NEURAL INFORMATION PROCESSING SYSTEMS
Published as a conference paper at NeurIPS 2024

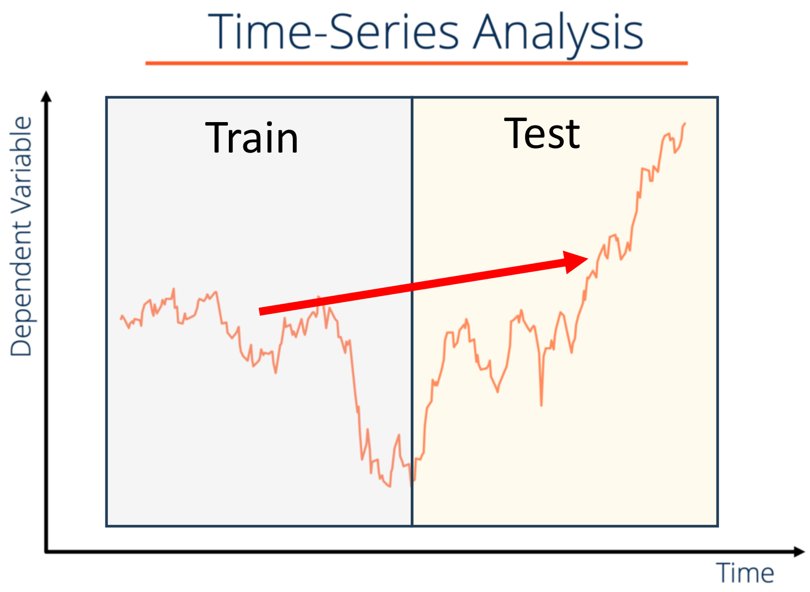
Frequency Adaptive Normalization For Non-stationary Time Series Forecasting

(Presenter)

Weiwei Ye, Songgaojun Deng, Qiaosha Zou, Ning Gui*

中南大學 CENTRAL SOUTH UNIVERSITY
UNIVERSITY OF AMSTERDAM
之江实验室 ZHEJIANG LAB

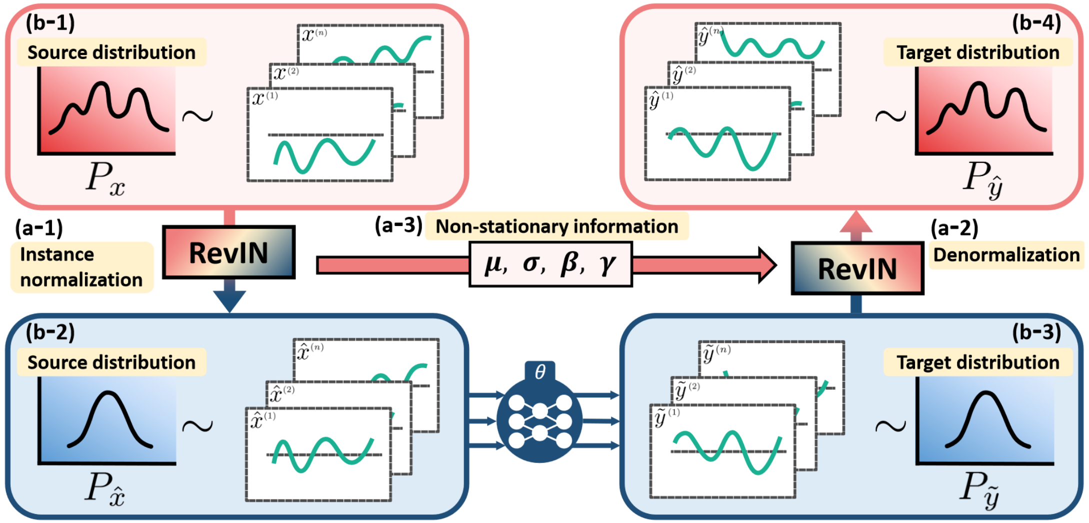
Non-stationary Time Series Forecasting



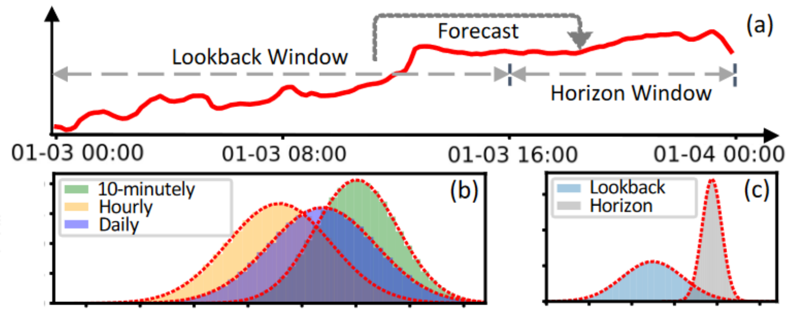
Distribution Shift of the training and testing datasets.

