2022 Fall

IE 313 Time Series Analysis

1. Introduction

Yongjae Lee

Department of Industrial Engineering

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Time series analysis

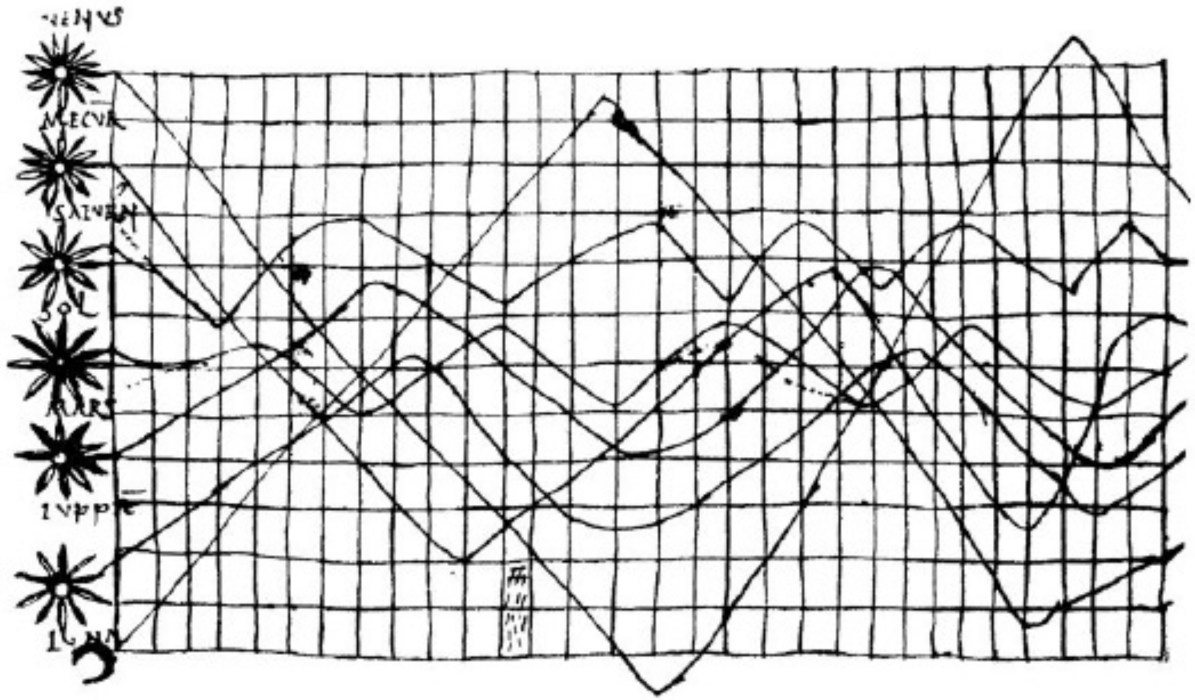
- Time series analysis**: the analysis of experimental data that have been observed at different points in time**
  - It leads to new and unique problems in statistical modeling in inference
  - The use of **many conventional statistical methods is severely restricted**
    - Many of these methods assume that the observations are independent and identically distributed (i.i.d.)
    - However, there is **obvious correlation due to the sampling of adjacent points in time**
  - Therefore, we have to deal with the mathematical and statistical questions posed by time correlations

Section 1.1

Examples of Time Series

The oldest known example of a time series plot

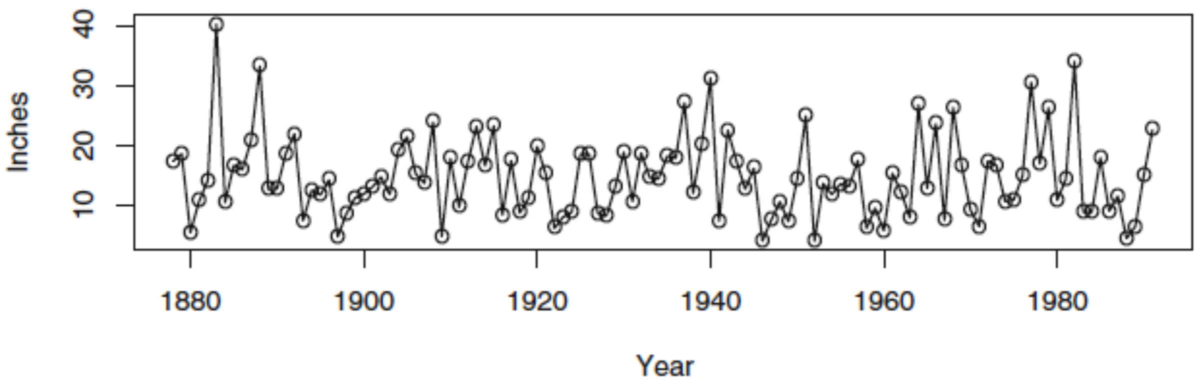
Exhibit 1.10 A Tenth-Century Time Series Plot



