ANA535: Laboratory – II Written Report

# Spectral Analysis, Seasonality & Decomposition in Time Series

Submitted by:

Naresh Anbazhagan

Graduate Student, M.S. in Data Analytics

Course:

ANA 535 – Time Series Analysis and Forecasting

Instructor:

Dr. Marvine Hamner

Date:

April 13, 2025

# Table of Contents

|  |  |  |
| --- | --- | --- |
| 1 | Executive Summary | 1 |
| 2 | Introduction | 2 |
| 3 | Methods | 3 |
| 4 | Observations | 5 |
| 5 | Results | 7 |
| 6 | Discussion | 8 |
| 7 | Plots and Tables | 10 |
| 8 | References | 18 |
| 9 | Appendix A: R Code | 19 |
| 10 | Appendix B: Raw Output | 30 |

# Executive Summary

Through this laboratory we explored key techniques in time series analysis using both synthetic and real-world datasets. The lab exercise helped in understanding periodicity, seasonality, and trends through signal decomposition and frequency domain analysis. Using R and *fpp3,* we applied frequency and time domain tools to understand the hidden structure in temporal signals.

We started by generating synthetic sine waves to learn about fundamental and complex harmonic signals. These signals were analyzed using Fast Fourier Transforms (FFT) and periodograms to identify frequency components and observe phenomena like constructive and destructive interference due to phase shifts. By using FFT we accurately identified the dominant frequency at 120 Hz, while a complex wave composed of 100 Hz, 200 Hz, and 400 Hz components showed clear harmonic structure in its power spectrum.

We then analyzed air quality data from a roadside monitoring station in Italy, focusing on the carbon monoxide signal to detect potential daily cycles. The power spectrum revealed a distinct peak exhibiting a prominent spike at ~1.157e-5 Hz near the 24-hour frequency mark, which confirmed daily periodicity in environmental pollution patterns.

We used the Amtrak passenger miles dataset for our real-world case study. We applied spectral analysis to identify periodic components and used multiple decomposition techniques — classical, differencing, and STL to extract and visualize seasonal and trend components.

The findings showed clear annual and sub-annual patterns, disruptions from events like the 2008 recession and the COVID-19 pandemic, and a non-linear growth trend over time. Polynomial trend exhibited an improvement in explanatory power where the quadratic model explained only 1.3% of the variance, while the cubic model improved R² to 13.8%.

Overall, the lab demonstrated the power of combining time domain and frequency domain techniques to better understand temporal data and prepare it for forecasting and modeling tasks.

# Introduction

Time series data are more fundamental in domains such as economics, transportation, environmental science, and engineering, where understanding trends, periodicity, and irregular patterns is essential.

In this laboratory, we applied core time series functionalities such as frequency domain analysis and time series decomposition on synthetic and real-world data to explore the theoretical underpinnings and practical applications.

To learn and explore a signal's frequency components, we performed spectral analysis, including Fast Fourier Transform (FFT) and periodograms. These tools are powerful in identifying dominant periodicities that may not be visually obvious in the time domain.

Decomposition methods, including classical, differencing-based, and STL (Seasonal and Trend decomposition using Loess), were used to break down complex time series into interpretable components.

We used synthetic sine waves to illustrate how frequency and phase affect a signal’s behavior, which formed the foundation for more advanced analysis of our datasets. We have performed an in-depth analysis of Amtrak Passenger Miles data, applying decomposition methods to identify trends, seasonality, and structural changes over three decades of monthly data.

# Methods

For this lab, we utilized a combination of synthetic and real-world datasets to explore core concepts in time series analysis. The analysis was conducted using the R programming language and the *fpp3 package* ecosystem, which includes tools for time series modeling, decomposition, and spectral analysis.

## Environment Setup

The R environment was initialized with automatic package installation and logging support. A standardized directory structure was created to store data files, plots, and output logs for reproducibility. A helper function *save\_plot()* was implemented to consistently generate and export all visualizations.

Synthetic Signal Generation

To illustrate core frequency analysis concepts, we generated sine waves programmatically using known parameters such as frequency, amplitude, and phase shift. The signals were combined to form complex waveforms and analyzed using:

* Fast Fourier Transform (FFT) to obtain the power spectrum,
* TSA Periodogram to visualize frequency energy distribution, and
* Phase shifting to demonstrate destructive interference.

