

# Not All Frequencies Are Created Equal: Towards a Dynamic Fusion of Frequencies in Time-Series Forecasting

Anonymous Authors

The Appendix provides supplementary information and additional details that support the primary discoveries and methodologies proposed in this paper. It is organized into several sections: Appendix A contains the proof of Theorem 2 and Theorem 3.

## A PROOFS

In this section, we will prove Theorem 2 and Theorem 3, which is used for the theoretical analysis.

### A.1 Proof of Theorem 2

Note that  $l$  is a convex regression loss function, which indicates that

$$l(f(x), y) = l\left(\sum_{h=1}^H w^h f^h(x^h), y\right) \leq \sum_{h=1}^H w^h l(f^h(x^h), y) \quad (1)$$

Then we take the expectation on both sides of the above equation

$$\mathbb{E}_{(x,y) \sim D} l(f(x), y) \leq \mathbb{E}_{(x,y) \sim D} \sum_{h=1}^H (w^h l(f^h(x^h), y)), \quad (2)$$

since expectation is a linear operator and the expected value of the product is equal to the product of the expected values plus the covariance, we can further decompose the right-hand side of the equation into

$$\begin{aligned} \mathbb{E}_{(x,y) \sim D} l(f, y) &\leq \sum_{h=1}^H \mathbb{E}_{(x,y) \sim D} [w^h l(f^h, y)] \\ &= \sum_{h=1}^H \mathbb{E}_{(x,y) \sim D} (w^h) \mathbb{E}_{(x,y) \sim D} (l(f^h, y)) \\ &\quad + \text{Cov}(w^h, l(f^h, y)) \end{aligned} \quad (3)$$

Next, we recap the Rademacher complexity measure for model complexity. We use complexity based learning theory to quantify the generalization error of the proposed model.

Given the historical sequence and the ground truth of the prediction sequence,  $\hat{E}(f^h)$  is the empirical error of  $f^h$ . Then for any hypothesis  $f$  in the finite set  $F$  and  $1 > \delta > 0$ , with probability at

least  $1 - \delta$ , we have

$$\mathbb{E}_{(x,y) \sim D} (f^h) \leq \hat{E}(f^h) + \mathfrak{R}_h(F) + \sqrt{\frac{\ln(1/\delta)}{2H}} \quad (4)$$

where  $\mathfrak{R}_h(F)$  is the Rademacher complexities.

Finally, it holds with probability at least  $1 - \delta$  that

$$\begin{aligned} \overline{\text{GError}}(f) &\leq \sum_{h=1}^H \mathbb{E}(w^h) \hat{E}(f^h) + \mathbb{E}(w^h) \mathfrak{R}_h(f^h) \\ &\quad + \text{Cov}(w^h, l(f^h, y)) + M \sqrt{\frac{\ln(1/\delta)}{2H}}. \end{aligned} \quad (5)$$

### A.2 Proof of Theorem 3

Let  $\overline{\text{GError}}(f_{dynamic})$ ,  $\overline{\text{GError}}(f_{static})$  be the upper bound of the generalization regression error of dynamic and static fusion method, respectively.  $\hat{E}(f^h)$  is the empirical error defined in Theorem 1. Theoretically, optimizing over the same function class results in the same empirical risk. Therefore,

$$\hat{E}(f_{static}^h) = \hat{E}(f_{dynamic}^h) \quad (6)$$

Additionally, the intrinsic complexity is also invariant

$$\mathfrak{R}_h(f_{static}^h) = \mathfrak{R}_h(f_{dynamic}^h). \quad (7)$$

Thus in this special case, it holds that

$$\sum_{h=1}^H \mathbb{E}(w_{dynamic}^h) \hat{E}(f^h) \leq \sum_{h=1}^H (w_{static}^h) \hat{E}(f^h), \quad (8)$$

and

$$\sum_{h=1}^H \mathbb{E}(w_{dynamic}^h) \mathfrak{R}_h(f^h) \leq \sum_{h=1}^H (w_{static}^h) \mathfrak{R}_h(f^h). \quad (9)$$

if  $\mathbb{E}(w_{dynamic}^h) = w_{static}^h$ .

