III Webinar 2021

26/08/2021 , 17h30

CT de Identificação de Sistemas e Ciência de Dados

https://www.youtube.com/watch?v=5bB6hOcFSJQ

Análise e previsão de séries temporais

Time series analysis and forecasting

Leandro dos Santos Coelho

Pontifícia Universidade Católica do Paraná (PUCPR), Escola Politécnica
Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS), Graduação em Engenharia de Controle e Automação (Mecatrônica)
Rua Imaculada Conceição, 1155, CEP 80215-901, Curitiba, PR, Brasil

Universidade Federal do Paraná (UFPR), Graduação e Pós-Graduação em Engenharia Elétrica, Campus Centro Politécnico
Av. Cel. Francisco H. dos Santos, 100, CEP 81530-000, Curitiba, PR, Brasil

Lattes (CNPq): *http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4792095Y4*

Google Scholar: *https://scholar.google.com/citations?user=0X7VkC4AAAAJ&hl=pt-PT*

Linkedin: *https://www.linkedin.com/in/leandro-dos-santos-coelho-07a08893*

Top H-Index for Scientists in Brazil: *http://www.guide2research.com/scientists/BR*

Email: lscoelho2009@gmail.com

Dynamical systems and time series forecasting

PPGEPS-PUCPR (Programa de Pós-graduação em Engenharia de Produção e Sistemas)
PPGEM-PUCPR (Programa de Pós-graduação em Engenharia Mecânica) (collaboration)
PPGEE-UFPR (Programa de Pós-graduação em Engenharia Elétrica)

2021

Hybrid multi-stage decomposition with parametric model applied to wind speed forecasting in Brazilian Northeast
S. R. Moreno, V. C. Mariani, L. S. Coelho
Renewable Energy, *https://www.sciencedirect.com/science/article/abs/pii/S0960148120316980*

Novel hybrid model based on echo state neural network applied to the prediction of stock price return volatility
G. T. Ribeiro, A. A. P. Santos, V. C. Mariani, L. S. Coelho
Expert Systems with Applications, *https://www.sciencedirect.com/science/article/abs/pii/S0957417421009003*

Hybrid wavelet stacking ensemble model for insulators contamination forecasting
S. F. Stefenon, M. H. D. M. Ribeiro, A. Nied, V. C. Mariani, L. S. Coelho, V. R. Q. Leihardt
IEEE Access, *http://dx.doi.org/10.1109/ACCESS.2021.3076410*

2020

Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network
S. R. Moreno, R. G. da Silva, V. C. Mariani, L. S. Coelho
Energy Conversion and Management, *https://www.sciencedirect.com/science/article/abs/pii/S0196890420304076*

Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting
G. T. Ribeiro, V. C. Mariani, L. S. Coelho
Engineering Applications of Artificial Intelligence, *https://www.sciencedirect.com/science/article/abs/pii/S0952197619300624*

A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting
R. G. da Silva, M. H. D. M. Ribeiro, S. R. Moreno, V. C. Mariani, L. S. Coelho
Energy, *https://www.sciencedirect.com/science/article/abs/pii/S0360544220322817*

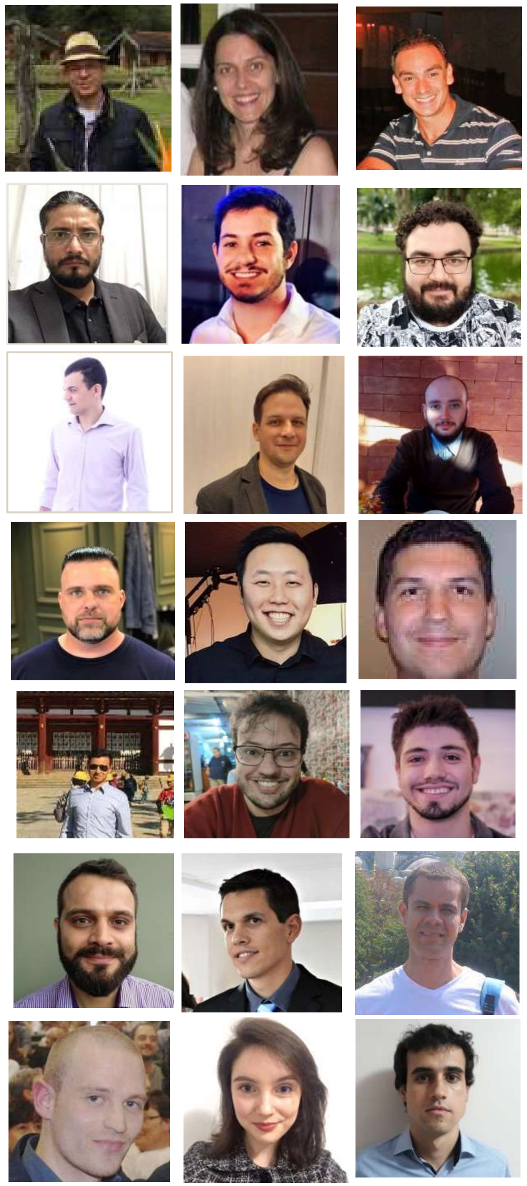
Multi-step ahead meningitis case forecasting based on decomposition and multi-objective optimization methods
M. H. D. M. Ribeiro, V. C. Mariani, L. S. Coelho
Journal of Biomedical Informatics, *https://www.sciencedirect.com/science/article/abs/pii/S1532046420302033*

Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables
R. G. da Silva, M. H. D. M. Ribeiro, V. C. Mariani, L. S. Coelho
Chaos, Solitons & Fractals, *https://www.sciencedirect.com/science/article/abs/pii/S0960077920304252*

Electrical insulators fault forecasting based on a wavelet neuro-fuzzy system
S. F. Stefenon, R. Z. Freire, L.S. Coelho, L. H. Meyer, R. B. Grenogi, W. G. Buratto, A. Nied
Energies, http://dx.doi.org/10.3390/en13020484

Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil
M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, L. S. Coelho
Chaos, Solitons & Fractals, *https://www.sciencedirect.com/science/article/abs/pii/S0960077920302538*

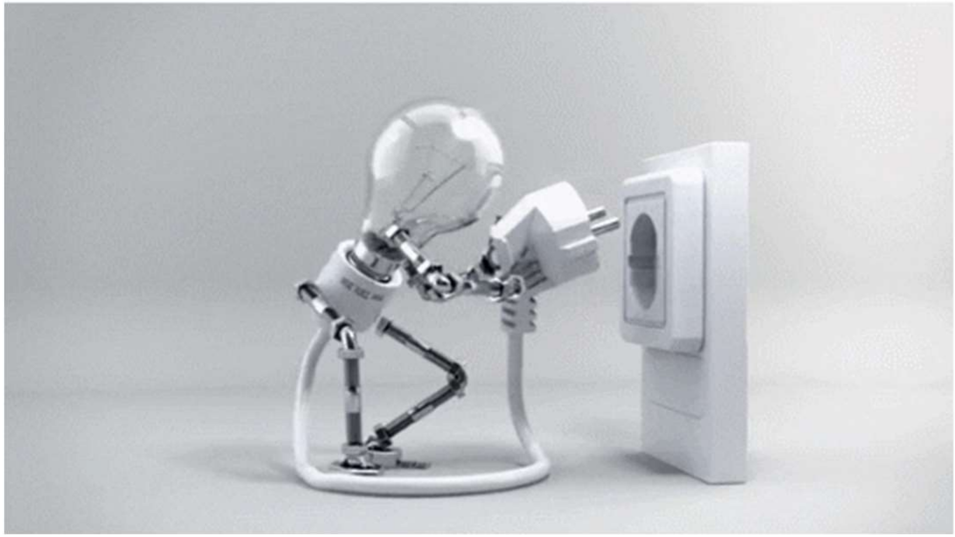
An R library for nonlinear black-box system identification
H. V. Ayala, M. C. Gritti, L. S. Coelho
SoftwareX, *http://dx.doi.org/10.1016/j.softx.2020.100495*



Agenda

- Time series **fundamentals**
- Time series **patterns**
- Time series **decomposition**
- Time series forecasting
  - Time series forecasting using **linear** models
  - Time series forecasting using **nonlinear** models
  - Time series forecasting using **machine learning**
- Performance analysis
- References

Time series fundamentals



https://www.theaidream.com/post/predict-electricity-consumption-using-time-series-analysis

What is a time series?

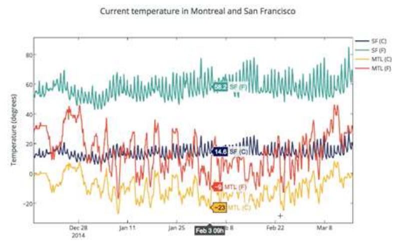
A **time series** is a sequence of data points recorded through time (collected at regular intervals).

The **time series analysis** is concerned with the systematic approaches to **extract** information from *time series*, i.e. from observations ordered in time.

The analysis is capable of giving us useful insights on how a variable changes over time.

Time series forecasting is the use of statistical methods to **forest** future behavior based on historical data.

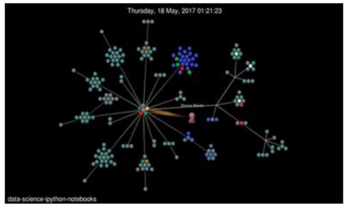
Time series data



Interactive visualizations

analysis
search
manipulate
encode

Data science



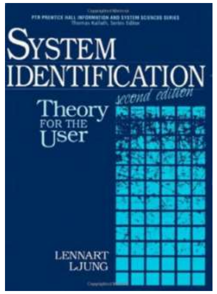
Machine learning

regression
classification
clustering

https://link.springer.com/chapter/10.1007%2F3-540-48022-6_7

https://www.daitan.com/innovation/fundamentals-of-time-series-data-and-forecasting-15e9490b2618/

System identification and time series forecasting

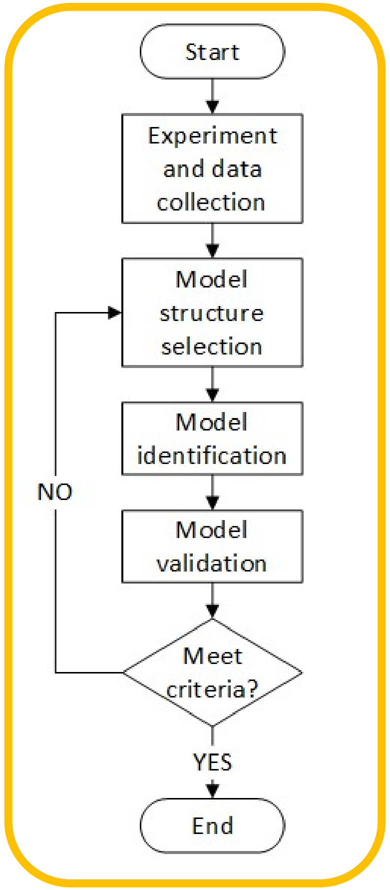


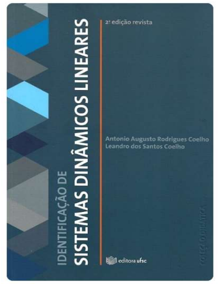
System identification is a **data driven modeling** approach related with obtaining a (mathematical) model from or for a time series.

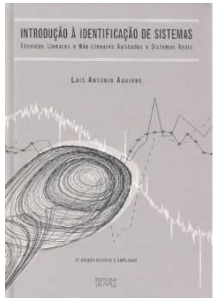
A system identification **procedure** is a set of rules for attaching a system (model) to time series data. In particular such **rules** may be algorithms.

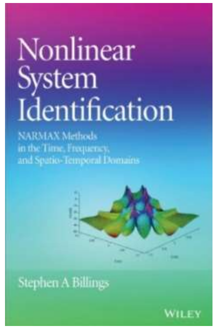
The **core part** of system identification consists of the rules (in particular algorithms) of attaching a "good" model from the model class to the preprocessed data.

Both construction and evaluation of the **quality of such rules** form the center part of the theory and of development of methods of system identification.







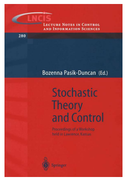


https://link.springer.com/chapter/10.1007%2F3-540-48022-6_7

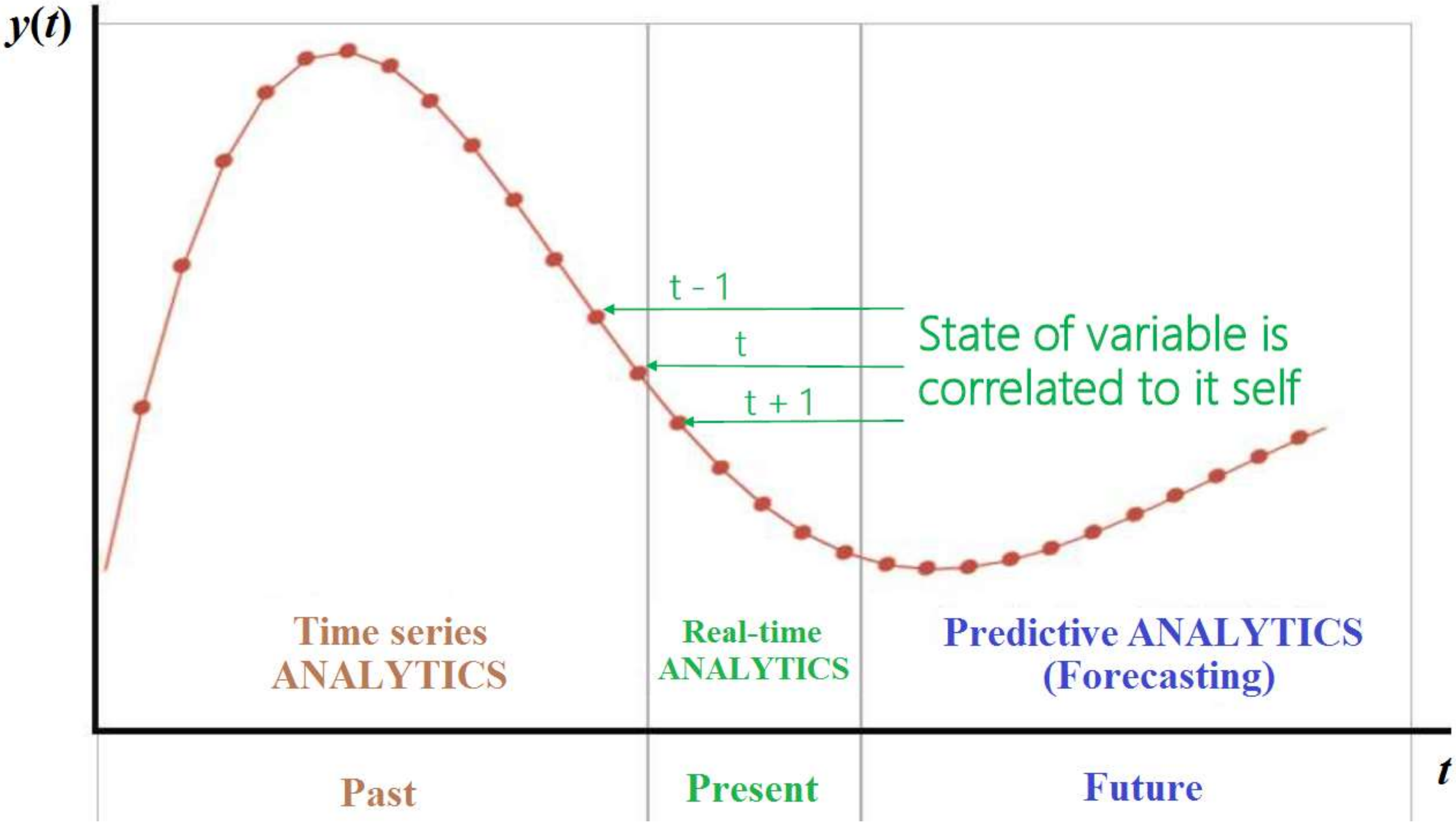
System Identification and Time Series Analysis: Past, Present, and Future

Manfred Deistler

Institut für Stochastik und Wirtschaftsmathematik
Ökonometrie und Systemtheorie



The time series connect past, present and future

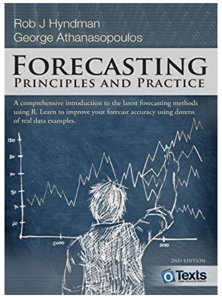


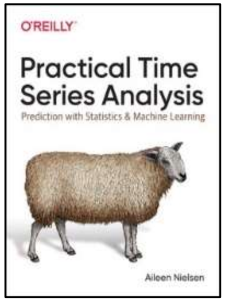
https://www.daitan.com/innovation/fundamentals-of-time-series-data-and-forecasting-15e9490b2618/

Difference between forecast and prediction?

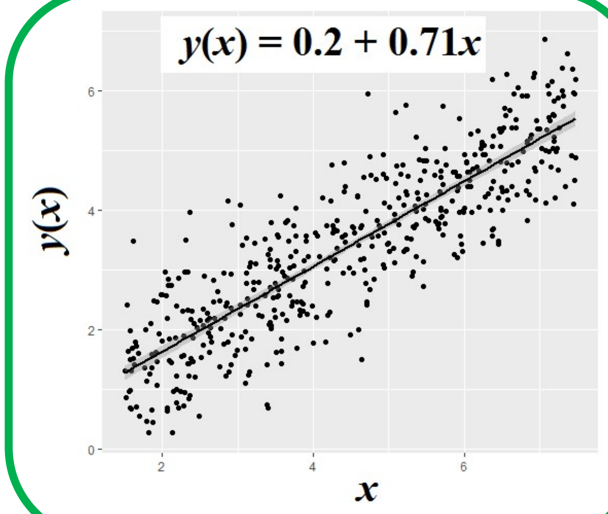
In **time series**, <u>forecasting</u> seems to mean to estimate a future values given past values of a time series (dynamical systems theory).

