Laboratory 2 Written Report

ANA 535 Forecasting

Periodicity in Time Series Data

Name KOHEI NISHITANI Student ID# 1122867

Date April 13, 2025

Table of Contents

Introduction

Spectral methods used for periodicity identification and management within data sets receive attention through the second laboratory session which extends basic time-series principles. The laboratory starts with creation of basic sine waves at specific frequencies and sampling rates to produce composed signals through wave combination. Through this procedure the Fast Fourier Transform (FFT) transforms time-domain waves into frequency-domain power spectral which produce specific frequency peaks. As part of the laboratory instruction researchers explain how moving a wave through π radians results in total destructive interference when combined with its original form. The evaluation of dominant frequencies in artificial and natural datasets makes use of periodograms for both verification and measurement purposes.

The real-world application takes the educated signals to explore genuine data sets that showcase:

- The recorded hourly CO along with NOx data shows clearly observable daily patterns in their measurement results.
- The monthly samples of Amtrak passenger information demonstrate either yearly or longer-term repeating patterns.

The educational content reveals time-series decomposition which breaks a sequence into Trend, Seasonal and Remainder components by using decompose() and lag-based differencing and X-11 approach. Using seasonal wave subtraction at particular lag intervals reduces regular cycles to make data more stationary and better fit the model. Lag=12 functions for yearly waves serve as an example in this method. The exercise ends with a demonstration of time regression analysis which acts as an alternative way to model continuing trends following seasonal adjustments.

This analysis platform includes several complementary statistical methods running on R software including FFT for finding cyclic patterns together with periodograms for validating frequency information along with differencing and decomposition for seasonal adjustments and polynomial regression for trending estimation. The discussed methods receive real dataset demonstration at the end of this laboratory through monthly Amtrak passenger miles and CO and Nox emission data which serves as base information for developing advanced forecasting solutions.

Background

The past laboratory session focused on preparing time-series data for analysis by addressing core steps like cleaning, sort and filtering data. The Amtrak passenger dataset is now ready for analysis but needed researchers to find and process seasonal patterns found in time-series data. Signals from actual measurements like pollution and passengers follow regular repeating periodic behaviour. The problem of unaccounted cycles affects our ability to treat the series as stationary for traditional forecasting tools. The lab activities focus on multiple main subject areas which described method and procedure section.

Data

Data from two sources enables the exploration of time-series analysis periodicity finding and adjustment through this exercise. The Amtrak dataset covers a time span from January 1991 to June 2024 holding about four hundred monthly records. The records contain a month field along with Ridership, PassengerMiles and RidersReported fields. Because monthly data shows recurring seasonal behaviour on annual cycles decomposition and differencing methods would make suitable choices for analysis. The other dataset Hourly COandNOx2 measurements from one year were recorded air quality dataset. The dataset consists of 8,760 rows that includes carbon monoxide measurements along with other measurements. Analysis starts with daily cycles which last 24 hours because this dataset rovides one-hour measurements. When using subsequent modeling methods which require stational data such as ARIMA it becomes necessary to identify and eliminate repeating patterns because doing so improving the achievement of stationary.

Methods and Procedures

All of exercises were conducted within Rstudio and using provided R script and libraries such as lubridate, tsbble, feast, forecast, and TSA. These packages provide essential functionality for time-series handling, spectral analysis, and decomposition. The R script for this laboratory, located in the appendix, is organized into four main segments reflecting the steps described below.

Initially, the script generates and visualize synthetic sine waves to illustrate fundamental concept of periodicity. A sampling frequency of 1,000HZ is used for a simple 120Hz wave, verifying that a FFT finds periodicities from sine wave. This reveals FFT can detect hidden frequency (Figure1).

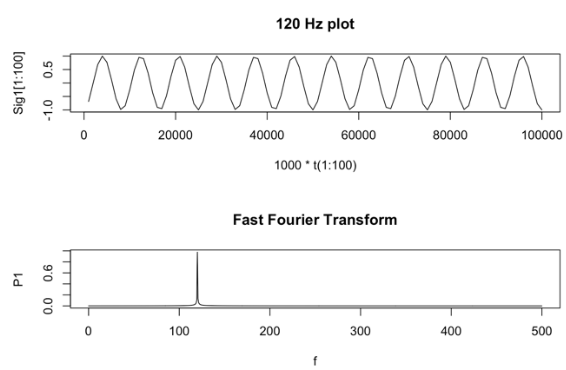


Figure 1

There are several methods to look at the energy in the frequency spectrum. One of the method is Periodogram from TSA library. It enables to find higher estimated spectral power (Figure. 2)

Table: Top 3 Frequencies from Periodogram

	freq	spec
184	0.1198	538.0955
185	0.1204	130.1534
183	0.1191	28.5253

Figure 2

As next, this exercise introduces phase shifts by adding pi to sine waves (Figure. 3) and demonstrates destructive interference in situation where the original and shifted signals align out of phase. This portion relies on time-domain plots to show how a shifted wave cancels the original wave if it is exactly half a cycle behind (Figure N).

