AMSI SUMMERRESEARCH SCHOLARSHIPS 2025–26

*Get a taste for Research this Summer*

# Clustering Time Series Using Spectral Density-Based Distances


Thi Hai Anh Luyen

Supervised by Dr. Gizem Intepe

Western Syndey University



Jane Street

AMSI Australian Mathematical Sciences Institute

**Abstract**

Clustering time series is widely used in statistics and data science, with applications in areas such as forecasting, energy and finance. A central issue is how similarity between series is defined. This report compares different distance measures for time series and highlights the limitations of traditional time-domain approaches. It introduces frequency-domain analysis using the Discrete Fourier Transform and explores spectral-density-based distance as an alternative for measuring similarity. This method provides a more effective way to identify common periodic patterns in time series.

# 1 Introduction

Time series data are sequences of observations recorded in chronological order, usually at evenly spaced points in time [1]. They represent how a phenomenon evolves over time and are used widely in many domains such as financial markets, communication engineering, weather forecasting and healthcare analytics. Analyzing such data is essential for understanding trends, predicting future behaviour, and detecting underlying patterns. One common approach to compare such data is clustering, which groups similar time series together to reveal hidden structures or common behaviours. For example, in stock index analysis, time series clustering can be used to identify structural similarities and classify assets based on their return characteristics [2].

A major challenge in time series clustering is identifying similarity between series. Traditional approaches often rely on time-domain distance measures, such as Euclidean distance and Manhattan distance which compare series point by point. While these measures are simple, they can fail to capture important characteristics when series have phase shifts, different lengths, or noise. Dynamic Time Warping (DTW) can address this issue but still focuses mainly on point-wise comparisons, which may fail to identify cyclic patterns.

To address these limitations, this report explores frequency-domain approaches. By transforming time series using the Discrete Fourier Transform (DFT), we can analyse the spectral density, which shows how variance is distributed across different frequencies. This helps to assess similarity based on shared periodic behaviours rather than exact timing of observations. As a result, frequency-domain methods are particularly effective for detecting seasonal patterns.

# 2 Statement of Authorship

- All text in this report is written by me, Thi Hai Anh Luyen.
- The core ideas and methodology discussed are based on existing research, which has been cited throughout the report.
- Any guidance or supervision provided by my advisor, Dr. Gizem Intepe, is acknowledged, including advice on structuring the report and clarifying concepts.

## 3 Distance Metric

In 1906, Maurice Fréchet introduced the idea of a distance function in a very general form, which included what is known as semi-metric [3]. Semi-metric relaxes one or more of the metric properties, usually the triangle inequality. Later in 1914, Felix Hausdorff developed the term metric space and extended the theory to provide a formal definition for modern distance metrics [4].

A distance metric is a function $d(x, y)$ that assigns a non-negative real number to each pair of elements $x, y$ in a set $X$, quantifying their dissimilarity [5]. A true distance metric has 4 characteristics:

1. **Non-negativity:**
$$d(x, y) \geq 0 \quad \forall x, y \in X$$
2. **Identity of Indiscernibles:**
$$d(x, y) = 0 \iff x = y$$
3. **Symmetry:**
$$d(x, y) = d(y, x) \quad \forall x, y \in X$$
4. **Triangle Inequality:**
$$d(x, y) \leq d(x, z) + d(z, y) \quad \forall x, y, z \in X$$

### 3.1 Euclidean Distance ($L_2$)

Euclidean distance is one of the most common distances, which computes straight-line distance between two time series of equal length. Given two time series (or vectors)

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad \mathbf{y} = (y_1, y_2, \ldots, y_n),$$

The Euclidean distance is defined as

$$d_E(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

When applied to time series, Euclidean distance compares each element of one series with the corresponding element at the same time index in the other series. It therefore measures the overall point-by-point difference between two time series and can be interpreted as the shortest straight-line path between them in feature space.

This measure is computationally simple and therefore widely used in clustering analysis. It works well in datasets with compact or well-defined clusters [6]. However, it assumes two time series to have the same length and both are perfectly aligned in time. Furthermore, because the mapping between each point in the series is fixed, this approach is highly sensitive to noise and misalignments in time, which means similar patterns occurring at slightly different time may be very dissimilar [7]. That problem was explained by Gianluca Rosso in 2014 where he showed that squaring the differences in Euclidean distance amplifies large deviations, meaning that a small number of noisy points or outliers can dominate the total distance and reduce the reliability of this specific measure [8]. Additionally, Euclidean distance becomes less effective with long and high-dimensional series, due to the curse of dimensionality [9].

