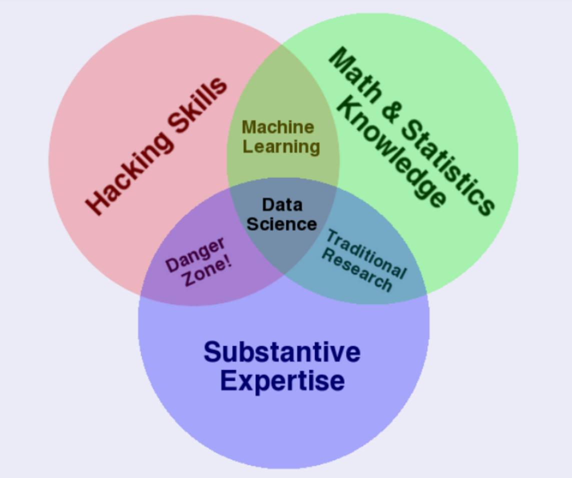
Time Series Data Analysis

Maximilian Toller

KDDM2

Time Series Data Analysis

Recall from earlier



What are *Time Series Data*?

What are *Time Series Data*?

Definition I

- Data recorded over time

- Examples:
  - Stock indices, e.g. Euro Stoxx 50, DAX, Dow Jones
  - Streaming data
  - Sensor data, e.g. monthly average rainfall

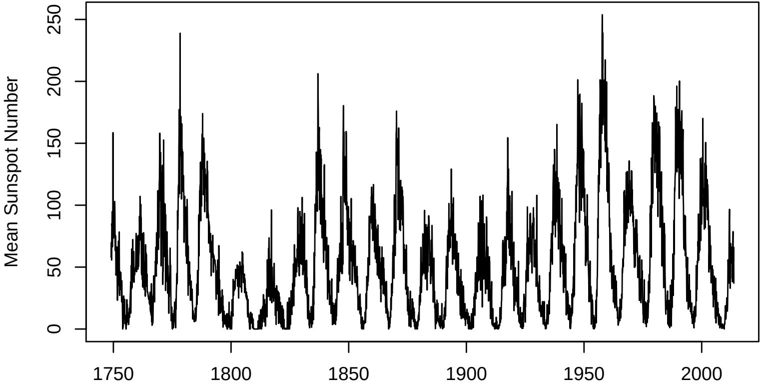
Definition II

- Single series (univariate): $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]$

- Many series (multivariate): $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \\ \vdots \\ \boldsymbol{X}^{(m)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} & X_2^{(1)} & \cdots & X_n^{(1)} \\ X_1^{(2)} & X_2^{(2)} & \cdots & X_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ X_1^{(m)} & X_2^{(m)} & \cdots & X_n^{(m)} \end{bmatrix}$

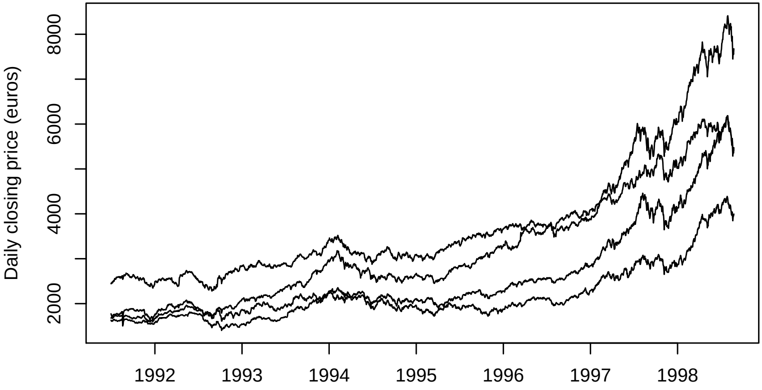
What are *Time Series Data*?

Example 1: Sunspot counts (monthly)



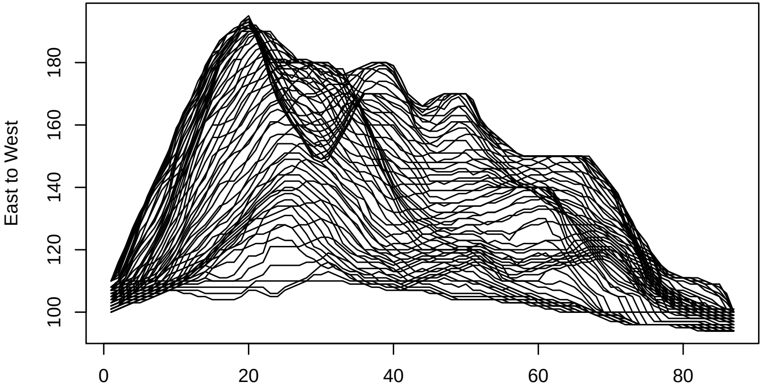
What are *Time Series Data*?

Example 2: EU stock market prices (daily)



What are *Time Series Data*?

Example 3: Volcano topography?



