Lecture 7 Introduction to Time Series

Shiwei Lan[1]

[1]School of Mathematical and Statistical Sciences
Arizona State University

STP598 Spatiotemporal Analysis
Fall 2020

Temporal data and models: challenges

When analyzing time series data, researchers in areas such as economics, climatology, epidemiology, and neuroscience are increasingly faced with challenges:

- highly multivariate, with many important predictors and response variables,
- non-stationary, hard to predict,
- often having single history, or missing data, and
- spatially correlated, as in multi-site signals or other spatially dependent multivariate data.

Time series data

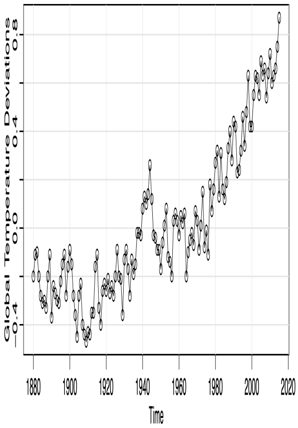

Fig. 1.2. Yearly average global temperature deviations (1880–2015) in degrees centigrade.

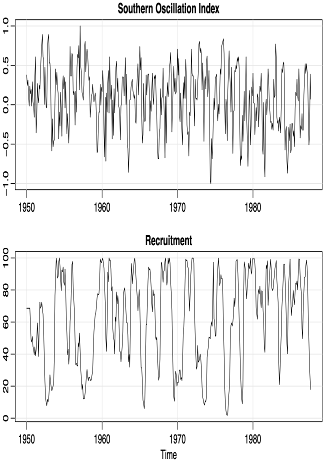

Fig. 1.5. Monthly SOI and Recruitment (estimated new fish), 1950-1987.

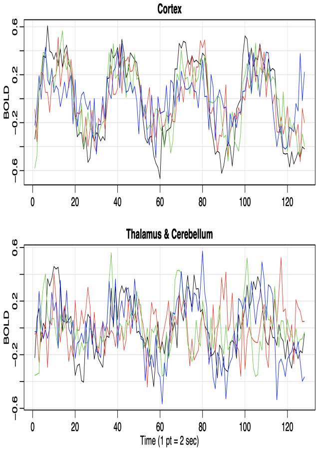

Fig. 1.6. fMRI data from various locations in the cortex, thalamus, and cerebellum; n = 128 points, one observation taken every 2 seconds.

Figure: Time Series Data: Non-stationary (left); cyclic (middle); and multivariate (right)

Time series analysis approaches

There are two separate, but not necessarily mutually exclusive methods for time series analysis:

- *time domain* approach views the investigation of lagged relationships as most important (e.g., how does what happened today affect what will happen tomorrow).
- *frequency domain* approach views the investigation of cycles as most important (e.g., what is the economic cycle through periods of expansion and recession).

In this course, we will focus on *time domain* approach.

Table of Contents

Statistical Models

- We consider a time series as a sequence of random variables, $x_1, x_2, x_3, \cdots$, denoted as $\{x_t\}$, indicating random value at time t.
- The collection of random variables $\{x_t\}$ is called a *stochastic process*. The observed values of a stochastic process is termed a *realization*. *Time series* $\{x_t\}$ is generically referred to as the process or a particular realization. How to model it?
- We could model x_t as a linear combination of *white noise* $\{w_t\}$, hence named *moving average model* **MA**(q):

$$x_t = \theta(B)w_t, \quad \theta(B) = \sum_{i=0}^{q} \theta_i B^i, \; w_t \sim wn(0, \sigma_w^2) \tag{1}$$

where B is the backward operator such that $B^i w_t = w_{t-i}$.

Autoregressive Moving Average (ARMA) Models

- Or we could model x_t as a linear combination of of its history, hence named *autoregressive model* **AR**(p):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t \tag{2}$$

Therefore it can be written as

$$\phi(B)x_t = w_t, \quad \phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i \tag{3}$$

- Or combining moving average and autoregression to obtain **ARMA**(p, q) model:

$$\phi(B)x_t = \theta(B)w_t \tag{4}$$

Drifted Models

- A model for analyzing trend such as seen in the global temperature data is the *random walk with drift* model

$$x_t = \delta + x_{t-1} + w_t \tag{5}$$

- The constant δ is called the *drift*. When $\delta = 0$, x_t is simply a *random walk*.
- The process can be rewritten as a cumulative sum of white noise variates:

$$x_t = \delta t + \sum_{j=1}^{t} w_j \tag{6}$$

- In general, we might want to write time series x_t in the simple additive format

$$x_t = s_t + v_t \tag{7}$$

where s_t denotes some unknown signal and v_t denotes a time series that may be white or correlated over time.