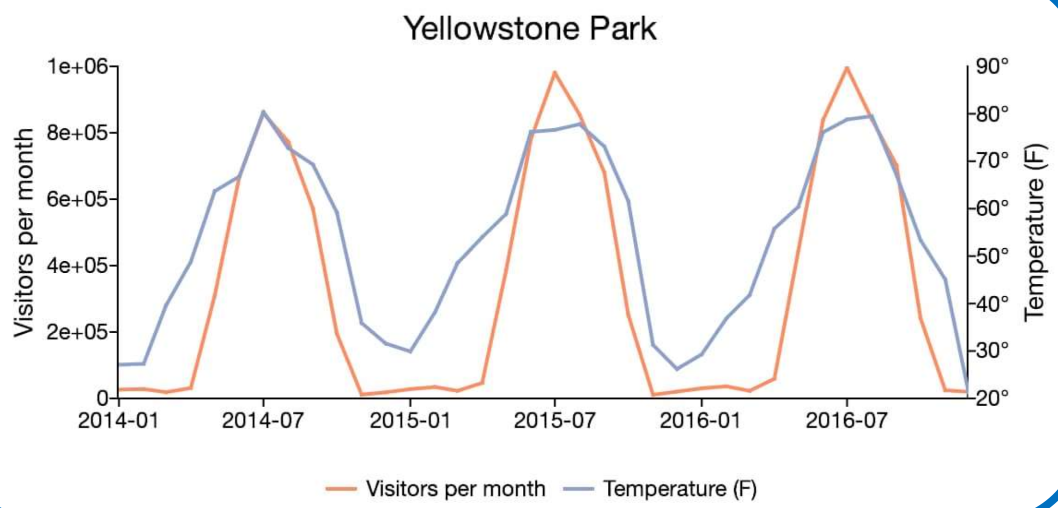
In **regression**, <u>prediction</u> seems to mean to estimate a value whether it is future, current or past with respect to the given data.





https://otexts.com/fpp2/



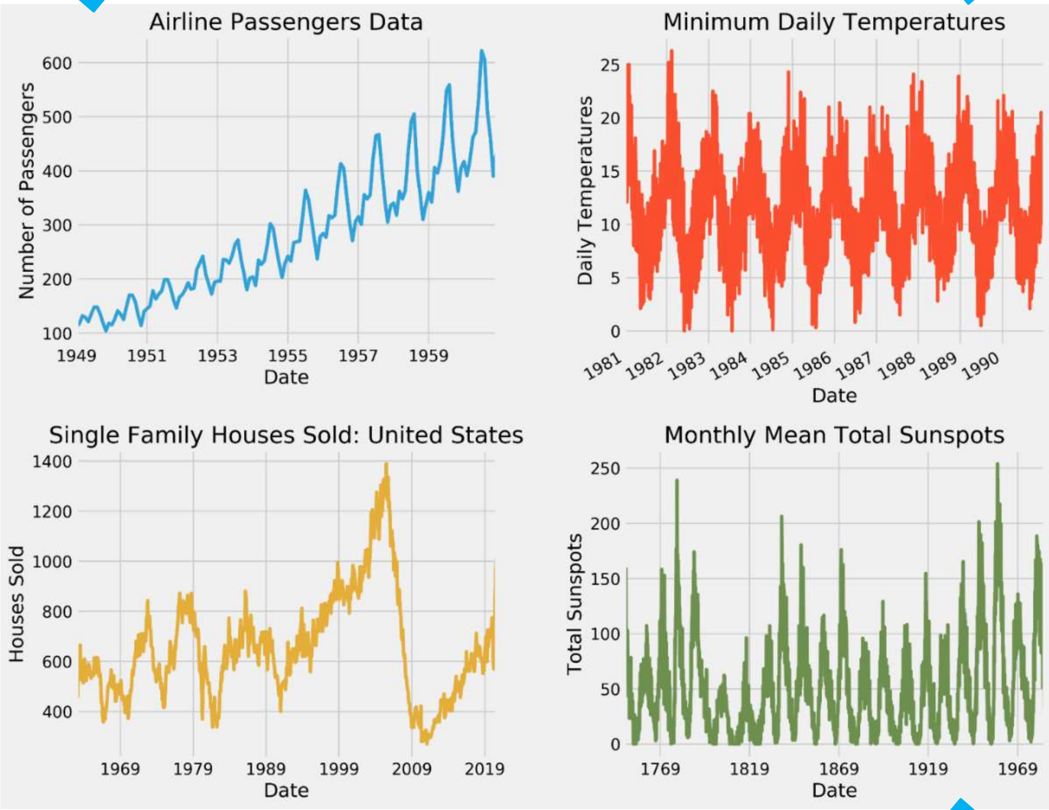


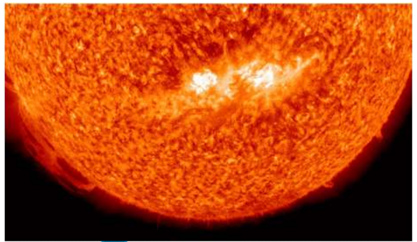
Forecasting would be a subset of prediction (supervised regression). Any time you predict into the future it is a forecast. All forecasts are predictions, but not all predictions are forecasts, as when you would use regression to explain the relationship between two variables.

https://stats.stackexchange.com/questions/65287/difference-between-forecast-and-prediction

Time series

Import libraries and data

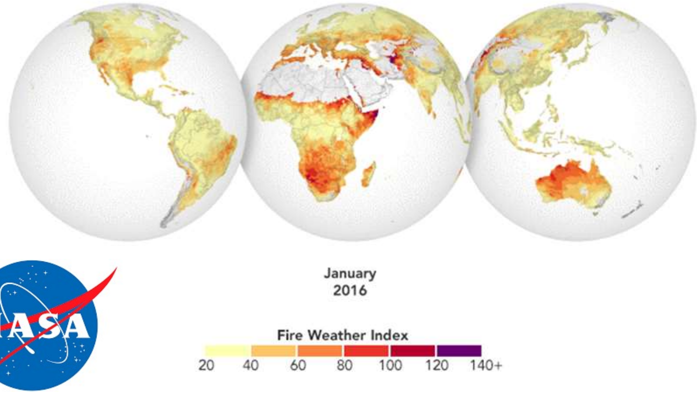




Examples of time series include weather data, rainfall measurements, temperature readings, energy data, heart rate monitoring, brain monitoring, quarterly sales, stock prices, automated stock trading, industry forecasts, and interest rates.

https://medium.com/quantjam/introduction-to-time-series-trend-decomposition-with-python-b54a29f8e038

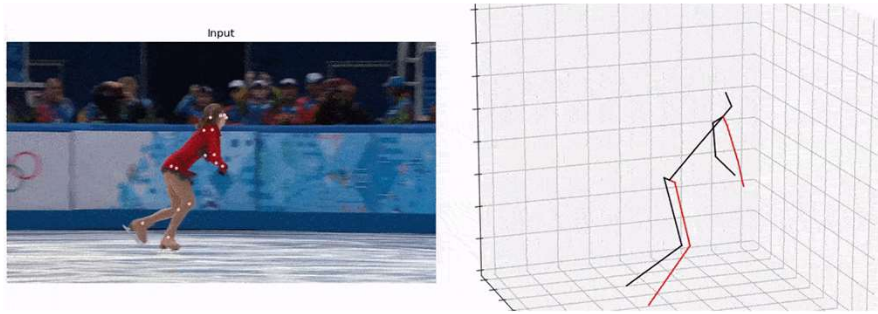
Applications related to time series



Forecasting fire

NASA researchers have created a model that analyzes various weather conditions, including rainfall, to predict the formation and spread of fires.

https://visibleearth.nasa.gov/images/92367/forecasting-fire/92367f



3D human pose estimation

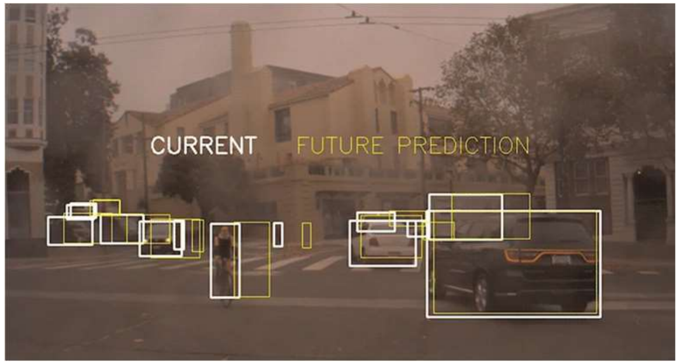
The idea of human pose estimation is detecting locations of people's joints, which form a "skeleton".

https://blog.usejournal.com/3d-human-pose-estimation-ce1259979306

Stock market forecasting

The time series analysis is a powerful tool for forecasting the trend or even future. The trend chart can provide adequate guidance for the investor.

https://www.theaidream.com/post/stock-market-forecasting-using-time-series-analysis



Autonomous vehicles

The key is to analyze temporal information in an image sequence in a way that generates accurate future motion predictions despite the presence of uncertainty and unpredictability.

https://blogs.nvidia.com/blog/2019/05/22/drive-labs-predicting-future-motion/

Types and models of time series analysis

✓ **Classification:** Identifies and assigns categories to the data.
✓ **Curve fitting:** Plots the data along a curve to study the relationships of variables within the data.
✓ **Descriptive analysis:** Identifies patterns in time series data, like trends, cycles, or seasonal variation.
✓ **Explanative analysis:** Attempts to understand the data and the relationships within it, as well as cause and effect.
✓ **Exploratory analysis:** Highlights the main characteristics of the time series data, usually in a visual format.
✓ **Forecasting:** Predicts future data. This type is based on historical trends. It uses the historical data as a model for future data, predicting scenarios that could happen along future plot points.
✓ **Intervention analysis:** Studies how an event can change the data.
✓ **Clustering/segmentation:** Splits the data into segments to show the underlying properties of the source information.

Aspects that come into play when dealing with time series.
Is it **stationary**? Is there a **seasonality**?
Is the target variable **autocorrelated**?

https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775
https://www.tableau.com/learn/articles/time-series-analysis

Exploratory data analysis (EDA) for time series

EDA is used by data scientists to analyze and investigate data sets and summarize their main characteristics.

❶ Importing required libraries and import time-series **data** sets. Clean the data.

❷ Get basic descriptive **statistics** and review the summary of time-series data

❸ Get inference from the **visualization** graphs of time series data

❹ Check the **behavior** of time series data (correlation plots, stationary check)

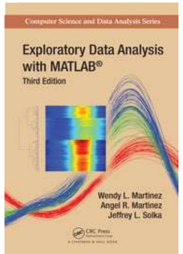
❺ Apply **transformation** functions to convert non-stationary to stationary

1972

1984

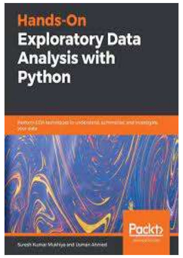
1985

https://www.mathworks.com/



1989-2000

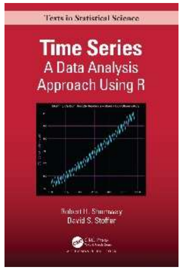
https://www.python.org/



S:1976, R:2001

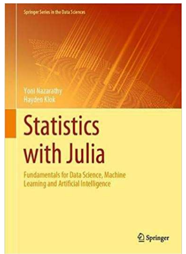
https://cran.r-project.org/

Stanford University



2012

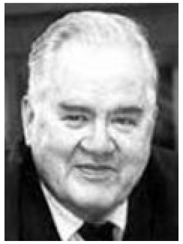
https://julialang.org/



https://stats.stackexchange.com/questions/65287/difference-between-forecast-and-prediction

Exploratory data analysis (EDA) for time series

Statistics, Probability, and **Data science**



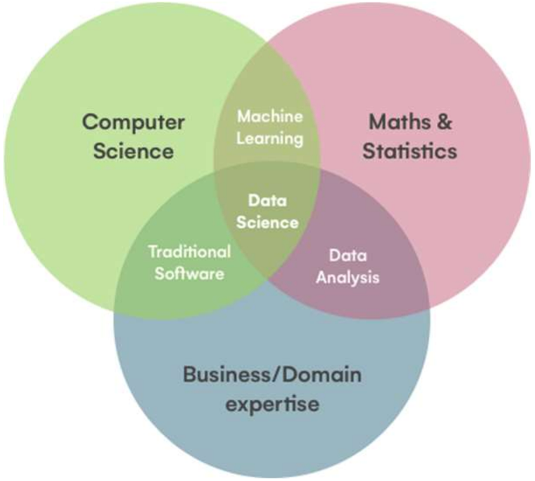
John Tukey
(1915-2000)

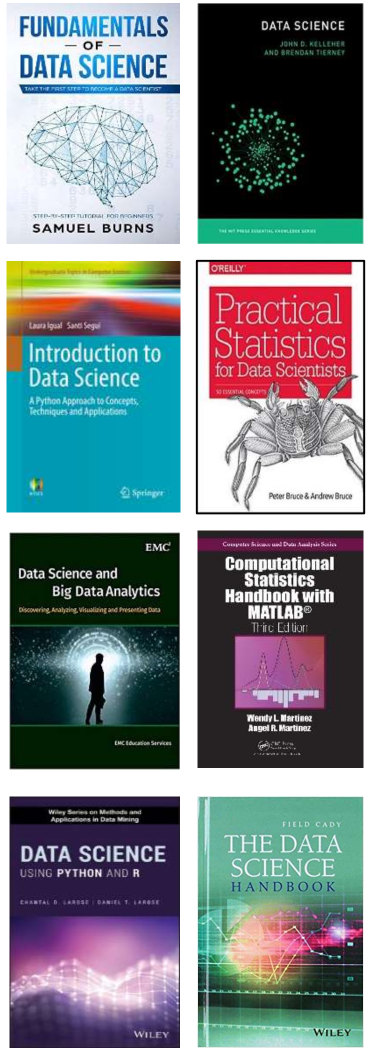
1962

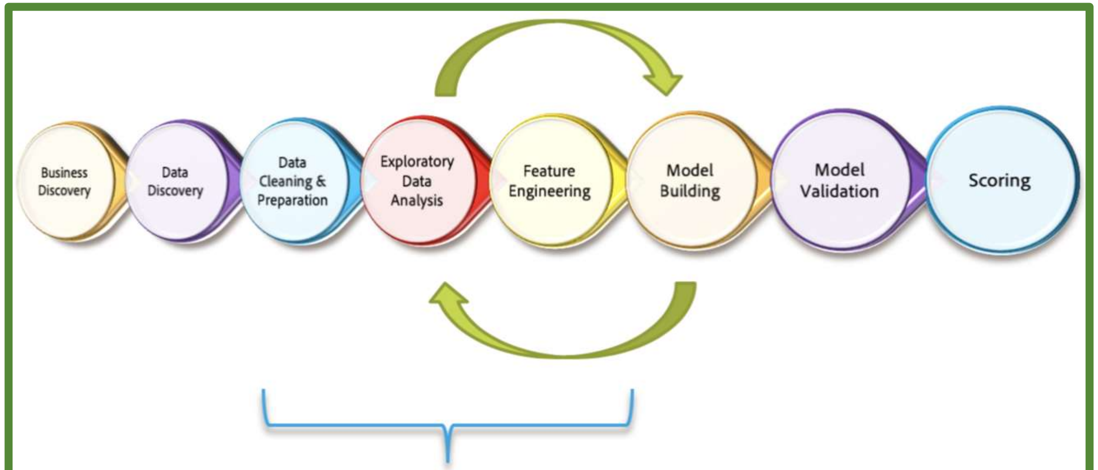
THE FUTURE OF DATA ANALYSIS

By John W. Tukey

PRINCETON UNIVERSITY







60-70% of effort is spent in data preparation out of which 40-50% is spent on EDA

https://towardsdatascience.com/the-data-science-process-a19eb7ebc41b

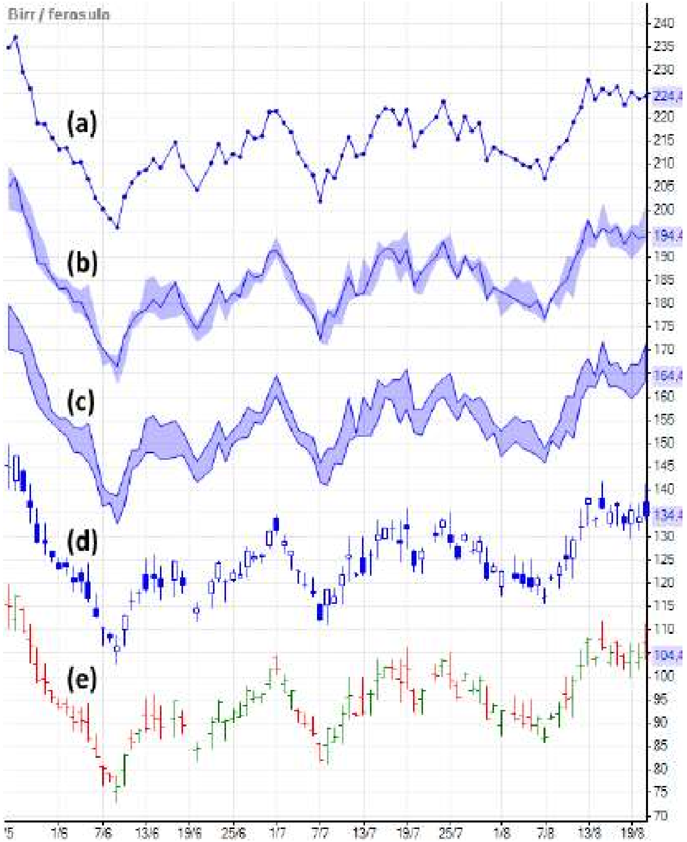
https://medium.com/adobetech/an-introductory-look-at-exploratory-data-analysis-on-adobe-experience-platform-1bfce7501d9a

https://towardsdatascience.com/exploratory-data-analysis-eda-a-practical-guide-and-template-for-structured-data-abfbf3ee3bd9

http://www.mat.ufrgs.br/~viali/estatistica/mat2274/material/textos/2237638.pdf

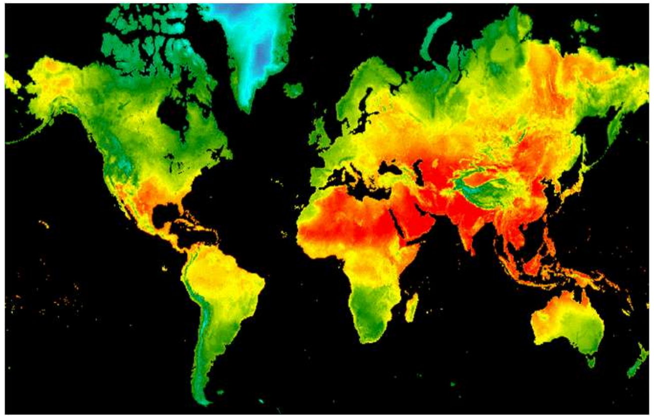
Exploratory data analysis (EDA) for time series

Data visualization



Ethiopian commodity market

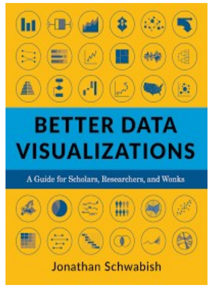
Uber

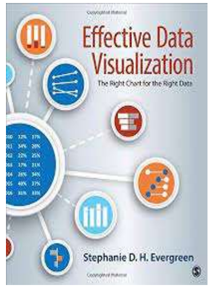


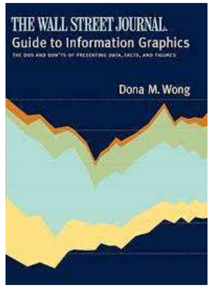
Google Earth Engine

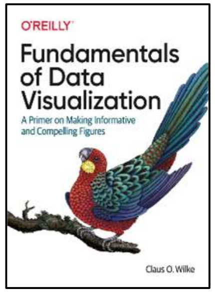
tableau SOFTWARE

Power BI







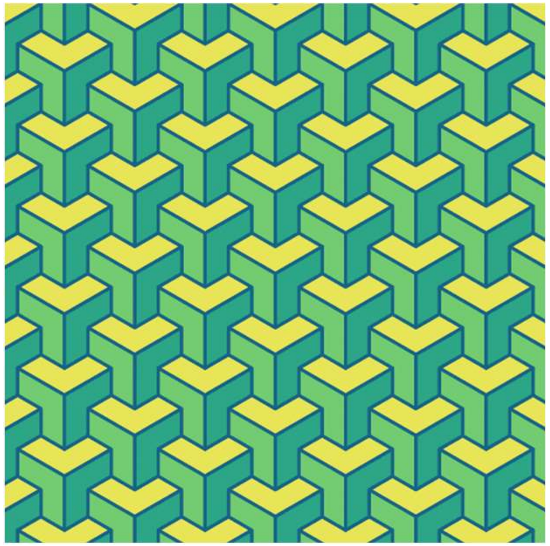


https://developers.google.com/earth-engine/guides/ic_visualization

https://ads9rca.wordpress.com/2017/01/10/anonymous-analytics/

https://www.researchgate.net/publication/268396263_Visualizing_the_Ethiopian_Commodity_Market/figures?lo=1&utm_source=google&utm_medium=organic

Time series patterns



https://giphy.com/explore/isometric-pattern

Leandro dos Santos Coelho

Time series patterns (components)

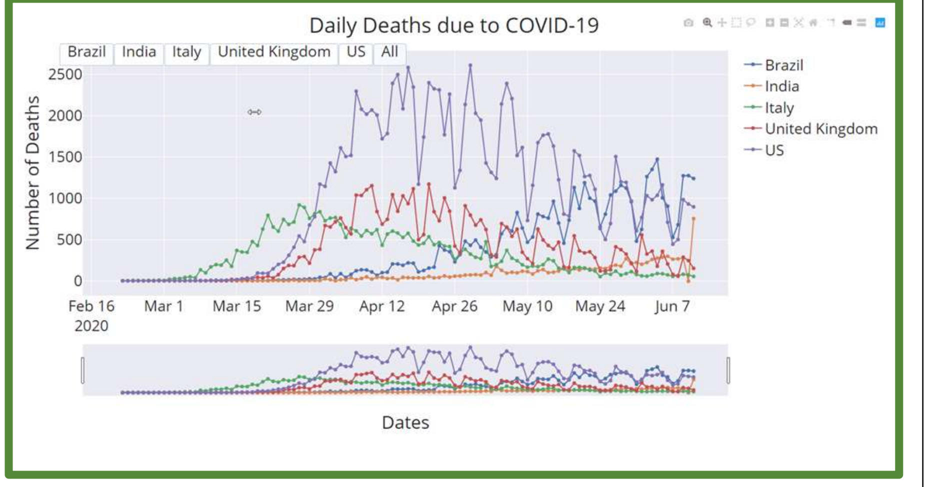
A useful abstraction for selecting forecasting methods is to break a time series down into systematic and unsystematic components.