Typical Time Series Data Tasks

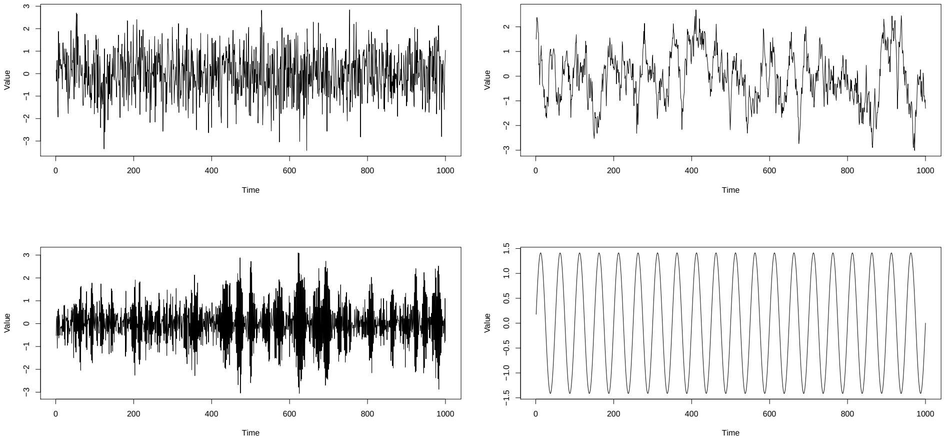
- Forecasting — What are future values of ***X***?
- Classification — Is ***X*** the sound of a cat or a dog?
- Representations — How can ***X*** be simplified?

Pitfalls

Why time series data analysis is hard

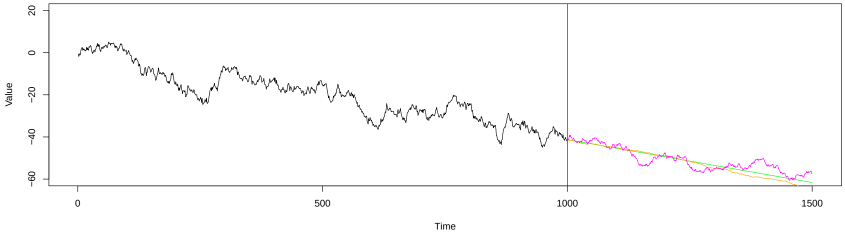
Pitfalls

#1: Time Series Data are not i.i.d.



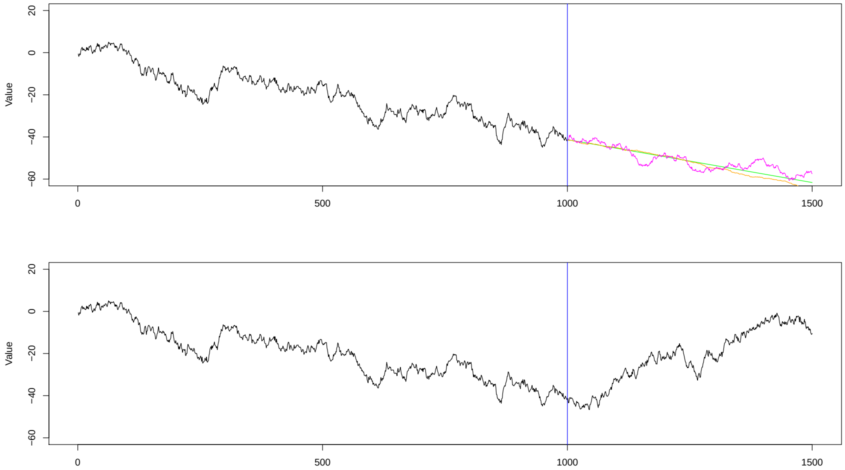
Pitfalls

#2: Lack of Invariates



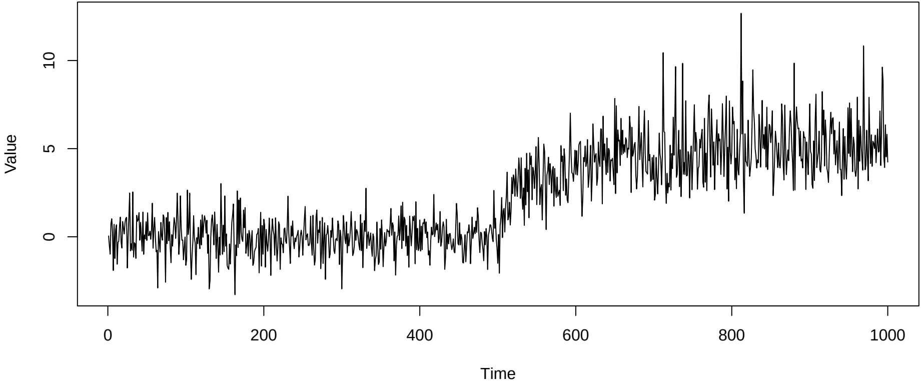
Pitfalls

#2: Lack of Invariates



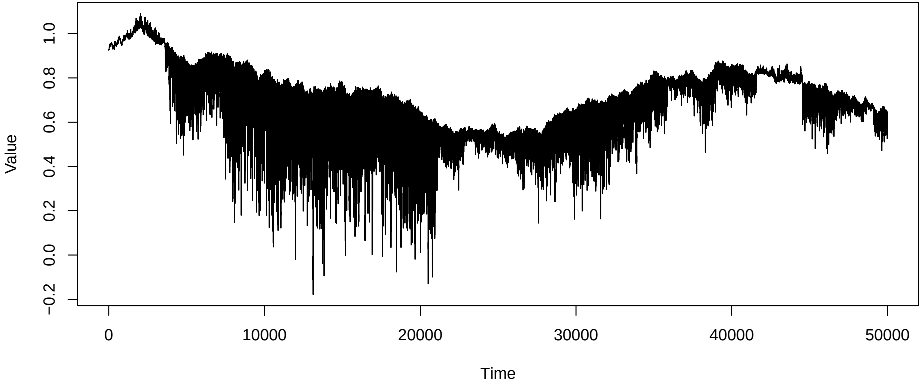
Pitfalls

#3: Concept Drift



Pitfalls

#4: Data Complexity



Stationarity

How to avoid falling into a pit

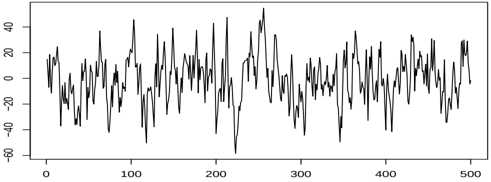
Advantages

If your data are stationary, then you . . .