Non-stationarity

(1) Reversible Instance Normalization



But it consider only the non-stationarity between the input instances.



[1] Kim, T., Kim, J., Tae, Y., Park, C., Choi, J. H., & Choo, J. (2021, May). Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*.

中南大學 CENTRAL SOUTH UNIVERSITY

(2) Handle the non-stationarity between the input and output

Dish-TS

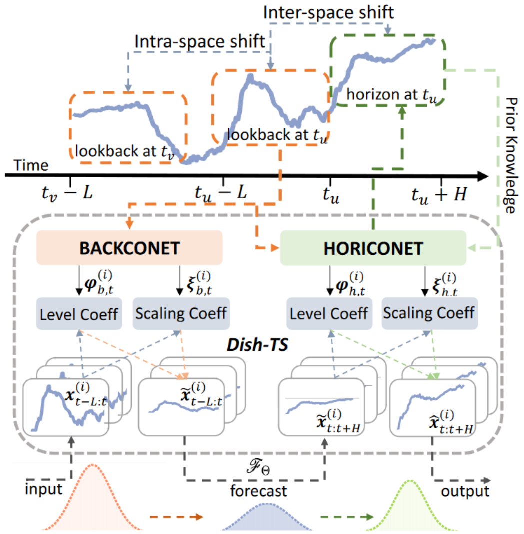
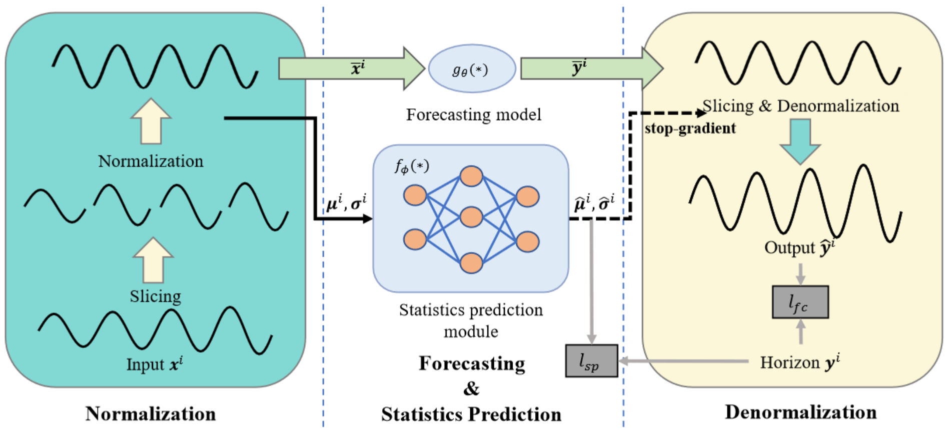


Figure 2: Overview of Paradigm *Dish-TS*.

