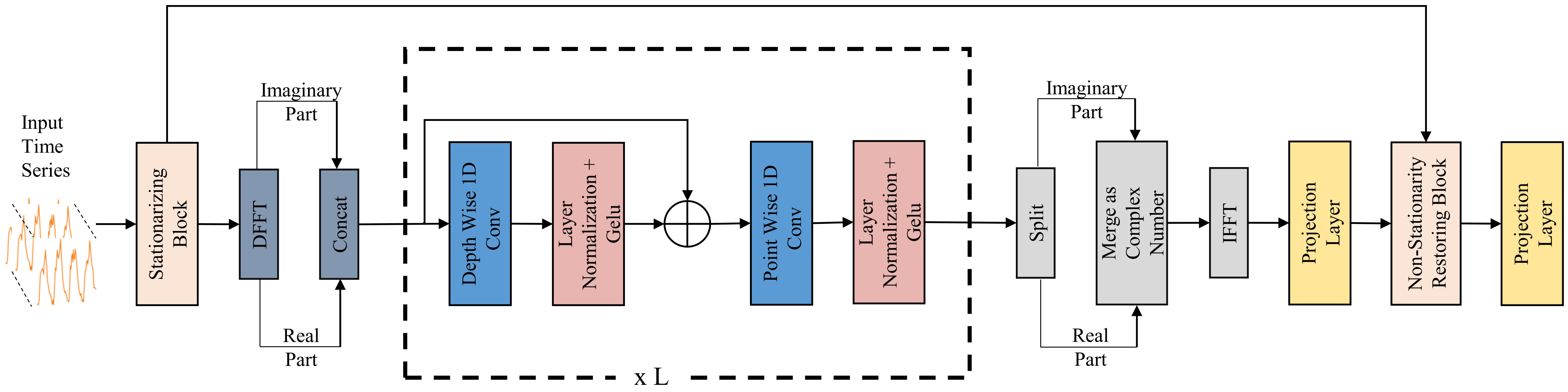


Fig 11: Overall Structure of Suggested Architecture.

Contributions

- Propose a deep learning forecasting framework to deal with both univariate and multivariate settings.
- 'Weak-stationarizing' and 'Non-stationarity Restoring' blocks to deal with non-stationarity of time series.
- Deal with the spectral components of time series and utilize ConvMixer [6] architecture to obtain quality forecasts.
- Achieve an average of average of 21% and up to 64.6% of relative performance improvements on 6 real-world datasets.

[6]. Trockman, A., & Kolter, J. Z. (2022). Patches Are All You Need?. *arXiv preprint arXiv:2201.09792*.

4. Experiments

Datasets

- **ETT [2] :** Data related to electricity transformers collected.
- **Electricity [7]:** Data related to electricity consumption of 321 customers.
- **Weather [8]:** Dataset of 21 different meteorological indicators collected.
- **Traffic [9]:** Data related to road occupancy rate measured by different sensors on San Francisco Bay area freeways.
- **ILI [10]:** Data of patients displaying influenza like illness.
- **Exchange [11]:** Data of daily exchange rate of eight different countries.

[7]. https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014 [8]. https://www.bgc-jena.mpg.de/wetter/ [9]. https://pems.dot.ca.gov/ [10]. https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

[11]. Lai, G., Chang, W. C., Yang, Y., & Liu, H. (2018, June). Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 95-104).