Measures of Dependence

- The marginal distribution functions of time series

$$F_t(x) = \Pr\{x_t \le x\} \tag{8}$$

- The corresponding marginal density functions, if exist,

$$f_t(x) = \frac{\partial F_t(x)}{\partial x} \tag{9}$$

- The **mean function** is defined as

$$\mu_{xt} = \mathrm{E}(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx \tag{10}$$

provided it exists. For simplicity we may denote μ_{xt} as μ_t.

Measures of Dependence

- The **autocovariance function** is defined as the second moment product

$$\gamma_x(s,t) = \mathrm{Cov}(x_s, x_t) = \mathrm{E}[(x_s - \mu_s)(x_t - \mu_t)] \tag{11}$$

for all s and t. For simplicity we may denote $\gamma_x(s,t)$ as $\gamma(s,t)$.

Example

Compute the autocovariances of: 1) a moving average $x_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$*; 2) a random walk* $x_t = \sum_{j=1}^t w_j$.

Measures of Dependence

- The **autocorrelation function (ACF)** is defined as

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} \tag{12}$$

- The ACF measures the linear predictability of the series at time t, say x_t, using only the value x_s.
- The **cross-covariance function** between two series x_t and y_t is

$$\gamma_{xy}(s,t) = \mathrm{Cov}(x_s, y_t) = \mathrm{E}[(x_s - \mu_{xs})(y_t - \mu_{yt})] \tag{13}$$

- There is also a scaled version of the cross-covariance function, **cross-correlation function (CCF)** given by

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}} \tag{14}$$

Stationary Time Series

- A **strictly stationary** time series is one for which the probabilistic behavior of every collection of values $\{x_{t_1}, \cdots, x_{t_k}\}$ is identical to that of the time shifted set $\{x_{t_1+h}, \cdots, x_{t_k+h}\}$. That is

$$\Pr\{x_{t_1} \le c_1, \cdots, x_{t_k} \le c_k\} = \Pr\{x_{t_1+h} \le c_1, \cdots, x_{t_k+h} \le c_k\} \tag{15}$$

  for all $k = 1, 2, \cdots$ and all time points $t_1, \cdots, t_k$, all numbers $c_1, \cdots, c_k$ and all time shifts $h = 0, \pm 1, \cdots$.
- A **weakly stationary** time series, x_t, is a finite variance process such that
  1. the mean value function, μ_t, is constant and does not depend on time t and
  2. the autocovariance function, $\gamma(s, t)$, depends on s and t only through their difference $|s - t|$.
- Two time series, x_t and y_t, are said to be **jointly stationary** if they are each stationary, and the cross-covariance function

$$\gamma_{xy}(h) = \mathrm{Cov}(x_{t+h}, y_t) = \mathrm{E}[(x_{t+h} - \mu_x)(y_t - \mu_y)] \tag{16}$$

  is a function only of lag h.

Stationary Time Series

- The **autocovariance function of a stationary time series** will be written as

$$\gamma(h) = \mathrm{Cov}(x_{t+h}, x_t) = \mathrm{E}[(x_{t+h} - \mu)(x_t - \mu)] \tag{17}$$

- The **autocorrelation function (ACF) of a stationary time series** will be written as

$$\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)} \tag{18}$$

- The **cross-correlation function (CCF)** of jointly stationary time series x_t and y_t is defined as

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\gamma_y(0)}} \tag{19}$$

Example

1) Plot ACF of a moving average $x_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$; *2) Is random walk a stationary time series?*

Stationary Time Series

- A **linear process**, x_t, is defined to be a linear combination of white noise variates w_t, given by

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \tag{20}$$

- We may show that the autocovariance function is given by

$$\gamma_x(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j \tag{21}$$

- A process, $\{x_t\}$, is said to be a **Gaussian process** if the n-dimensional vectors $x = (x_{t_1}, \cdots, x_{t_n})'$, for every collection of distinct time points $t_1, \cdots, t_n$, and every positive integer n, have a multivariate normal distribution.