SAN

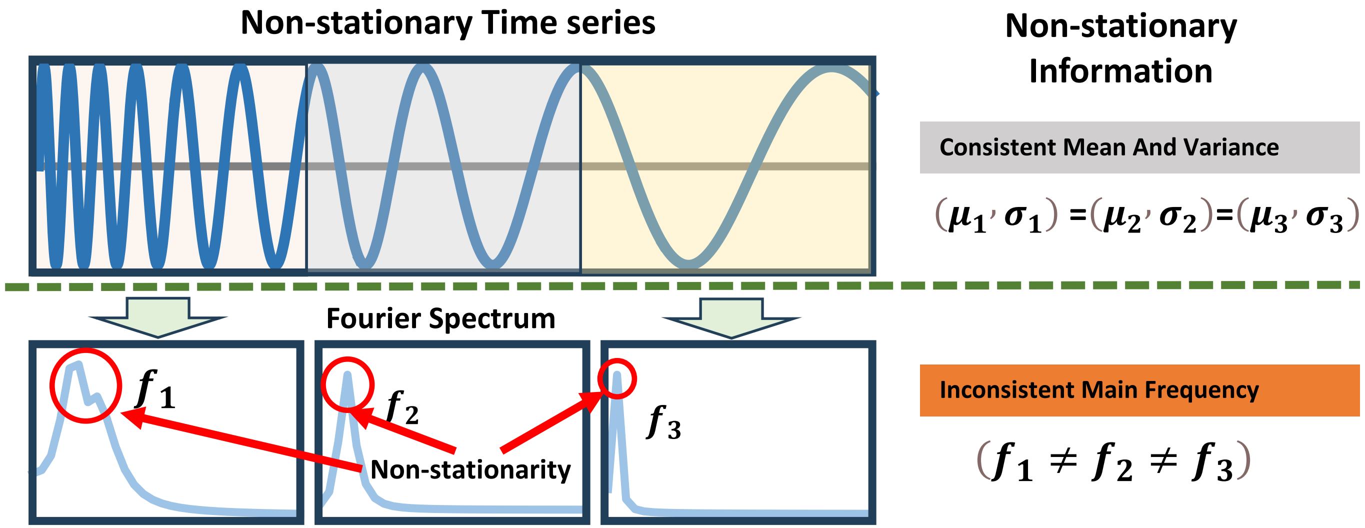


Handle the non-stationarity between input and output series through analysis and prediction of the internal statistics, which focus on most salient trend, rather than seasonality.

[2] Fan, W., Wang, P., Wang, D., Wang, D., Zhou, Y., & Fu, Y. (2023, June). Dish-ts: a general paradigm for alleviating distribution shift in time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 6, pp. 7522-7529).
[3] Liu, Z., Cheng, M., Li, Z., Huang, Z., Liu, Q., Xie, Y., & Chen, E. (2024). Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective. *Advances in Neural Information Processing Systems*, *36*.

Problem and Motivation

A frequency-based non-stationarity scenario



(1) Previous statistics-based methods failed to distinguish this type of non-stationarity.

(2) Previous Fourier-based methods select main frequencies randomly or fixedly.

Contributions

(1) We propose FAN, which adeptly addresses both trend and seasonal non-stationary patterns within time series data.

2) We explicitly address pattern evolvement with a simple MLP that predicts the top K frequency signals of the horizon series and applies these predictions to reconstruct the output.

3) We apply FAN to four general backbones for time series forecasting across eight real-world popular benchmarks. The results demonstrate that FAN significantly improves their predictive effectiveness

FAN

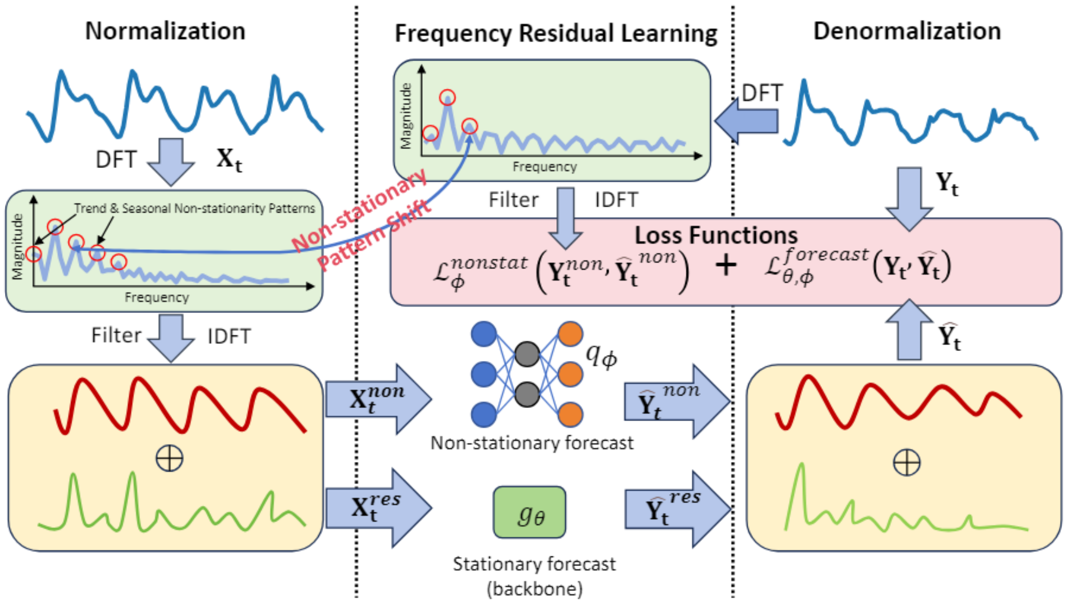


Figure 2: An overview of FAN which consists of normalization, frequency residual learning, denormalization steps, and incorporates a prior loss for non-stationary patterns.