Results for Multivariate Settings

	Models	Suggested		Autoformer [4]		SCINet [5]		Informer [3]		LogTrans [12]		Reformer [13]		LSTNet [11]	
Datasets	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	**0.183**	**0.259**	0.255	0.339	0.413	0.470	0.365	0.453	0.768	0.642	0.658	0.619	3.142	1.365
	192	**0.245**	**0.303**	0.281	0.340	0.433	0.481	0.533	0.563	0.989	0.757	1.078	0.827	3.154	1.369
	336	**0.307**	**0.348**	0.339	0.372	0.633	0.580	1.363	0.887	1.334	0.872	1.549	0.972	3.160	1.369
	720	**0.405**	**0.404**	0.422	0.419	0.864	0.680	3.379	1.388	3.048	1.328	2.631	1.242	3.171	1.368
Electricity	96	**0.154**	**0.249**	0.201	0.317	0.212	0.321	0.274	0.368	0.258	0.357	0.312	0.402	0.680	0.645
	192	**0.166**	**0.261**	0.222	0.334	0.242	0.345	0.296	0.386	0.266	0.368	0.348	0.433	0.725	0.676
	336	**0.177**	**0.275**	0.231	0.338	0.248	0.354	0.300	0.394	0.280	0.380	0.350	0.433	0.828	0.727
	720	**0.231**	**0.326**	0.254	0.361	0.270	0.368	0.373	0.439	0.283	0.376	0.340	0.420	0.957	0.811
Exchange	96	**0.082**	**0.203**	0.197	0.323	0.309	0.412	0.847	0.752	0.968	0.812	1.065	0.829	1.551	1.058
	192	**0.149**	**0.283**	0.300	0.369	1.354	0.783	1.204	0.895	1.040	0.851	1.188	0.906	1.477	1.028
	336	**0.243**	**0.368**	0.509	0.524	1.656	0.888	1.678	1.036	1.659	1.081	1.357	0.976	1.507	1.031
	720	**0.509**	**0.559**	1.447	0.941	1.272	0.855	2.478	1.310	1.941	1.127	1.510	1.016	2.285	1.243

Table 1: Results for Multivariate Setting

Results for Multivariate Settings

	Models	Suggested		Autoformer [3]		SCINet [4]		Informer [2]		LogTrans [12]		Reformer [13]		LSTNet [11]	
Datasets	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	96	**0.516**	**0.316**	0.613	0.388	0.690	0.440	0.719	0.391	0.684	0.384	0.732	0.423	1.107	0.685
	192	**0.499**	**0.307**	0.616	0.382	0.708	0.453	0.696	0.379	0.685	0.390	0.733	0.420	1.157	0.685
	336	**0.525**	**0.327**	0.622	0.337	0.752	0.474	0.777	0.420	0.733	0.408	0.742	0.420	1.216	0.730
	720	**0.557**	**0.337**	0.660	0.408	0.812	0.494	0.864	0.472	0.717	0.396	0.755	0.423	1.481	0.805
Weather	96	**0.206**	**0.230**	0.266	0.336	**0.190**	0.258	0.300	0.384	0.458	0.490	0.689	0.596	0.594	0.587
	192	**0.242**	**0.264**	0.307	0.367	**0.235**	0.298	0.598	0.544	0.658	0.586	0.752	0.638	0.560	0.587
	336	**0.283**	**0.299**	0.359	0.395	0.292	0.343	0.578	0.523	0.797	0.652	0.639	0.596	0.597	0.587
	720	**0.341**	**0.343**	0.419	0.428	0.377	0.401	1.059	0.741	0.869	0.675	1.130	0.792	0.618	0.599
ILI	24	**2.564**	**1.034**	3.483	1.287	11.293	2.576	1.677	4.480	4.480	1.444	4.400	1.382	6.026	1.770
	36	**2.165**	**0.945**	3.103	1.148	10.817	2.468	1.467	4.799	4.799	1.467	4.783	1.448	5.340	1.668
	48	**2.323**	**0.994**	2.669	1.085	10.982	2.467	1.469	4.800	4.800	1.468	4.832	1.465	6.080	1.787
	60	**2.293**	**0.998**	2.770	1.125	10.967	2.479	1.564	5.278	5.278	1.560	4.882	1.483	5.548	1.720

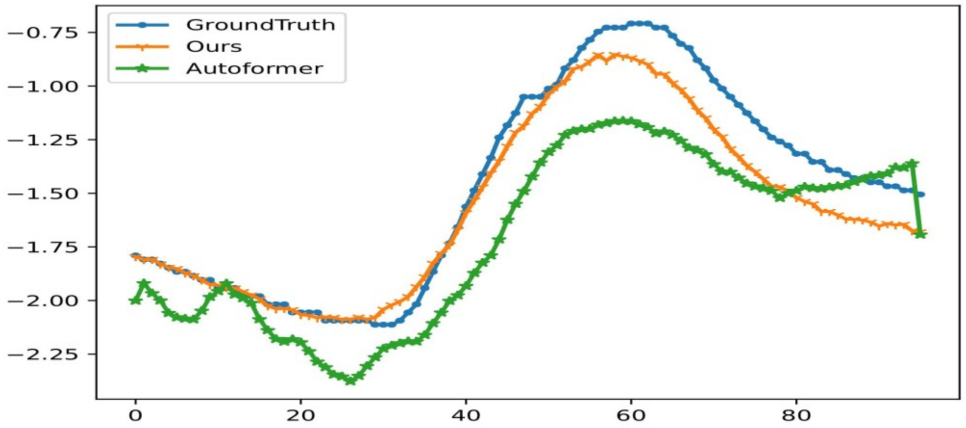
Table 2: Results for Multivariate Setting (Continued)