- It is from the 10th or 11th century showing the inclinations of the planetary orbits

Example 1: Annual rainfall in Los Angeles

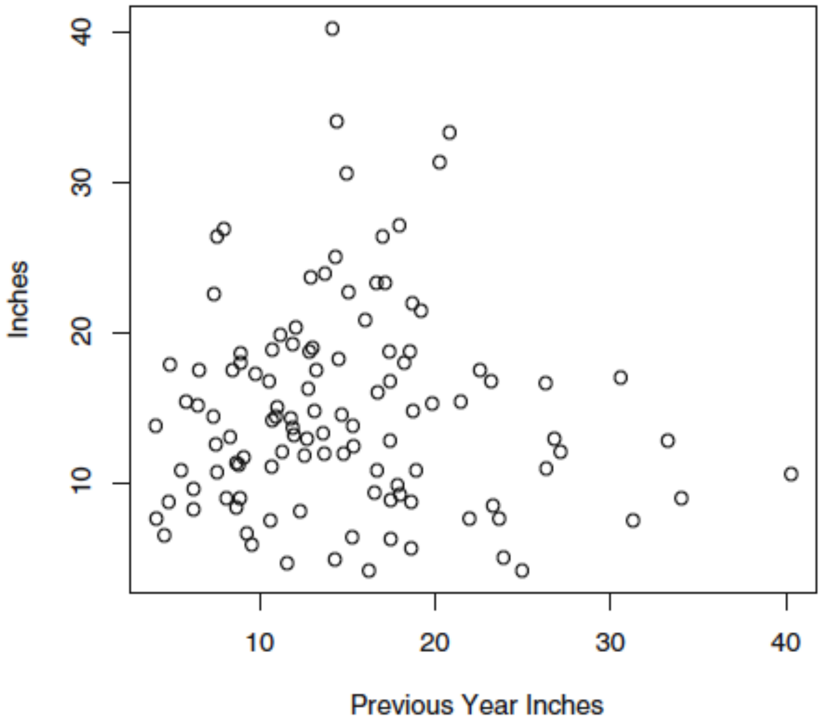
Exhibit 1.1 Time Series Plot of Los Angeles Annual Rainfall



- **Rough examination**
  - Considerable variation in rainfall amount over the years
  - Exceptionally high in 1983
- **Quick question**
  - Relationship between the amount of rainfall in consecutive years?

Example 1: Annual rainfall in Los Angeles

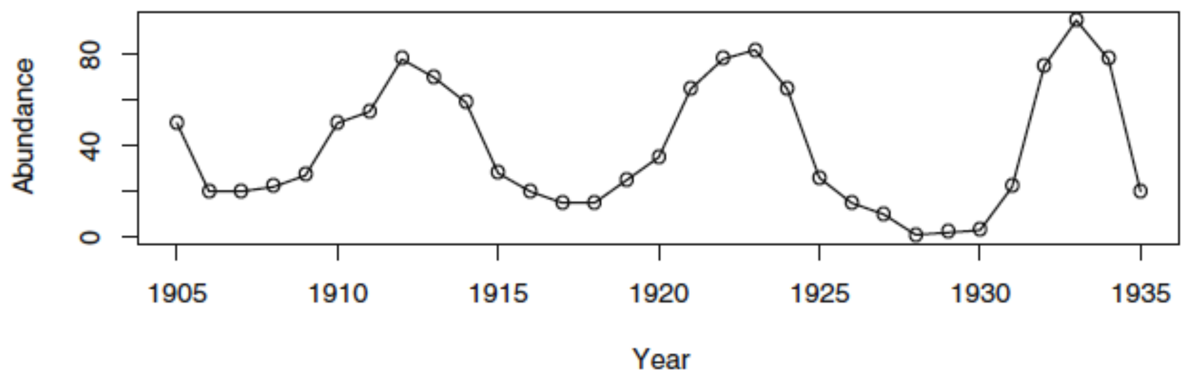
Exhibit 1.2 Scatterplot of LA Rainfall versus Last Year's LA Rainfall



- **Quick question**
  - Relationship between the amount of rainfall in consecutive years? **Probably NO**