Forecast/Denormalization

$$\hat{\mathbf{Y}}_t^{non} = q_\phi(\mathbf{X}_t^{non}, \mathbf{X}_t) = \mathbf{W}_3\,\mathrm{ReLU}\left(\mathbf{W}_2\,\mathrm{Concat}(\mathrm{ReLU}\left(\mathbf{W}_1\mathbf{X}_t^{non}\right), \mathbf{X}_t)\right)$$

$$\hat{\mathbf{Y}}_t^{res} = g_\theta(\mathbf{X}_t^{res})$$

$$\hat{\mathbf{Y}}_t = \hat{\mathbf{Y}}_t^{res} + \hat{\mathbf{Y}}_t^{non}$$

Loss functions

$$\phi, \theta = \arg\min_{\phi,\theta} \sum_t \left(\mathcal{L}_\phi^{nonstat}(\mathbf{Y}_t^{non}, \hat{\mathbf{Y}}_t^{non}) + \mathcal{L}_{\theta,\phi}^{forecast}(\mathbf{Y}_t, \hat{\mathbf{Y}}_t)\right)$$

Normalization

$$\mathbf{Z}_t = \mathrm{DFT}(\mathbf{X}_t) \quad \text{and} \quad \mathcal{K}_t = \mathrm{TopK}(\mathrm{Amp}(\mathbf{Z}_t)) \quad \text{and} \quad \mathbf{X}_t^{non} = \mathrm{IDFT}(\mathrm{Filter}(\mathcal{K}_t, \mathbf{Z}_t))$$

$$\mathbf{X}_t^{res} = \mathbf{X}_t - \mathbf{X}_t^{non}$$

Experiment

Table 2: Forecasting errors with and without FAN. The bold values indicate the best performance.

Methods		DLinear		+FAN		FEDformer		+FAN		Informer		+FAN		SCINet		+FAN	
Metrics		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
ETTm2	96	0.203	0.080	**0.198**	**0.078**	0.208	0.082	**0.194**	**0.074**	0.226	0.091	**0.198**	**0.077**	0.206	0.079	**0.198**	**0.078**
	168	0.220	0.093	**0.219**	**0.093**	0.249	0.116	**0.220**	**0.093**	0.251	0.112	**0.219**	**0.092**	0.226	0.094	**0.218**	**0.093**
	336	0.245	0.114	**0.241**	**0.113**	0.282	0.143	**0.272**	**0.131**	0.283	0.140	**0.245**	**0.114**	0.262	0.122	**0.241**	**0.113**
	720	0.270	0.142	**0.264**	**0.139**	0.308	0.174	**0.275**	**0.145**	0.347	0.212	**0.287**	**0.154**	0.297	0.153	**0.264**	**0.139**
Electricity	96	0.277	0.195	**0.269**	**0.184**	0.298	0.183	**0.243**	**0.148**	0.376	0.277	**0.250**	**0.153**	0.296	0.188	**0.261**	**0.168**
	168	0.272	0.183	**0.268**	**0.178**	0.305	0.191	**0.251**	**0.154**	0.371	0.269	**0.257**	**0.156**	0.306	0.196	**0.258**	**0.163**
	336	0.294	0.197	**0.289**	**0.192**	0.312	0.194	**0.272**	**0.167**	0.377	0.273	**0.273**	**0.167**	0.330	0.214	**0.278**	**0.175**
	720	0.333	0.233	**0.325**	**0.227**	0.330	0.213	**0.300**	**0.189**	0.401	0.311	**0.306**	**0.194**	0.352	0.240	**0.312**	**0.204**
Exchange	96	**0.164**	**0.052**	0.167	0.053	0.260	0.112	**0.186**	**0.062**	0.532	0.412	**0.189**	**0.066**	0.218	0.085	**0.169**	**0.055**
	168	0.219	0.090	**0.217**	**0.088**	0.312	0.163	**0.222**	**0.090**	0.582	0.491	**0.257**	**0.128**	0.266	0.126	**0.221**	**0.093**
	336	**0.288**	**0.155**	0.297	0.162	0.456	0.338	**0.336**	**0.198**	0.721	0.847	**0.333**	**0.191**	0.337	0.203	**0.303**	**0.167**
	720	0.453	0.352	**0.406**	**0.292**	0.669	0.661	**0.436**	**0.329**	0.889	1.210	**0.513**	**0.474**	0.502	0.430	**0.439**	**0.345**
Traffic	96	0.387	0.504	**0.334**	**0.403**	0.348	0.383	**0.326**	**0.371**	0.350	0.428	**0.314**	**0.364**	0.399	0.471	**0.344**	**0.393**
	168	0.588	0.804	**0.334**	**0.414**	0.366	0.422	**0.336**	**0.391**	0.366	0.457	**0.324**	**0.383**	0.377	0.443	**0.348**	**0.403**
	336	0.380	0.504	**0.346**	**0.437**	0.383	0.452	**0.348**	**0.414**	0.414	0.555	**0.356**	**0.427**	0.384	0.459	**0.360**	**0.426**
	720	0.407	0.532	**0.372**	**0.472**	0.391	0.465	**0.372**	**0.454**	0.656	1.002	**0.397**	**0.482**	0.401	0.490	**0.377**	**0.454**
Weather	96	0.249	0.180	**0.214**	**0.173**	0.368	0.299	**0.252**	**0.187**	0.299	0.221	**0.221**	**0.175**	0.265	0.199	**0.215**	**0.170**
	168	0.284	0.237	**0.254**	**0.210**	0.409	0.358	**0.304**	**0.240**	0.363	0.320	**0.258**	**0.215**	0.305	0.245	**0.256**	**0.208**
	336	0.344	0.304	**0.298**	**0.275**	0.463	0.459	**0.366**	**0.321**	0.439	0.437	**0.323**	**0.297**	0.341	0.310	**0.304**	**0.270**
	720	0.380	0.358	**0.345**	**0.340**	0.495	0.526	**0.441**	**0.432**	0.496	0.524	**0.368**	**0.360**	0.383	0.371	**0.340**	**0.322**

