Wavelet Transform for Frequency-Domain Learning Non-Stationary Time Series Analysis

February 13, 2025

Riccardo De Sanctis[1] **Matteo De Sanctis**[1]

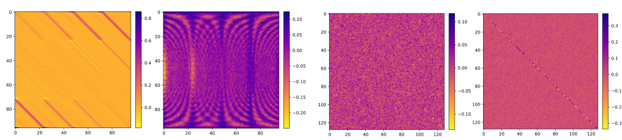
Abstract

Time series forecasting (TSF) is a complex task, requiring models to effectively capture both **temporal dependencies** (intra-series) and **channel-wise relationships** (inter-series). Many real-world time series are **non-stationary**, with evolving statistical properties that make modeling even more challenging.
Recent research trends have moved toward simplified architectures, with increased use of **frequency-domain transformation (FT)** (Liu & Wang, 2024; Yi et al., 2023a), primarily using the **Fourier Transform**, which, while effective for stationary signals, lacks **time-localization**. In this project, we integrate the **Wavelet Transform (WT)** (Rhif et al., 2019) into **frequency-based TSF models** (Yi et al., 2023b), leveraging its ability to capture both **global frequency** and **local time variations** to improve forecasting performance on **non-stationary time series**.
Code, results and visualizations are available on the GitHub repository.

1. Introduction

Time series forecasting is essential in fields like finance, healthcare, energy and geo-sciences, for informed decision-making. While models like **RNNs**, **GNNs**, and **Transformers** capture temporal dependencies, recent trends favor **simpler architectures**, such as **multi-layer perceptrons** (MLPs).

To overcome **MLP limitations** (e.g., point-wise mappings, information bottlenecks), frequency-domain modeling approaches like FreMLPs (Yi et al., 2023b) have been proposed. These models leverage **Fourier Transform** not merely to improve the original architecture, but to <u>*learn*</u> both **channel-wise** and **time-wise** dependencies <u>*directly in the frequency domain*</u>, being so capable to capture **global dependencies** and achieve **energy compaction**, focusing on key spectral components (Figure 1).



(a) **Left**: time domain. **Right**: frequency domain. (b) **Left**: time domain. **Right**: frequency domain.

Figure 1. Visualization of learned patterns from MLPs in both the time and frequency domains (Yi et al., 2023b). (a) *Global view*: Clearer global periodic patterns in the frequency domain. (b) *Energy compaction*: More distinct diagonal dependencies and key patterns in the frequency domain.

While **Fourier-based models** are effective for stationary signals, they lack **time-localization**, making them less suitable for **non-stationary** time series, where statistical properties change and evolve over time. To address this, the **Wavelet Transform** provides both **frequency and time-localized information**, making it a promising alternative for addressing non-stationary time series forecasting.

In this project, we aim to enhance **frequency-domain MLPs** (Yi et al., 2023b) by integrating the **Wavelet Transform**, evaluating its impact on **non-stationary datasets** from the paper. Our implementation achieves *promising results*, sometimes *surpassing* the paper's base model, highlighting WT's *effectiveness* for non-stationary time series.

2. Related Work

We build on FreTS (Figure 2), the frequency learning architecture of (Yi et al., 2023b) which applies Frequency-domain MLPs across inter-series and intra-series scales. It consists of two stages: (i) **Domain Conversion**, transforming time-domain signals into frequency-domain representations; and (ii) **Frequency learning**, where FreMLPs learn real and imaginary parts of frequency components.
The approach leverages Fourier transform to identify periodic and trend patterns, and **learning in the frequency spectrum** helps in capturing a greater number of them.

[1]Sapienza University of Rome Email: Matteo De Sanctis <desanctis.1937858@studenti.uniroma.it>, Riccardo De Sanctis <desanctis.1937859@studenti.uniroma.it>.

Deep Learning and Applied AI 2024, Sapienza University of Rome, 2nd semester a.y. 2023/2024.

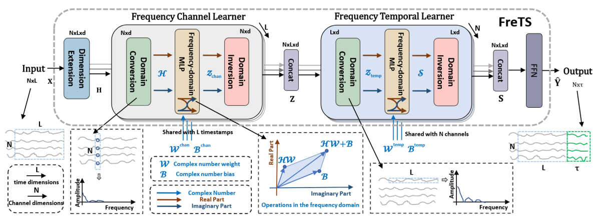
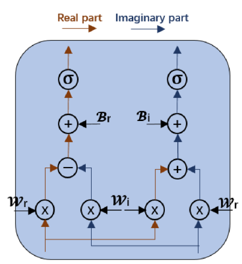


Figure 2. Overview of FreTs: the Frequency Channel Learner focuses on inter-series dependencies, the Frequency Temporal Learner on temporal dependencies.

Frequency Channel Learner allows to capture interactions and correlations between different variables. It models the N channels dependencies by operating at each timestamp and sharing the same weights across L timestamps.

Frequency Temporal Learner focuses on learning the temporal patterns in the freq. domain. FreMLPs model L timestamps by sharing weights between N channels.