Systematic: Components of the time series that have **consistency** or **recurrence** and can be **described** and **modeled**.

Non-systematic: Components of the time series that **cannot** be **<u>directly</u>** modeled.

A given time series is thought to consist of **systematic components**

- ☞ level
- ☞ trend
- ☞ seasonality
- ☞ noise (non-systematic)



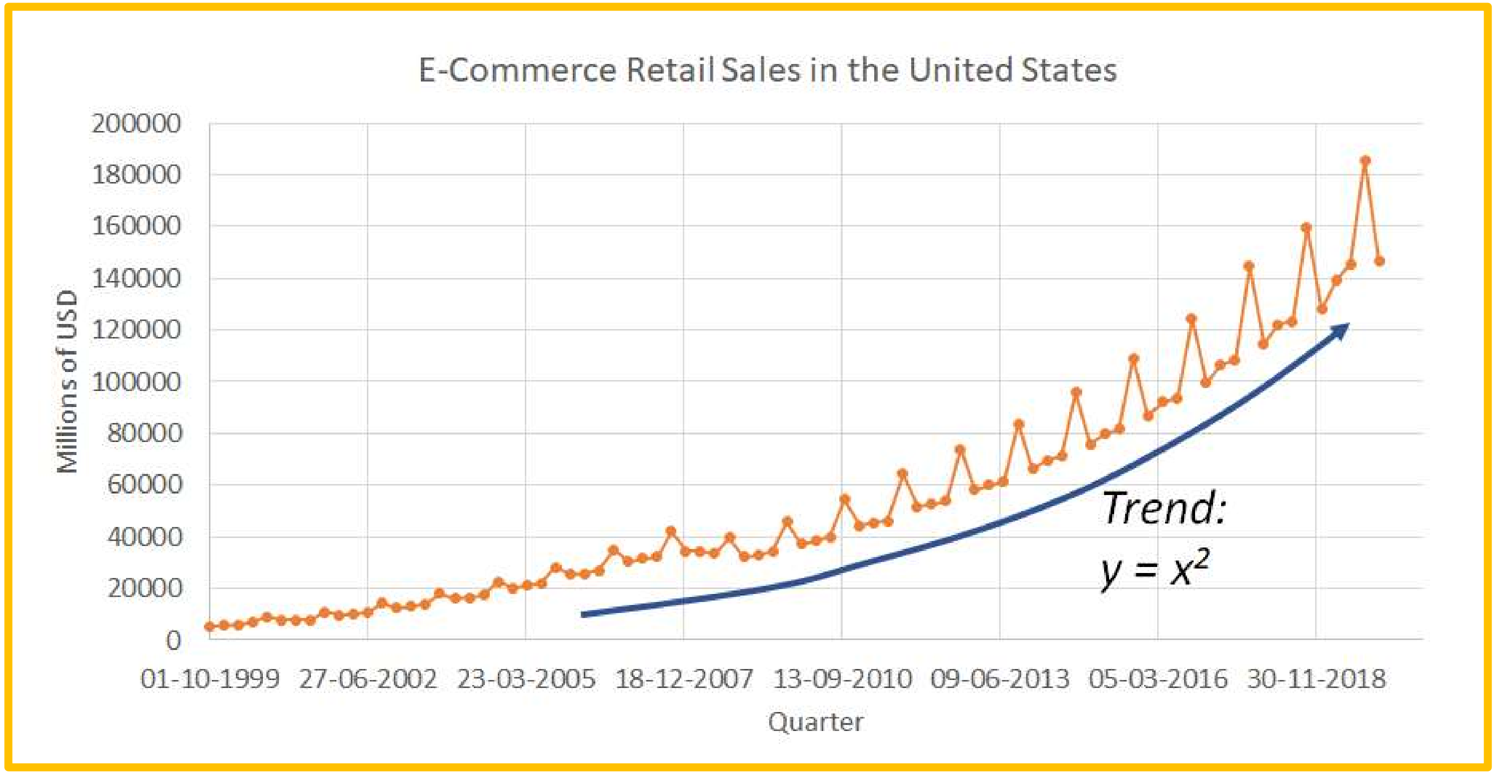
https://towardsdatascience.com/how-to-create-interactive-visualisations-in-python-4af42cf83ba4

https://machinelearningmastery.com/introduction-to-time-series-forecasting-with-python/

Time series patterns

❶ **Trend** exists when there is a long-term **increase** or **decrease** in the data over a period that persists over a long time. The trend can be linear or non-linear.

The **trend** component



https://timeseriesreasoning.com/contents/time-series-decomposition/

Time series patterns

❶ **Trend** exists when there is a long-term increase or decrease in the data. The trend can be linear or non-linear.

❶ Trend

There are **two types** of classification in the trend:

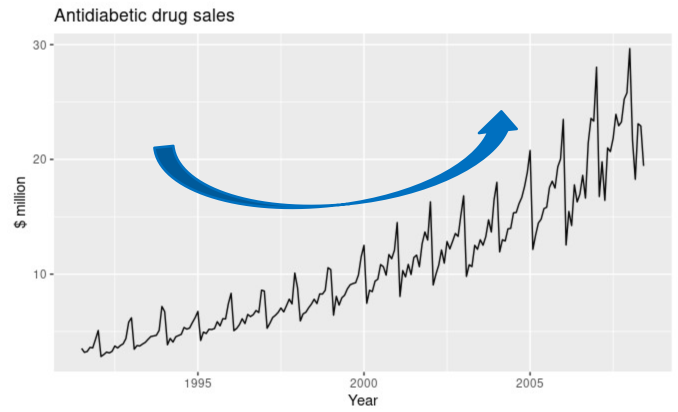
Deterministic trend

These are trends in which the value of the time-dependent variable increases or decreases **consistently**.

Non-deterministic trend (stochastic trend)

These are trends in which the value of the time-dependent variable increases or decreases **inconsistently**.

Monthly sales of **antidiabetic drugs** in Australia



There is also a strong seasonal pattern that increases in size as the level of the series increases.

The sudden drop at the start of each year is caused by a government subsidization scheme that makes it cost-effective for patients to stockpile drugs at the end of the calendar year.

https://sanjayjsw05.medium.com/time-series-analysis-complete-tutorial-for-beginners-part-1-58523000a0a3

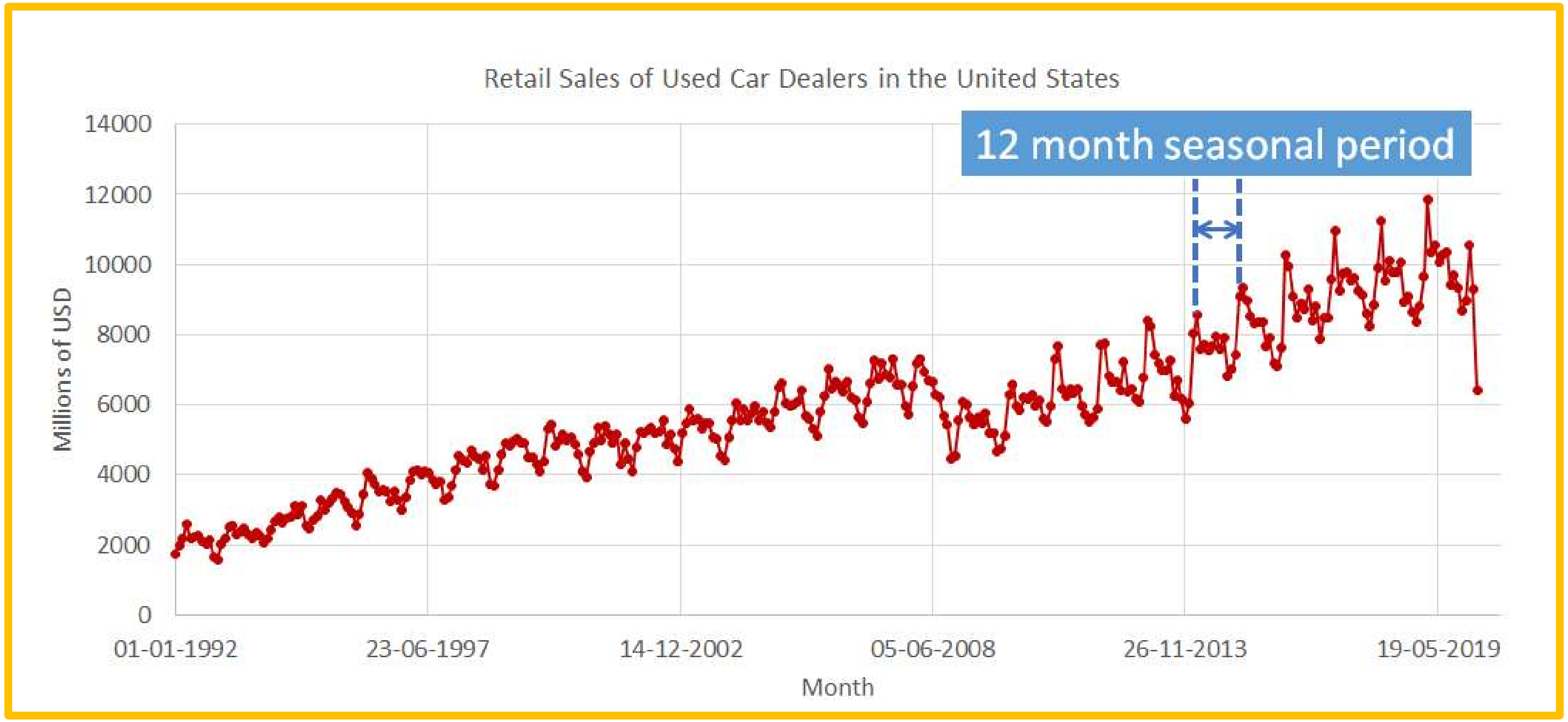
https://otexts.com/fpp2/tspatterns.html

Time series patterns

❷ **Seasonality** (seasonal variations) patterns occur when a time series is affected by **seasonal factors** such as the time of the year or the day of the week.

Seasonality is always of a **fixed** and **known frequency**.

The **seasonal** component



https://timeseriesreasoning.com/contents/time-series-decomposition/

Time series patterns

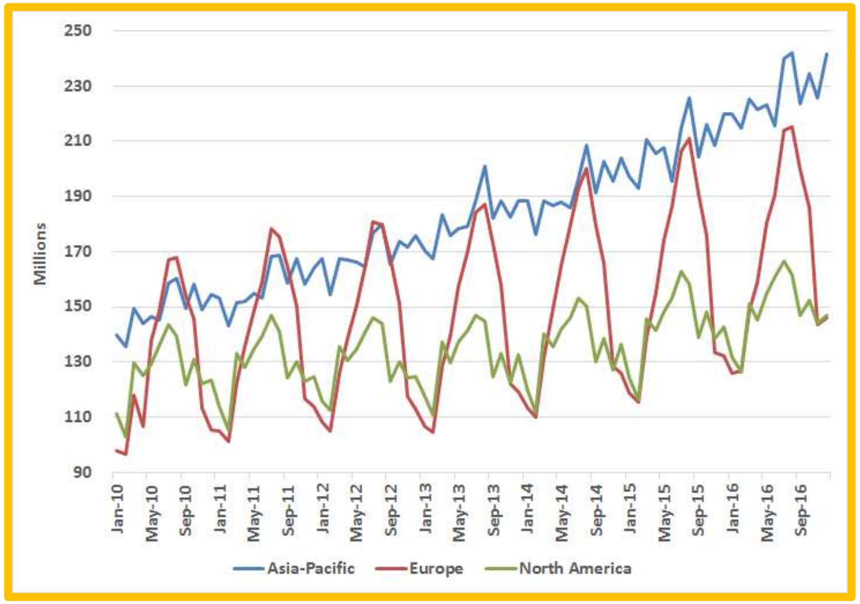
Seasonality (Seasonal variations)

❷ **Seasonality** (seasonal variations) patterns occur when a time series is affected by **seasonal factors** such as the time of the year or the day of the week. Seasonality is always of a **fixed** and **known frequency**.

It is the regular pattern of up and down fluctuations in a time series. It may be a short-term variation occurring due to seasonal factors.

Recurring patterns, e.g., produced by **humans habits**.

Monthly passenger traffic by region (2010-2016)



The main sources of seasonality are

- ☞ Climate
- ☞ Institutions
- ☞ Social habits and practices
- ☞ Calendar

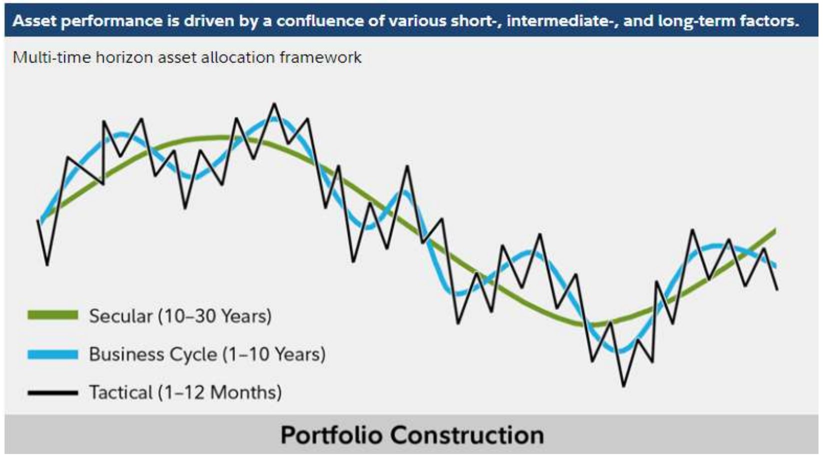
https://blog.aci.aero/airport-markets-and-seasonal-variations/
https://sanjayjsw05.medium.com/time-series-analysis-complete-tutorial-for-beginners-part-1-58523000a0a3
https://otexts.com/fpp2/tspatterns.html

Time series patterns

❸ Cyclicity (Cyclical fluctuations)

It can be defined as a medium-term variation caused by circumstances that repeat in **irregular** intervals. The cycle rises and falls without a fixed frequency. The duration of these fluctuations is usually of at least 2 years.

A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency.



Examples

☞ The great recession from 2007-2008 (business cycle)

☞ The Spanish flu pandemic in 1918

☞ The COVID-19 crisis in 2020-2021

Many people confuse cyclic behavior with seasonal behavior, but they are really quite different. If the fluctuations are **not of a fixed frequency** then they are cyclic. If the frequency is unchanging and associated with some aspect of the calendar, then the pattern is seasonal. In general, the average length of cycles is longer than the length of a seasonal pattern, and the magnitudes of cycles tend to be more variable than the magnitudes of seasonal patterns.

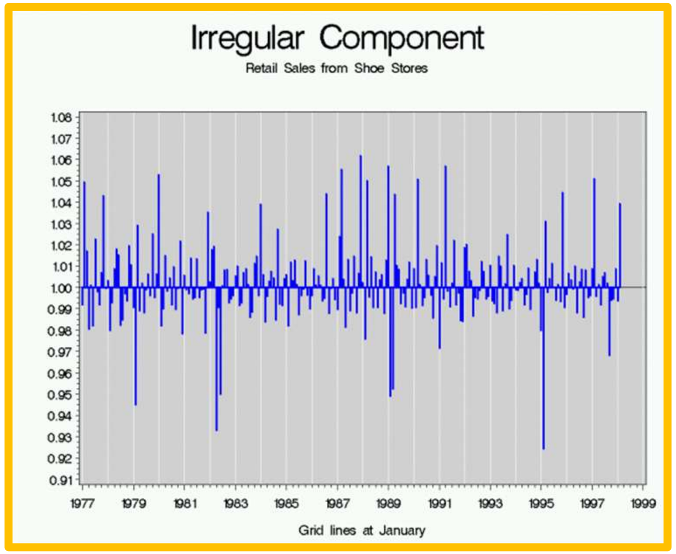
https://www.fidelity.com/viewpoints/investing-ideas/business-cycle-investing
https://sanjayjsw05.medium.com/time-series-analysis-complete-tutorial-for-beginners-part-1-58523000a0a3
https://otexts.com/fpp2/tspatterns.html

Time series patterns

❹ Irregularity (Irregular variations)

It refers to variation occurs due to unpredictable factors and also do not repeat in particular patterns.

Anything not included in the trend-cycle or the seasonal effects (or in estimated trading day or holiday effects). Its values are unpredictable in regards to timing, impact, and duration. It can arise from sampling error, non-sampling error, unseasonable weather, natural disasters, strikes, etc.



Examples

☞ Economical variations caused due to natural disasters.

☞ Irregular component from retail sales from shoe stores.

http://www.catherinechhood.net/safaqgeneral.html

https://sanjayjsw05.medium.com/time-series-analysis-complete-tutorial-for-beginners-part-1-58523000a0a3

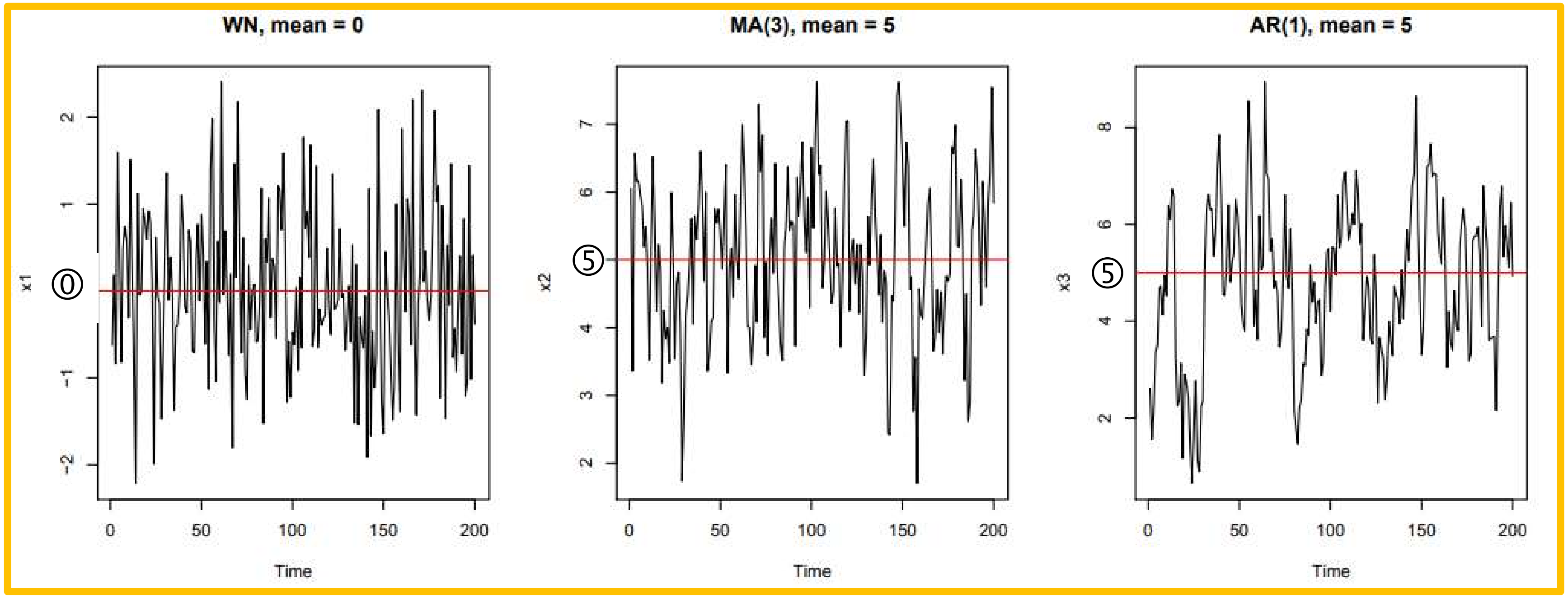
Time series patterns

* Stationary time series and non-stationary time series

A stationary series (process): a random (stochastic) process with a **constant** mean, variance and covariance.