Estimation of Correlation

- If a time series is stationary, the mean function $\mu_t = \mu$ is constant and estimated by the *sample mean*

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{22}$$

- Its variance can be computed

$$\begin{aligned}
\mathrm{Var}(\bar{x}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{t=1}^{n} x_t\right) = \frac{1}{n^2}\mathrm{Cov}\left(\sum_{t=1}^{n} x_t, \sum_{s=1}^{n} x_s\right) \\
&= \frac{1}{n^2}\left(n\gamma_x(0) + (n-1)\gamma_x(1) + \cdots + \gamma_x(n-1)\right. \\
&\qquad \left. + (n-1)\gamma_x(-1) + \cdots + \gamma_x(1-n)\right) \\
&= \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma_x(h)
\end{aligned}$$

Estimation of Correlation

- The **sample autocovariance function** is defined as

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}) \tag{23}$$

  with $\hat{\gamma}(-h) = \hat{\gamma}(h)$ for $h = 0, 1, \cdots, n-1$.
- The variances of linear combinations of the variates x_t can be estimated $\widehat{\mathrm{Var}}(\sum_{j=1}^{n} a_j x_j) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \hat{\gamma}(j-k)$.
- The **sample autocorrelation function** is defined as

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \tag{24}$$

- If x_t is iid with finite fourth moment, then $\hat{\rho}_x(h) \xrightarrow{d} N(0, 1/\sqrt{n})$.

Estimation of Correlation

- The estimators of the cross-covariance function, $\gamma_{xy}(h)$ can be given by **sample cross-covariance function**

$$\hat{\gamma}_{xy}(h) = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(y_t - \bar{y}) \tag{25}$$

  where $\hat{\gamma}_{xy}(-h) = \hat{\gamma}_{yx}(h)$ deermines the function for negative lags.

- The estimators of the cross-correlation, $\rho_{xy}(h)$ can be given by **sample cross-correlation function**

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}} \tag{26}$$

- For x_t and y_t independent linear processes, we have $\hat{\rho}_{xy}(h) \xrightarrow{d} N(0, 1/\sqrt{n})$ if at least one of the process is independent of white noise.

Vector-Valued Series

- A vector time series $x_t = (x_{t1}, \cdots, x_{tp})'$ contains as its components p univariate time series.
- For the stationary case, we define the mean vector $\mu = (\mu_{t1}, \cdots, \mu_{tp})' = \mathrm{E}(x_t)$.
- the $p \times p$ autocovariance matrix

$$\Gamma(h) = \mathrm{E}[(x_{t+h} - \mu)(x_t - \mu)'] \tag{27}$$

- The elements of the matrix $\Gamma(h)$ are the cross-covariance functions

$$\gamma_{ij}(h) = \mathrm{E}[(x_{t+h,i} - \mu_i)(x_{tj} - \mu_j)] \tag{28}$$

- Their sample estimates are $\bar{x} = n^{-1} \sum_{t=1}^{n} x_t$ and $\hat{\Gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})'$ respectively.

Multidimensional Series

- The *autocovariance function* of a stationary multidimensional process, x_s, can be defined as a function of the multidimensional lag vector, $h = (h_1, \cdots, h_r)'$

$$\gamma(h) = \mathrm{E}[(x_{s+h} - \mu)(x_s - \mu)'], \quad \mu = \mathrm{E}(x_s) \tag{29}$$

- The *multidimensional sample autocovariance function* is defined as

$$\hat{\gamma}(h) = (S_1 \cdots S_r)^{-1} \sum_{s_1} \cdots \sum_{s_r} (x_{s+h} - \bar{x})(x_s - \bar{x}) \tag{30}$$

  where $s = (s_1, \cdots, s_r)'$ and the range of the summation for each argument is $1 \le s_i \le S_i - h_i$ for $i = 1, \cdots, r$.
- The mean is computed over the r-dimensional array

$$\bar{x} = (S_1 \cdots S_r)^{-1} \sum_{s_1} \cdots \sum_{s_r} x_{s_1, \cdots, s_r} \tag{31}$$

- The multidimensional sample autocorrelation function follows $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$.