- . . . may compute mean, standard deviation and histograms
- . . . can use the past to predict the future.
- . . . will not encounter concept drifts.
- . . . might be able to reduce data complexity.

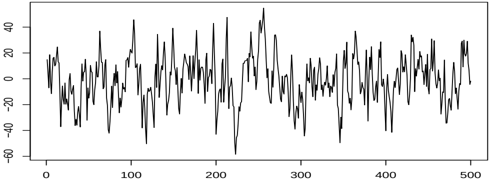
Stationarity

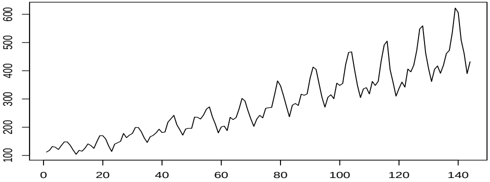
A short quiz



Stationarity

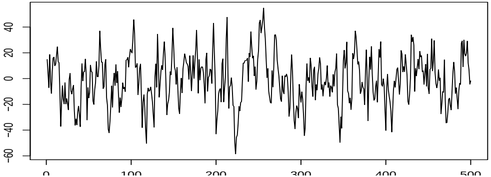
A short quiz

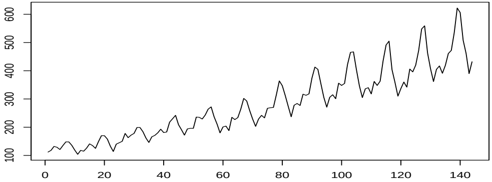


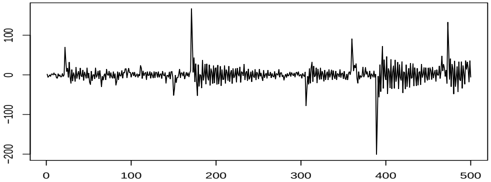


Stationarity

A short quiz

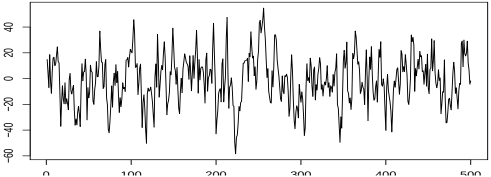


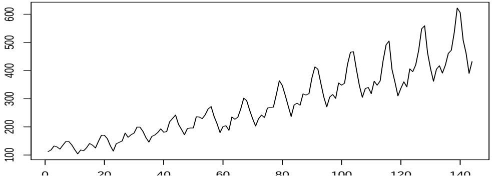


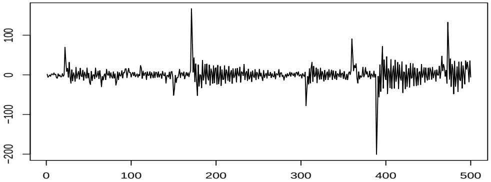


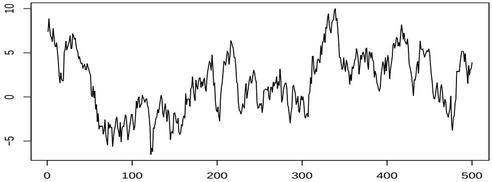
Stationarity

A short quiz









About the quiz. . .

- Hard to judge visually
- Reason: Data are finite, self-similarity hard to see
- Easier: Model stochastic process
- But what are stationary stochastic processes?

What are Stochastic Processes? I

- A collection of random variables that are ordered in some way
- $\{X(t) : t \in T\}$ where T is set of indices
- In time series case $T = \mathbb{Z}$, typical notation $\{X_t : t \in \mathbb{Z}\}$

What are Stochastic Processes? II

- Examples: $(\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),\ t \in \mathbb{Z})$
  - $X_t = \varepsilon_t$ White noise
  - $X_t = X_{t-1} + \varepsilon_t$ Random walk
  - $X_t = \phi X_{t-1} + \varepsilon_t$ Autoregressive process
  - $X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$ ARMA process

- Important: *t* is not fixed and can be set to any integer

Strict Stationarity

- A property of stochastic processes
- $F_X(X_t, \ldots, X_{t+k})$ cumulative distribution function of joint probability
- Time series $\{X_t\}$ is strictly stationary if

$$F_X(X_t, \ldots, X_{t+k}) = F_X(X_{t+\tau}, \ldots, X_{t+\tau+k}) \quad \forall t, k, \tau \in \mathbb{Z}$$

- $\rightarrow$ It doesn't matter *when* a subset was observed.