Results for Univariate Settings

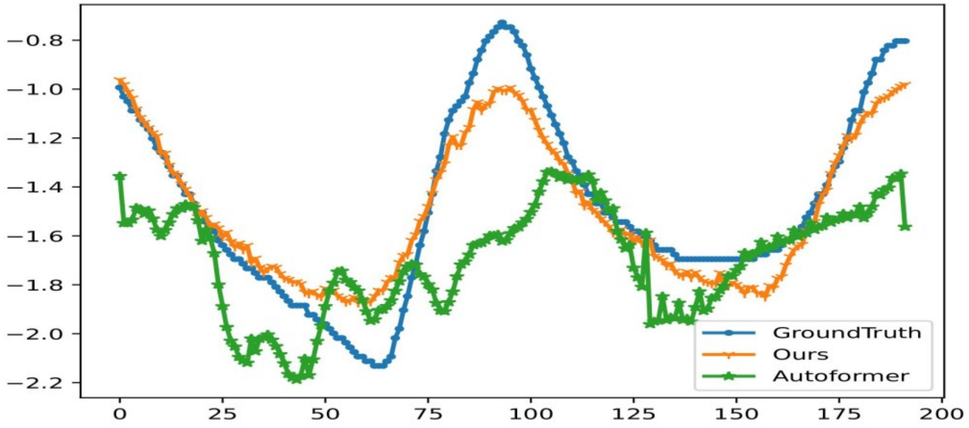
	Models	Suggested		Autoformer [3]		SCINet [4]		Informer [2]		LogTrans [12]		DeepAR [14]		ARIMA [1]	
Datasets	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTM2	96	0.071	0.190	**0.065**	**0.189**	0.0821	0.217	0.088	0.225	0.082	0.217	0.099	0.237	0.211	0.362
	192	**0.104**	**0.237**	0.118	0.256	0.187	0.341	0.132	0.283	0.133	0.284	0.154	0.310	0.261	0.406
	336	**0.134**	**0.277**	0.154	0.305	0.171	0.324	0.180	0.336	0.201	0.361	0.277	0.428	0.317	0.448
	720	**0.180**	**0.326**	0.182	0.335	0.198	0.346	0.300	0.435	0.268	0.407	0.332	0.468	0.366	0.487
Exchange	96	**0.092**	**0.228**	0.241	0.387	0.207	0.362	0.591	0.615	0.279	0.441	0.417	0.515	0.112	0.245
	192	**0.184**	**0.348**	0.273	0.403	0.395	0.497	1.183	0.912	1.950	1.048	0.813	0.735	0.304	0.404
	336	**0.326**	**0.451**	0.508	0.539	0.659	0.640	1.367	0.984	2.438	1.262	1.331	0.962	0.736	0.598
	720	1.036	0.791	**0.991**	**0.768**	0.875	1.872	1.872	1.072	2.010	1.247	1.894	1.181	1.871	0.935