Example 2: Annual abundance of Canadian hare

Exhibit 1.5 Abundance of Canadian Hare

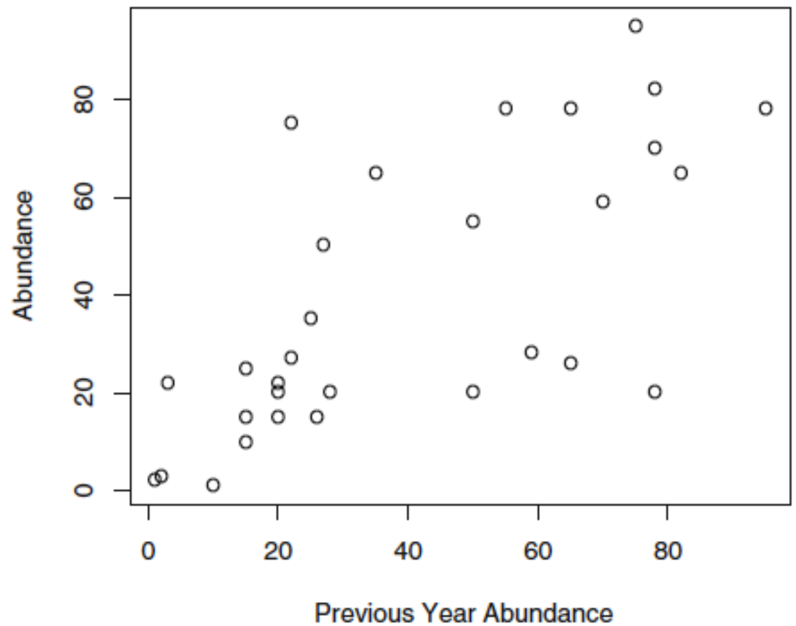


- **Rough examination**
  - Neighboring values are closely related

- **Quick question**
  - Relationship between the abundance in consecutive years?

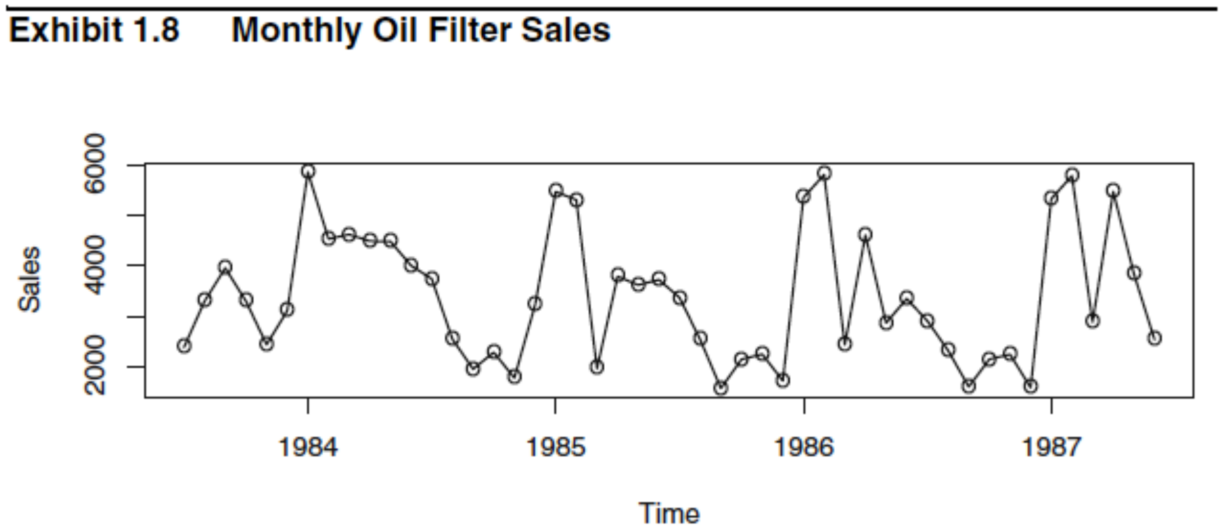
Example 2: Annual abundance of Canadian hare

Exhibit 1.6 Hare Abundance versus Previous Year's Hare Abundance



- **Quick question**
  - Relationship between the abundance in consecutive years? **Probably YES**

