CH5350: Applied Time-Series Analysis

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

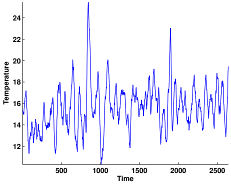
Introductory Talk

Time-Series

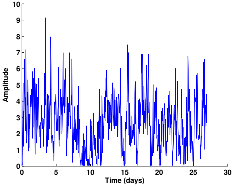
A time-series simple refers to an **ordered** collection of data (usually in time)

- e.g., yearly wages, annual production, daily temperature, hourly satellite images
- Measurements could be a function of other dimensions (e.g., frequency, space)
- Data may be collected at regular or irregular intervals
- Many variables could be recorded simultaneously (multivariate data)

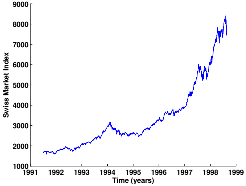
Examples



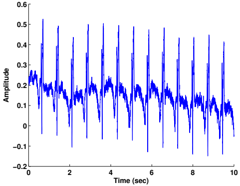
(a) Temperature



(b) Wind speed



(c) Swiss SMI

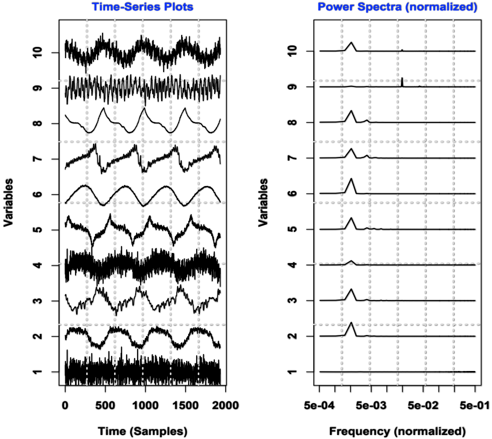


(d) ECG

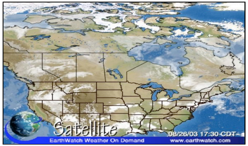
Time-series can exhibit different features

- Non-stationarity (e.g., trend, random walk)
- Oscillatory (periodicity)
- Seasonality
- Non-linearity