### 3.2 Manhattan Distance ($L_1$)

Given two time series (or vectors)

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad \mathbf{y} = (y_1, y_2, \ldots, y_n),$$

Manhattan distance, also known as $L_1$ norm, is defined as

$$d_M(x, y) = \sum_{i=1}^{n} |x_i - y_i|.$$

Geometrically, it follows a stepwise path rather than a straight-line path between two points like Euclidean distance.

This measure also performs well with compact clusters and since it contains no squaring, it is less sensitive than Euclidean distance when dealing with outliers and noisy data [10][11]. Manhattan therefore is more robust to noise and random fluctuations. In high-dimensional datasets, Manhattan distance still suffers from the curse of dimensionality, but its impact is less severe than for Euclidean distance [9]. This is also because the $L_1$ norm sums the absolute differences linearly, preserving better contrast between points and maintaining a more meaningful distance relationships in high-dimensional spaces. This was demonstrated by Aggarwal, Hinneburg, and Keim (2001) using the relative contrast as a measure to compute how meaningful distances are in high-dimensional spaces [9]. According to the results, for Manhattan distance, differences between points increase as dimensionality grows, meaning distances remain meaningful in high dimensions, whereas for Euclidean distance, differences remain roughly the same, indicating distances are moderately meaningful but do not improve in high dimensions.

### 3.3 Cosine Distance

Unlike the two distances above, cosine distance measures the dissimilarities between two series based on their orientation rather than magnitude. Given two time series (or vectors)

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad \mathbf{y} = (y_1, y_2, \ldots, y_n),$$

Cosine similarity is defined as

$$\text{sim}_{\cos}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=0}^{n-1} x_i y_i}{\sqrt{\sum_{i=0}^{n-1} x_i^2} \sqrt{\sum_{i=0}^{n-1} y_i^2}}.$$

The cosine distance is then defined as

$$\begin{aligned} d_{\cos}(\mathbf{x}, \mathbf{y}) &= 1 - \mathrm{sim}_{\cos}(\mathbf{x}, \mathbf{y}) \\ &= 1 - \frac{\sum_{i=0}^{n-1} x_i y_i}{\sqrt{\sum_{i=0}^{n-1} x_i^2} \sqrt{\sum_{i=0}^{n-1} y_i^2}} \end{aligned}$$

Cosine distance is particularly helpful when time series differ in magnitudes but have similar overall trends, as it measures similarity based on the direction rather than their absolute values [12].

However, this measure has limitations. It is not appropriate when magnitudes carry meaningful information, such as peak heights, and when time series contain zero vectors as the measure is undefined for zero-magnitude vectors. Additionally, cosine distance also assumes point-to-point alignment, so temporal misalignment between series can lead to misleading results.

### 3.4 Correlation Distance

Similar to cosine distance, correlation distance measures the similarity between two time series by focusing on the shape and pattern of their fluctuations instead of absolute magnitude. Given two time series (or vectors)

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad \mathbf{y} = (y_1, y_2, \ldots, y_n),$$

the Pearson correlation coefficient is defined as

$$\mathrm{corr}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}},$$

where $\bar{x}$ and $\bar{y}$ are the mean values of $\mathbf{x}$ and $\mathbf{y}$, respectively. The correlation distance is then

$$\begin{aligned} d_{\mathrm{corr}}(\mathbf{x}, \mathbf{y}) &= 1 - \mathrm{corr}(\mathbf{x}, \mathbf{y}) \\ &= 1 - \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}. \end{aligned}$$

Correlation distance works well when the overall pattern of a series is more important than the actual values. It is especially useful for long time series, because it captures the similarity in shape across all points, unlike standard $L_2$ distance, which can sometimes fail to highlight pattern similarities.

On the other hand, correlation distance has some drawbacks. It does not work well for series that are nearly constant or have very little variation, since the correlation can become unstable or undefined [13]. In these cases, even small amounts of noise can have a big effect, which may lead to misleading similarity results [14].