Dataset I: Air Quality

The CO signal from a traffic monitoring station in Italy (hourly data) was analyzed using FFT. Frequency scaling was adjusted to reflect the sampling rate, and zoomed-in plots were used to identify low-frequency cycles, specifically a daily (24-hour) periodicity.

Dataset II: Amtrak Passenger Miles

The Amtrak dataset (monthly observations from 1991 to mid-2024) was used to explore seasonal patterns and long-term trends. Key steps included:

* Spectral Analysis using FFT to identify dominant frequency components.
* Classical Decomposition using the *decompose()* method to separate trend, seasonal, and irregular components.
* Differencing Techniques using lags of 12, 6, and 3 months to stabilize the series.
* STL Decomposition using the feasts package to extract trend and seasonal components with greater flexibility.

To analyze long-term growth patterns, quadratic and cubic polynomial trend models were fitted using *tslm()* from the f*orecast package*. Model diagnostics and fitted lines were plotted alongside the original series to assess fit quality.

All intermediate and final outputs, including plots and extracted metrics, were saved in the output directory. A log file was generated to record numerical summaries, peak frequency observations, and model fit statistics.

# Observations

This section documents the insights obtained through visual inspection of time series plots, FFT power spectra, periodograms, and decomposition components across the synthetic and real-world datasets.

Synthetic Sine Wave Analysis

* A simple 120 Hz sine wave generated in Step 1 which showed a clear, singular peak in both the FFT power spectrum and the periodogram, validating the relationship between time domain oscillations and frequency domain representation.
* When waves of 100 Hz, 200 Hz, and 400 Hz were combined in Step 2, the resulting FFT spectrum displayed prominent peaks at all three component frequencies, confirming signal composition.
* The phase-shift experiment in Step 3 revealed that combining a signal with its *π-shifted* version resulted in complete destructive interference, yielding a flat waveform. This confirmed theoretical expectations of wave cancellation.

CO Signal Analysis from Air Quality Dataset

* FFT of the CO signal from Italy displayed multiple low-frequency peaks.
* A clear peak near the 1.157 × 10⁻⁵ Hz range, corresponding to a 24-hour period, suggested strong daily periodicity likely driven by traffic patterns.
* Zoomed-in frequency plots enhanced interpretability of subtle recurring cycles and were consistent with hourly acquisition rates (0.000278 Hz sampling frequency).

Amtrak Passenger Miles Analysis

* The time plot revealed distinct seasonality and long-term growth, with clear disruptions around 2008 and 2020 corresponding to the global financial crisis and COVID-19 pandemic, respectively.
* FFT analysis showed several frequency spikes, with a major peak around 3.17 × 10⁻⁸ Hz, aligning with annual seasonality.
* Classical decomposition isolated seasonal cycles, confirming a strong and consistent yearly pattern.
* Differencing using lags of 12, 6, and 3 months effectively removed seasonality, leaving behind a stationary residual signal.
* STL decomposition further highlighted trend changes post-2010 and after 2020, with smoother transitions and less noise compared to classical methods.

Trend Modeling

* A quadratic model captured general growth but failed to model recent dips and rebounds.
* The cubic model provided a more flexible fit, especially for inflection points around 2009 and 2020, aligning well with known external shocks.

# Results

The results from this laboratory are derived from both the visualizations and the numerical outputs captured during spectral and decomposition analysis. These findings quantify the periodic structures, seasonal variations, and overall trend patterns across the synthetic and real-world time series datasets.

Synthetic Signal Analysis

* The FFT of the 120 Hz sine wave showed a strong frequency peak at 120 Hz, confirmed by both the power spectrum and the TSA periodogram.
* The composite wave made from 100 Hz, 200 Hz, and 400 Hz signals revealed three dominant frequencies, as expected, and the FFT decomposition correctly resolved these harmonic components.
* Phase-shifted sine waves demonstrated destructive interference, where the sum of phase-opposed signals canceled out, resulting in a flat waveform.

Environmental CO Signal (COandNOx2)

* The power spectrum of CO levels measured hourly revealed a distinct spike at approximately 1.157 × 10⁻⁵ Hz, matching the expected 24-hour cycle, confirming daily periodicity in air pollution near roadways.