Strict Stationarity vs iid data

iid

- all data same distr. and independent
- $P(X_t|X_1, X_2, \ldots, X_{t-1}, \boldsymbol{\theta}, t) = P(X_t|\boldsymbol{\theta}) \quad X_t \sim \mathcal{D}(\boldsymbol{\theta}) \quad \forall t \in \mathbb{Z}$
- No dependence between observations, i.e. $\mathrm{Cov}(X_t, X_r) = 0 \quad t \neq r$
- iid implies strict stationarity

strict stationarity

- all data same joint distr.
- $P(X_t|X_1, X_2, \ldots, X_{t-1}, \boldsymbol{\theta}, t) = P(X_t|X_1, X_2, \ldots, X_{t-1}, \boldsymbol{\theta}) \quad \forall t \in \mathbb{Z}$
- Observations may be dependent, i.e. $\mathrm{Cov}(X_t, X_r)$ is not always 0.
- Strict stationarity does not imply iid

Problems with Strict Stationarity

- It is hard to test for strict stationarity
- Very few real-world dataset exhibit strict stationarity

Weak Stationarity I

- A time series process $\{X_t\}$ is weakly stationary if $\forall t \in \mathbb{Z}$
  - $\mathbb{E}[X_t] = \mu$
  - $\mathbb{E}[|X_t|^2] < \infty$
  - $\mathrm{Cov}[X_t, X_{t+h}] = \mathrm{Cov}[X_0, X_h] = \gamma_X(h) \quad \forall h \in \mathbb{Z}$

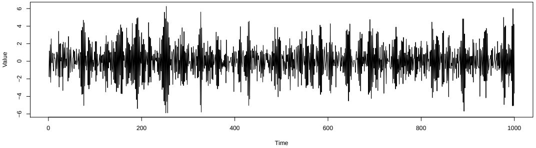
Weak Stationarity II

- In simpler terms [1]:
  - Mean is constant (never changes)
  - Variance is finite (very few outliers)
  - Autocovariance (self-similarity) does not depend on time

[1] A simple tutorial on expected values: https://tinyurl.com/y4u3a7pp

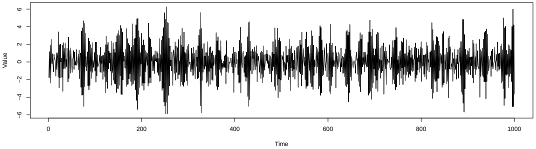
Stationarity

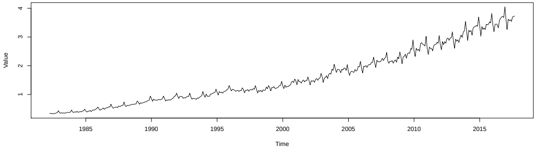
Weak Stationarity: Constant Mean?



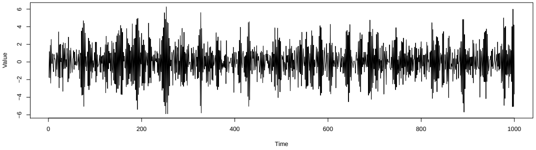
Stationarity

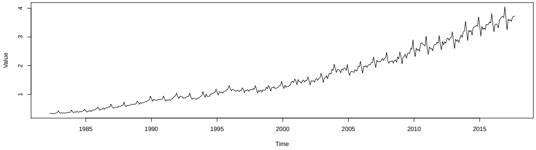
Weak Stationarity: Constant Mean?

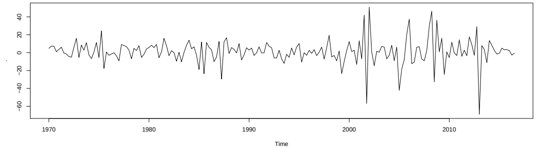




Stationarity
Weak Stationarity: Constant Mean?

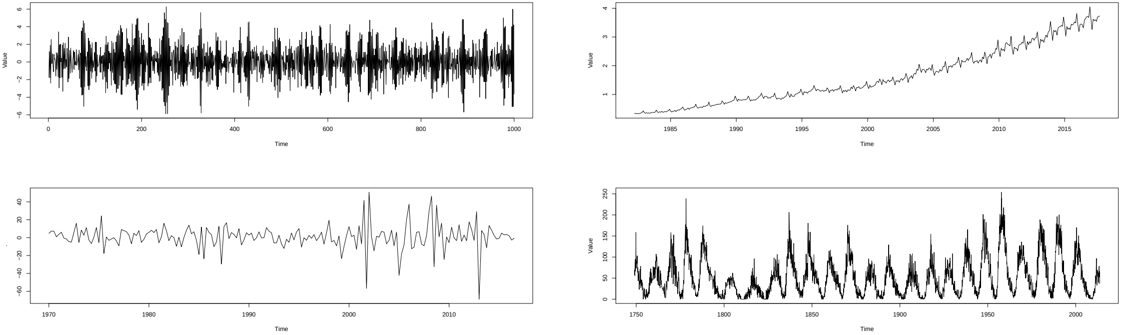




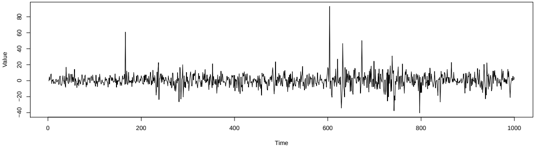


Stationarity

Weak Stationarity: Constant Mean?

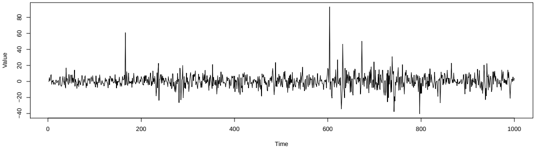


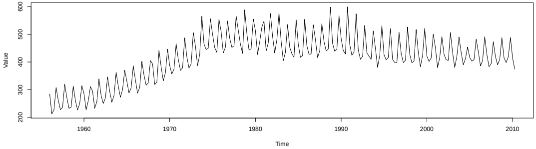
Weak Stationarity: Finite Variance?



Stationarity

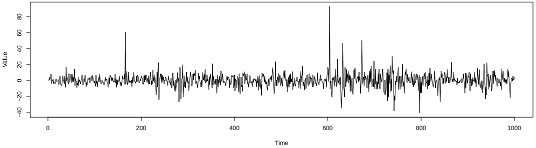
Weak Stationarity: Finite Variance?