Table 3: Results for Univariate Setting

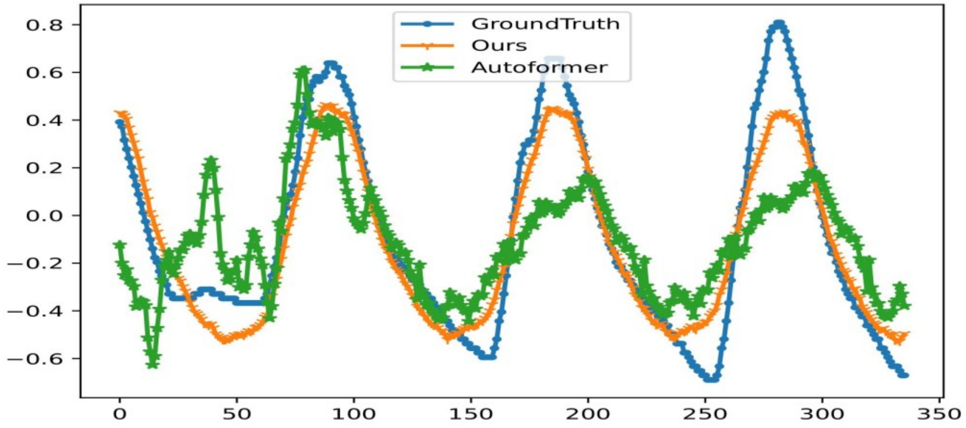
Visualization of Forecasts



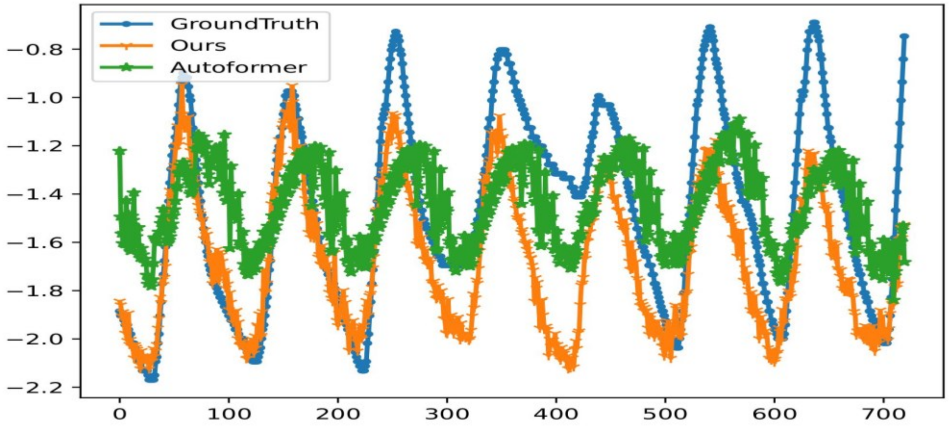
(a)



(b)



(c)



(d)

Fig 12: Visualizing the forecasting results for ETTm2 Dataset (a) Horizon Length=96 (b) Horizon Length=96 (c) Horizon Length=96 (d) Horizon Length=96