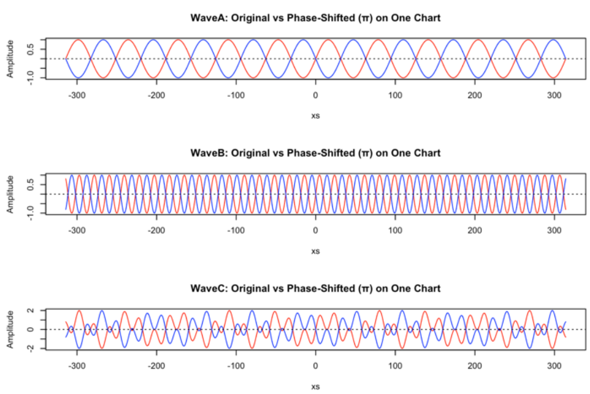


Figure. 3

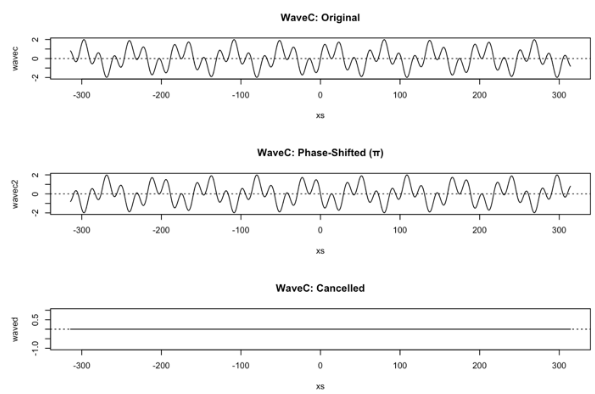


Figure. 4

Lastly, this exercise also introduces the concept of superposition of waves and demonstrate how FFT identifies the frequency from combined and complex wave. As an example, this exercise generated three types of waves with different frequency and amplitude, then combine it to one wave (Figure. 5). Because this complex and combined wave is composed by three different frequencies, FFT can detect the frequencies by showing power spectral(Figure 6).

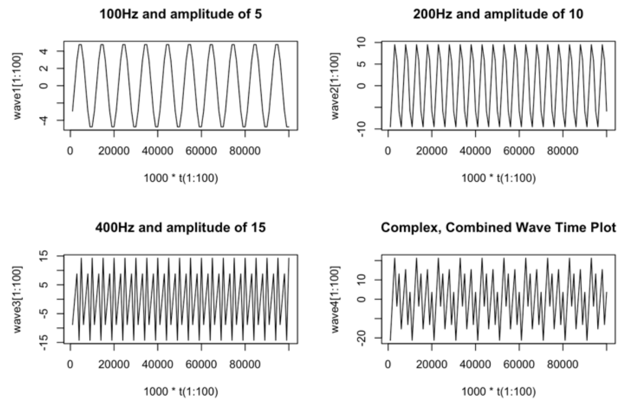


Figure. 5

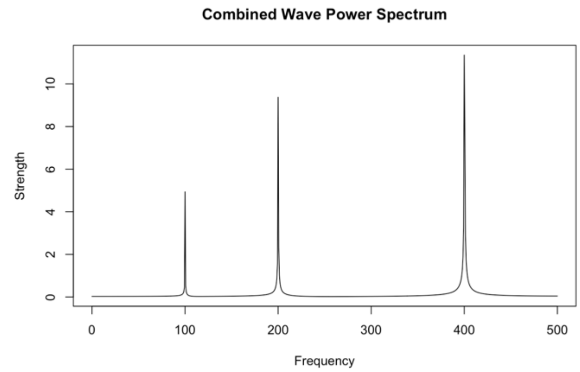
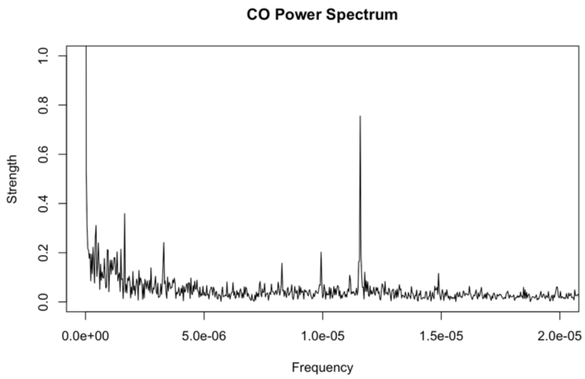