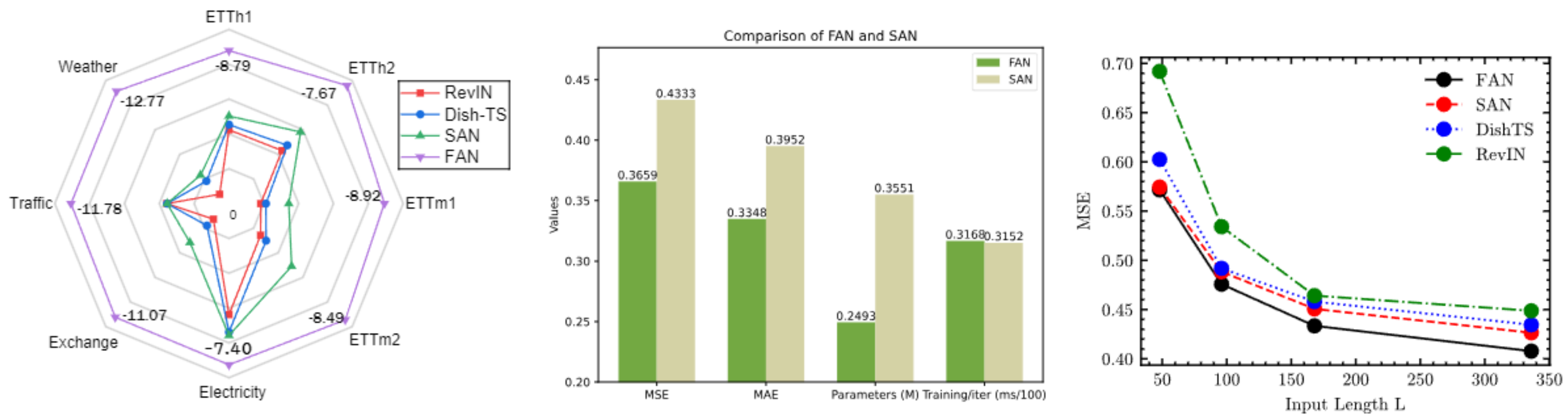
Table 3: The MSE performance averaged across all steps. Bold values indicate the best performance.

Models	DLinear				FEDformer				Informer				SCINet			
Methods	FAN	SAN	Dish-TS	RevIN	FAN	SAN	Dish-TS	RevIN	FAN	SAN	Dish-TS	RevIN	FAN	SAN	Dish-TS	RevIN
ETTh1	**0.441**	0.454	0.465	0.477	**0.443**	0.530	0.565	0.591	**0.465**	0.624	0.714	0.688	**0.442**	0.454	0.489	0.472
ETTh2	0.135	**0.134**	0.136	0.149	0.149	**0.148**	0.217	0.183	**0.164**	0.201	0.259	0.199	**0.136**	0.139	0.160	0.149
ETTm1	0.395	**0.390**	0.405	0.419	**0.400**	0.416	0.489	0.491	**0.397**	0.427	0.504	0.485	0.395	**0.393**	0.424	0.443
ETTm2	**0.105**	0.106	0.108	0.113	0.111	**0.106**	0.125	0.121	**0.106**	0.114	0.153	0.130	**0.105**	**0.105**	0.122	0.112
Electricity	**0.193**	0.200	0.201	0.207	**0.164**	0.169	0.181	0.180	**0.167**	0.191	0.219	0.190	0.177	0.175	0.207	**0.164**
Exchange	**0.149**	0.172	0.265	0.190	**0.170**	0.192	0.333	0.267	**0.168**	0.265	0.472	0.238	**0.162**	0.174	0.281	0.183
Traffic	**0.432**	0.514	0.591	0.652	0.408	**0.395**	0.433	0.424	**0.400**	0.515	0.446	0.894	**0.419**	0.431	0.489	0.442
Weather	**0.249**	0.250	0.269	0.272	0.295	**0.272**	0.562	0.280	**0.254**	0.256	0.322	0.275	**0.242**	**0.242**	0.250	0.251

- Main Results: our proposed FAN effectively enhances the performance of backbone models, on the ETTm2, Electricity, Exchange, Traffic, and Weather datasets, the average MSE performance improvements are rather significant: 10.81%, 21.49%, 51.27%, 21.97%, and 21.55\% respectively.