Visualization of Forecasts

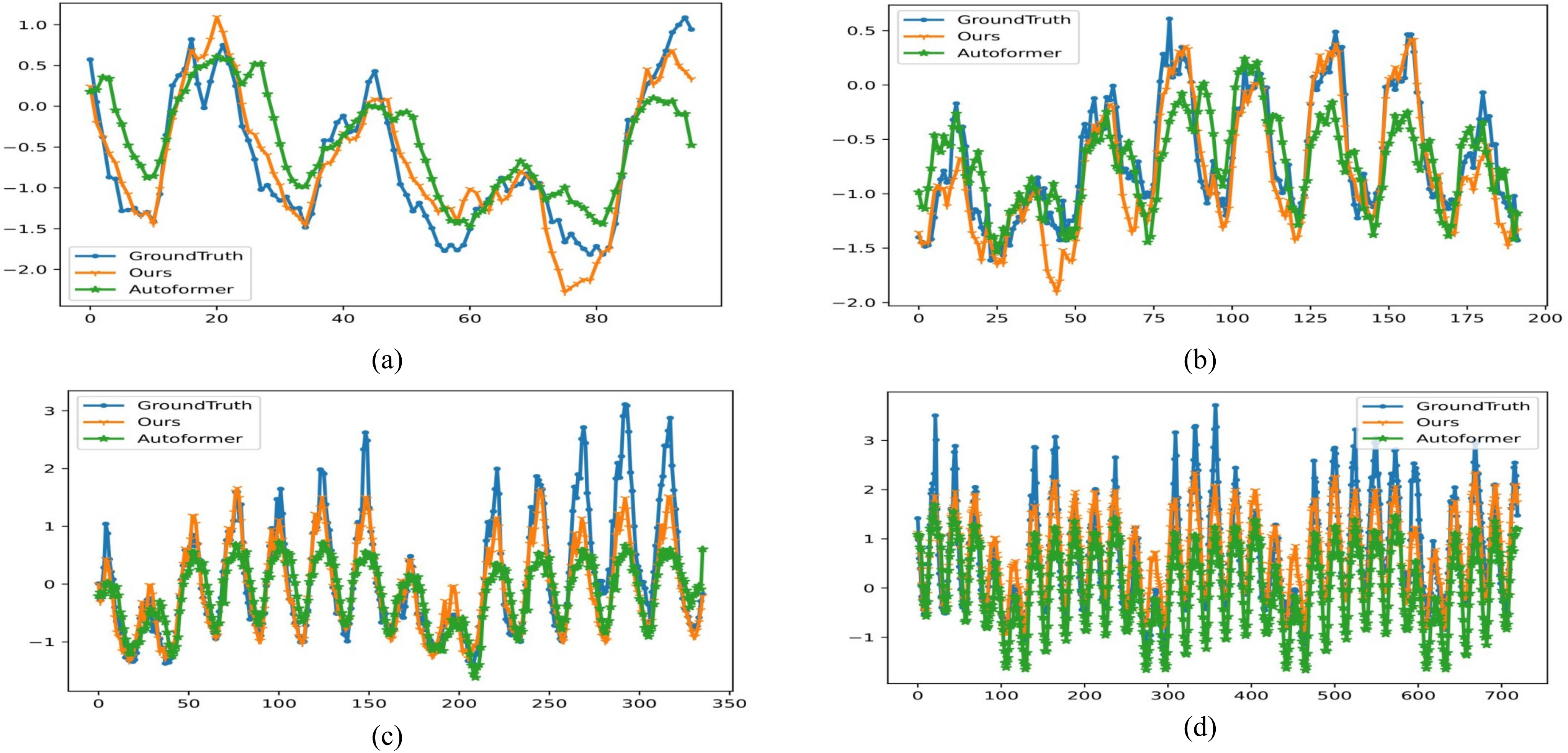
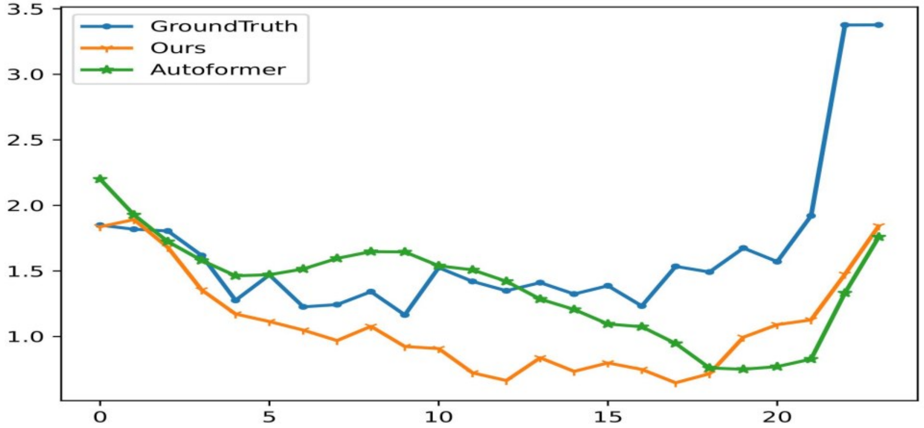
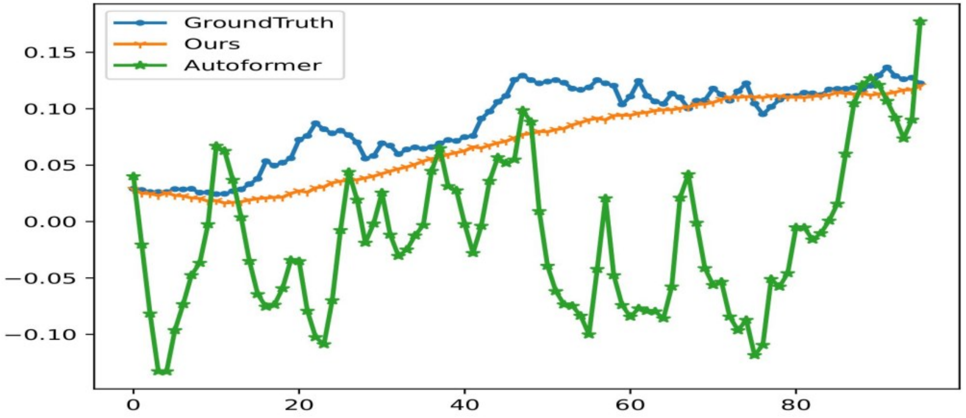