Since the covariance and the correlation coefficient have the same sign, when  $r(w^h, l^h) \leq 0$ , the covariance  $\text{Cov}(w^h, l^h)$  is also less than or equal to zero. Therefore, it holds that

$$\overline{\text{GError}}(f_{dynamic}) \leq \overline{\text{GError}}(f_{static}) \quad (10)$$

with probability at least  $1 - \delta$ , if we have

$$\mathbb{E}(w_{dynamic}^h) = w_{static}^h \quad (11)$$

and

$$r(w_{dynamic}^h, l(f^h)) \leq 0 \quad (12)$$

for all frequencies  $h$ .

Permission to make digital or hard copies of all or part of this work for personal or

**Unpublished working draft. Not for distribution.** distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ACM MM, 2024, Melbourne, Australia

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## B DATASET DESCRIPTIONS

In this paper, we conducted tests using eight real-world datasets. These datasets include:

- ETT contains two sub-datasets: ETT1 and ETT2, collected from two electricity transformers at two stations. Each of them has two versions in different resolutions (15 minutes and 1h). ETT dataset contains multiple series of loads and one series of oil temperatures.
- Weather covers 21 meteorological variables recorded at 10-minute intervals throughout the year 2020. The data was collected by the Max Planck Institute for Biogeochemistry’s Weather Station, providing valuable meteorological insights.
- Exchange-rate, which contains daily exchange rate data spanning from 1990 to 2016 for eight countries. It offers information on the currency exchange rates across different time periods.
- ECL contains the electricity consumption of 370 clients for short-term forecasting while it contains the electricity consumption of 321 clients for long-term forecasting. It is collected since 01/01/2011. The data sampling interval is every 15 minutes.
- Solar-Energy is about the solar power collected by the National Renewable Energy Laboratory. We choose the power plant data points in Florida as the data set which contains 593 points. The data is collected from 01/01/2006 to 31/12/2016 with the sampling interval of every 1 hour.

We follow the same data processing and train-validation-test set split protocol used in iTransformer, where the train, validation, and test datasets are strictly divided according to chronological order to make sure there are no data leakage issues. As for the forecasting settings, the lookback length is set to 96, while their prediction length varies in {96, 192, 336, 720}. The details of the datasets are provided in Table 1.

## C INFLUENCE OF DYNAMIC FUSION

We conduct an ablation study to investigate the influence of dynamic fusion. We replace the learnable weight vector with a fixed weight vector and name this modified model FreSF. Predictions are carried out on the ETT(ETTh1, ETTh2, ETTm1, ETTm2), Weather, and Exchange-rate datasets. We visualize the prediction results (with a prediction length  $S = 96$ ) for both FreSF and FreDF in Figure 1. In the ETTh1 and ETTm1 datasets, static fusion manages to capture the overall trends and detailed shifts, but the predicted values deviate significantly from the actual figures. This issue can be effectively addressed by employing dynamic fusion. Similarly, in the ETTh2 and ETTm2 datasets, static fusion’s predictions are not quite accurate at the extreme points, a problem that dynamic fusion can effectively solve. Likewise, when dynamic fusion is applied to the Weather and Exchange-rate datasets, the prediction curve aligns more closely with the ground truth.

## D ABLATION ON THE INFLUENCE OF OPERATION SEQUENCE

We conduct an ablation study about the influence of operation sequence. Our model initially conducts an inverse fast Fourier

transform(iFFT) on the spectrum  $\mathcal{M}_{out}^{l,m}$  obtained from the transfer function for each frequency  $\mathcal{M}_{in}^{l,m}$ . We then dynamically fusion the predicted results  $Z^{l,m}$  for each frequency. In this section, we first dynamically fuse the spectrums  $\mathcal{M}_{out}^{l,m}$  obtained from the transfer function for each frequency, followed by an iFFT on the result of this fusion. We present the result in Table 3

## E ABLATION ON THE INFLUENCE OF FDBLOCK NUMBER: $L$

To further figure out the impact of different numbers of FDBlock, we perform experiments with three different  $L$  ranging from 1 to 3. As shown in Table 4. And  $L = 1$  is a generally good choice.