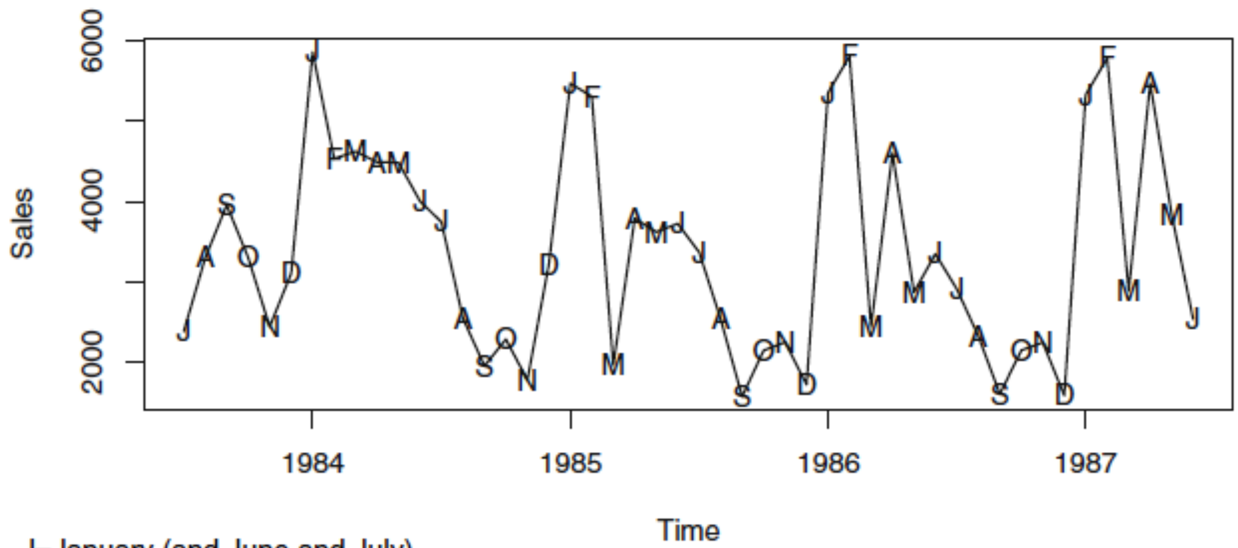
Example 3: Monthly oil filter sales



- **Rough examination**
  - Quite fluctuating
- **Quick question**
  - Is there any seasonality in this data?

Example 3: Monthly oil filter sales

Exhibit 1.9 **Monthly Oil Filter Sales with Special Plotting Symbols**



J=January (and June and July),
F=February, M=March (and May), and so forth

- **Quick question**
  - Is there any seasonality in this data? **Probably YES**

Section 1.2

A Model-Building Strategy

How to find appropriate models for time series?

- Box and Jenkins (1976) developed a multi-step model-building strategy

  1. **Model specification** (or identification)
  2. **Model fitting**
  3. **Model diagnostics**

Step 1. Model specification

- **First, the classes of time series models are selected that may be appropriate for a given observed series**

These are also called exploratory data analysis

  – Look at the time plot of the series
  – Compute many different statistics from the data
  – Any knowledge of the subject matter in which the data arise (biology, business, ecology, …)

  – Note that the model chosen at this stage is ***tentative*** and subject to revision later

  – **Principle of parsimony**: the model used should require the smallest number of parameters that will adequately represent the time series

  (This may not be true anymore. We have seen the success of deep learning)

Step 2. Model fitting

- **Second, parameter values must be estimated from the observed series**
  - Find the best possible estimates of unknown parameters within a given model
  - Should consider criteria such as
    - Least squares
    - Maximum likelihood

Step 3. Model diagnostics

- **Third, assess the quality of the model that we have specified and estimated**
  - How well does the model fit the data?
  - Are the assumptions of the model reasonably well satisfied?

  - If no inadequacies are found, the modeling may be assumed to be complete
  - And the model may be used, for example, to forecast future values

  - Otherwise, we choose another model in the light of the inadequacies found
    - That is, we return to Step 1

Section 1.3

Traditional methods vs AI methods

Model-driven approach

- Traditional models are mostly model-driven
  - Assume specific form of utility functions
  - Assume specific distribution of random variables
  - Pros
    - Easy to utilize domain knowledge
    - Easy to generalize the results
  - Cons
    - Almost impossible to model everything when the problem or the environment is too complex

Data-driven approach

- AI models are mostly data-driven
  - No specific assumption on random variables
  - No specific assumption on functional forms
  - Pros
    - Can incorporate complex (usually non-linear) relationships between multiple variables
  - Cons
    - Need enough data
    - Hard to interpret the results

Model-driven vs data-driven

- Throughout this course, we will learn both
  - Model-driven time series models
    - AR, MA, ARMA, ARIMA, …
  - Data-driven time series models
    - RNN, LSTM, Attention-based models (Transformers), …
  - Something in between
    - LDM, HMM, …
- We will see the difference in the two approaches and learn how we can take the advantages from both