- Comparison with other normalization methods: It is evident that FAN generally outperforms the baseline models （MSE improvements around 7.76%~37.90%）.

Experiment-further analysis



(a) Stationarity Analysis (b) Model Efficiency (c) Various Input length

Figure 4: Comparison with other normalization methods. (a) ADF test after normalization, the smaller the value, the higher the stationarity. (b) Model efficiency comparison with SAN, including MSE/MAE, parameters (in millions), and training time per iteration (ms/100). (c) Performance in MSE vs. input length on the ETTm2 dataset.

(1) Compared to previous normalization methods, our model achieves greater stationarity across all datasets, particularly incases with larger seasonal patterns (Traffic, ETTh1, ETTm1).

(2) FAN and SAN have similar training iteration times, but FAN has 29.79% less parameters. Moreover, FAN achieves a 15.56% improvement in MSE and a 15.30% improvement in MAE.

(3) compared to other models, as the input length increases, among these normalizations, the enhancement of increases the most, this demonstrates that the instance-wise DFT is capable of extracting more seasonal patterns from the longer input windows.

Instance-wise selection of frequencies

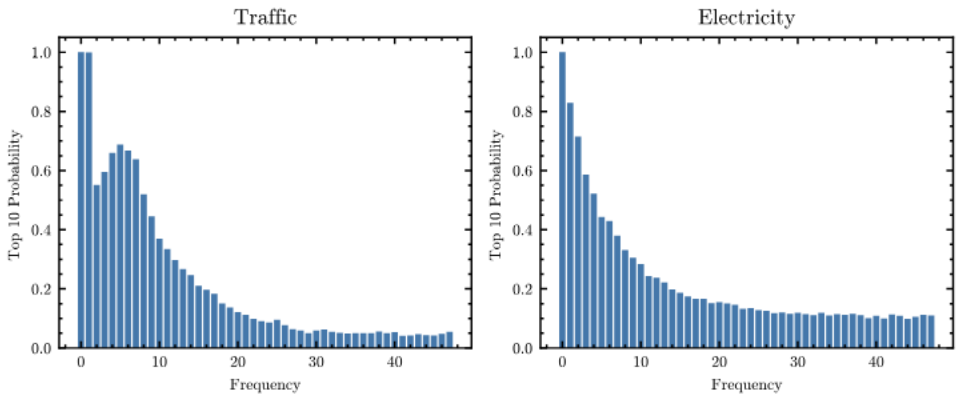


Figure 6: Top 10 selection propablity density on Traffic and Electricity datasets.

Table 5: MSE Performance between instance-wise (FAN) and global selection (Fixed) on SCINet backbone.

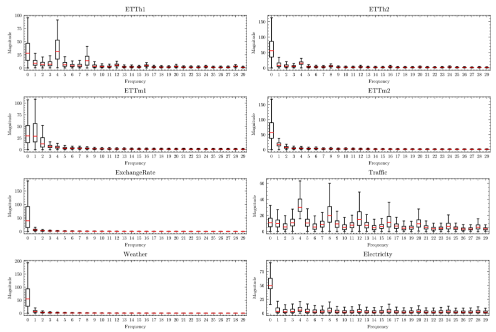
Electricity					
Steps	96	168	336	720	Avg.Imp.
FAN	**0.162**	**0.165**	**0.173**	**0.194**	18.50%
Fixed	0.176	0.192	0.231	0.265	-
Traffic					
Steps	96	168	336	720	Avg.Imp.
FAN	**0.393**	**0.403**	**0.426**	**0.454**	10.29%
Fixed	0.446	0.457	0.469	0.496	-

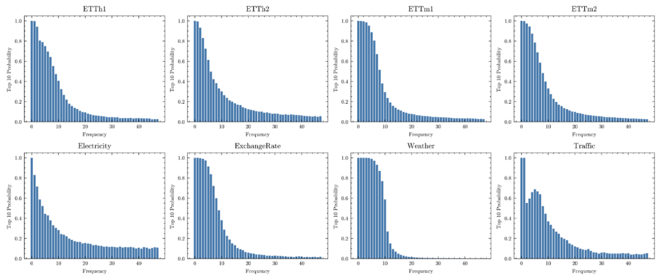
As shown in Table 5, by selecting instance-wise predominant frequencies, FAN achieves an average improvement of 18.50% and 10.29% on the Electricity and Traffic datasets respectively. This highlights instance-wise frequency selection rather than assuming fixed frequency patterns.