Figure. 6

Results

One objective of this exercise was to examine the COandNOx2 dataset, which consists of hourly measurements of carbon monoxide over the course of a full year. The goal was to investigate what kind of periodicities exist on this dataset. An FFT applied to this hourly time series revealed a prominent spike near the 24-hour mark (Figure 7), indicating a pronounced daily repeat. Closer inspection of frequencies around 1.16×10^{-5} Hz (the reciprocal of 86 400 seconds) confirms that the dominant periodic component aligns closely with a one-day cycle. A periodogram from the TSA package further supports this finding by showing maximum spectral power in precisely those intervals(Figure 8). These observations suggest that hourly pollution measurements exhibit a strong daily pattern, implying that additional processing—such as subtracting a fitted 24-hour seasonal component—would be required to approach stationarity and facilitate accurate forecasting.



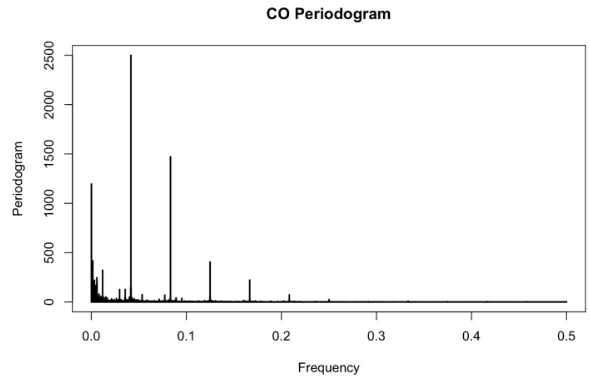


Figure. 7

Figure. 8

In contrast to the hourly CO data, the Amtrak dataset is sampled monthly, covering period from 1991 to 2024. The decomposition plot(Figure 9), which segregates trend, seasonality, and reminder reveals an annual pattern in the seasonal component.

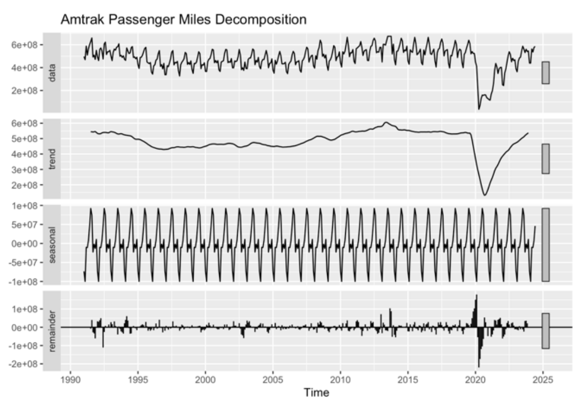


Figure. 9

This exercise also introduced two methods how to reduce or remove this seasonal effect. First, using diff() command from forecast package, differencing at lag=12, and optionally at shorter lags like 6 or 3, demonstrated how subtracting the previous year’s readings stabilizes the data and reduces annual variation(Figure 10).

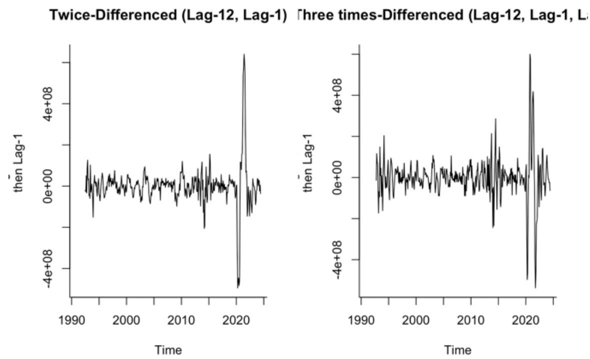


Figure 10

Secondly, classical decomposition approach calculates separate trend and seasonal components using decompose() function. By subtracting the seasonal wave from original time-series data, this output shows seasonally adjusted version whose recurring peak subtracted. Bottom left Zoomed-in plot(Figure 11) shows seasonally adjusted data and it reveals no strong periodicity, that suggesting that differencing or direct subtraction effectively removes periodic cycle.