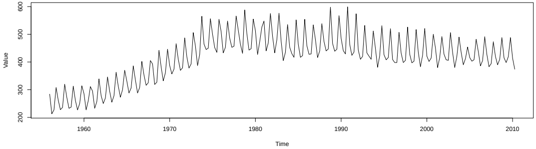


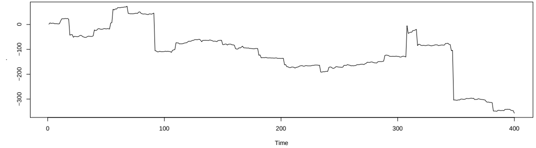


Stationarity

Weak Stationarity: Finite Variance?

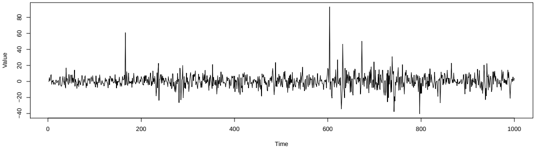


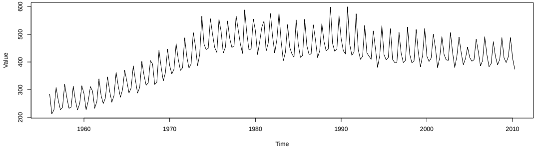


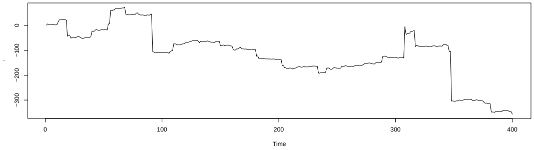


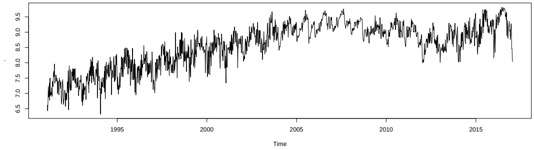
Stationarity

Weak Stationarity: Finite Variance?



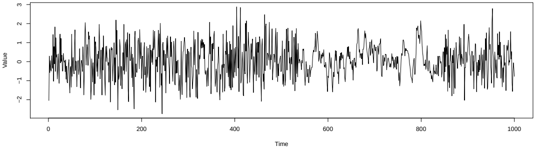






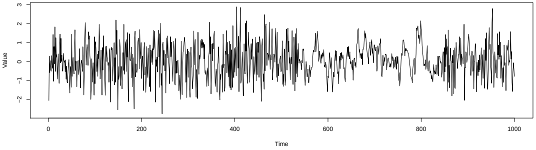
Stationarity

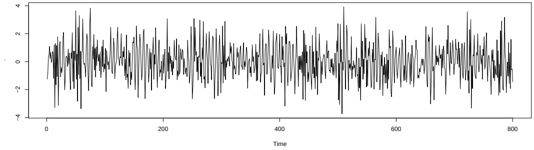
Weak Stationarity: Time-Independent Autocovariance?



Stationarity

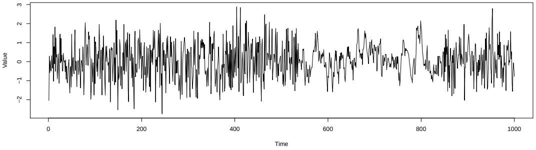
Weak Stationarity: Time-Independent Autocovariance?

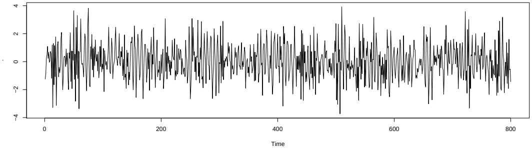


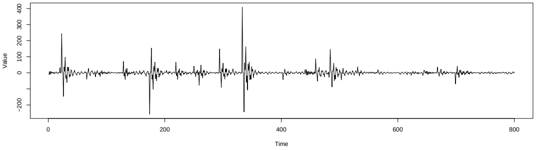


Stationarity

Weak Stationarity: Time-Independent Autocovariance?

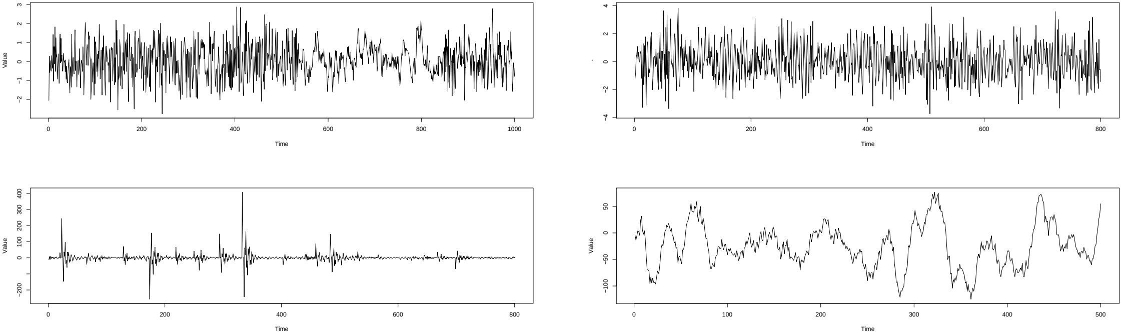






Stationarity

Weak Stationarity: Time-Independent Autocovariance?