WN: **W**hite **N**oise
MA: **M**oving **A**verage
AR: **A**uto**R**egressive

Examples of **stationary** time series.

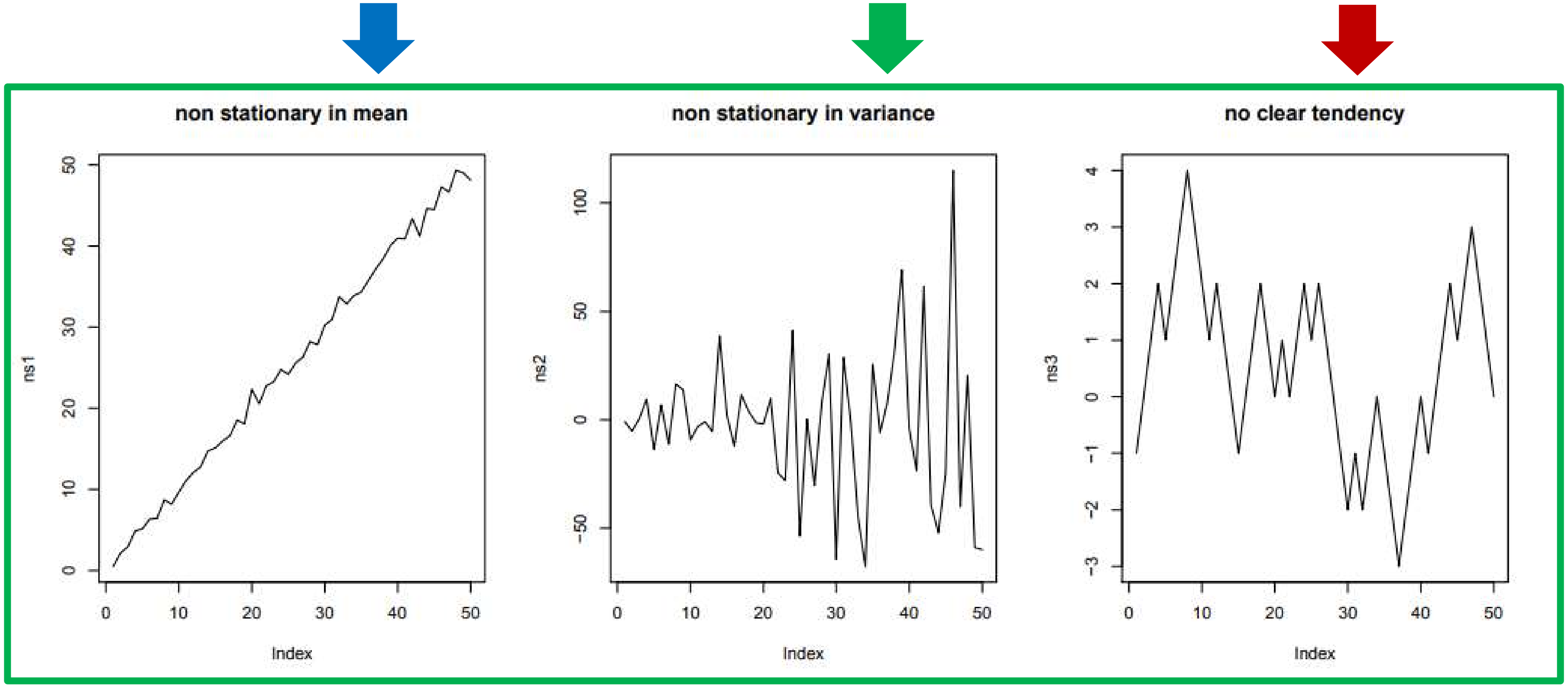


http://web.vu.lt/mif/a.buteikis/category/practical-econometrics/practical-econometrics-ii-ii/

Time series patterns

* Stationary time series and non-stationary time series

Examples of non-stationary time series.



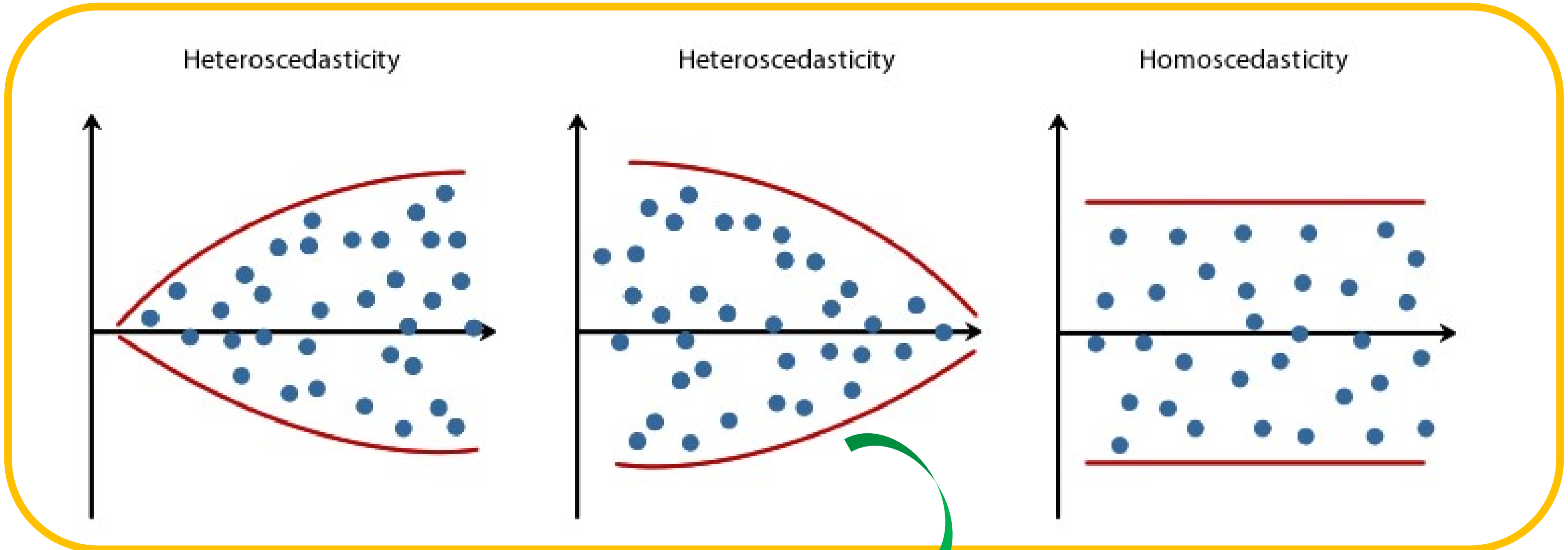
http://web.vu.lt/mif/a.buteikis/category/practical-econometrics/practical-econometrics-ii-ii/

Time series patterns

Stationary time series and non-stationary time series

Heteroscedasticity is what you have in your data when the *conditional* ***variance*** in your data is **not** constant.

Conditional variance is the variability that you see in the dependent variable y for each value of the explanatory variables $\boldsymbol{X}$, or each value of time period $\boldsymbol{t}$ (in case of time series data).



GARCH (**G**eneralized **A**uto**R**egressive **C**onditional **H**eteroskedasticity) model

https://www.albert.io/blog/key-assumptions-of-ols-econometrics-review/

https://towardsdatascience.com/heteroscedasticity-is-nothing-to-be-afraid-of-730dd3f7ca1f

Tests for identifying stationarity in time series

statsmodels

✓ **DF (Dickey-Fuller) test**

`from statsmodels.tsa.stattools import adfuller`

This is one of the statistical tests for checking the presence of **stationarity**.

Null hypothesis: Time series is non-stationary.

Alternative hypothesis: Time series is stationary.

https://la.mathworks.com/help/econ/assess-stationarity-of-a-time-series.html

https://cran.r-project.org/web/packages/tseries/tseries.pdf

✓ **ADF (Augmented Dickey-Fuller) Test**

It is used for checking the presence of a **unit root** in time series (AR model).

Null hypothesis: The series has a unit root which means series is non-stationary.

Alternative hypothesis: The series has no unit root which means series is stationary.

✓ **KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Test**

`from statsmodels.tsa.stattools import kpss`

It is another statistical test used for checking the presence of **stationarity**.

Null hypothesis: The series stationary.

Alternative hypothesis: The series has a unit root which means series is non-stationary.

It has to be noted that the null and alternative hypothesis is the opposite for ADF and KPSS test.

Dickey, D.A. and W.A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association, vol. 74, no. 166, pp. 427-431. *http://www.jstor.org/pss/2286348*

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54: 159-178. *https://www.sciencedirect.com/science/article/abs/pii/030440769290104Y*

https://machinelearningmastery.com/time-series-data-stationary-python/

https://nwfsc-timeseries.github.io/atsa-labs/sec-boxjenkins-aug-dickey-fuller.html

https://sanjayjsw05.medium.com/time-series-analysis-complete-tutorial-for-beginners-part-1-58523000a0a3

https://towardsdatascience.com/detecting-stationarity-in-time-series-data-d29e0a21e638

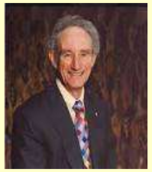
Time series patterns

* Complex patterns: Chaos theory

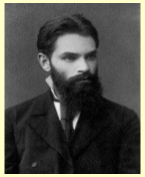
Chaos theory is related for a dynamical system whose apparently random states of disorder and irregularities are actually governed by underlying patterns and deterministic laws that are **highly sensitive** to initial conditions.

Chaos theory patterns:

- ☞ interconnectedness
- ☞ constant feedback loops
- ☞ repetition
- ☞ self-similarity
- ☞ fractals
- ☞ self-organization



Lord Robert May
(1936-2020)



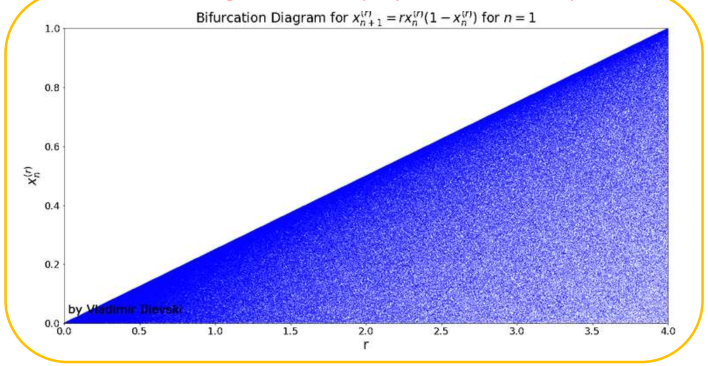
Aleksandr Mikhailovich Lyapunov
(1857-1918)

To characterize the attractor:

✓Dimension quantifies the self-similarity of a geometrical object (Higuchi, Katz, Petrosian,...)

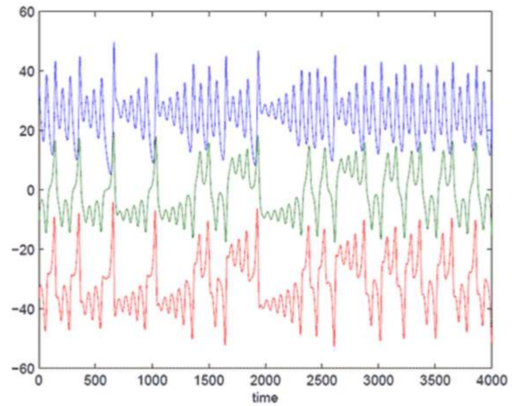
✓ A positive maximal **Lyapunov exponent** is a strong signature of chaos

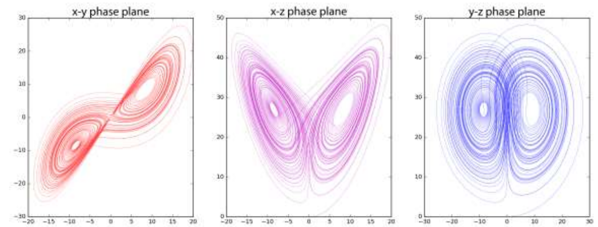
Logistic map (1838, 1976)



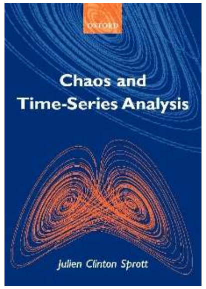
Lorenz attractor (1963)

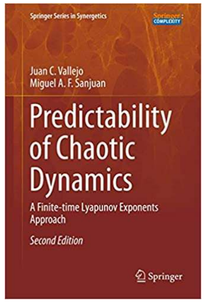
$$\frac{dx}{dt} = \sigma(y - x) \qquad \frac{dy}{dt} = x(\rho - z) - y \qquad \frac{dz}{dt} = xy - \beta z$$

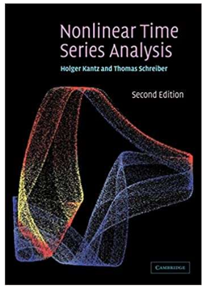


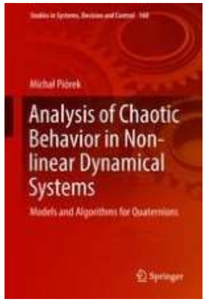


weather prediction









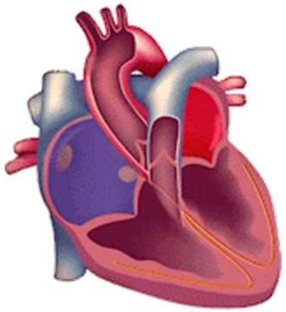
http://www.bodybannao.com/doctor_template/weichihhu/clean/index.html
https://en.wikipedia.org/wiki/Electrocardiography
https://www.aedsuperstore.com/resources/ekg-quick-reference-guide/

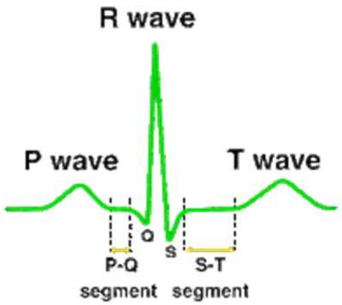
Time series patterns

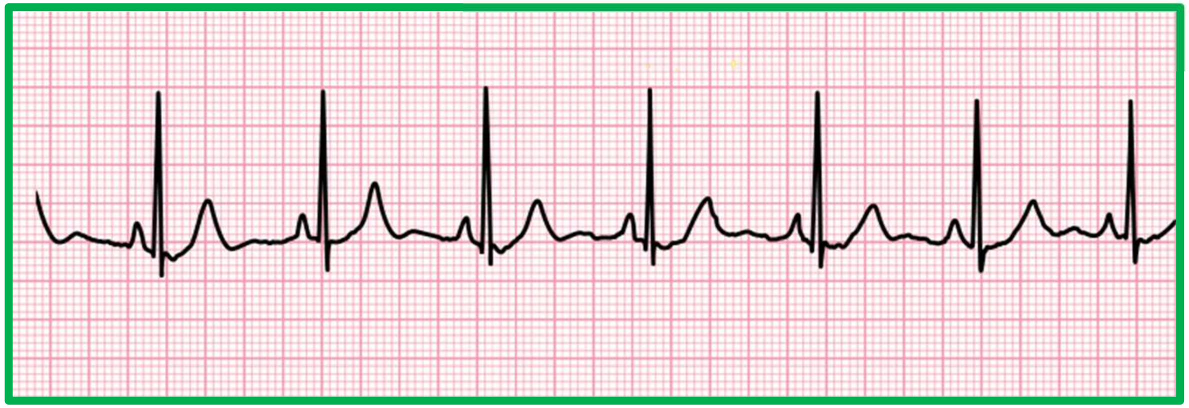
* Complex patterns

Electrocardiogram (ECG)

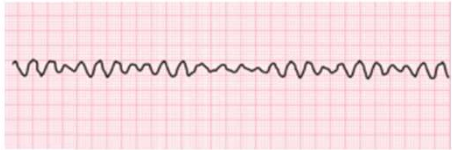
anomaly detection



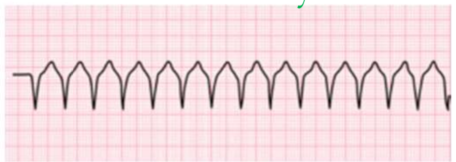




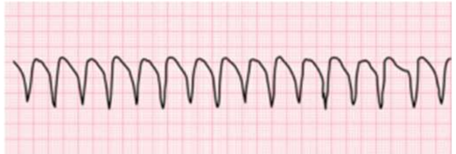
ventricular fibrillation



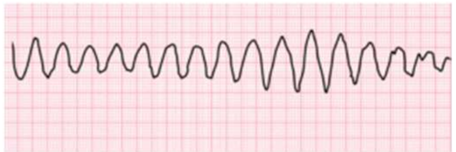
ventricular tachycardia



ventricular tachycardia monomorphic



ventricular tachycardia polymorphic



Wolff-Parkinson-White syndrome

wandering atrial pacemaker

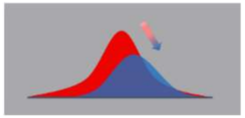
ventricular tachycardia torsade de Pointes

http://www.bodybannao.com/doctor_template/weichihhu/clean/index.html
https://en.wikipedia.org/wiki/Electrocardiography
https://www.aedsuperstore.com/resources/ekg-quick-reference-guide/

P wave: represents the depolarization of the atria
QRS complex: represents the depolarization of the ventricles
T wave: represents the repolarization of the ventricles.