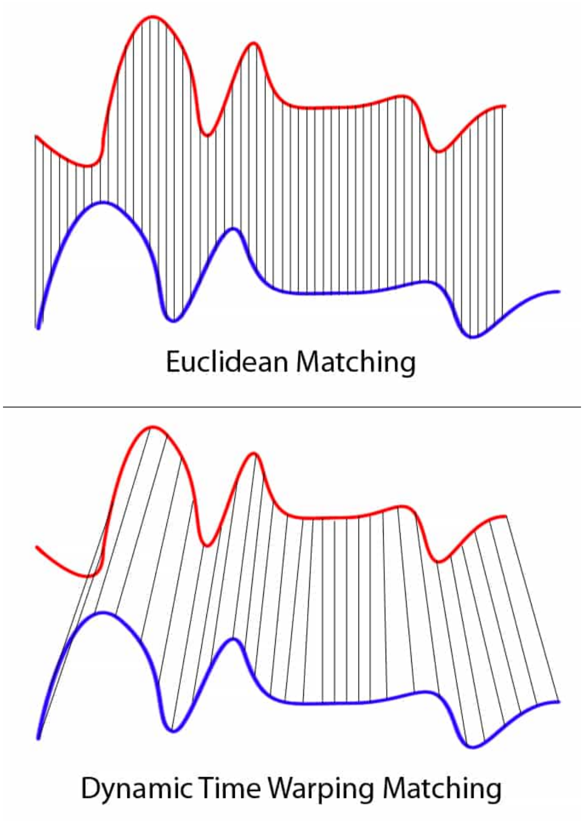


Figure 1: Euclidean Matching vs Dynamic Time Warping Matching

### 3.5 Dynamic Time Warping

Point to point measures such as Euclidian, Manhattan, cosine and correlation distances can only be used in equal-length and well aligned time series. When dealing with uneven-length or temporarily misaligned time series, these measures are no longer suitable. In such cases, an elastic similarity measure such as Dynamic Time Warping (DTW) can be used (Figure 1).

Dynamic Time Warping, introduced by Donald J. Berndt and James Clifford in 1994, is one of the most researched and used elastic distance measures [15]. It compares two time series by "stretching" or "compressing" the time axis so that similar patterns line up, which means one point from one series can match to multiple points in the other one. One boundary condition is required such that the first and the last points must be matched.

However, because DTW does not satisfy the triangle inequality requirement, it does not qualify as an actual distance metric even though it is commonly referred to as a distance measure in practice [16]. Furthermore, since it computes cumulative distance for each point to all points in the other series, this approach is computationally expensive, therefore more suitable for relatively short time series or small datasets [16]. In addition, DTW is sensitive to noise, which can lead to over-warping [17].

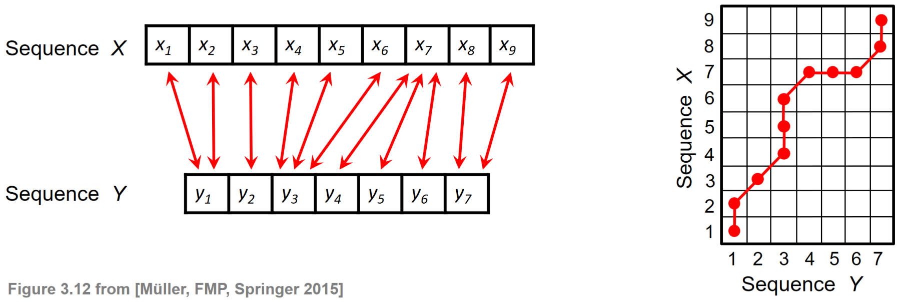


Figure 2: Matching Two Sequences with DTW

#### 3.5.1 DTW algorithm

Given two time series

$$\mathbf{x} = (x_1, x_2, \ldots, x_N), \quad \mathbf{y} = (y_1, y_2, \ldots, y_M),$$

DTW constructs an $N \times M$ cost matrix $D(i, j)$. The matrix is initialized as

$$\begin{aligned} D(0,0) &= 0, \\ D(i,0) &= \infty, \quad \text{for } i = 1, \ldots, N, \\ D(0,j) &= \infty, \quad \text{for } j = 1, \ldots, M. \end{aligned}$$

Each cell of the matrix is computed by

$$D(i,j) = d(x_i, y_j) + \min\{D(i-1, j-1),\, D(i-1, j),\, D(i, j-1)\},$$

where $d(x_i, y_j)$ denotes a local distance measure, usually Euclidean or Manhattan distance

$$d(x_i, y_j) = \sqrt{(x_i - y_j)^2}.$$

or

$$d(x_i, y_j) = |x_i - y_j|.$$

The DTW distance is defined as

$$\text{DTW}(\mathbf{x}, \mathbf{y}) = D(N, M).$$

The optimal alignment (warping path) is obtained by tracing back from $D(N, M)$ to $D(0, 0)$ (Figure 2).