More examples

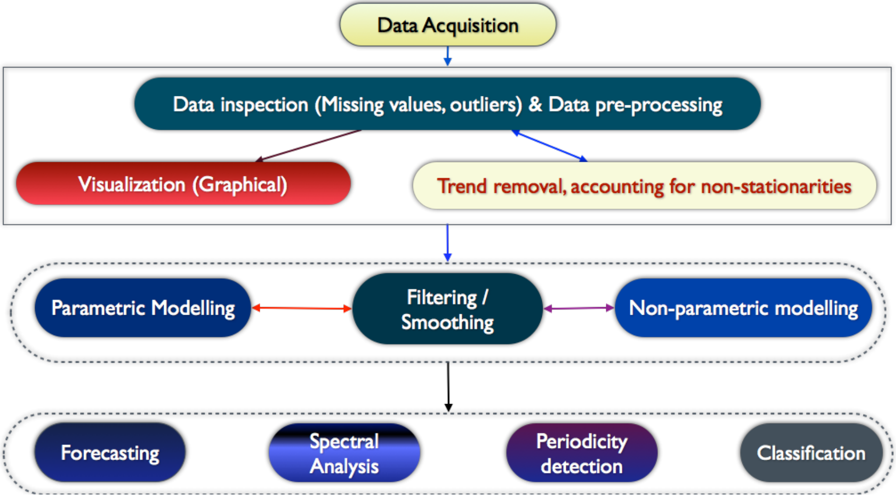


3-D Satellite Image of
Canadian weather condition

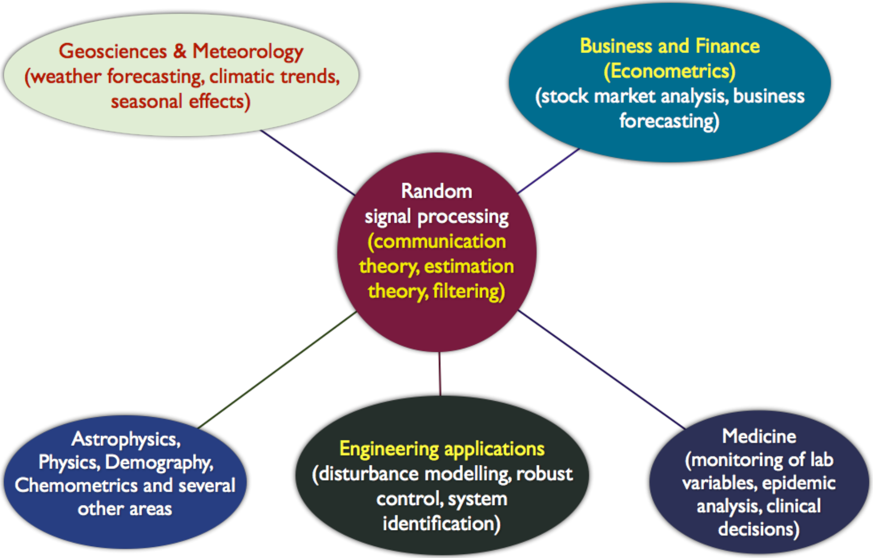


Time-series could be **multivariate**, **multidimensional** and also **transformed** into another domain

Systematic analysis of time-series data



Applications



Challenges in time-series analysis

Randomness & Uncertainty

1. Lack of a precise mathematical function to describe the process of interest (leading to a **probabilistic** framework)
2. To be able to draw inferences on the ensemble from a single realization
3. Estimation of "unknowns" from (uncertain) observations / knowns.

Sources of uncertainty and randomness

1. **Uncertainty in process knowledge:**
   - Process characteristics are seldom known accurately and/or completely. e.g., atmospheric process, roll of a die, grade of a student, etc.
   - From a prediction point of view, therefore, only outcomes with chances can be stated at best.

Uncertainty and randomness ... contd.

2. **Uncertainty in measurements:**
   - Every sensor introduces its characteristics into measurements. Hence, in reality
     $$\text{Observation} = \text{Truth} + \text{Perception (of the sensor)}$$
   - Sensor characteristics are usually not fully understood and therefore have to be treated as random variables
3. **Unknown causes:** In several situations, the causal variables are unknown and/or known, but cannot be measured or quantified.

The "realization" challenge

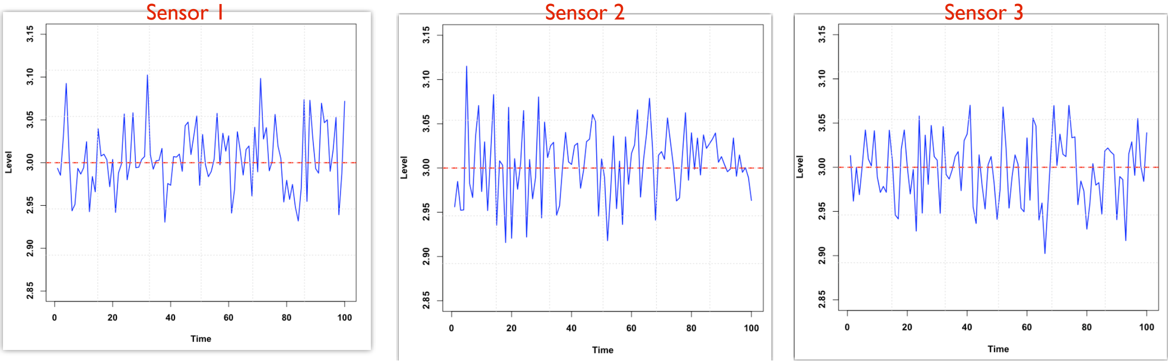
A time-series is an ordered collection of random variables. Therefore, a single time-series is one of the many possible combinations.

A single time-series is said to be a **realization** of the random process.

The collection of all possible realizations is said to be the **ensemble**.

Example: Liquid level measurement

Consider measuring liquid level in a storage tank. Neglecting all other losses, the level is constant. A single time-series (record) is a consequence of using one sensor to observe the process. Readings from three different sensors are shown (true value shown in red).



The "realization" challenge . . . contd.

In practice, we have only a single realization. The challenge is to be able to infer the truth from this single realization.