Time series patterns

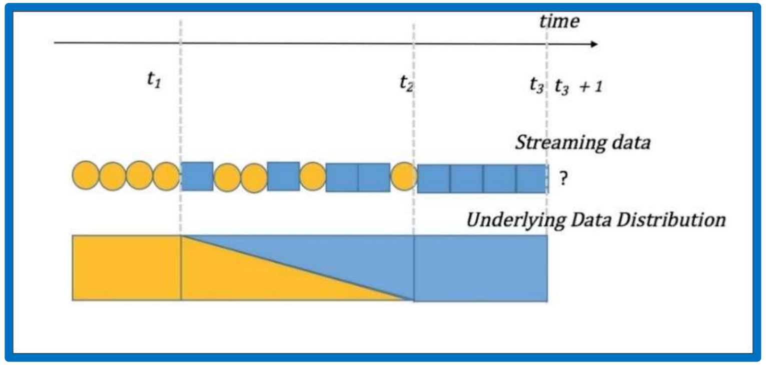
Concept drift

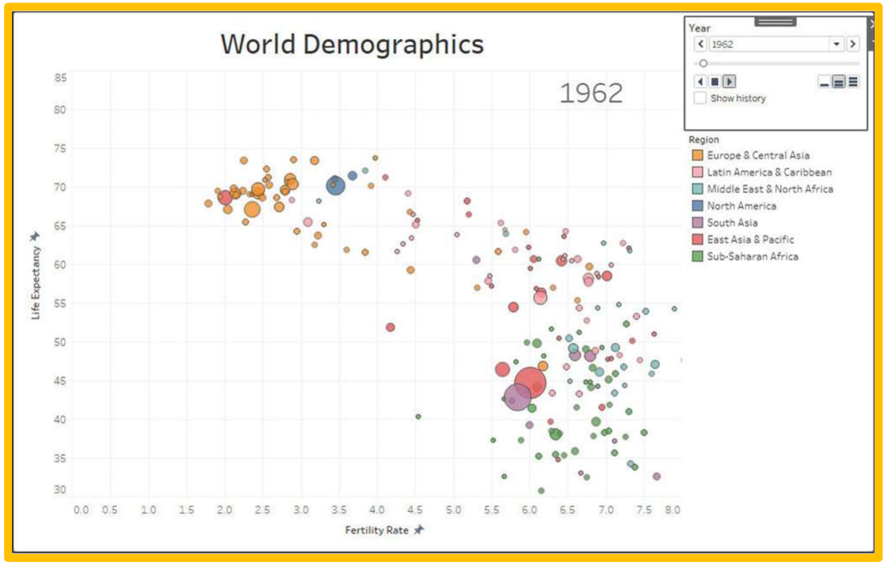


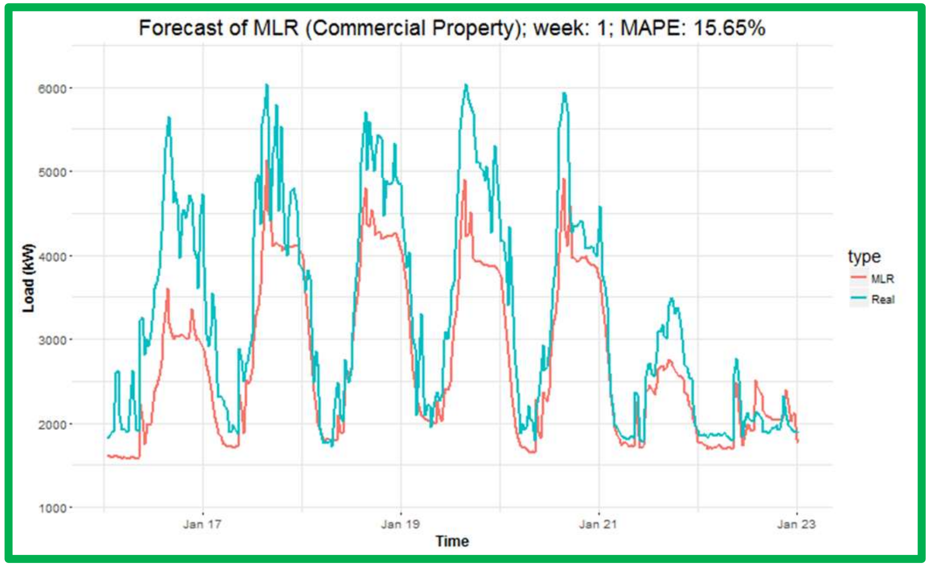
Heraclitus, the Greek philosopher said, "**Change** is the only **constant** in life."

Data can **change** over time. This can result in **poor and degrading predictive performance** in predictive models that **assume a static relationship** between input and output variables.

The term **concept** refers to the quantity to be predicted.



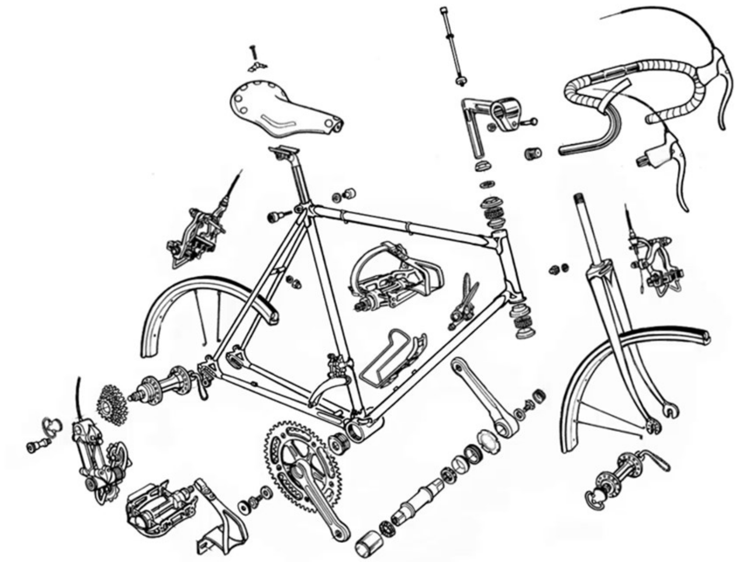




https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.pinterest.com%2Fpin%2F684758318328613568%2F&psig=AOvVaw0YO-oEoURdsA-TrQcUXVgy&ust=1629716898891000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCODqxtq-xPICFQAAAAAdAAAAABAy

https://www.r-bloggers.com/2016/12/forecast-double-seasonal-time-series-with-multiple-linear-regression-in-r/

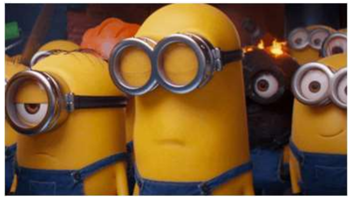
Time series decomposition



Time series decomposition involves thinking of a series as a **combination** of level, trend, seasonality, and noise **components**.

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwifflegif.com%2Ftags%2F379422-parts-gifs&psig=AOvVaw1g0ww7pIXG4X0Vz9EEJZDD&ust=1629592376818000&source=images&cd=vfe&ved=0CAwQjhxqFwoTCNCUqOjuwPICFQAAAAAdAAAAABAD

Some terminologies



Decomposition

A proper time series analysis demands the data to be in a stationary way but we rarely get data with stationary characteristics. Hence, the non-stationary time series converted to a stationary series by removing its trend and seasonality. This process of **removing trend and seasonality is** called **Time series decomposition**.

Differencing

Differencing is a decomposition process through which trend and seasonality are eliminated. Here, we usually take the difference of observation with particular instant with previous instant.

Transformation

Level: The average value in the series.

Transformations can help to stabilize the variance of a time series. Transformation is the easiest way to remove trends from a time series by converting the data into different scales using various operations like logarithms, square roots, etc. However, it is rarely implemented over differencing due to the possibility of a loss of information.

https://www.google.com/url?sa=i&url=https%3A%2F%2Fgiphy.com%2Fexplore%2Fgif-campaa-2&psig=AOvVaw32IdiNPef9Ca_IjvfbAurA&ust=1629630608333000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCND6yJz9wfICFQAAAAAdAAAAABAR

https://sanjayjsw05.medium.com/time-series-analysis-complete-tutorial-for-beginners-part-1-58523000a0a3

Pattern decomposition

There are **two main methods** among other methods to decompose seasonality:

- ✓ linear models with additive or multiplicative models, and
- ✓ season-trend decomposition using **LOESS**

Local regression: LOESS (locally estimated scatterplot smoothing)

The seasonal_decompose model uses moving averages to decompose seasonality trends.

The **additive** models has following format:

Time Series = Level + Trend + Seasonality + Residual

linear

The **multiplicative** time series model has the following format:

Time Series = Level * Trend * Seasonality * Residual

nonlinear

https://github.com/anejad/QuantJam_Notebooks

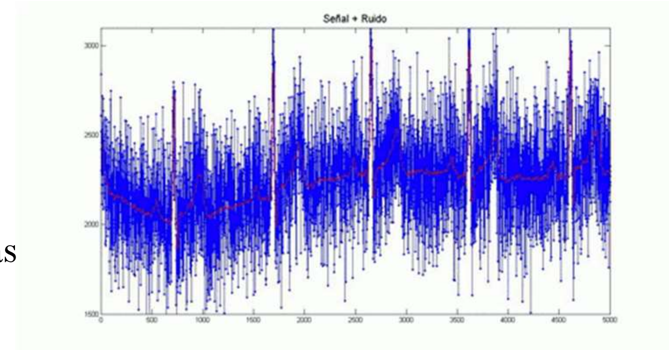
https://medium.com/quantjam/introduction-to-time-series-trend-decomposition-with-python-b54a29f8e038

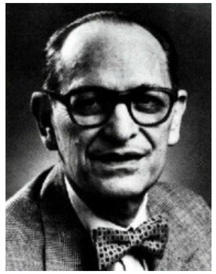
https://medium.com/quantjam/introduction-to-time-series-trend-decomposition-with-python-b54a29f8e038

LOESS

LOcally Estimated Scatterplot Smoothing

LOESS (1979) which is commonly referred to as **Savitzky-Golay smoothing filter (1964).**





Abraham Savitzky
(1919-1999)

Marcel Jules Edouard Golay
(1902-1989)

A Savitzky-Golay filter is a **digital filter** that can be applied to a set of digital data points for the purpose of **smoothing** the data, that is, to increase the precision of the data without distorting the signal tendency. This is achieved, in a process known as convolution, by fitting successive sub-sets of adjacent data points with a low-degree polynomial by the method of linear least squares.

❶ Select a **window** (say, five points) around that point
❷ Fit a **polynomial** to the points in the selected window
❸ **Replace** the data point in question with the corresponding value of the **fitted polynomial**.

William S. Cleveland rediscovered the method in **1979** and gave it a distinct name: **LOWESS** (**LO**cally **W**eighted **S**catterplot **S**moothing)

2001: Data science lecture

William Swain Cleveland
(1943 -)

Savitzky, A., Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry. 36(8): 1627-1639. *https://pubs.acs.org/doi/10.1021/ac60214a047*

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, Journal of the American Statistical Association, 74(368): 829-836. *https://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10481038*

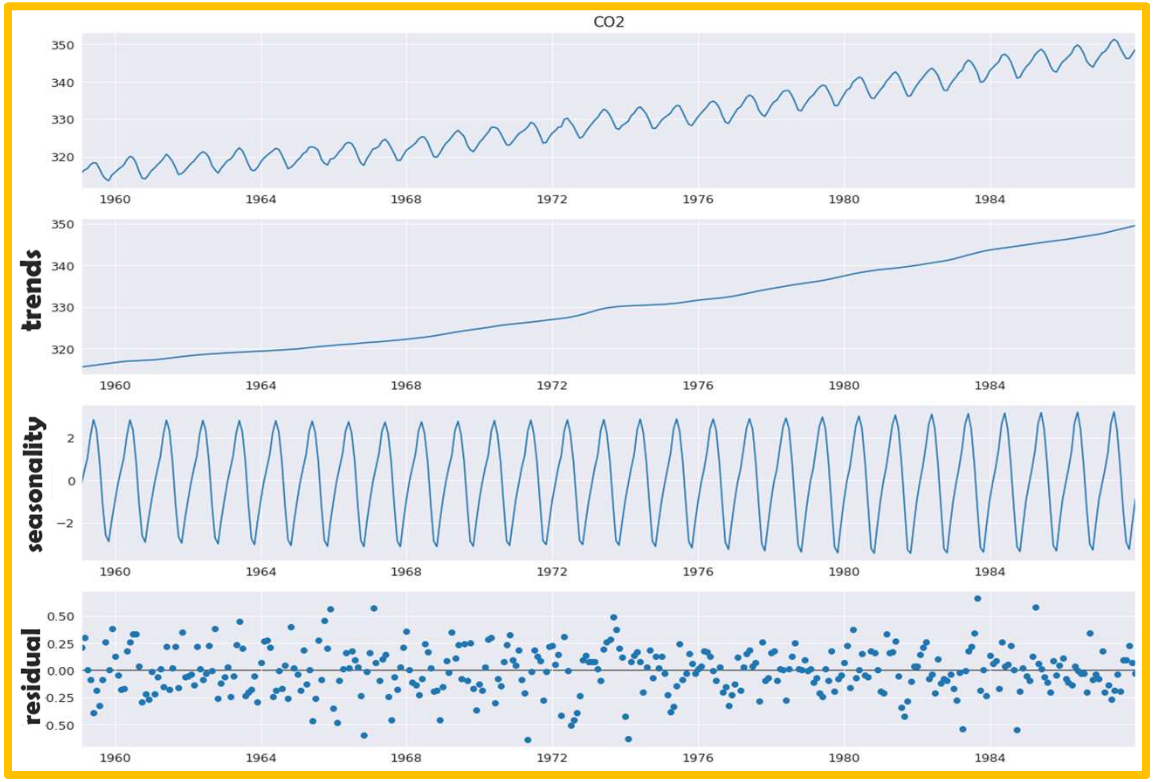
https://github.com/anejad/QuantJam_Notebooks

https://medium.com/quantjam/introduction-to-time-series-trend-decomposition-with-python-b54a29f8e038

Some important terminologies

Decomposition

STL (**S**easonal and **T**rend decomposition using **L**oess)

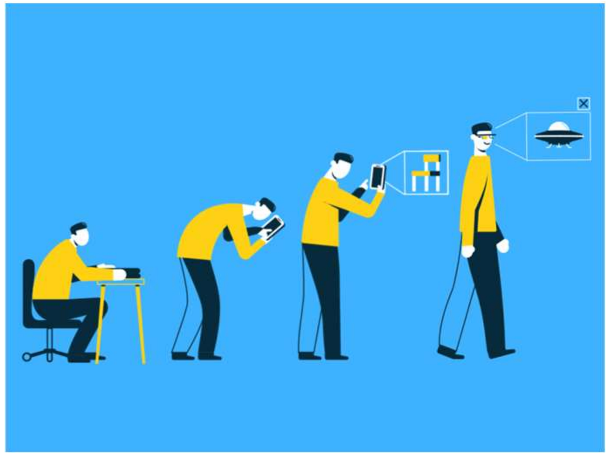


Cleveland et al. (1990) uses CO_2 data. This monthly data (January 1959 to December 1987) has a clear trend and seasonality across the sample.

y = trends + seasonality + remainder (residual, error)

$$y_t = f(S_t, T_t, R_t)$$

where $y_t =$ data at period t
$T_t =$ trend-cycle component at period t
$S_t =$ seasonal component at period t
$R_t =$ remainder component at period t



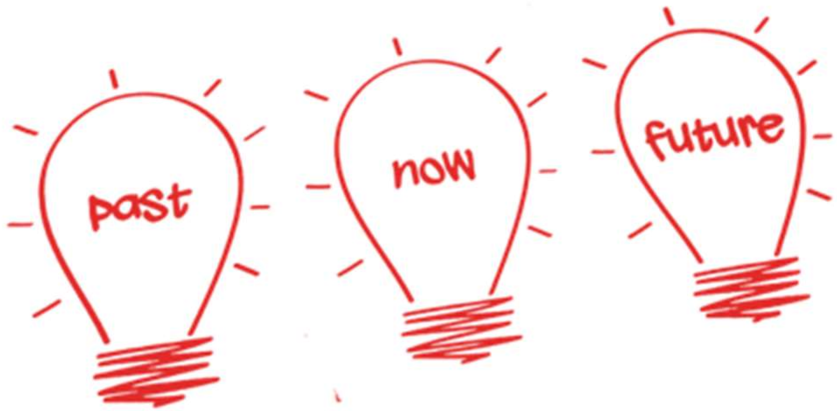
Time series must
- Be seasonal
- Have at least two full periods

https://math.unm.edu/~lil/Stat581/6-decomposition.pdf

https://otexts.com/fpp2/components.html

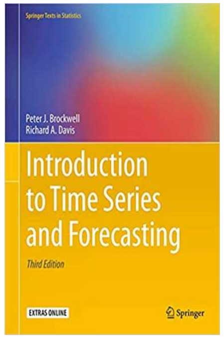
Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1), pp. 3-33. *http://bit.ly/stl1990*

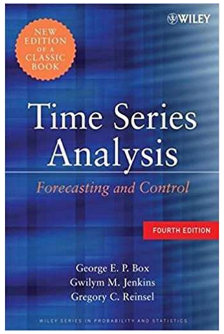
Time series forecasting

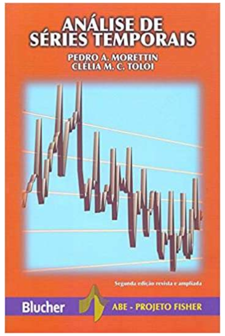


https://giphy.com/explore/isometric-pattern

Classical (statistical) models







✓ Simple exponential smoothing

$$y(t) = \alpha \,.\, x(t) + (1- \alpha) \,.\, y(t\text{-}1)$$

✓ Double exponential smoothing or Holt linear method

$$y(t) = \alpha \,.\, x(t) + (1- \alpha) \,.\, \{ y(t\text{-}1) + b(t\text{-}1)\}$$
$$b(t) = \beta \,.\, \{y(t) - y(t\text{-}1)\} + (1- \beta) \,.\, b(t\text{-}1)$$

✓ Triple exponential smoothing or Holt-Winters method

$$y(t+h) = l(t) + h.b(t) + s(t+h\text{-}L[k+1])\}$$
$$l(t) = \alpha \,.\, \{ x(t) - s(t\text{-}L)\} + (1- \alpha) \,.\, \{ l(t\text{-}1) + b(t\text{-}1)\}$$
$$b(t) = \beta \,.\, \{ l(t) - l(t\text{-}1)\} + (1- \beta) \,.\, b(t\text{-}1)$$
$$s(t) = \gamma. \{ x(t) - l(t\text{-}1) - b(t\text{-}1)\} + (1- \gamma) \,.\, \{s(t\text{-}L) \}$$

additive seasonality

α , β , γ tuning

Additive time series model
$y(t) = T(t) + S(t) + R(t)$
Multiplicative time series model
$y(t) = T(t) \cdot S(t) \cdot R(t)$
With the help of logarithms it is possible to pass from the multiplicative model to the additive model.