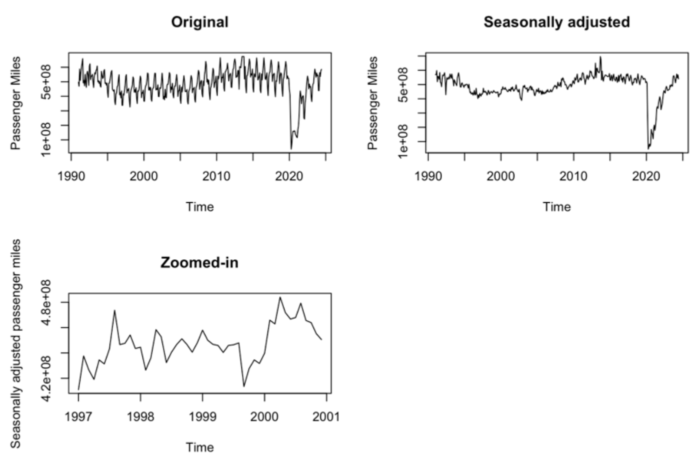


Figure 11

Lastly, X-11 method from seasonal package converts time-series to tsibble format(Figure 12). And this package generates components of raw data, season-adjusted series, and trend, easily extract those components and visualize in one plot(Figure 13).

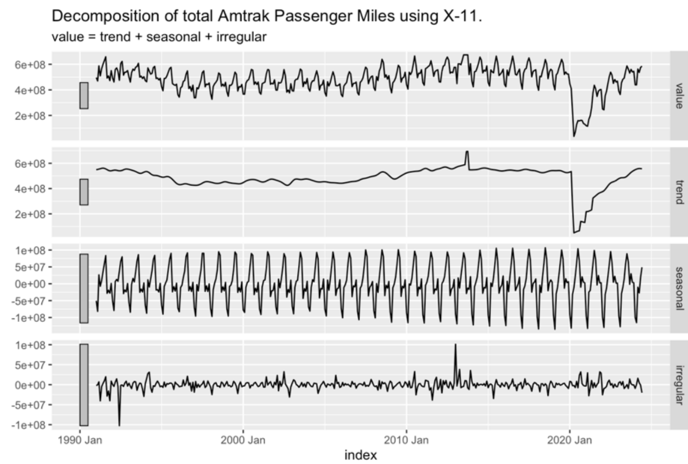


Figure 12

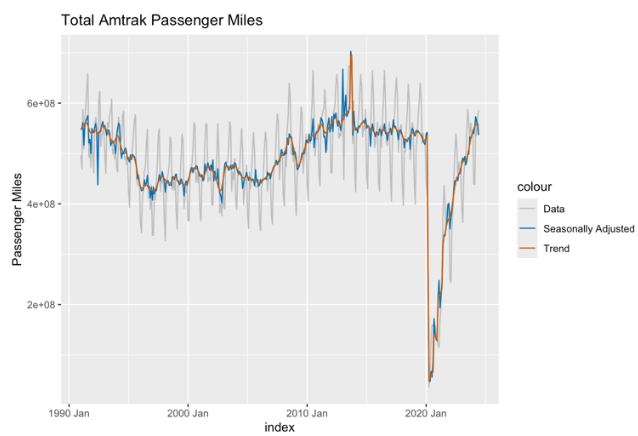


Figure 13

Conclusions

This exercise demonstrated several methods for detecting and removing periodicities in time-series data. Synthetic waves confirmed the effectiveness and usefulness of FFT and periodograms for revealing hidden cycle. This exercise also engages with real-world dataset, it showed strong daily and annual patterns respectively, it enables to achieve stationary to run forecasting accordingly.

References

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on April 8 2025.

Peter J. Brockwell and Richard A. Davis (1996) Introduction to time series and forecasting

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

S. De Vito, E. Massera, M. Piga, L. Martinotto, G. Di Francia. (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.

References

```
library(xlsx)
library(fpp3)
library(dplyr)
library(tidyverse)
library(ggplot2)
library(tsibble)
library(tsibbledata)
library(fable)
library(feasts)
```