There are no good tests

- ~~Visual inspection~~
- Stationarity Tests (unreliable)
  - Augmented Dickey-Fuller test
  - Priestley-Subba-Rao test
  - Wavelet spectrum test
- $\rightarrow$ From data alone, we can't decide if data are stationary!
  - Use domain knowledge
  - *Assume* that data are (or can be made) stationary.

How to make data stationary

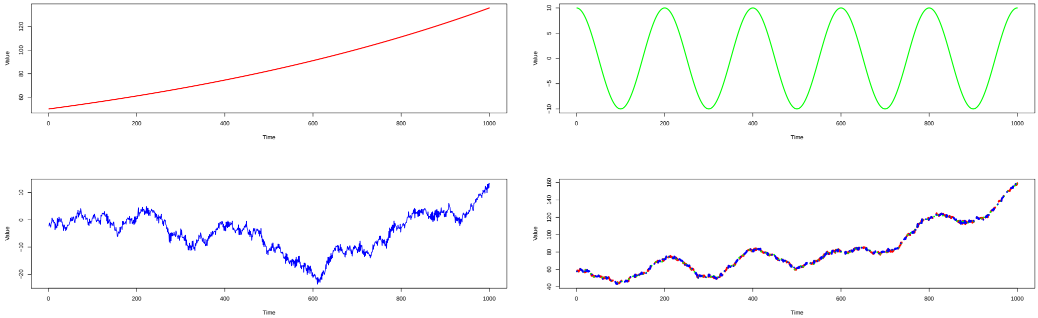
Removing trends & seasonality

The Additive Model

- Commonly time series data has 3 components
  - Trend
  - Seasonality
  - Randomness
- Additive model: $X_t = D_t + S_t + Y_t$ (there are other models)

How to make data stationary

The Additive Model: Example

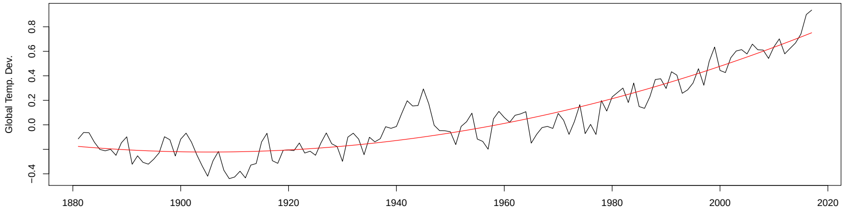


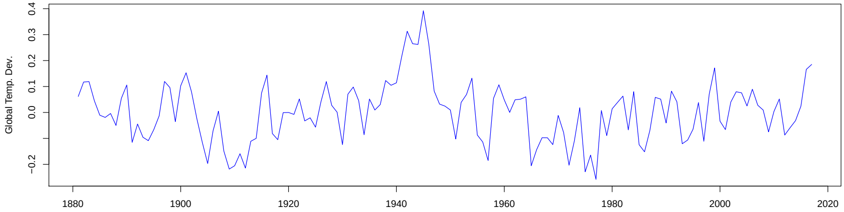
Detrending

- Linear filter
  - Choose filter coefficients
  - Requires $S_t = 0$, but removes any polynomial
- Regression
  - Choose basis function and use least squares estimator
  - Problem: Which basis function?
- Differencing
  - Compute $\nabla^1[X](t) = X_t - X_{t-1}$ (pad with 0)
  - Recursively repeat until data seem stationary

How to make data stationary

Detrending





Removing Seasonality

- Low-pass filtering
  - Estimate frequency f of $\{S_t\}$
  - Apply low-pass filter with f as cutoff
- Lagged differencing
  - Estimate season length $s = 1/f$ of $\{S_t\}$
  - Compute $X_t - X_{t-s}$ (pad with 0)
- Problem: How to estimate f or s?

How to make data stationary

Season Length Estimation

Spectral analysis

- Compute periodogram with
  $\hat{\mathcal{S}}_X(\omega) = \frac{1}{n}\left|\sum_{j=1}^{n} X_j \exp(-\mathrm{i}2\pi j\omega)\right|^2$
- $s \approx \frac{1}{\mathrm{argmax}_\omega \hat{\mathcal{S}}_X(\omega)}$
- Inconsistent estimator

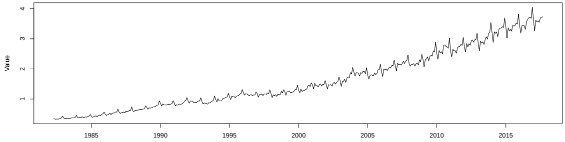
Autocorrelation analysis

- Estimate autocor. function with
  $\hat{\gamma}_X(h) = \frac{1}{n}\sum_{j=1}^{n-|h|}(X_j - \bar{X})(X_{j-h} - \bar{X})$
- $s = \mathrm{argmin}_h\{h :$
  $\left|\nabla^1[\hat{\gamma}_X](\tau)\right| < \varepsilon, \nabla^2[\hat{\gamma}_X](\tau) < 0\}$
- Hard to find local maximum