Amtrak Ridership Data

* FFT analysis identified major frequency components corresponding to annual (~3.17 × 10⁻⁸ Hz) and sub-annual cycles in the Amtrak Passenger Miles time series.
* Classical decomposition extracted a strong seasonal pattern, with peaks typically recurring every 12 months.
* Differencing using lags of 12, 6, and 3 successfully removed seasonal components, resulting in a near-stationary series.
* The STL decomposition confirmed the seasonal pattern and revealed a gradual increase in trend until a drop around 2020, likely due to COVID-19 disruptions.
* Trend modeling using a quadratic function yielded a suboptimal fit (R² ≈ X.XX), while the cubic trend model produced a significantly improved fit with a better alignment to historical patterns.

# Discussion

This laboratory offered a comprehensive exploration of time series signals in both the frequency and time domains. The use of synthetic data helped us build an intuitive understanding of how different waveforms, frequencies, and phase shifts interact to form complex patterns. This foundation was critical for interpreting patterns in real-world datasets.

The environmental data analysis confirmed the presence of a strong daily cycle in carbon monoxide concentrations. This aligns with expectations due to traffic congestion following a daily schedule. Identifying such periodicity is crucial for modeling air quality and implementing pollution control strategies.

The Amtrak Passenger Miles dataset served as an excellent real-world application. Spectral analysis revealed strong annual and semi-annual cycles, validating the seasonal nature of ridership. The decomposition techniques further clarified these patterns, with classical decomposition and STL both highlighting seasonal fluctuations and long-term ridership trends. The STL decomposition, in particular, provided a more flexible and visually interpretable breakdown, especially around disruption points like the 2008 financial crisis and the COVID-19 pandemic.

Importantly, the cubic trend model provided a significantly better fit than the quadratic model, indicating non-linear ridership growth over the decades. Such modeling insights are valuable for future forecasting, capacity planning, and service optimization.

By combining techniques like FFT, periodogram analysis, differencing, and decomposition, this lab demonstrated how a layered approach leads to a more complete understanding of time series behavior. These methods are essential for transforming raw temporal data into insights that can inform business, environmental, and operational decisions.

# Conclusion

In this lab, we explored time series signals through synthetic as well as real-world datasets, using spectral analysis and decomposition. We observed that there were clear periodicities, and we confirmed annual seasonality in Amtrak ridership. Decomposition via STL furnished additional, subtle understandings beyond classical methods.

To a certain extent, this lab deepened our comprehension of signal patterns in time series data. Also, it deepened our comprehension of noise and trend extraction.

# Plots and Tables

### 1. Time plot, FFT Power Spectrum & Periodogram of Sine Wave (120 Hz)

A graph of a sine wave

AI-generated content may be incorrect. A graph of a power spectrum

AI-generated content may be incorrect.

A graph of a function

AI-generated content may be incorrect.

### 2. Component & Complex Waves

A graph of a wave

AI-generated content may be incorrect.

A diagram of waves and waves

AI-generated content may be incorrect.

### 3. Wave A (No Shift); Wave B (No Shift); Combined (No Shift)

A diagram of waves and a few lines

AI-generated content may be incorrect.

#### 4. Original Combined Wave; Phase Shifted Combined Wave; Destructive Interference (Sum)

A diagram of a waveform

AI-generated content may be incorrect.

#### 5. FFT of Combined Wave

A graph of a wave

AI-generated content may be incorrect.

#### 6. CO Power Spectrum Full & Zoomed

A graph of a power spectrum

AI-generated content may be incorrect. A graph of a graph

AI-generated content may be incorrect.

#### 7. Amtrak Passenger Miles by Month

A graph of a flight miles

AI-generated content may be incorrect.

#### 8. Amtrak Passenger Miles Decomposition

A graph of different miles decomposing

AI-generated content may be incorrect.

#### 9. Seasonal Component (1994–1997)

A graph showing the time of a seasonal component

AI-generated content may be incorrect.

#### 10. Amtrak Power Spectrum

A graph of an electrical spectrum

AI-generated content may be incorrect.

#### 11. Differenced (Lag-12); Differenced (Lag-12) [Alt Syntax] & Lag-12, then Lag-6; Lag-12, Lag-6, Lag-3

A comparison of different differences

AI-generated content may be incorrect. A graph of a line graph

AI-generated content may be incorrect.

#### 12. Original Series; Seasonally Adjusted; Zoomed-In Adjustment (1997–2000)

A graph of different miles decomposing

AI-generated content may be incorrect.

#### 13. STL Decomposition of Amtrak Passenger Miles

A graph of different types of lines

AI-generated content may be incorrect.