Brown, R. G. (1959). Statistical forecasting for inventory control. McGraw/Hill, NY, USA.
Holt, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. *https://www.sciencedirect.com/science/article/abs/pii/S0169207003001134?via%3Dihub*
Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342. *https://pubsonline.informs.org/doi/abs/10.1287/mnsc.6.3.324*

① Overall smoothing
② Trend smoothing
③ Seasonality smoothing

Time series forecasting using linear models

Forecasting time series using linear models with eXogenous inputs (X)

1952

Z-Transform (correlation of x[n] with z^n):

$$X(z) = \sum_{n=0}^{\infty} x[n] z^{-n}$$

Example: ARMAX models

AR X MA

$$\boxed{\mathbf{A}(z).y(t) = \boldsymbol{B}(z).u(t\text{-}nk) + \boldsymbol{C}(z).e(t)}$$

z (or *q*)

Example of order tuning and delay:

$na = 2$ ***A*** polynomial order

$nb = 2$ ***B*** polynomial order

$nc = 2$ ***C*** polynomial order

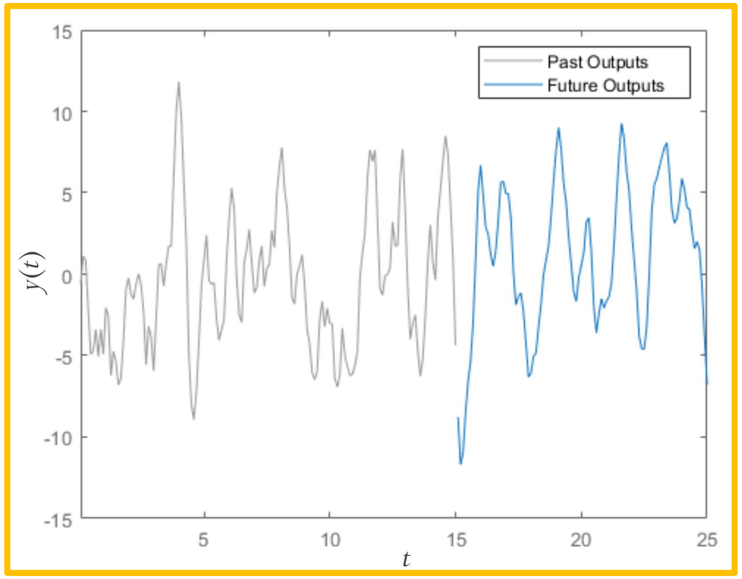
$nk = 1$ Input delay (transport delay, deadtime)

lags

$\mathbf{A}(z) = 1 - 1.512\, z^{-1} + 0.7006\, z^{-2}$

$\mathbf{B}(z) = -0.2606\, z^{-1} + 1.664\, z^{-2}$

$\mathbf{C}(z) = 1 - 1.604\, z^{-1} + 0.7504\, z^{-2}$



Ragazzini, J. R., Zadeh, L. A. (1952). The analysis of sampled-data systems. Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry. 71 (5): 225–234. *https://ieeexplore.ieee.org/document/6371274?signout=success*

https://www.mathworks.com/help/ident/ref/armax.html

Forecasting time series using linear models

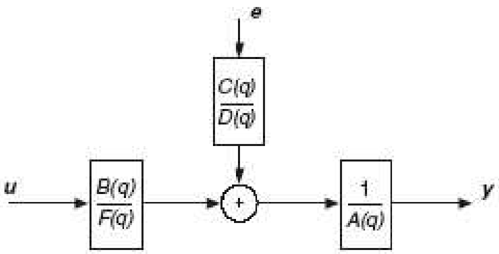
AR $\hat{y}(t) = -a_1 y(t-1) - a_2 y(t-2)$

MA $\hat{y}(t) = c_1 e(t-1) + c_2 e(t-2)$

ARMA $\hat{y}(t) = -a_1 y(t-1) - a_2 y(t-2) + c_1 e(t-1) + c_2 e(t-2)$

ARMAX $\hat{y}(t) = -a_1 y(t-1) - a_2 y(t-2) + c_1 e(t-1) + c_2 e(t-2) + b_1 u(t-1) + b_2 u(t-2)$

$$\boxed{\mathbf{y} = \mathbf{X}^T \hat{\boldsymbol{\theta}}}$$



z **or** *q*

Data: 3 data and 2 unknowns

$$\mathbf{X} = \begin{bmatrix} 2 & -0.8 \\ -1.2 & 2 \\ -1 & 0.85 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1.1 \\ 0.95 \\ -0.2 \end{bmatrix}$$

Find Least Squares solution to:

$$\begin{bmatrix} 2 & -0.8 \\ -1.2 & 2 \\ -1 & 0.85 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1.1 \\ 0.95 \\ -0.2 \end{bmatrix}$$

Form variance/covariance matrix and cross correlation vector

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 6.44 & -4.85 \\ -4.85 & 5.36 \end{bmatrix} \qquad \mathbf{X}^T\mathbf{y} = \begin{bmatrix} 1.26 \\ 0.85 \end{bmatrix}$$

Invert variance/covariance matrix

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 0.4870 & 0.4404 \\ 0.4404 & 0.5848 \end{bmatrix}$$

Least squares solution

$$\hat{\boldsymbol{\theta}} = [0.988 \quad 1.052]^T$$

$$\boxed{\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}}$$

Others

ARI

The ARI (**A**uto**R**egressive **I**ntegrated) model is an AR model with an **integrator** in the noise channel.

$$A(q)y(t) = \frac{1}{1-q^{-1}}e(t)$$

ARIX

(**A**uto**R**egressive **I**ntegrated with Extra **I**nput)
Estimate the parameters of an ARIX model. An ARIX model is an ARX model with integrated noise.

$$A(q)y(t) = B(q)u(t-nk) + \frac{1}{1-q^{-1}}e(t)$$

https://slideplayer.com/slide/4813947/
https://www.ni.com/pt-br/support/documentation/supplemental/06/selecting-a-model-structure-in-the-system-identification-process.html
https://www.mathworks.com/help/ident/ref/arx.html

Forecasting time series using linear models

ARIMA models / Box-Jenkins

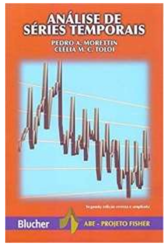
(model to **non-stationary** time series)

$$\Delta^D y_t = \sum_{i=1}^{p} \phi_i \Delta^D y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{m=1}^{M} \beta_m Xm,t + \epsilon_t$$

$$\epsilon_t \sim N\left(0, \sigma^2\right)$$

☞ **ACF** How to detect autocorrelation?
Durbin-Watson test
Bruesch-Godfrey test
☞ **ACF and PACF**
De-trending the data
Identifying the significant terms

USP



PUC-Rio

p: The number of lag observations included in the model, also called the lag order (AR order).

d: The number of times that the raw observations are differenced, also called the degree of differencing.

q: The size of the moving average window, also called the order of moving average **(MA order)**.

ACF: Auto Correlation Function
PCF: Partial Autocorrelation Function

✓ Identification of an **AR** model is often best done with the **PACF**.
✓ Identification of an **MA model** is often best done with the **ACF** rather than the PACF.

Box-Jenkins method

A. Model form selection
  1. Evaluate stationarity
  2. Selection of the differencing level (*d*) – to fix stationarity problems
  3. Selection of the AR level (*p*)
  4. Selection of the MA level (*q*)
B. Parameter estimation
C. Model checking

Variants
❶ **V**ector **A**uto **R**egression (**VAR**)
❷ **V**ector **M**oving **A**verage (**VMA**)
❸ **V**ector **A**utoregressive **M**oving **A**verage (**VARMA**)
❹ Seasonal **ARIMA** (**SARIMA**)
❺ **SARIMA** with e**X**ogeneous inputs (**SARIMAX**)

https://github.com/seanabu/seanabu.github.io/blob/master/_posts/2016-03-22-time-series-seasonal-ARIMA-model-in-python.markdown

https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/

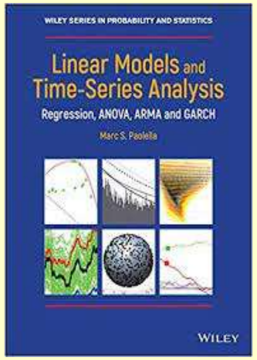
Time series models

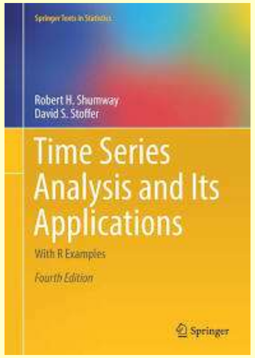
stationary models

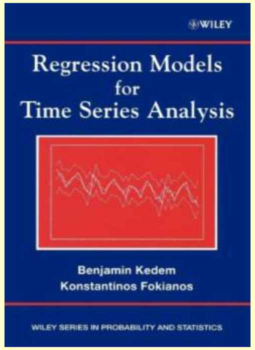
- ARMAX: AutoRegressive Moving Average model with Exogenous inputs
- ARMA: AutoRegressive Moving Average model
- ARX: AutoRegressive model with Exogenous inputs
- AR: AutoRegressive model
- MA: Moving Average model

non-stationary models

- ARIMA: AutoRegressive Integrated Moving Average model
- ARCH, GARCH
  - AutoRegressive Conditional Heteroskedasticity
  - Generalized AutoRegressive Conditional Heteroskedasticity

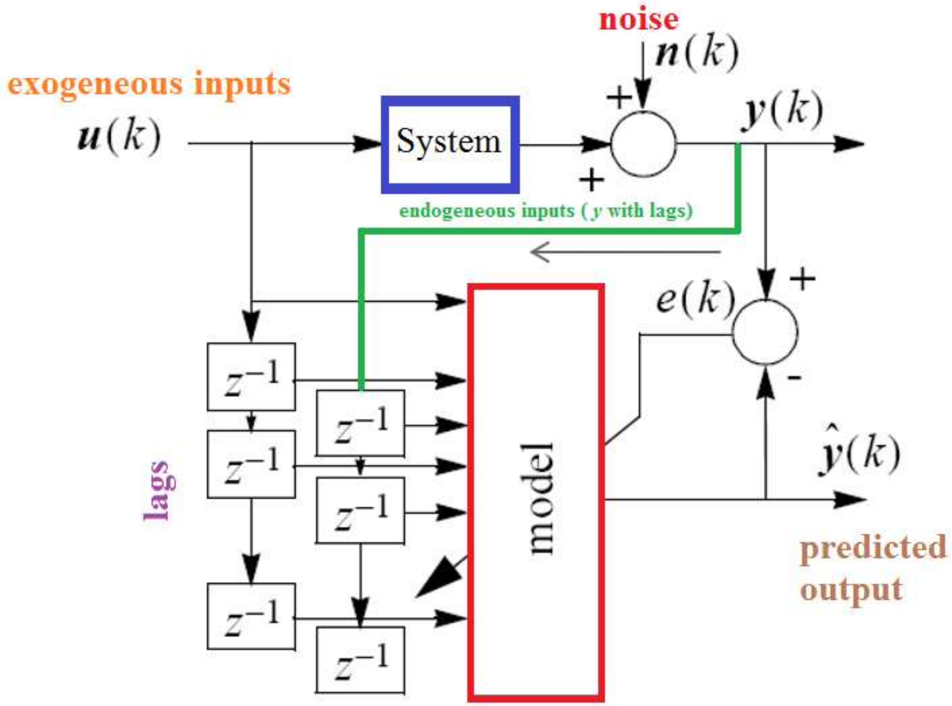




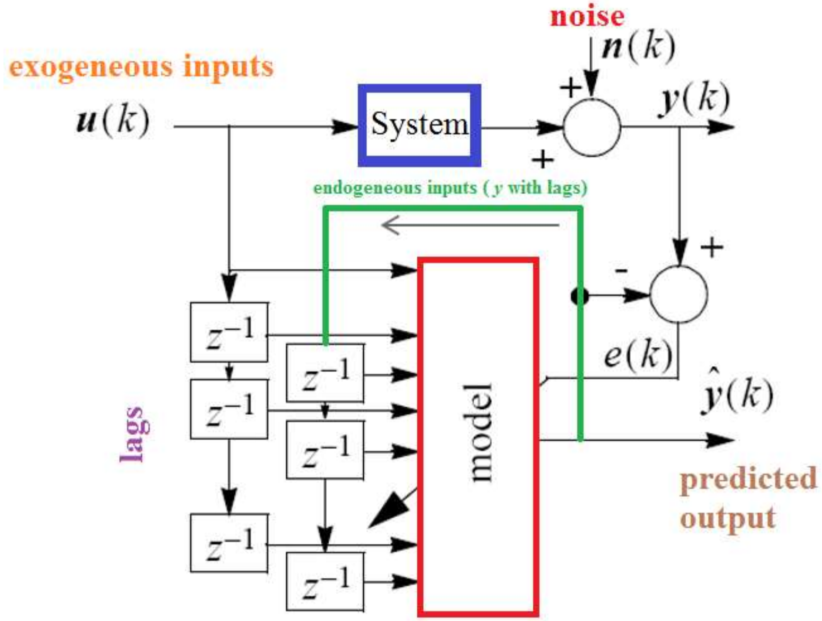


https://www.sciencedirect.com/science/article/pii/S1474034614000834

Series-parallel and parallel models



series-parallel model
(one step ahead forecasting)



parallel model
(multistep ahead forecasting)

J.V. Gorp (2000). Nonlinear identification with neural networks and fuzzy logic, PhD thesis, Vrije Universiteit Brussel, Dept. ELEC, 2000. page 62

IEEE TRANSACTIONS ON NEURAL NETWORKS. VOL. 1. NO. 1. MARCH 1990

Identification and Control of Dynamical Systems Using Neural Networks

KUMPATI S. NARENDRA FELLOW, IEEE, AND KANNAN PARTHASARATHY

http://homepages.vub.ac.be/~jschouk/johan/phdthesis/25_phdjurgenvangorp.pdf

http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=80202

Why cross-validation different with time series?

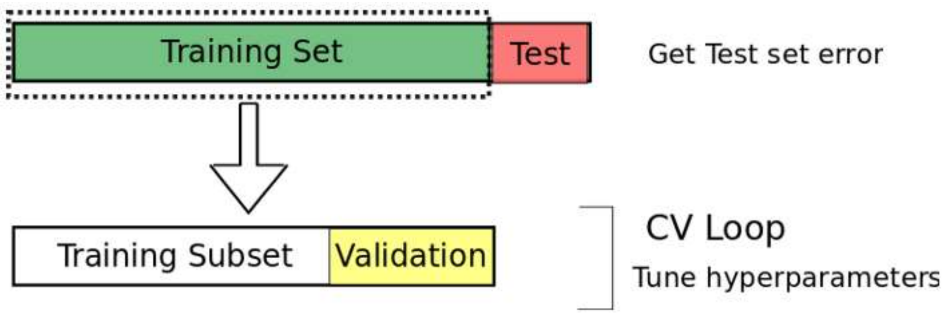
When dealing with time series data, traditional cross-validation (like *k*-fold) should not be used for two reasons:

1. Temporal dependencies

With time series data, particular care must be taken in splitting the data in order to prevent data leakage. In order to accurately simulate the "real world forecasting environment, in which we stand in the present and forecast the future" (Tashman 2000), the forecaster must withhold all data about events that occur chronologically after the events used for fitting the model.

So, rather than use *k*-fold cross-validation, for time series data we utilize **hold-out cross-validation** where a subset of the data (*split temporally*) is reserved for validating the model performance. For example, see **Figure 1** where the test set data comes chronologically after the training set. Similarly, the validation set comes chronologically after the training subset.

Figure 1



Tashman, L. J. Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting, 16(4):437–450, 2000. *https://www.sciencedirect.com/science/article/abs/pii/S0169207000000650*

https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9

Why cross-validation different with time series?

2. Arbitrary choice of test set 1/2

You may notice that the choice of the test set in **Figure 1** (Temporal dependencies) is fairly arbitrary, and that choice may mean that our test set error is a poor estimate of error on an independent test set.

To address this, we use a method called **Nested Cross-Validation.** Nested CV contains an outer loop for error estimation and an inner loop for parameter tuning (see **Figure 2 in the next slide**).

Advantageous:

A nested cross-validation procedure provides an almost unbiased estimate of the true error. (Varma and Simon 2006)

Varma, S., Simon, R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7(1):91, Feb 2006.
https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-7-91

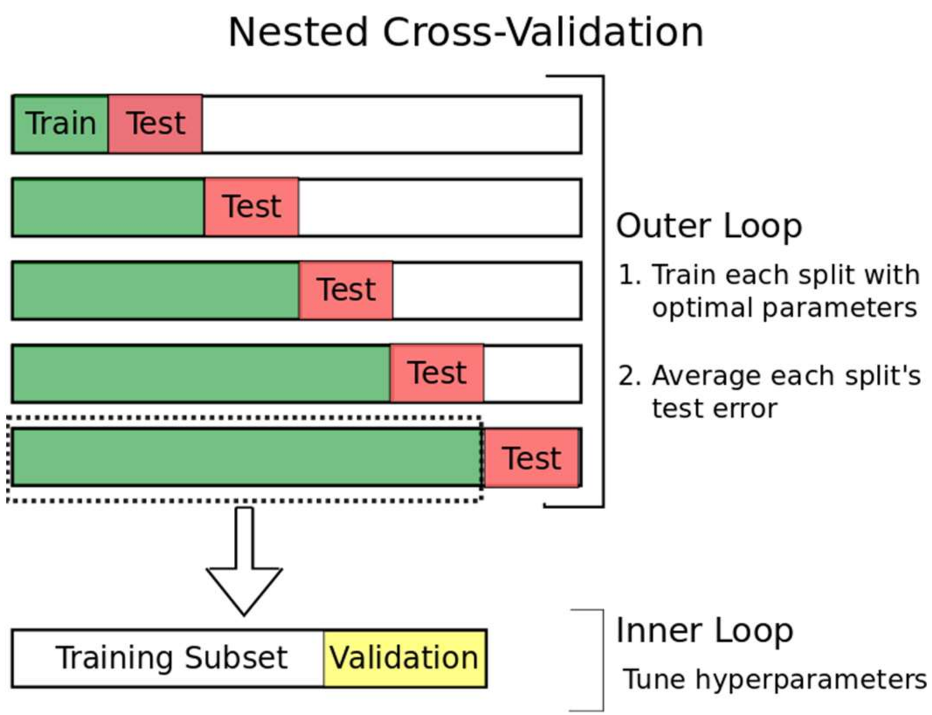
https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9

Why cross-validation different with time series?

2. Arbitrary choice of test set 2/2

The inner loop works exactly as discussed before: the training set is split into a training subset and a validation set, the model is trained on the training subset, and the parameters that minimize error on the validation set are chosen. However, now we add an outer loop which splits the dataset into multiple different training and test sets, and the error on each split is averaged in order to compute a robust estimate of model error.

Figure 2



https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9

Time series forecasting using nonlinear models

In general, a nonlinear black-box model can be considered as a concatenation mapping from previously observed data to a regressor space, which is followed by a nonlinear function expansion-type mapping to the space of the system's output.

Forecasting time series using nonlinear models

Forecasting Response of Nonlinear AR (or ARX or ARMAX) Models

A time series nonlinear AR (or ARX or ARMAX) model has the following structure:

$$y(t) = f(y(t-1),\ y(t-2),\ \ldots,\ y(t-N)) + e(t)$$

where f is a nonlinear function with inputs $R(t)$, the model regressors.

The **regressors** can be the time-lagged variables $y(t\text{-}1)$, $y(t\text{-}2)$,… , $y(t\text{-}N)$ and their nonlinear expressions, such as $y(t\text{-}1)^2$, $y(t\text{-}1)y(t\text{-}2)$, $abs(y(t\text{-}1))$.

Suppose that time series data from your system can be fit to a 2nd-order linear-in-regressor model with the following **polynomial** regressors:

$$R(t) = [y(t-1), y(t-2), y(t-1)^2, y(t-2)^2, y(t-1)y(t-2)]^T$$

https://www.mathworks.com/help/ident/ug/forecasting-response-of-dynamic-systems.html

Forecasting time series using nonlinear models

Forecasting Response of Nonlinear AR Models

The **nonlinear AR (or ARX or ARMAX)** model has the form:

$$y(t) = f(y(t-1),\ y(t-2),\ \ldots,\ y(t-N)) + e(t)$$

$$R(t) = [y(t-1), y(t-2), y(t-1)^2, y(t-2)^2, y(t-1)y(t-2)]^T$$

$$y(t) = w_1 y(t-1) + w_2 y(t-2) + w_3 y(t-1)^2 + w_4 y(t-2)^2 + w_5 y(t-1)y(t-2) + c + e(t)$$

where the objective is estimate the model parameters (**weights**) $\boldsymbol{W}$ and **constant** (offset) $\boldsymbol{c}$.

https://www.mathworks.com/help/ident/ug/forecasting-response-of-dynamic-systems.html

NARMAX model

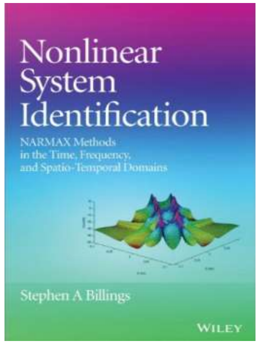
NARMAX: Nonlinear Auto-Regression Moving Average model with eXogenous inputs

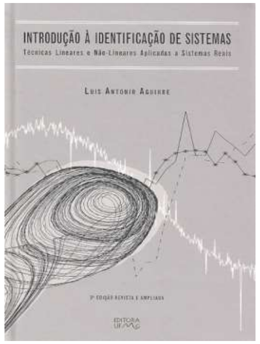
A popular structure selection approach used for identification of **NARX** (Nonlinear AutoRegressive eXogenous) **or NARMAX** (Nonlinear AutoRegressive, Moving Average eXogenous) models is the class of **Orthogonal Least Squares (OLS)** type algorithms, introduced in **Billings et al. (1988)**, which can be considered as wrapper method.