Fig 13: Visualizing the forecasting results for Electricity Dataset (a) Horizon Length=96 (b) Horizon Length=96 (c) Horizon Length=96 (d) Horizon Length=96

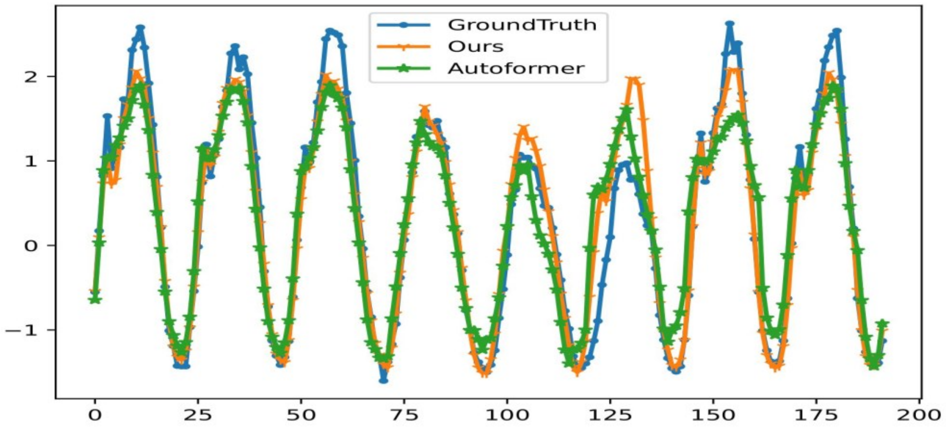
Visualization of Forecasts



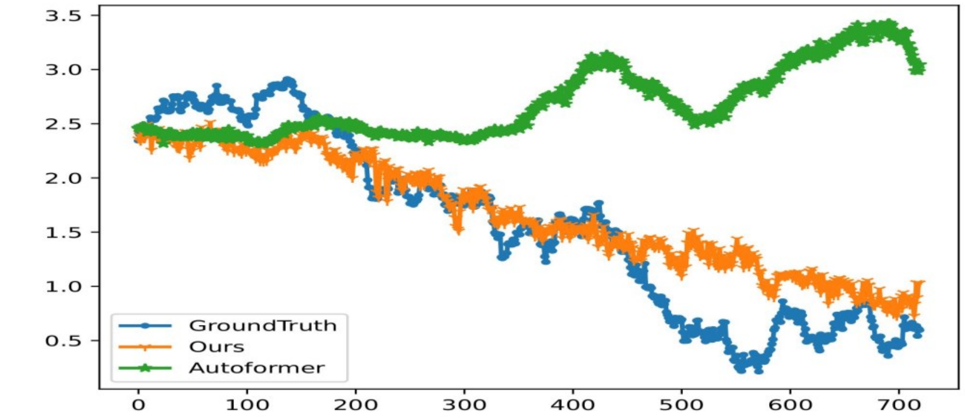
(a)



(b)



(c)



(d)

Fig 14: Visualizing the forecasting results for Illness, Weather, Traffic and Exchange Datasets (a) Illness with Horizon 24 (b) Weather with horizon 96 (c) Traffic with horizon 192 (d) Exchange with horizon 720

Ablation Study

	Models	With WS and NSR Blocks				Without WS and NSR Blocks			
	Variation	Spectral		Time		With Skip Connection		Without Skip Connection	
Datasets	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	24	0.253	0.313	0.256	0.314	**0.232**	**0.300**	0.260	0.330
	48	**0.308**	**0.340**	0.321	0.349	0.327	0.360	0.368	0.401
	96	**0.338**	**0.354**	0.341	0.360	0.342	0.373	0.462	0.474
	288	**0.402**	**0.397**	0.404	0.369	0.406	0.404	0.541	0.530
	672	**0.474**	**0.439**	0.477	0.441	0.489	0.460	0.717	0.635
ECL	96	**0.163**	**0.260**	0.183	0.274	0.183	0.282	0.310	0.398
	192	**0.177**	**0.276**	0.188	0.280	0.196	0.294	0.332	0.413
	336	**0.194**	**0.295**	0.202	0.295	0.215	0.320	0.310	0.388
	720	**0.238**	**0.330**	0.248	0.339	0.242	0.338	0.325	0.399
Exchange	96	**0.086**	**0.207**	0.092	0.215	0.298	0.417	1.048	0.830
	192	**0.153**	**0.283**	0.154	0.285	0.364	0.482	1.591	1.031
	336	**0.243**	**0.368**	0.252	0.378	0.397	0.475	1.984	1.114
	720	0.920	0.715	**0.827**	**0.684**	0.947	0.730	2.579	1.220