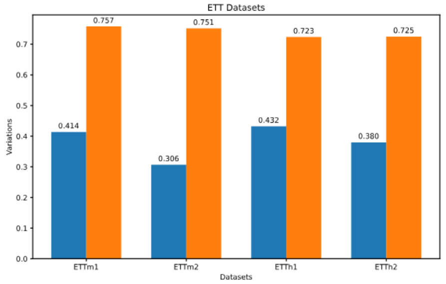
As shown in Table 5, by selecting instance-wise predominant frequencies, FAN achieves an averageimprovement of 18.50% and 10.29% on the Electricity and Traffic datasets respectively. Thishighlights instance-wise frequency selection rather than assuming fixed frequency patterns.

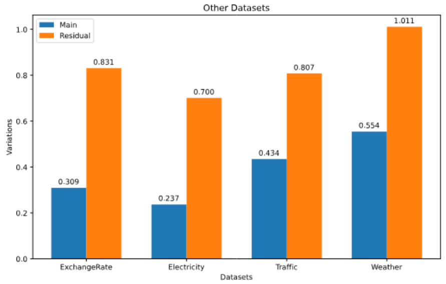
Other analysis

Dataset Analysis









Theoretical Analysis

C.2 Variance Over Spectrum

Along with the time series spectral theory [35], a time series with smaller variance in the spectrum is more stationary, in this section, we try to prove the proposed FAN can reduce the variance over spectrum, thus enhance the stationarity of the input data. Hence, we prove that, given an univariate time series real value vector $\mathbf{x} \in \mathbb{R}^T$, after removing main frequency components $\mathbf{z}[k] \in \mathcal{K}$, the variance on spectrum can be reduced $\mathrm{Var}(\mathbf{a}^{res}) < \mathrm{Var}(\mathbf{a})$.

Here, the marginal distribution of the amplitude vector (the spectrum) $\mathbf{a}$ is represented as a joint Rayleigh distribution with different scale parameters:

$$\begin{aligned} f(\mathbf{a}) &= \int f(\mathbf{a}, \mathbf{p}) d\mathbf{p} \\ &= \prod_{i=1}^{L} \frac{a}{\sigma_i^2} \cdot \exp\left(-\frac{a^2}{2\sigma_i^2}\right) \end{aligned} \tag{12}$$

Note that although we assume that the frequency components are independent with each other, this assumption is actually widely used [16] since it is quite possible that a specific component changes independently, e.g., the daily weekly changes while the monthly periodicity stays the same. Following the principle of additivity of variance for independent variables [13], the variance of the amplitude vector $\mathbf{a}$ can be expressed as follows:

$$\mathrm{Var}(\mathbf{a}) = \sum_{i}^{L} \frac{4-\pi}{2} \sigma_i^2 \tag{13}$$

after removing frequencies $k \in \mathcal{K}$, the joint distribution actually becomes:

$$f(\mathbf{a}^{res}) = \prod_{i=1, i \notin \mathcal{K}}^{L} \frac{a}{\sigma_i^2} \cdot \exp\left(-\frac{a^2}{2\sigma_i^2}\right) \tag{14}$$

thus, the variance of the whole distribution after removing top K-amplitude signals reduces to a smaller number, since the independent variance of of each dimension is positive, which is:

$$\mathrm{Var}(\mathbf{a}^{res}) = \sum_{i=1, i \notin \mathcal{K}}^{L} \frac{4-\pi}{2} \sigma_i^2 < \mathrm{Var}(\mathbf{a}) \tag{15}$$

Fourier Spectrum Analysis

C.4 Fourier Spectrum Empirical Analysis

The variance in the Fourier spectrum is an important indicator reflecting stationarity [20]. The closer the frequency components are to each other, the smaller the variance between the components, thus the stronger the stationarity [35]. Therefore, we compare the changes in frequency domain components for different methods and present the results in Fig. [10]. In Fig. [10], after FAN's normalization step, the distribution exhibits alignment of the input and output, and the range of the distribution mean has decreased to 8, compared with previous methods which are round 80, 70, 70 respectively for SAN, Dish-TS and RevIN. However, other methods still show significant differences between the input and output distributions, with the range of the frequency domain amplitude distribution reaching up to 80, indicating the presence of strong non-stationary signals. This highlights the effectiveness of our method in handling non-stationarity, especially for seasonal periodic signals, which previous methods have not successfully considered.