Stephen A. Billings
Department of Automatic Control and Systems Engineering
University of Sheffield

Luis A. Aguirre
Engenharia Elétrica
UFMG

This approach enables the identification of the model structure for Linear In-the-Parameter (LIP) models by evaluating the relevance of individual model terms in an orthogonal space in a forward regression manner. For the measure of quality, the error reduction ratio is used (see, e.g., **Billings, 2013**).





Billings, S.A. (2013). Nonlinear system identification – NARMAX methods in time, frequency, and spatiotemporal domains. Wiley.
Billings, S.A., Korenberg, M.J., Chen, S. (1988). Identification of non-linear output affine system using an orthogonal least-squares algorithm. Int. J. of Systems Science, 19(8), 1559-1568.
Korenberg, M., Billings, S.A., Liu, Y.P., McIlroy, P.J. (1988). Orthogonal parameter estimation algorithm for non-linear stochastic systems. Int. J. Control 48, 193-210.

https://www.sciencedirect.com/science/article/pii/S2405896315028566

NARMAX model

The initial structure of the NARMAX polynomial is determined by *d*, *Nu*, *Ny*, *Ne*, and *l*.

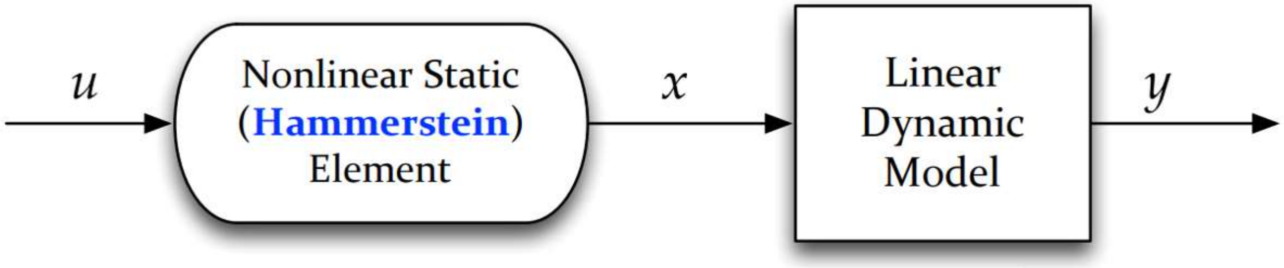
A clear disadvantage of polynomial models is the enormous number of terms a general nonlinear polynomial may have.

The significance of a model term is expressed using its corresponding **Error Reduction Ratio (ERR)**. This value is calculated for every term as part of the NARMAX estimation process.

The ERR is an indication of the reduction in the model's prediction error that occurs when the model term considered is introduced into the model. This reduction is expressed in proportion to the maximum error (a constant) that results from removing all the terms from the model. The value of the ERR is therefore proportional to the significance of the term to which it corresponds.

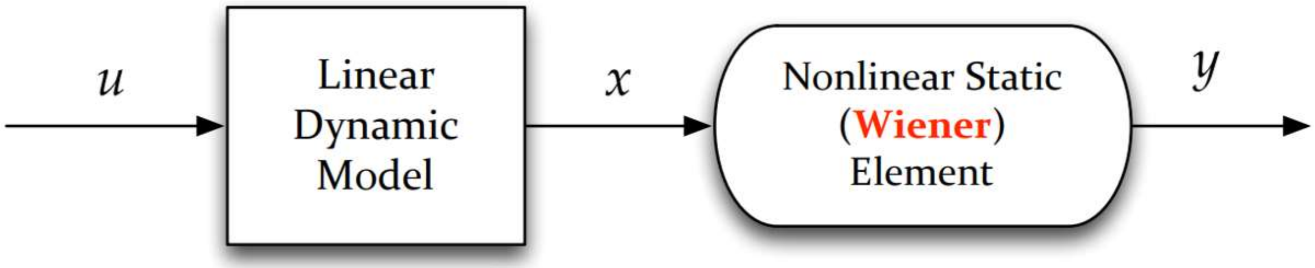
https://www.sciencedirect.com/science/article/pii/S0952197617301082

Block-oriented models



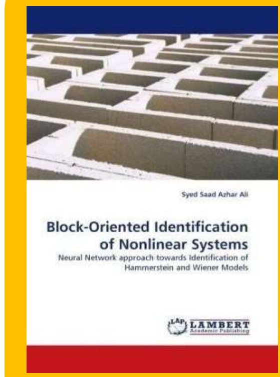
Hammerstein model

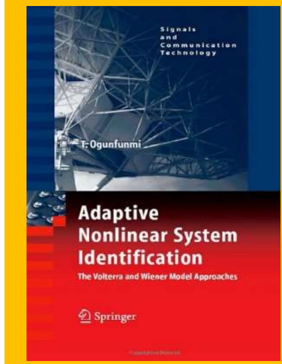
K.S. Narendra, P.G. Gallman An iterative method for the identification of nonlinear systems using a Hammerstein model, IEEE Trans. Aut. Control, Vol. AC-11, pp. 546-550, July 1966. *https://www.sciencedirect.com/science/article/pii/S0005109817303990#fig2a?*

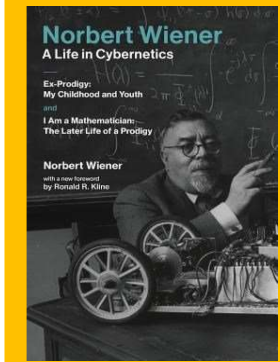


Wiener model

M. Shoukens, K. Tiels Identification of block-oriented nonlinear systems starting from linear approximations: A survey, Automatica, Vol. 85, pp. 272-292, 2017. *https://www.sciencedirect.com/science/article/pii/S0005109817303990#fig2a?*







https://www.sciencedirect.com/science/article/pii/S0005109817303990#fig2a?

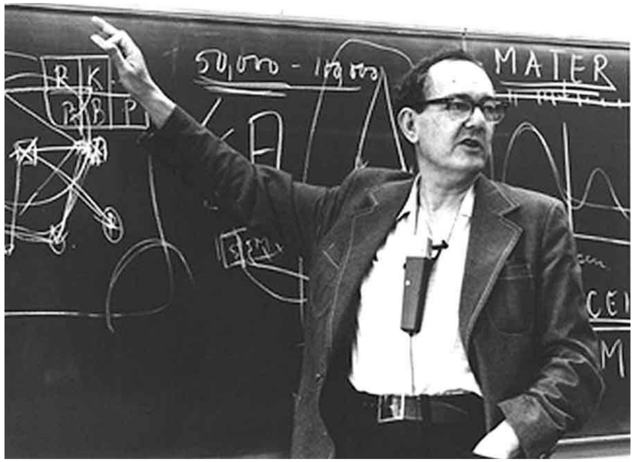
https://www.control.isy.liu.se/student/tsrt62/oh/mos6.pdf

Time series forecasting using machine learning

Machine learning

Learning is any process by which a system improves performance from experience.

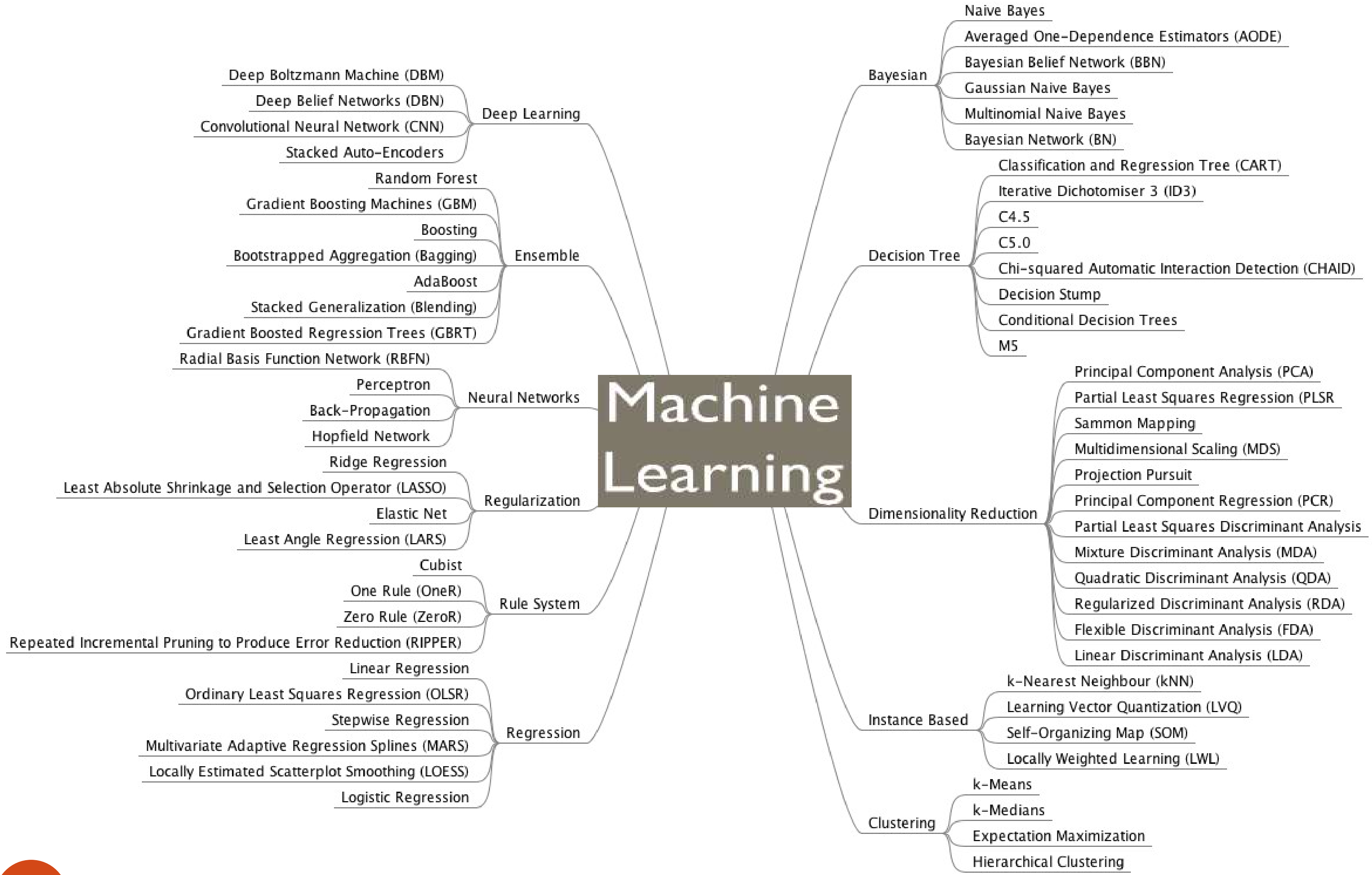
Machine Learning is concerned with computer programs that automatically improve their performance through **experience**.



Herbert Alexander Simon
(1916-2001)

TURING AWARD 1975

Models and algorithms



http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Steps

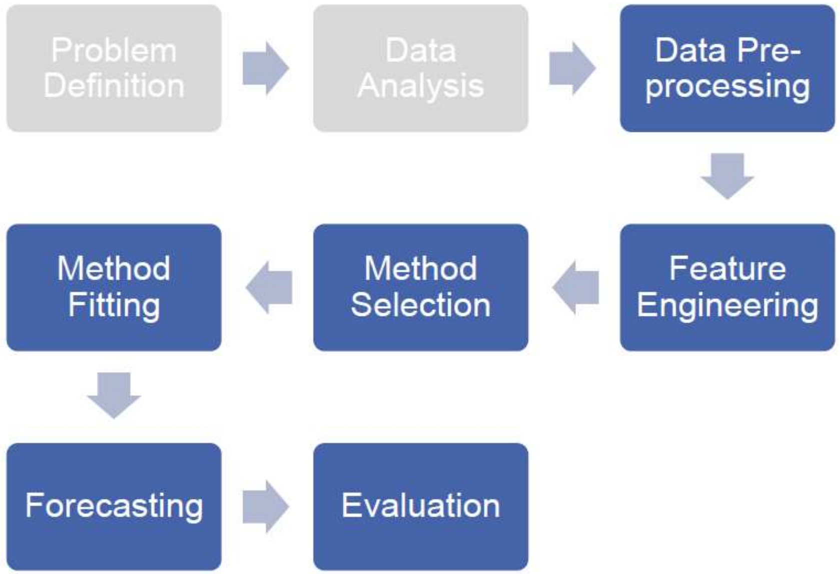
Data pre-processing

Feature engineering

Method selection

Method (model) fitting

Evaluation



https://pt.slideshare.net/mlvlc/12-evaluating-machine-learning-algorithms-i?from_action=save

Feature engineering

- Feature engineering, also known as feature creation, is the process of constructing new features from existing data to train a machine learning model.
- Feature engineering is the process of using domain knowledge of the data to create variables that make machine learning algorithms work.
- Feature engineering starts with retaining existing good features, creating additional useful features and discarding irrelevant features from the dataset.
- Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data (Jason Brownlee).
- Transforming data to create model inputs.

Feature engineering = creating features of the appropriate granularity for the task

https://www.slideshare.net/gabrielspmoreira/feature-engineering-getting-most-out-of-data-for-predictive-models-tdc-2017

https://www.slideshare.net/odsc/feature-engineering

https://towardsdatascience.com/four-unavoidable-tips-for-the-art-of-feature-engineering-in-machine-learning-ab4d1ce0cbda

Feature transformation

Some machine learning models assume that the variables are normally distributed. Other models may benefit from a more homogeneous spread of values across the value range. If variables are not normally distributed, we can apply a mathematical transformation to enforce this distribution. Typically used mathematical transformations are:

- Logarithm transformation – $\log(x)$ → Compresses the range of large numbers and expand the range of small numbers.
- Reciprocal transformation – $1 / x$
- Square/Cubic root transformation – $\mathrm{sqrt}(x)$, $x^{(1/3)}$
- Power/Polynomial transformation – x^2, x^3
- Exponential transformation – $\exp(x)$
- Hyperbolic tangent transformation – $\tanh(x)$

Dealing with **heteroskedasticity**:

- Box-Cox transformation (1964)

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln y_i & \text{if } \lambda = 0 \end{cases}$$

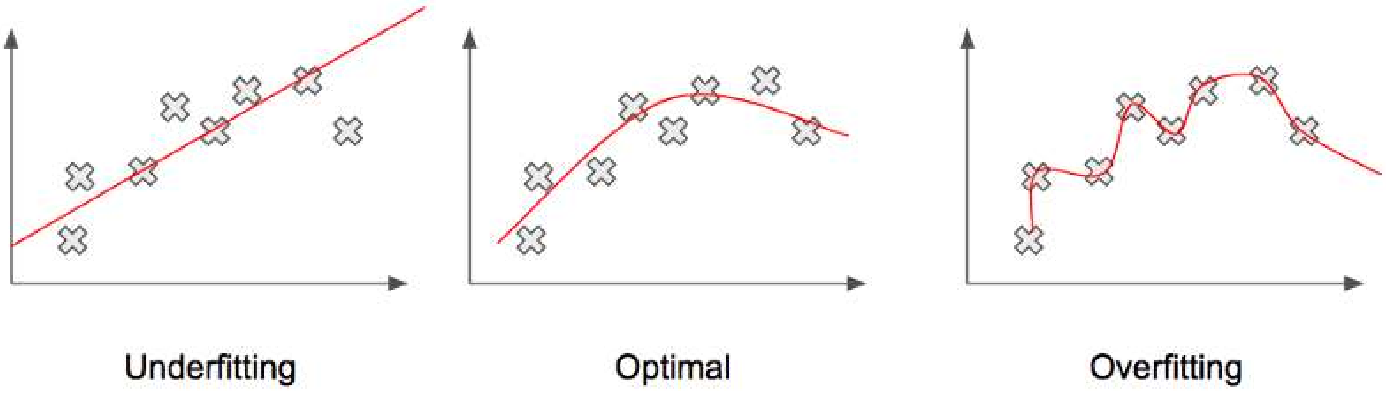
λ = 1 (no transformation)
λ = 0 (log)
λ = 0.5 (square root)
λ = −1 (inverse)

- Yeo-Johnson transformation (2000)

$$y_i^{(\lambda)} = \begin{cases} ((y_i + 1)^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y_i + 1) & \text{if } \lambda = 0, y \geq 0 \\ -[(-y_i + 1)^{(2-\lambda)} - 1]/(2 - \lambda) & \text{if } \lambda \neq 2, y < 0 \\ -\log(-y_i + 1) & \text{if } \lambda = 2, y < 0 \end{cases}$$

https://www.trainindatablog.com/feature-engineering-for-machine-learning-comprehensive-overview/

Model validation and tuning lifecycle



A model is said to be **underfitting** (High **Bias**) when it performs poorly on the training data.
An **overfitting** model, (High **Variance**) on the other hand, performs well on the training data but does not perform well on the evaluation data.