Table 4: Impact of processing the time series in spectral domain and the impact of usage of 'Weak-stationarizing' (WS) and 'Non-stationarity Restoring' (NSR) blocks

Ablation Study

Number of Layers		1		2		3		4		5	
Datasets	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTM1	24	0.253	0.313	0.322	0.347	0.252	0.314	0.257	0.316	**0.251**	**0.311**
	48	0.308	0.340	0.316	0.345	**0.305**	**0.338**	0.316	0.346	0.310	0.342
	96	**0.338**	**0.354**	0.344	0.361	0.347	0.363	0.339	0.358	0.353	0.365
	288	0.402	0.397	**0.402**	**0.395**	0.406	0.397	0.406	0.398	0.415	0.404
	672	0.474	0.439	0.471	0.435	0.478	0.442	**0.468**	**0.434**	0.479	0.443
ECL	96	0.163	0.260	0.158	0.256	0.156	0.254	0.156	0.253	**0.154**	**0.251**
	192	0.177	0.277	0.171	0.271	0.167	0.267	0.168	0.267	**0.166**	**0.265**
	336	0.194	0.295	0.185	0.286	0.194	0.295	0.184	0.287	**0.183**	**0.285**
	720	0.238	0.330	0.220	0.319	0.218	0.318	**0.213**	**0.313**	0.222	0.321
Exchange	96	0.086	0.207	**0.085**	**0.204**	0.086	0.207	0.085	0.205	0.087	0.209
	192	**0.153**	**0.283**	0.154	0.285	0.163	0.289	0.156	0.287	0.154	0.283
	336	**0.243**	**0.368**	0.244	0.374	0.249	0.374	0.252	0.379	0.243	0.373
	720	0.921	0.715	**0.887**	**0.698**	0.932	0.719	1.008	0.209	0.903	0.707