FreMLP The frequency-domain MLPs in FreTS are redesigned for the complex numbers of frequency components. As both the input and the parameters are complex, according to the complex number multiplication rule, each layer performs a separate computation of the real and imaginary parts of frequency components, as seen in the side image. In our project work, attention layer is build on top of this block.

Dimension extension block Inspired by word embeddings, this block enhances model's capability by expanding the input representation, allowing for a more expressive hidden encoding of the input lookback window, capturing more semantic information.

3. Method

A significant part of our implementation effort focused on ensuring the proper functionality of the two Frequency Learners while carefully handling tensor dimensionalities within the WT framework, which we introduce below.

3.1. Wavelet Transform

The Wavelet Transform (Rhif et al., 2019) is a time-frequency transformation that decomposes a signal into scaled and translated versions of a wavelet function (time series of coefficients). Each coefficient describes the time evolution of the signal in the corresponding frequency band. Unlike the Fourier transform, which represents the signal in terms of infinite sinusoidal functions, the WT uses localized wavelet functions to capture both global frequency information and local time-dependent variations, identifying the locations containing observed frequency content, instead of only extracting pure frequencies from the signal.

The **Continuous Wavelet Transform** (CWT) of a signal $x(t)$ is defined as:

$$WT(a,b) = \int_{-\infty}^{\infty} x(t)\Psi^*_{a,b}(t)dt \tag{1}$$

or in a cleaner way $WT(a,b) = \langle x(t), \Psi_{a,b}(t)\rangle$ if one is accustomed with the bra-ket notation.

Ψ is the wavelet basis function (*Haar, Daubechies, Morlet, etc.*), and $\Psi^*_{a,b}(t)$ is the complex-conjugate, scaled and shifted version of $\Psi(t)$: $\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\Psi(\frac{t-b}{a})$.

The wavelet coefficients a and b, represent the signal at scale a (frequency—which controls the wavelet's dilation or compression, affecting frequency resolution), and with translation b (time—shifting the wavelet across time, capturing local features). The term $\frac{1}{\sqrt{|a|}}$ ensures energy normalization across different scales.

The intuition is that high-scale wavelets (large a) capture low-frequencies (long-duration components), and low-scale wavelets (small a) capture high-frequency (short-duration details). Having this multi-resolution property (and time-frequency localization) makes WT ideal for analyzing non-stationary signals where frequency components evolve over time. However, working at every possible scale a and translation b is redundant, making it computationally expensive for discrete data.

DTCWT The direct implementation of the theoretical framework of CWT is achieved with **discrete wavelet transform**, approximating CWT efficiently. Among them, the **Dual-Tree Complex Wavelet Transform (DTCWT)** offers several advantages over traditional wavelet transforms thanks to its complex representation, including *approximate shift invariance* (more stable coefficients on shifted signals), *improved directional selectivity in two dimensions* (having six directional filters), and *better phase information representation* (better separating amplitude and phase). Therefore, we opted for DTCWT using the pytorch_wavelets library.

3.2. Attention Module

We also design an **attention layer** operating in the **complex domain** of **WT spectral representations**, following the same principles as **FreMLP**. **Key, Value, and Query computations** adhere to complex number multiplication rules, with one key difference: since **Softmax** is not defined in the complex domain, we apply it to the **magnitude** of the attention scores. The **Tanh activation** is the non-linearity. Implementation details and code here.

3.3. Stationarity tests

We selected inherently non-stationary datasets—**ECG**, **COVID-19**, **Solar**, and **Bitcoin**—and confirmed their non-stationarity using **ADF** and **KPSS** tests, justifying the use of **Wavelet Transform** to address time-localized variations. The ADF test examines the presence of a unit root, while the KPSS test checks for trend-stationarity,

4. Results

We present here the key and most valuable results on non-stationary time series, even though we experimented with both stationary and non-stationary ones to analyze potential differences. To ensure fair comparisons with (Yi et al., 2023b), we tested both their **FreTS** model (using Fourier Transform) and our **WT** model under the same settings. We also evaluated the impact of an **attention** module and increased **depth** (D—which represents amount of stacked layers), we show the best depth results.
We also tested each model with different configurations of embedding size, epochs and layers.

ECG For the ECG dataset slight improvements with WT are obtained (Table 1), yet not in the same scale as the improvements obtained with the FreTS architecture against other architectures (as shown in sec.4 of (Yi et al., 2023b), ~ 0.002 for MAE & RMSE). However, deep enough layers achieve similar **improvements, in the same scale**.

Covid As for the Covid results (Table 2) we see slight depth-1 improvements, in the same scale of the paper improvements (MAE ~ 0.03, RMSE ~ 0.04). Here the attention module greatly reinforces the model's inductive bias.

Solar Solar results (Table 3) show yet another **interesting outcome**, deeper WT model, *differently than deeper FreTS*, achieve best enhancements. Such improvements are partially greater than the paper ones (MAE $\sim$ 0.003, RMSE $\sim$ 0.02).