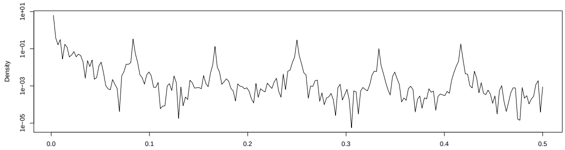
How to make data stationary

Season Length Estimation: Example

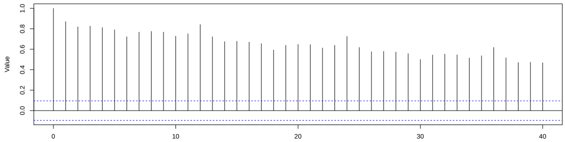
Data



Periodogram



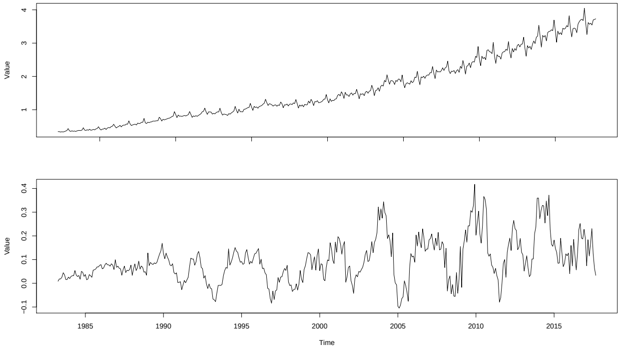
ACF



How to make data stationary

Lagged Differencing

- Easy, once season length $s = 12$ is known
- Simply compute $X_t - X_{t-12}$



Forecasting

What are future values of ***X***

Definition

- Given past $\boldsymbol{X}_p = \{X_1, \ldots, X_n\}$, estimate future $\boldsymbol{X}_f = \{X_{n+1}, X_{n+2}, \ldots\}$
- In statistics:
  - Make estimations $\hat{\boldsymbol{X}}_f = \{\hat{X}_{n+1}, \hat{X}_{n+2}, \ldots\}$
  - Analyze expected estimation error $\mathbb{E}[(\boldsymbol{X}_f - \hat{\boldsymbol{X}}_f)^2]$
- Problem: We do not know $\boldsymbol{X}_f$
- Solution: Use $\{X_1, \ldots, X_r\}$, $r < n$ to estimate $\{X_{r+1}, \ldots, X_n\}$
- Question: What are good estimators?

Methods

- There are many different forecasting methods
  - Regression
  - ARIMA
  - Exponential smoothing
  - Artificial neural networks
  - ...
- All assume stationarity
- What else do they have in common?

The only reasonable forecast I

- Data can be split
  - Constant (deterministic)
  - Changing (non-deterministic)

- One **cannot** forecast non-determinism

The only reasonable forecast II

- *In every case of a reasonable forecast, we first extract something that is constant from the process, and extend it to future.*

- Only reasonable forecast:
  - Invariate (trend, seasonality) $\rightarrow$ stays the same, extrapolate
  - Variate (random components) $\rightarrow$ expected value, use e.g. ARIMA, neural network, ...

Live Demonstration

Classification

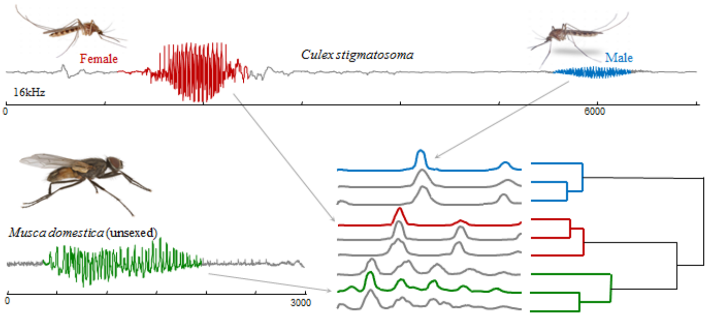
Does ***X*** belong to *A* or *B*?

Definition

- Setting
  - Time Series Database $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \\ \vdots \\ \boldsymbol{X}^{(m)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} & X_2^{(1)} & \dots & X_n^{(1)} \\ X_1^{(2)} & X_2^{(2)} & \dots & X_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ X_1^{(m)} & X_2^{(m)} & \dots & X_n^{(m)} \end{bmatrix}$
  - Classes $\boldsymbol{\mathcal{C}} = \{\mathcal{C}_1, \dots, \mathcal{C}_k\}$
  - Set of labels $\boldsymbol{L} = \{L_1, \dots, L_m\}$,
  - Query $\boldsymbol{Q} = \{Q_1, \dots, Q_n\}$
- Goal: Given database and labels, infer label of query

Classification

Example



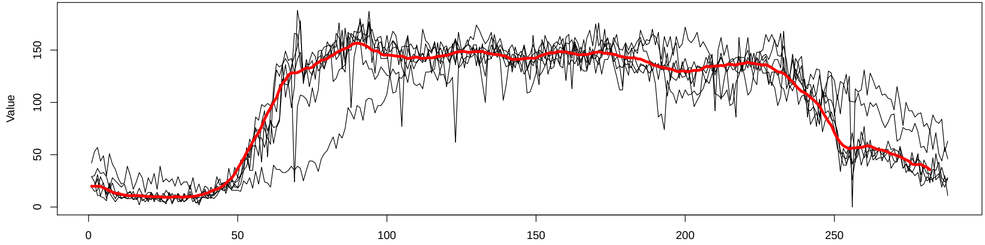
Insect classification by clustering audio snippet time series. Adapted from *Insect Detection and Classification Based on Wingbeat Sound* by Yanping Chen 2014, retrieved from `http://alumni.cs.ucr.edu/~ychen053/`. Copyright 2014 by Yanping Chen.

k-Nearest neighbor

- Needs distance function $d(\cdot)$
- Approach:
  1. Compute $d(\boldsymbol{X}^{(j)}, \boldsymbol{Q})$ for all j
  2. Class of query is label of kth smallest j (kth nearest neighbor)
- Distance functions
  - Dynamic time warping
  - Longest common subsequence
  - Edit distance with $\mathbb{R}$ penalty

Shapelets

- Approach:
  1. For each class, find/infer a representative member = shapelet
  2. Class of query is label of most similar shapelet



Representations

How can $\boldsymbol{X}$ be simplified?