## 4 Previous Studies on Frequency-Domain Clustering

So far, all the distance measures we have discussed are used in the time domain. Distances like Euclidean distance or DTW compare time series point by point, but they do not reveal underlying repeating cycles. In many real-world time series such as climate, health or sale data, they can contain periodic behaviour that cannot be spotted easily from the original time series. To identify these patterns, it is useful to analyse the data in the frequency domain, which tells us which cycles are present and how strong they are.

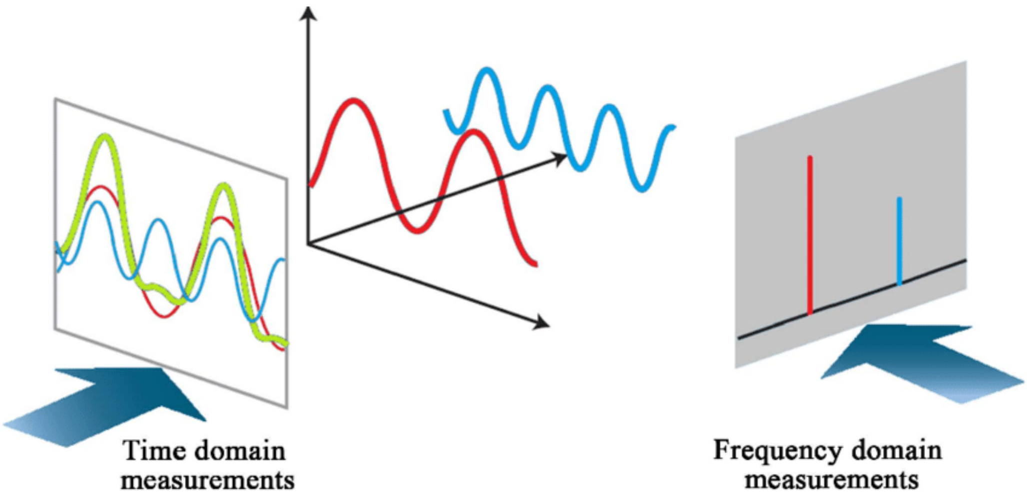


Figure 3: Time-domain vs frequency-domain graphs

Several studies have highlighted the advantages of frequency-domain clustering. Two notable works are ”Fuzzy clustering of time series in the frequency domain” by Maharaj & D'Urso (2011) and ”Robust Clustering for Time Series Using Spectral Densities and Functional Data Analysis” by Rivera-García et al. (2017).

### 4.1 Maharaj & D'Urso (2011)

Maharaj and D'Urso introduced a fuzzy clustering method based on features of the frequency-domain, which are periodograms and cepstral coefficients. Different from traditional clustering where each series belongs to one cluster, fuzzy clustering allows one series to belong to multiple clusters with different membership strengths. This approach is specifically suitable for time series whose patterns gradually change or overlap [22].

Through simulations, it was found cepstral-based fuzzy clustering (frequency-domain features) outperformed methods based on autocorrelation (time-domain features), wavelet coefficients (time–frequency features), and also other spectral representations such as the normalized and log periodogram. These results show that frequency-domain features are particularly effective at capturing similarities in periodic behaviour that may not be clearly visible in the original time domain [22].

### 4.2 Rivera-García et al. (2017)

In this study, Rivera-García and his colleagues used a functional data approach for spectral clustering. They transformed each time series into its estimated spectral density and treated as a function. Then, they applied clustering algorithms directly on these functional representations, which allowed the method to consider for the whole frequency structure of the series instead of just looking at individual time points. Their study applied this method to real-world oscillatory signals, such as electricity load profiles, and the results showed that it grouped the series much better than the standard time-domain clustering. This work highlights the practical advantage of spectral density-based distances in identifying patterns in real datasets [23].

### 4.3 Motivation

Overall, these studies suggest that spectral or frequency-domain features provide a more effective way to capture underlying periodic behaviour than traditional time-domain distance measures. This motivates the focus of this report on spectral density-based distances for clustering time series with cyclical patterns.