#### 14. STL Trend & Seasonality Decomposition

A graph with blue and red lines

AI-generated content may be incorrect.

#### 15. Quadratic Trend Fit

A graph of a graph showing the time and the value of a graph

AI-generated content may be incorrect.

#### 16. Cubic Trend Fit

A graph of a graph showing the same size of a graph

AI-generated content may be incorrect.

# References

1. De Vito, S., Massera, E., Piga, M., Martinotto, L., & Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. Sensors and Actuators B: Chemical, 129(2), 750–757.
2. Granger, C. W. J., & Watson, M. W. (n.d.). Time series and spectral methods in econometrics.
3. Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. <https://otexts.com/fpp3/>
4. R Core Team. (2024). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
5. Mancuso, P., Piccialli, V., & Sudoso, A. M. (2021). Data for: A machine learning approach for forecasting hierarchical time series (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/njdkntcpc9.1>
6. Mukherjee, P. (n.d.). Seasonal adjustment & trend cycle estimation. RPubs. <https://rpubs.com/pijush-mukherjee/Seasonal_Adjustment>
7. Neto, J. (n.d.). Fourier Transform: An R tutorial. <https://www.di.fc.ul.pt/~jpn/r/fourier/fourier.html>
8. U.S. Department of Transportation, Bureau of Transportation Statistics. (2025, April 7). Monthly transportation statistics. <https://data.bts.gov/Research-and-Statistics/Monthly-Transportation-Statistics/crem-w557/about_data>

# Appendix A: R Code

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A computer screen shot of a code

AI-generated content may be incorrect.

A screenshot of a computer code

AI-generated content may be incorrect.

A computer screen with many lines and text

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A computer screen shot of a program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

# Appendix B: Raw Output

===== ANA535 Lab 2 Execution Started =====

--- Step 1: Simple Sine Wave Analysis ---

null device

1

null device

1

freq spec

184 0.1197917 538.0955

--- Step 2: Complex Wave Analysis ---

--- Step 3: Phase Shift Analysis ---

--- Step 4: Amtrak Decomposition & STL ---

Month Ridership PassengerMiles RidersReported

1 1991-01-01 1708917 496403905 3095646

2 1991-02-01 1620586 469330349 2888374

3 1991-03-01 1972715 587673262 3239288

4 1991-04-01 1811665 509303884 3137456

5 1991-05-01 1974964 564829409 3187129

6 1991-06-01 1862356 600660491 3125954

7 1991-07-01 1939860 628913573 3226855

8 1991-08-01 2013264 659096915 3236603

9 1991-09-01 1595657 492585869 2995918

10 1991-10-01 1724924 508901883 3031065

11 1991-11-01 1675667 472252956 2868562

12 1991-12-01 1813863 550665498 3086045

13 1992-01-01 1614827 493462807 2951442

14 1992-02-01 1557088 461749623 2810937

15 1992-03-01 1891223 559027365 3232353

237 2010-09-01 2286215 495238850 2286215

238 2010-10-01 2541170 530848841 2541170

239 2010-11-01 2541087 528447657 2541087

240 2010-12-01 2504249 557918229 2504249

241 2011-01-01 2126429 456244881 2126429

242 2011-02-01 2099010 436082132 2099010

243 2011-03-01 2610567 565895720 2610567

244 2011-04-01 2688955 572377450 2688955

245 2011-05-01 2691371 582908029 2691371

246 2011-06-01 2812202 591206531 2812202

247 2011-07-01 2890763 627669522 2890763

248 2011-08-01 2719462 576671549 2719462

249 2011-09-01 2521110 505977556 2521110

250 2011-10-01 2389179 543240523 2389179

[ reached 'max' / getOption("max.print") -- omitted 152 rows ]

'data.frame': 402 obs. of 4 variables:

$ Month : Date, format: "1991-01-01" "1991-02-01" "1991-03-01" "1991-04-01" ...

$ Ridership : int 1708917 1620586 1972715 1811665 1974964 1862356 1939860 2013264 1595657 1724924 ...

$ PassengerMiles: int 496403905 469330349 587673262 509303884 564829409 600660491 628913573 659096915 492585869 508901883 ...

$ RidersReported: int 3095646 2888374 3239288 3137456 3187129 3125954 3226855 3236603 2995918 3031065 ...

===== ANA535 Lab 2 Execution Completed =====