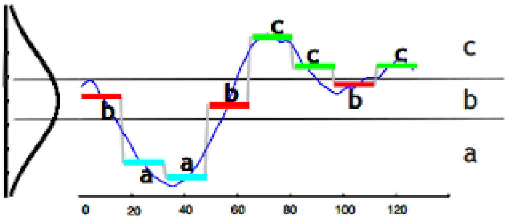
Stochastic models

- ARIMA
  - **A**utoreg**R**essive model: $X_t = \phi_0 + \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t$
  - **M**oving **A**verage model: $X_t = \theta_0 + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$
  - Differencing (reverse **I**ntegration) to achieve stationarity
- GARCH
  - $X_t = \sigma_t \varepsilon_t$
  - $\sigma_t^2 = \phi_0 + \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{q} \theta_j \sigma_{t-j}^2$
  - $\rightarrow$ random residuals follow an ARMA process

Representations

Symbolic **A**ggregate Appro**X**imation

- Piecewise-aggregate approximation
- Distribution-based breakpoints
- Symbolization

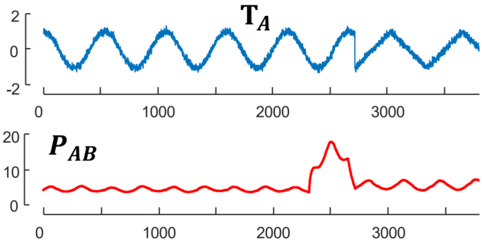


Lin, Jessica, et al. “A symbolic representation of time series, with implications for streaming algorithms.”Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. ACM, 2003.

Representations

Matrix Profile

- Sliding window $W_X(i,m) = \{X_i, \ldots, X_{i+m}\}$
- z-normalize all windows
- $\mathrm{MP}(i) = \min_j(d(W_X(j,m)), W_X(i,m))$

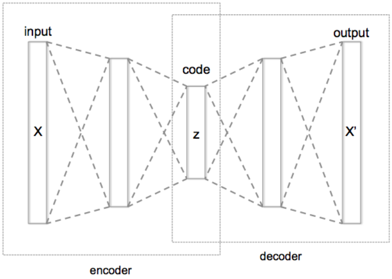


Yeh, Chin-Chia Michael, et al. "Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets." 2016 IEEE 16th international conference on data mining (ICDM). IEEE, 2016.

Representations

Autoencoders

- Train ANN to reproduce input (segments)
- Put bottleneck in the middle
- Query segment and look at bottleneck
- $\rightarrow$ Training segments need to be iid
- $\rightarrow$ Overfit? Cannot interpret.

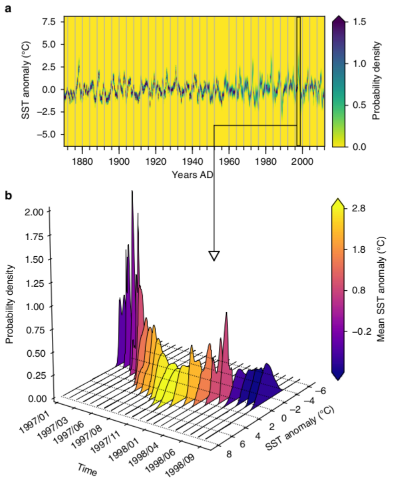


https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png

Representations

Probability Density Representation

- Hybrid representation
- Only for multivariate timeseries
- Use $\underline{X}_t = [X_t^{(1)}, \ldots, X_t^{(m)}]$ to estimate probability density function (PDF)
- Analyze relations between PDFs
- $\rightarrow$ Very powerful, directly interpretable



Goswami, Bedartha, et al. "Abrupt transitions in time series with uncertainties." Nature communications 9.1 (2018): 48.

Final Remarks

Tools I

- Descriptive statistics (if stationary)
  - Central tendency: mean, median, mode, . . .
  - Data dispersion: variance, median absolute deviation, IQR, . . .
  - Distribution: histogram, periodogram, . . .
- Windows
  - Sliding window
  - Disjoint window
  - Window functions (tapering)

Tools II

- Machine Learning
  - Tensorflow
  - Keras
- Visualization
  - Control chart
  - Scatter plot

Final Remarks

Software

- R 😎
  - http://www.statmethods.net/advstats/timeseries.html
  - https://cran.r-project.org/web/views/TimeSeries.html
  - https://github.com/robjhyndman/
- Python 🙂
  - Prophet
  - TS-Fresh
  - Pandas, NumPy, scikit-learn, Statsmodels

Final Remarks

Software

- MatLab/Octave 😐
  - TSA
  - Signal
- Java 😐
  - JMotif
  - Weka

Final Remarks

Software

- C# ☹
  - Cronos
  - Can call R . . .
- Javascript, PHP, Perl, . . .
- C++, Fortran, . . .

Further Reading

- Bagnall, Anthony, et al. "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances." Data Mining and Knowledge Discovery 31.3 (2017): 606-660.
- Fu, Tak-chung. “A review on time series data mining.” Engineering Applications of Artificial Intelligence 24.1 (2011): 164-181.
- Szegedy, Christian, et al. “Intriguing properties of neural networks.” arXiv preprint arXiv:1312.6199 (2013).
- Hyndman, Rob and Athanasopoulos, George “Forecasting: Principles and Practice.” https://otexts.com/fpp2/

The End

Thank you for your attention!