Bitcoin Since this dataset is unique to our project, we lack direct comparison of FreTS against the other architectures presented in the paper. However, results show very good improvements, with deep models achieving the best performance, even outperforming the base model by a considerable margin (compared to ECG, where stats are similar).

Table 1. ECG Dataset results: 20 Epochs - 16 Embedding size.

ECG	FreTS	WT	WTAtt	Multi FreTS D-9	Multi WT D-8
MSE	0.01005	**0.00991**	*0.00996*	0.00933	**0.00931**
MAE	0.0669	**0.0664**	*0.0665*	0.0642	**0.0641** ($\Delta \approx 0.003$)
RSE	0.7370	**0.7319**	*0.7337*	0.7101	**0.7095**
RMSE	0.1002	**0.0995**	*0.0998*	0.0966	**0.0965** ($\Delta \approx 0.004$)

Table 2. Covid Dataset results: 20 Epochs - 128 Embedding size.

Covid	FreTS	WT	WTAtt	Multi FreTS	Multi WT
MSE	0.3359	*0.3230*	**0.3043**	-	-
MAE	0.3712	0.3766	**0.3627**($\Delta \approx 0.01$)	-	-
RSE	0.8728	*0.8558*	**0.8307**	-	-
RMSE	0.5796	*0.5683*	**0.5516** ($\Delta \approx 0.03$)	-	-

Table 3. Solar Dataset results: 20 Epochs - 1 Embedding size.

Solar	FreTS	WT	WTAtt	Multi FreTS D-3	Multi WT D-3
MSE	0.0253	**0.0246**	-	0.0256	**0.0238**
MAE	0.1054	**0.1015**	-	0.1033	**0.0946** ($\Delta \approx 0.01$)
RSE	0.6507	**0.6421**	-	0.6539	**0.6321**
RMSE	0.1590	**0.1569**	-	0.1598	**0.1544** ($\Delta \approx 0.005$)

Table 4. Bitcoin Dataset results: 20 Epochs - 1 Embedding size.

BTC	FreTS	WT	WTAtt	Multi FreTS D-3	Multi WT D-3
MSE	0.0094	**0.0083**	-	0.0096	**0.0076**
MAE	0.0694	**0.0640**	-	0.0712	**0.0603** ($\Delta \approx 0.01$)
RSE	0.3746	**0.3517**	-	0.3784	**0.3364**
RMSE	0.0972	**0.0912**	-	0.0982	**0.0873** ($\Delta \approx 0.01$)

5. Discussion and Conclusions

It's worth to note that due to GPU constraints, we sometimes ran the models with **smaller embeddings** or **different input/pred lenghts**, or were unable to run them.
Depth often enhances performance, and ***WT shows enhancements compared to the Fourier transform***, as deeper WT models tend to outperform deeper FreTS models.
Additionally, as model performance improves with the larger paper settings, the relative scale of our improvements may diminish. Nonetheless, our results still demonstrate notable WT improvements and remain highly promising.
Here we presented the best scenarios for WT, however, this is not always the case, as performance varies depending on several factors, sometimes even leading to worse results. A full account of results can be found here.

Future research includes adding **residual connections** for deeper models, exploring more **advanced architectures**, and *expanding the range of wavelets*. Further evaluation on diverse datasets and deeper *analysis of frequency-domain learning*, investigating the behavior of WT in cases where performance lags behind Fourier-based approaches. Another key direction is **modeling time-channel dependencies jointly**, enabling more expressive cross-channel representations, as channels can influence each other across time. *Attention mechanisms* over this joint distribution could further enhance the model's ability to capture complex multivariate temporal interactions.

Overall, this project was highly valuable to us, allowing us to explore a personally meaningful research question and gain valuable insights. Our results, demonstrating **improvements over the baseline method**, suggest that this approach is moving in the right direction; they hence *highlight the potential of time-localized frequency transformations in enhancing the effectiveness of frequency-domain learning for non-stationary time series*, especially in the case of **deeper models** being leveraged.

Bibliography.

References

Liu, X. and Wang, W. Deep time series forecasting models: A comprehensive survey. *Mathematics*, 12(10), 2024. ISSN 2227-7390. doi: 10.3390/math12101504. URL https://www.mdpi.com/2227-7390/12/10/1504.

Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., and Sang, Y. Wavelet transform application for/in non-stationary time-series analysis: A review. *Applied Sciences*, 9(7), 2019. ISSN 2076-3417. doi: 10.3390/app9071345. URL https://www.mdpi.com/2076-3417/9/7/1345.

Yi, K., Zhang, Q., Cao, L., Wang, S., Long, G., Hu, L., He, H., Niu, Z., Fan, W., and Xiong, H. A survey on deep learning based time series analysis with frequency transformation, 2023a. URL https://arxiv.org/abs/2302.02173.

Yi, K., Zhang, Q., Fan, W., Wang, S., Wang, P., He, H., Lian, D., An, N., Cao, L., and Niu, Z. Frequency-domain mlps are more effective learners in time series forecasting, 2023b.