## 5 Spectral Analysis of Time Series

Frequency-domain analysis can decompose a time series into sinusoid components (sine and cosine) at different frequencies, thereby uncovering intrinsic cycles in the series. This is a key motivation behind using spectral methods; rather than focusing solely on raw temporal values, the frequency domain reveals repeating patterns across the series [18]. Specifically, it can quantify how much variance is contributed by each frequency component, allowing dominant cycles such as seasonal or long-term trends to be clearly identified. As a result, frequency-domain representations offer a complementary perspective to time-domain measures [19]

### 5.1 Discrete Fourier Transform (DFT)

To move from the time domain to the frequency domain, we use Fourier Transform (FT), which converts a time series into a representation based on its frequency components. There are several forms of the Fourier Transform based on time-domain properties, including the Fourier Series (FS), Fourier Transform (FS), Discrete Fourier Transform (DFT) and Discrete-Time Fourier Transform (DTFT) [20]. In this report, we focus on the Discrete Fourier Transform (DFT), since time series data are discrete and finite in practice.

Mathematically, for a time series

$$x = (x_0, x_1, \ldots, x_{N-1})$$

The DFT is given by

$$X_k = \sum_{t=0}^{N-1} x_t \, e^{-i2\pi kt/N}, \quad k = 0, 1, \ldots, N-1.$$

Each coefficient $X_k$ is a complex number that represents the contribution of the frequency indexed by $k$ to the original time series. The exponential term $e^{-i2\pi kt/N}$ corresponds to a sinusoidal wave with frequency $k/N$ and can be separated into real and imaginary parts using Euler's formula:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Applying to the exponentional term we get

$$e^{-i2\pi kt/N} = \cos\left(\frac{2\pi kt}{N}\right) - i\sin\left(\frac{2\pi kt}{N}\right).$$

Euler's theorem plays an important role in Fourier Transform as it lets us combine sine and cosine waves as a single complex exponential, avoiding losing some parts of the signal when using only sine or cosine presentations.

After applying Euler's formula, the DFT produces a list of complex numbers $X_k$ presenting each frequency components inside the series. From each frequency, we can determine the magnitude and the phase:

$$|X_k| = \sqrt{\mathrm{Re}(X_k)^2 + \mathrm{Im}(X_k)^2}.$$

$$\phi_k = \arctan\left(\frac{\mathrm{Im}(X_k)}{\mathrm{Re}(X_k)}\right).$$

When the signal matches the frequency $k$ well, the magnitude of $X_k$ is large, indicating that the corresponding frequency is strongly present in the data; otherwise, it is small or close to zero, meaning weak or absent periodic components.

### 5.2 Spectral Density and Periodogram

After identifying the frequency components, a periodogram can be used to measure the power of each corresponding frequency component. Power, or spectral density, shows how much variance each frequency contributes to the overall signal. Strong peaks in the periodogram indicate dominant repeating patterns.

For a time series of length $N$, the periodogram at frequency index $k$ is given by:

$$I(f_k) = \frac{1}{N}|X_k|^2, \quad k = 0, 1, \ldots, N-1,$$

This transforms the original time series into a frequency sequence, where each entry represents the strength of a specific repeating cycle.

#### 5.2.1 Smoothed Periodogram

Raw periodograms can often be noisy, especially when the dataset is small. Peaks observed in a raw periodogram may sometimes appear prominent due to random fluctuations rather than true underlying patterns. To address this issue, the smoothed periodogram was introduced [21]. This method averages the spectral estimates over nearby frequencies, which reduces the variance of the spectral density estimate. As a result, the smoothed periodogram can lower the variability of the spectral estimate, which then makes it easier to identify real cycles. Although smoothing slightly flattens sharp peaks, this small bias is acceptable because it improves the overall stability and clarity of the analysis.

## 6 Conclusion

This report has shown that traditional time-domain distance measures often fall short when they are used in clustering time series with strong periodic or repeating patterns. Methods like Euclidean distance and Dynamic Time Warping compare series point by point, which can make it easy to lose important repeating patterns. Many studies have shown that frequency-domain features, such as spectral density and cepstral coefficients, can be used to define measures between time series. This helps to capture similarities in cyclical patterns better than standard time-domain distances.

By looking at time series in the frequency domain, dominant cycles and repeating patterns become much easier to detect and compare across different datasets. This makes clustering based on spectral density distances especially useful for data with clear seasonal or oscillatory behaviour, like electricity demand, climate records, or biological signals. Overall, spectral density-based distances provide a more meaningful and practical way to group time series with strong cyclical components.

## References

[1] Bangert, P 2021, Machine learning and data science in the oil and gas industry, Elsevier, Amsterdam.