Table 5: Impact of varying number of ConvMixer Layers

Efficiency Analysis

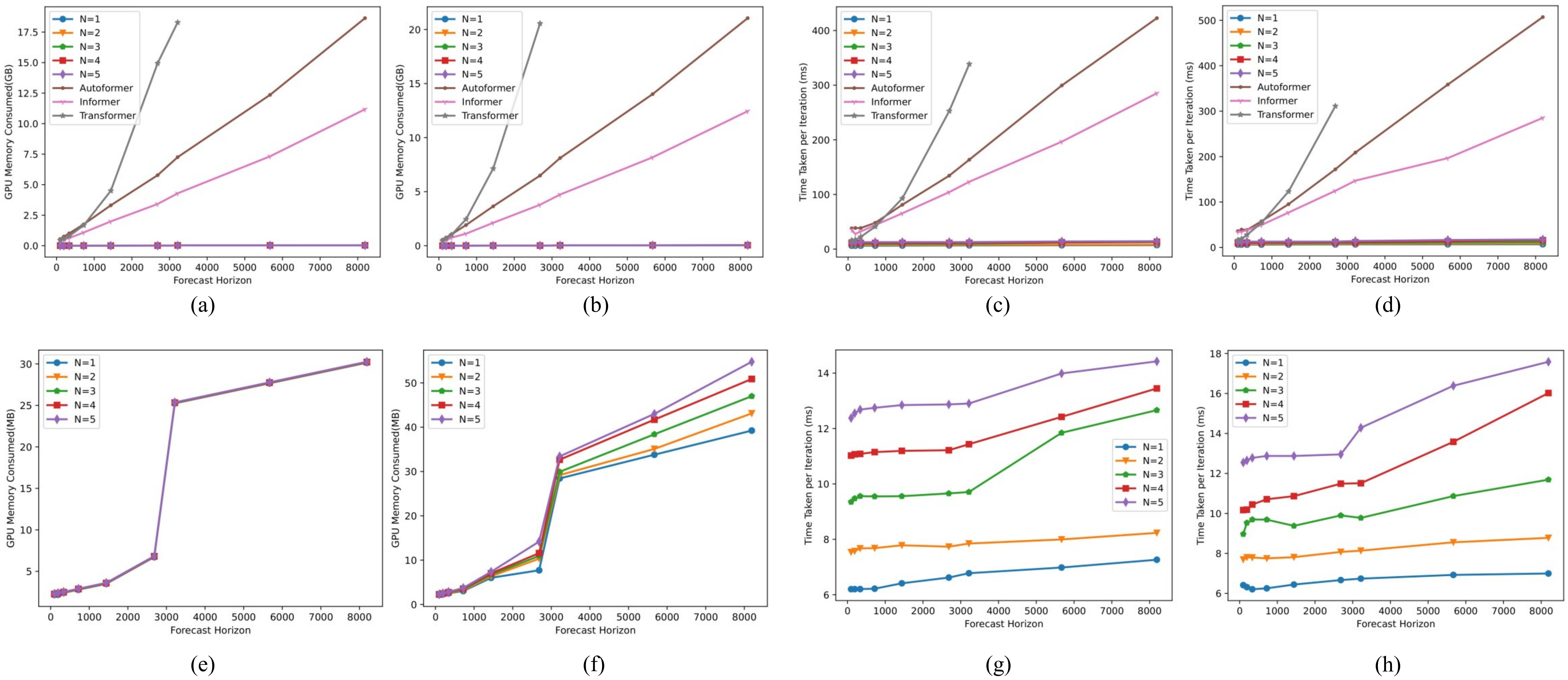


Fig 15: Run-time requirements (RTR) and Memory Consumption (MC) analysis. (a):MC for varying forecast horizon length (FHL), (b): MC for varying lookback window length (LWL), (c): RTR for varying FHL, (d): RTR for varying LWL, (e): MC for varying FHl (suggested), (f): MC for varying LWL (suggested), (g): RTR for varying FHL (suggested), (h): RTR for varying LWL (suggested)

5. Conclusion

Conclusion and Future Works

- To obtain good quality forecasts:
  - Non-stationary property of real-life time series data needs to be considered.
  - Inter-series and Intra-series intricacies should be address.
- Suggested 'Weak-stationarizing' and 'Non-stationarity' restoring blocks deal with non-stationary property.
- Mixer architecture realizes the intricacies.
- State-of-the-art results have been achieved.
- The results on non-seasonal datasets need improvements.
- The suggested blocks can be modified to be used alongside existing works.

Paper Submitted to:

Conference Website: https://www.cikm2022.org/

Excellence in Research in Australia (ERA) ranking: A

Created by Australian deans and the Australian Computing Research and Education Association of Australasia (CORE). The rankings range from A (=best) to C (=worst).

Qualis Ranking: A1

This conference ranking has been published by the Brazilian ministry of education and uses the H-index as performance measure for conferences. Based on the H-index percentiles, the conferences are grouped into performance classes that range from A1 (=best), A2, B1, ..., B5 (=worst).

References

1. Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. Journal of statistical software, 27, 1-22.
2. Liu, M., Zeng, A., Lai, Q., & Xu, Q. (2021). Time Series is a Special Sequence: Forecasting with Sample Convolution and Interaction. arXiv preprint arXiv:2106.09305
3. Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, *34*.
4. Liu, M., Zeng, A., Lai, Q., & Xu, Q. (2021). Time Series is a Special Sequence: Forecasting with Sample Convolution and Interaction. arXiv preprint arXiv:2106.09305.
5. Koopmans, L. H. (1995). The spectral analysis of time series. Elsevier.

References

6. Trockman, A., & Kolter, J. Z. (2022). Patches Are All You Need?. arXiv preprint arXiv:2201.09792

7. https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014

8. https://www.bgc-jena.mpg.de/wetter/

9. https://pems.dot.ca.gov/

10. https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

11. Lai, G., Chang, W. C., Yang, Y., & Liu, H. (2018, June). Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 95-104).

References

12. Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y. X., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. Advances in Neural Information Processing Systems, 32.

13. Kitaev, N., Kaiser, Ł., & Levskaya, A. (2020). Reformer: The efficient transformer. arXiv preprint arXiv:2001.04451.

14. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 36(3), 1181-1191.

Thank You!