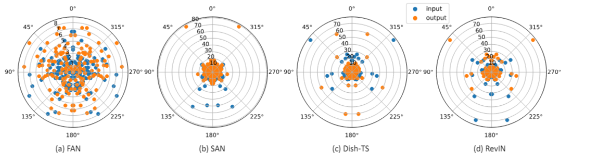
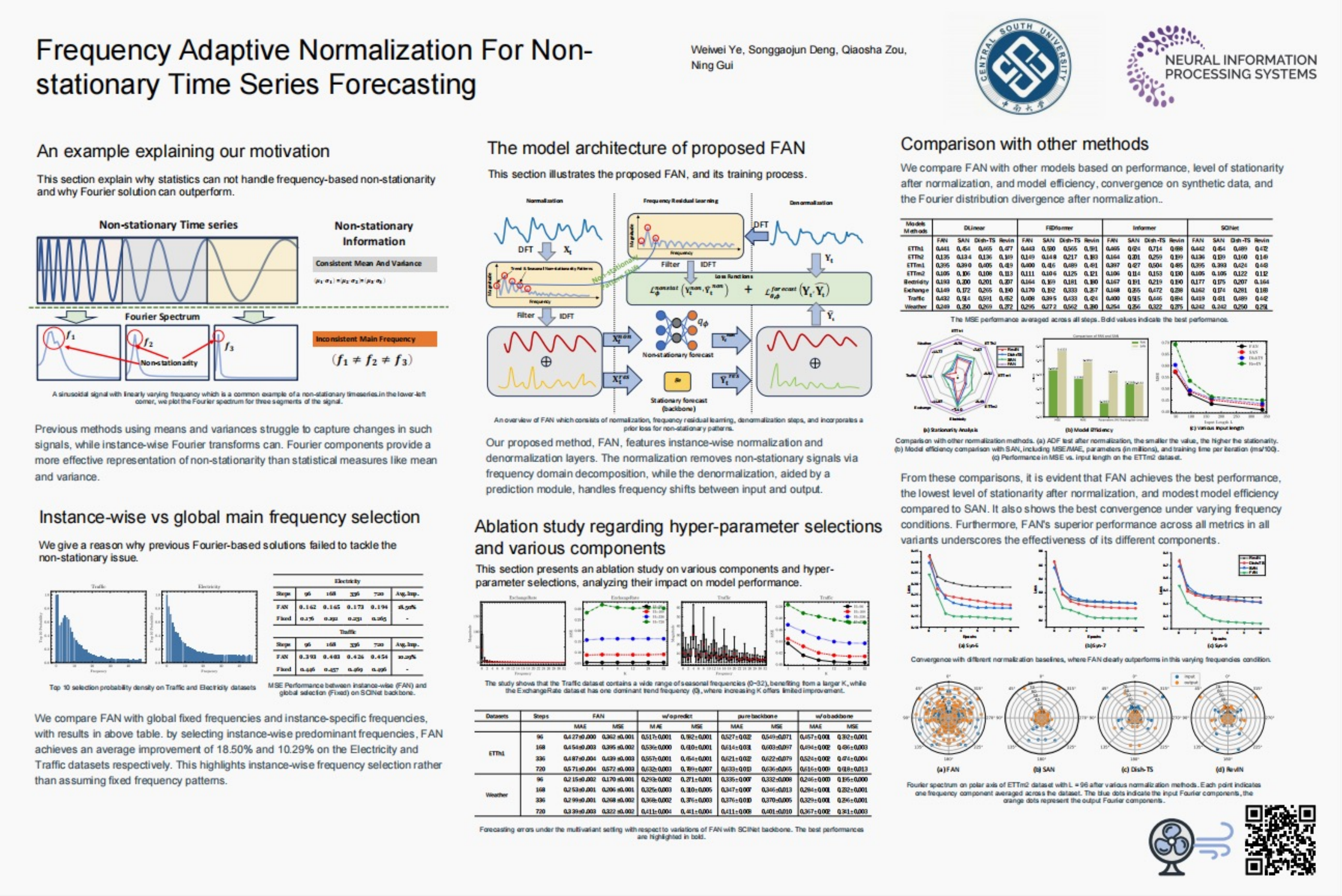


Figure 10: Fourier spectrum on polar axis of ETTm2 dataset with $L = 96$ after various normalization methods. Each point indicates one frequency component averaged across the dataset. The blue dots indicate the input Fourier components, the orange dots represent the output Fourier components. FAN remove top 5 Fourier components, and SAN slice in 12.

CENTRAL SOUTH UNIVERSITY



Poster Session: Wed 11 Dec 4:30 p.m. PST — 7:30 p.m. PST

Email: wwye155@gmail.com, ninggui@gmail.com

中南大學
CENTRAL SOUTH UNIVERSITY

Thank you!