[2] Liu, J & Tanizaki, H 2025, 'On prediction of stock return and volatility using clustering techniques: taking an example of the Japanese stock market', Open Journal of Social Sciences, vol. 13, no. 10, pp. 532–545.

[3] Fréchet, M 1906, 'Sur quelques points du calcul fonctionnel (On some points of functional calculus)', Rendiconti del Circolo Matematico di Palermo, vol. 22, no. 1, pp. 1–72.

[4] Hausdorff, F 1914, Grundzüge der Mengenlehre (Foundations of set theory), Veit & Comp., Leipzig.

[5] Marie, M 2018, Encyclopedia of distances, Springer, Berlin.

[6] Prompook, T, Jittanon, S, Phumeesut, K et al. 2025, 'Impact of distance measures in adaptive K-means clustering on load profiles and spatial patterns of distributed substations in Thailand', Scientific Reports, vol. 15, no. 1, p. 21123.

[7] Ding, H, Trajcevski, G, Scheuermann, P, Wang, X & Keogh, E 2008, 'Querying and mining of time series data', Proceedings of the VLDB Endowment, vol. 1, no. 2, pp. 1542–1552.

[8] Rosso, G 2026, 'Outliers emphasis on cluster analysis: the use of squared Euclidean distance and fuzzy clustering to detect outliers in a dataset', arXiv preprint.

[9] Aggarwal, CC, Hinneburg, A & Keim, DA 2001, 'On the surprising behaviour of distance metrics in high dimensional space', Database Theory — ICDT 2001, vol. 1973, pp. 420–434.

[10] Gere, A 2023, 'Recommendations for validating hierarchical clustering in consumer sensory projects', Current Research in Food Science, vol. 6, p. 100522.

[11] Tummalapalli, V 2025, 'Understanding distance metrics in KNN imputation: theoretical insights and applications', Journal of Mathematical & Computer Applications, vol. 4, no. 4, pp. 1–4.

[12] Temsiri Prompook, T, Jittanon, S & Phumeesut, K 2025, 'Impact of distance measures in adaptive K-means clustering on load profiles and spatial patterns of distributed substations in Thailand', Scientific Reports, vol. 15, no. 1.

[13] Aggarwal, CC & Reddy, CK 2014, Data clustering: algorithms and applications, Chapman and Hall/CRC, Boca Raton, FL.

[14] Kim, Y, Kim, T-H & Ergün, T 2015, 'The instability of the Pearson correlation coefficient in the presence of coincidental outliers', Finance Research Letters, vol. 13, pp. 243–257.

[15] Berndt, DJ & Clifford, J 1994, 'Using dynamic time warping to find patterns in time series', in Proceedings of the KDD-94 Workshop on Knowledge Discovery in Databases, AAAI Press, pp. 359–370.

[16] Lemire, D 2009, 'Faster retrieval with a two-pass dynamic-time-warping lower bound', Pattern Recognition, vol. 42, no. 9, pp. 2169–2180.

[17] Keogh, EJ & Pazzani, MJ 2001, 'Derivative dynamic time warping', in V Kumar & RL Grossman (eds), Proceedings of the First SIAM International Conference on Data Mining (SDM 2001), SIAM, Philadelphia, pp. 1–11.

[18] Luo, Y, Lyu, Z & Huang, X 2023, 'TFDNet: Time-frequency enhanced decomposed network for long-term time series forecasting', arXiv preprint.

[19] Zhao, Q, Lai, F & Mo, X 2025, 'Exploring time series analysis in the frequency domain with complex-valued spectral attention and bidirectional variable Mamba', Journal of Supercomputing, vol. 81, p. 875.

[20] Kulkarni, SR 2002, Frequency domain and Fourier transforms, lecture notes for ELE 201: Introduction to Electrical Signals and Systems, Princeton University.

[21] Jenkins, G.M. & Priestley, M.B., 1957. The Spectral Analysis of Time-Series. Journal of the Royal Statistical Society: Series B (Methodological), 19(1), pp.1–12.

[22] Maharaj, EA & D'Urso, P 2011, 'Fuzzy clustering of time series in the frequency domain', Information Sciences, vol. 181, no. 7, pp. 1187–1211.

[23] Rivera-García, D, Ruiz-Medina, MD & Álvarez-Esteban, PC 2017, 'Robust clustering for time series using spectral densities and functional data analysis', arXiv preprint.