Two questions that repeatedly arise:

1. How good are the estimates (of model parameters, statistical properties, etc.) from a single realization?
2. Is a single estimate sufficient?

   Example: The average reading from sensor 1 is 3.005 and from sensor 3, is 2.996. What can we say about the true level?

A unifying question

Under what conditions can the inferences drawn from a single record of data be meaningful and useful?

1. **Stationarity:** Invariance property of the process w.r.t. time (or space).
2. **Ergodicity:** Ability to replace ensemble averages with time (or spatial) averages.

Approaches in TSA

Approaches depend on objective:

1. **Forecasting:** The most prevalent objective of TSA - applies to every field.
   - Tests for within-series and across-series predictability - **auto-correlation** and **cross-correlation** functions.
   - Develop **difference equation models** relating present to past - "**auto-regressive**" models.
   - Imagine the series to be a result of an unpredictable **shock wave** passing through a filter - "**moving average**" models.
   - Build **mixed** models that can also include other effects such as trends, random walk non-stationarities, etc. - (seasonal) **ARIMA** models.

Approaches in TSA . . . contd.

2. **Detection of periodicities:** A common goal of many engineering, meteorological, astronomical and biomedical applications.
   - **Frequency-domain or spectral representations** - extensive use of Fourier transforms (other transforms are also used in advanced analysis)
   - Notion of a "random periodic process"

It is necessary to distinguish between a **periodic deterministic** and **periodic random** process

Periodic processes

- **Periodic deterministic process**
  - Accurately predictable after one period.
    e,g., ideal spring-mass system without any damping/friction and excited with an impulse
- **Periodic random process (harmonic processes):**
  - The process has an underlying periodicity coupled with some randomness
  - Randomness could be in amplitude and/or in phase (if the signal is viewed as a sum of sine waves)
- The **power spectrum / power spectral density** is a powerful concept that allows us to study both these classes in a single framework.

Estimation

Estimators are at the heart of TSA. They produce estimates of "unobserved" or "hidden" quantities / variables from observations.

We shall learn how to:

- Characterize "goodness" of estimators
- Estimate statistical properties / parameters / signals
- Report estimation results and test hypotheses (on random processes)

Scope of this course

Course deals with largely basic and a few advanced concepts. The objective is to equip the learner with foundations of time-series analysis and estimation.

- **Linear** random processes
- **Stationary** and **non-stationary** processes.
- Mostly **univariate** and to a lesser extent, **bivariate** analysis
- **Time-domain** predictive models (ARMA, ARIMA and SARIMA models)
- **Frequency-domain (spectral)** analysis (deterministic and stochastic)
- **Estimation** theory (MoM, LS, MLE and Bayesian estimators and their properties)

A few remarks

- Emphasis is on **concepts** rather than on the rigour.
- Idea is to introduce and illustrate ideas first through examples.
- Student is expected to take this first-level treatment to a higher level by self-inquiry.
- TSA requires **both theory and skill**, i.e., it is both a science and an art.
- A computational tool is indispensable to a good understanding and practice.

R: Software for TSA

R® is an integrated software package for data manipulation, calculation and graphical display

- Powerful graphical (plotting) and statistical analysis tools.
- An expression language and like many other languages, case sensitive
- R provides the user with a very wide range of data structures (objects)
  - Symbolic, vector, array, expressions, functions, lists, data frames, factors, ...
  - Each such data object can have attributes, names, dimensions, classes, etc.
- Writing user-defined and specialized add-on packages is easy
- Use of **RStudio**® makes the use of R easy and useR-friendly!

Bibliography

- Box, G. E., G. M. Jenkins, and G. C. Reinsel (2008). *Time Series Analysis: Forecasting and Control*. New York, USA: John Wiley & Sons, Inc.
- Brockwell, P. (2002). *Introduction to Time-Series and Forecasting*. New York, USA: Springer-Verlag.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ, USA: Princeton University Press.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London, UK: Academic Press.
- Shumway, R. and D. Stoffer (2006). *Time Series Analysis and its Applications*. New York, USA: Springer-Verlag.
- Tangirala, A. K. (2014). *Principles of System Identification: Theory and Practice*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group.