## F MOTIVATION EXPERIMENTS

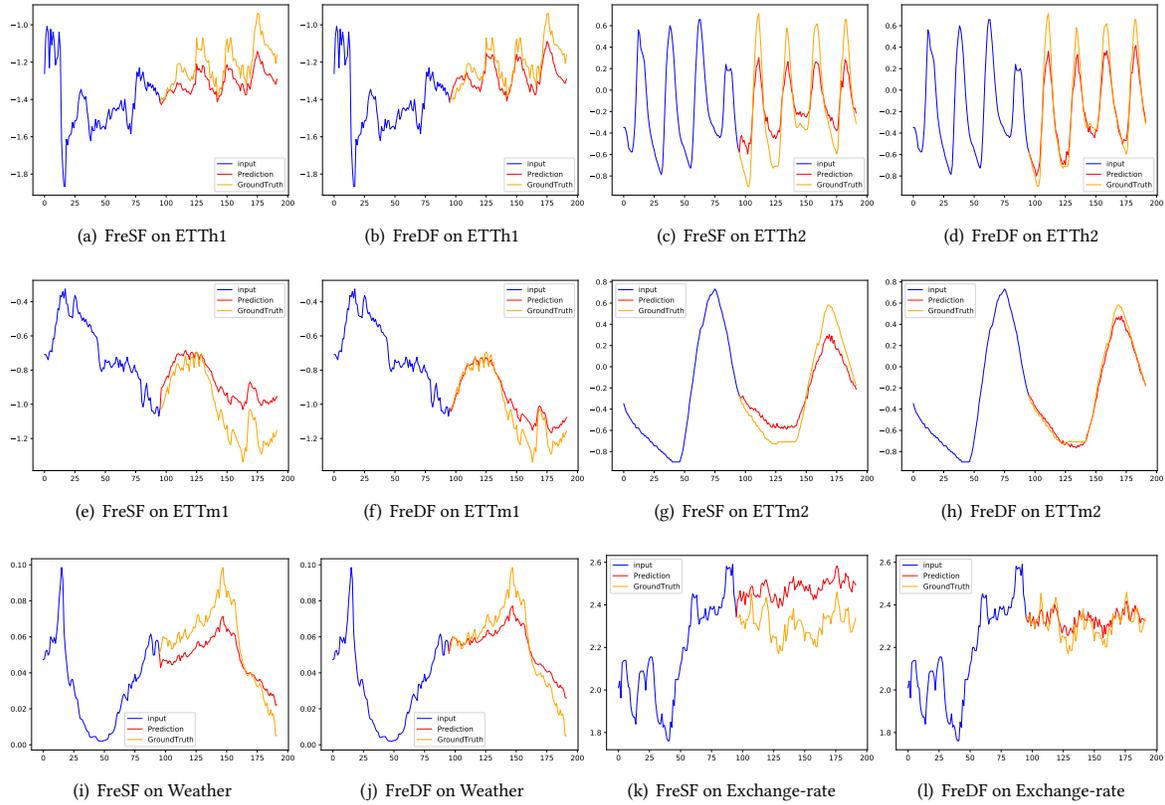
We present the full results of motivation experiments in Table 2. After eliminating low-frequency, we observe enhanced accuracy in the ETTm2 and ETTh2 datasets. Similarly, eliminating mid-frequency signals led to improved results in the ETTh2 and Exchange-rate datasets. The elimination of high-frequency results in more precise results in ETTh1, ETTm2, ETTh1, ETTh2, Exchange-rate, and weather datasets. The most significant improvement is seen when eliminating high-frequency signals. Nevertheless, removing low-frequency and mid-frequency signals can also improve prediction performance on some datasets.

## G LOOKBACK WINDOW: MODERNTCN VS US

At the 2024 International Conference on Learning Representations (ICLR), a groundbreaking approach, ModernTCN, was presented for time series forecasting. However, it’s worth mentioning that the lookback window length used in their study consistently surpasses 96. In our case, all experiments utilize a lookback window length of 96, which makes a direct comparison not feasible.

**Table 1: Detailed dataset descriptions.** *Dim* denotes the variate number of each data set. *Dataset Size* denotes the total number of time points in (Train, Validation, Test) split, respectively. *Prediction Length* denotes the future time points to be predicted, and four prediction settings are included in each data set. *Frequency* denotes the sampling interval of time points.

Dataset	Dim	Prediction Length	Dataset Size	Frequency	Information
ETTh1,ETTh2	7	{96, 192, 336, 720}	(8545, 2881, 2881)	Hourly	Electricity
ETTM1,ETTM2	7	{96, 192, 336, 720}	(34465, 11521, 11521)	15min	Electricity
Exchange	8	{96, 192, 336, 720}	(5120, 665, 1422)	Daily	Economy
Weather	21	{96, 192, 336, 720}	(36792, 5271, 10540)	10min	Weather
ECL	321	{96, 192, 336, 720}	(18317, 2633, 5261)	Hourly	Electricity
Solar-Energy	137	{96, 192, 336, 720}	(36601, 5161, 10417)	10min	Energy



**Figure 1: Visualization comparing FreSF and FreDF.**

**Table 2: The comparison of prediction results using different frequencies on six datasets. The best Forecasting results in bold. The lower MSE/MAE indicates the more accurate prediction result.**

Models	W all frequency		W/O low-frequency		W/O mid-frequency		W/O high-frequency	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETm1	0.0221	0.0954	0.0839	0.1827	0.0723	0.1683	<b>0.0116</b>	<b>0.0583</b>
ETm2	0.0601	0.1544	<b>0.0185</b>	<b>0.0850</b>	0.0852	0.1468	0.0511	0.1149
ETTh1	0.0475	0.1352	0.1333	0.2351	0.0543	0.1432	<b>0.0409</b>	<b>0.1193</b>
ETTh2	0.5288	0.4734	0.2559	0.3178	0.1396	0.2217	<b>0.1041</b>	<b>0.1848</b>
Exchange	0.0436	0.1218	0.0482	0.1292	<b>0.0156</b>	0.0745	0.0157	<b>0.0733</b>
Weather	0.0011	0.0217	0.0015	0.0207	0.0041	0.0379	<b>0.0011</b>	<b>0.0186</b>

**Table 3: Ablation on the influence of operation sequence.**

Methods	Metric	Weather				ECL				Exchange-rate			
		96	192	336	720	96	192	336	720	96	192	336	720
Dynamic Fusion on $Z^{l,m}$ (Our)	MSE	<b>0.157</b>	<b>0.205</b>	<b>0.259</b>	<b>0.341</b>	<b>0.150</b>	<b>0.161</b>	<b>0.176</b>	<b>0.217</b>	<b>0.082</b>	<b>0.172</b>	<b>0.316</b>	<b>0.835</b>
	MAE	<b>0.208</b>	<b>0.246</b>	<b>0.287</b>	<b>0.339</b>	<b>0.242</b>	<b>0.253</b>	<b>0.268</b>	<b>0.311</b>	<b>0.199</b>	<b>0.294</b>	<b>0.405</b>	<b>0.687</b>
Dynamic Fusion on spectrums $\mathcal{M}_{out}^{l,m}$	MSE	0.161	0.209	0.262	0.345	0.154	0.163	0.176	0.221	0.083	0.174	0.318	0.839
	MAE	0.214	0.253	0.292	0.347	0.242	0.253	0.270	0.313	0.199	0.295	0.406	0.670

**Table 4: Ablation on the number of FDBlock.**

FDBlock	Metric	ETTh2		ETm2		Weather	
		MSE	MAE	MSE	MAE	MSE	MAE
$L = 1$	96	<b>0.292</b>	<b>0.341</b>	<b>0.175</b>	<b>0.257</b>	<b>0.157</b>	<b>0.208</b>
	192	<b>0.376</b>	0.391	0.242	0.300	<b>0.205</b>	<b>0.246</b>
	336	0.419	0.428	<b>0.303</b>	<b>0.341</b>	0.260	0.287
	720	<b>0.420</b>	<b>0.439</b>	<b>0.405</b>	<b>0.396</b>	<b>0.341</b>	<b>0.339</b>
$L = 2$	96	0.295	0.342	0.176	0.257	0.157	0.209
	192	0.377	<b>0.390</b>	0.241	0.300	0.206	0.246
	336	<b>0.415</b>	<b>0.426</b>	0.304	0.343	<b>0.259</b>	<b>0.287</b>
	720	0.422	0.440	<b>0.406</b>	0.398	0.342	0.339
$L = 3$	96	0.293	0.341	0.176	0.258	0.158	0.208
	192	0.378	0.394	<b>0.241</b>	<b>0.299</b>	0.207	0.248
	336	0.416	0.426	0.304	0.345	0.262	0.289
	720	0.421	0.440	0.407	0.397	0.343	0.340