Generally, for a **more complex**/capable model: ↓ **bias**, ↑ **variance**

i.e., due to an under-complex model for a real-world task

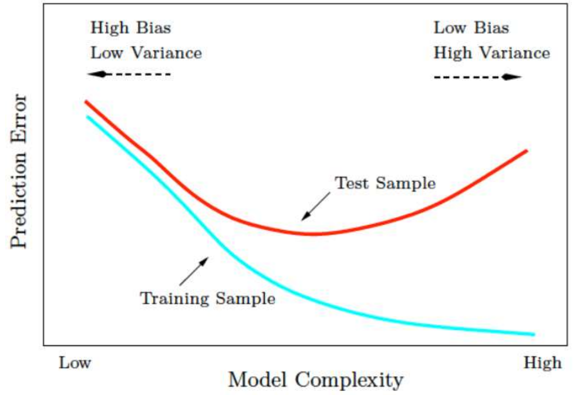
i.e., change expected if we use a different training set

$$E = \underbrace{systematic\ error}_{} + \underbrace{dependence\ on\ specific\ sample}_{} + \underbrace{random\ nature\ of\ process}_{}$$

Bias: average prediction's deviation from the truth

Variance: sensitivity of prediction to specific training sample

Irreducible error or **Bayes' rate**: due to "noise"



https://www.freecodecamp.org/news/how-to-get-a-grip-on-cross-validations-bb0ba779e21c/

Regression/Forecasting models

Linear regression
Polynomial regression
Ridge regression, LASSO regression, Elasticnet

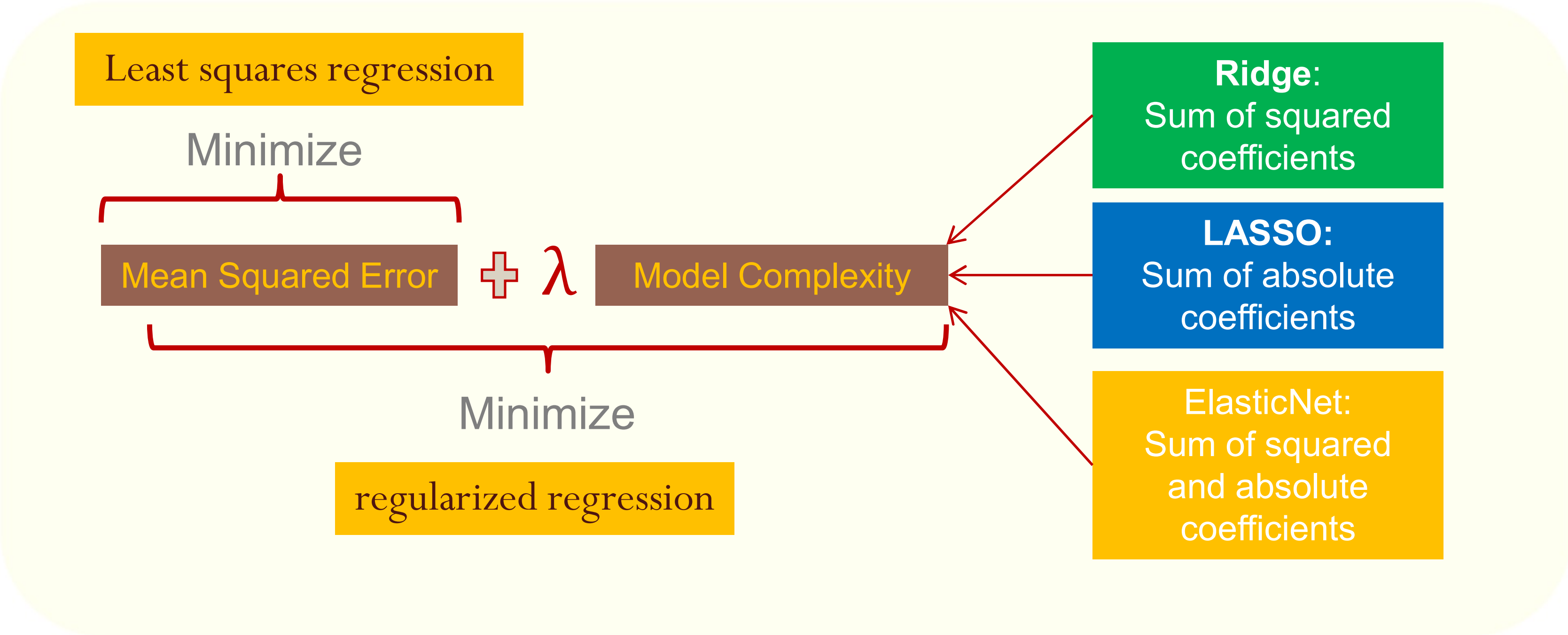
Quantile regression
Support vector regression (SVR)
Nearest neighboor (kNN)

Gaussian process
Decision tree regression
Random forest regression
Extra trees regression
Cubist
Gradient boosting machine, CatBoost, XGBoost, LightGBM

Artificial neural networks (shallow networks and deep learning)

LASSO: **L**east **A**bsolute **S**hrinkage and **S**election **O**perator

Regularized regression



✓ Any regularized regression approach tries to balance model performance and model complexity

✓ **λ – regularization parameter**, to be estimated

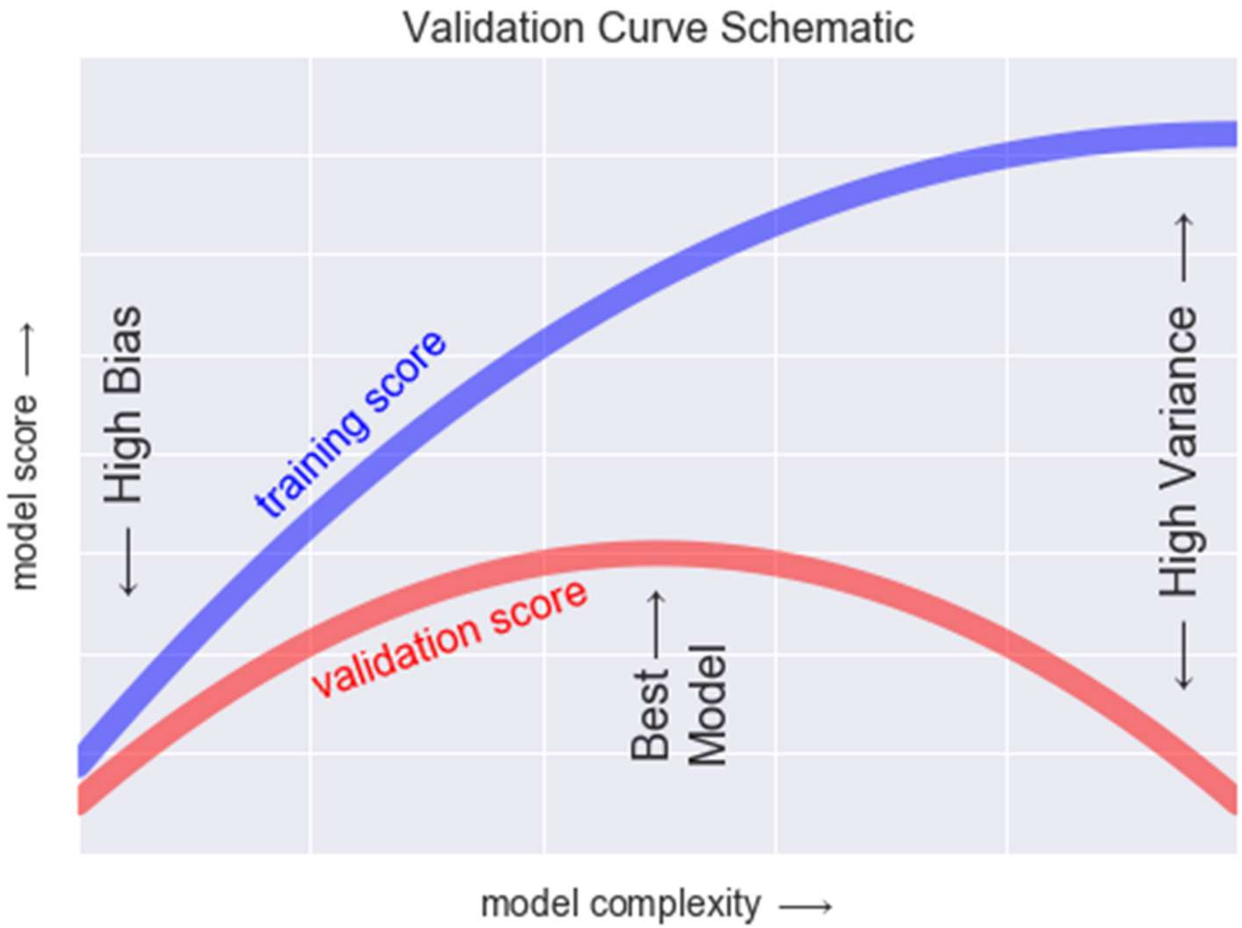
$\lambda = \infty$ Null model zero-coefficients (maximum possible penalty)

$\lambda = 0$ Least squares solution (no penalty)

http://maximr-ml.blogspot.com/2016/09/train-nnls-model.html

LASSO: **L**east **A**bsolute **S**hrinkage and **S**election **O**perator

Model validation and tuning lifecycle



https://jakevdp.github.io/PythonDataScienceHandbook/05.03-hyperparameters-and-model-validation.html

Performance analysis

Performance analysis

Performance metrics

R-squared (R^2) coefficient of determination
Adjusted R-squared (R^2)
Mean Squared Error (MSE)
Root Mean Squared Error (RMSE)
Mean Absolute Error (MAE)
Mean Absolute Percentage Error (MAPE)
Symmetric Mean Absolute Percentage Error (sMAPE)
Weighted Mean Absolute Percentage Error (WMAPE)
Mean Absolute Scaled Error (MASE)

Cross-validation

Normality tests of the residuals

Shapiro-Walk test https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html
D'Agostino1s K^2 test https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html
Anderson-Darling test https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.anderson.html

Performance analysis

Regression diagnostics and specification tests

✓ **Heteroscedasticity tests**

For these test the null hypothesis is that all observations have the same error variance, i.e. errors are homoscedastic. The tests differ in which kind of heteroscedasticity is considered as alternative hypothesis.

✓ **Autocorrelation tests**

This group of test whether the regression residuals are not autocorrelated. They assume that observations are ordered by time.

✓ **Non-linearity tests**

✓ **Tests for Structural change, Parameter stability**

Test whether all or some regression coefficient are constant over the entire data sample.

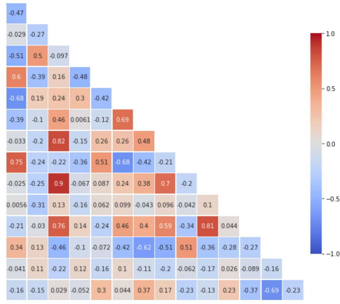
✓ **Multicollinearity tests**

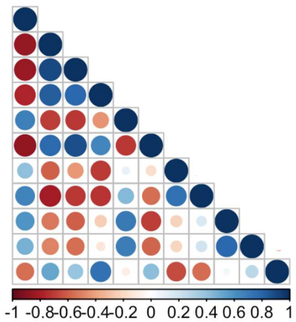
✓ **Normality and distribution tests**

https://www.statsmodels.org/stable/diagnostic.html

Multicollinearity

Correlation matrix / Correlation plot is used to detect the presence of multicollinearity. A correlation plot can be used to identify the correlation or bivariate relationship between two independent variables whereas **Variation Inflation Factor (**VIF) is used to identify the correlation of one independent variable with a group of other variables.





Hence, it is preferred to use **VIF** for better understanding.

VIF = 1 → No correlation
VIF = 1 to 5 → Moderate correlation
VIF >10 → High correlation

$$VIF = \frac{1}{(1-R^2)}$$

https://towardsdatascience.com/how-to-detect-and-deal-with-multicollinearity-9e02b18695f1 *https://bdtd.ibict.br/vufind/*

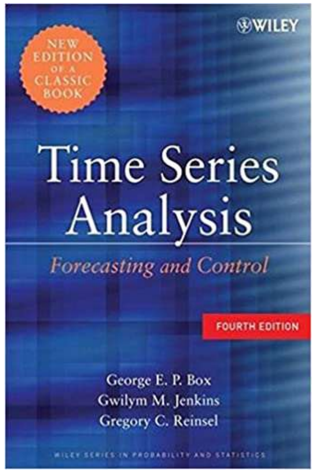
References

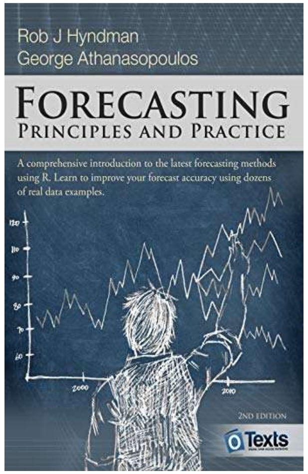
https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php?option=com_plogin&ym=3&pds_handle=&calling_system=primo&institute=CAPES&targetUrl=http://www.periodicos.capes.gov.br&Itemid=155&pagina=CAFe&

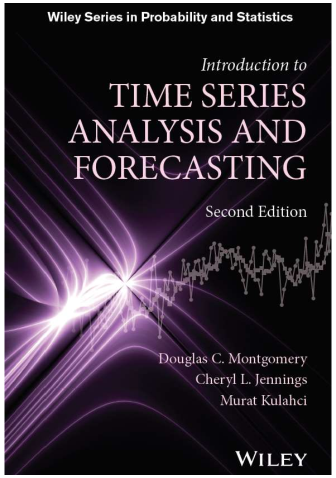
To be continued ...

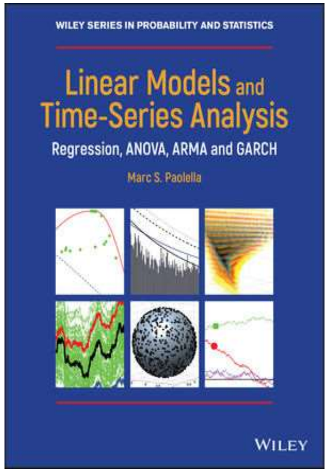
Books

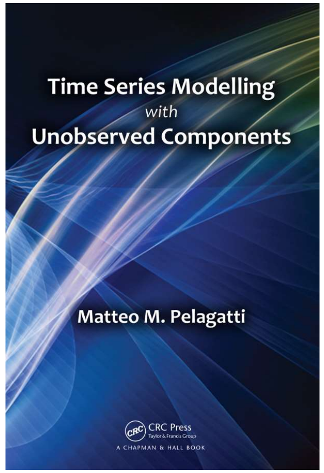
Time series forecasting

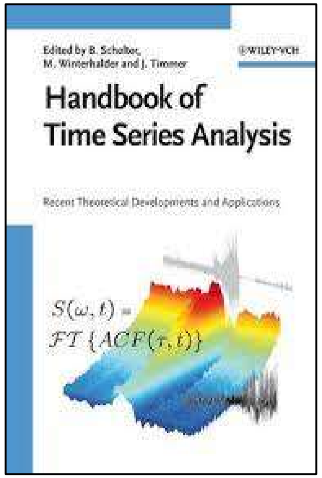


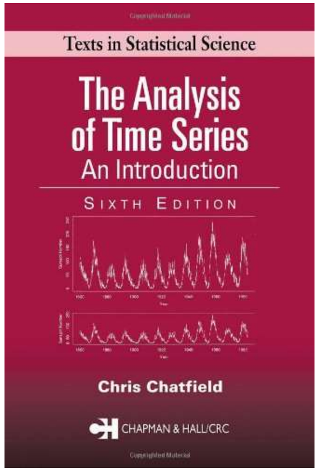


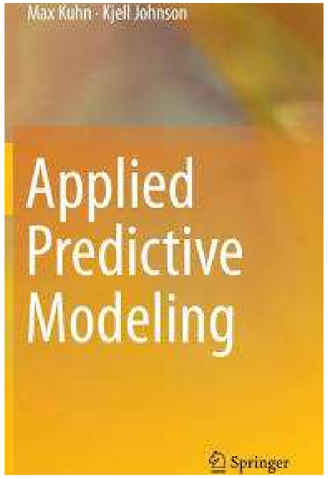


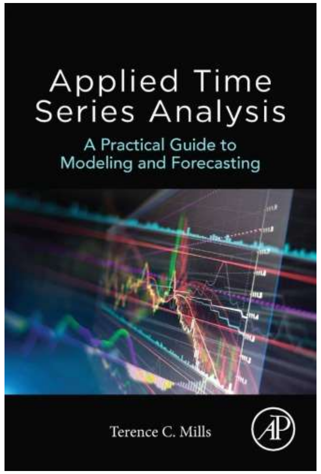


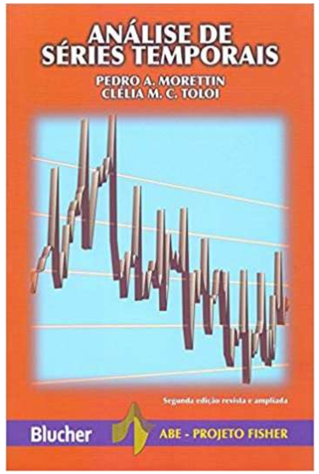






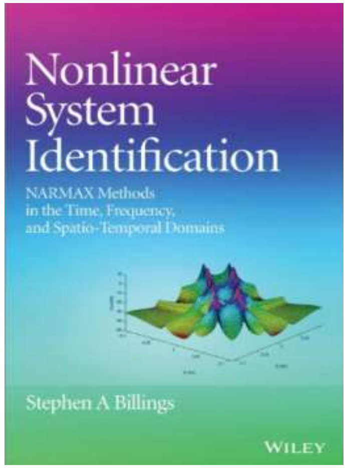


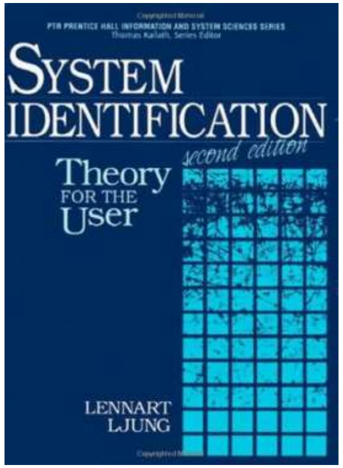


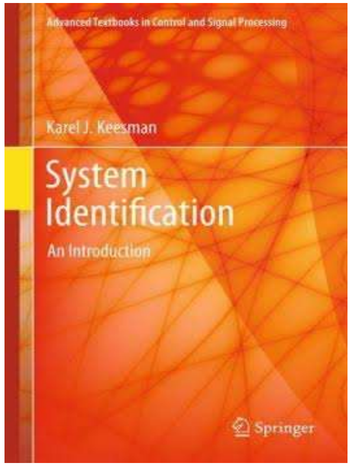


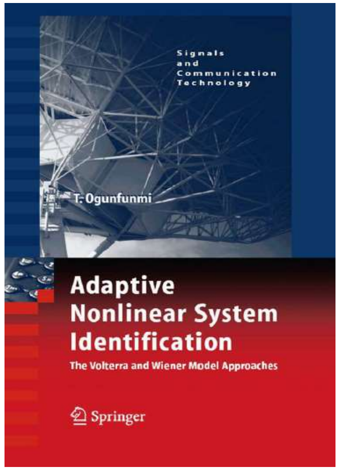
Books

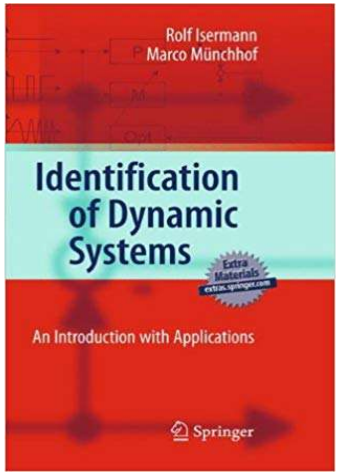
System identification

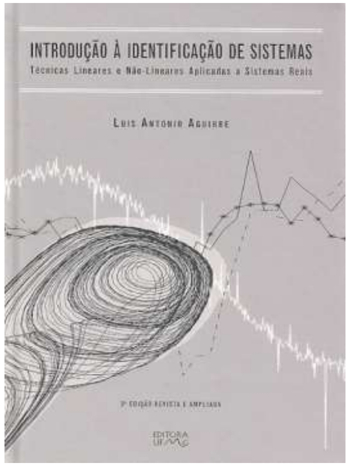


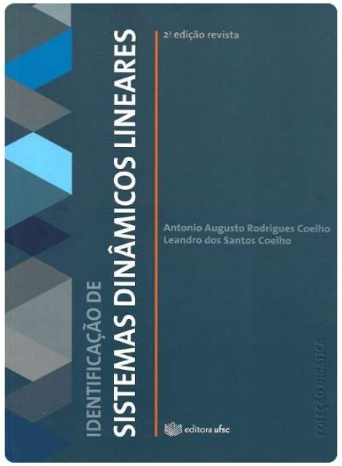


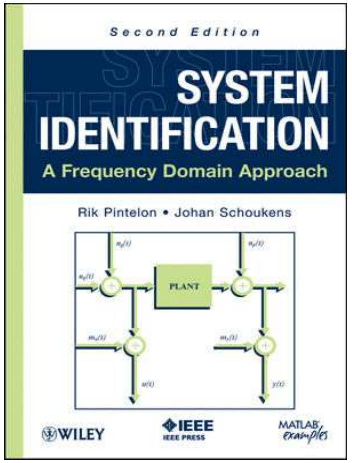


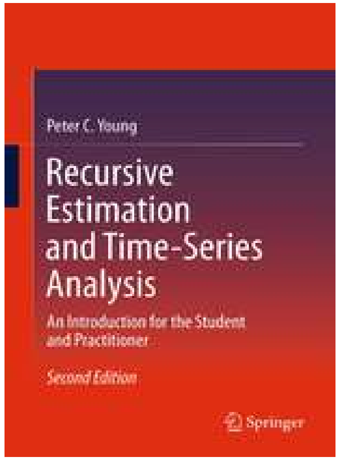


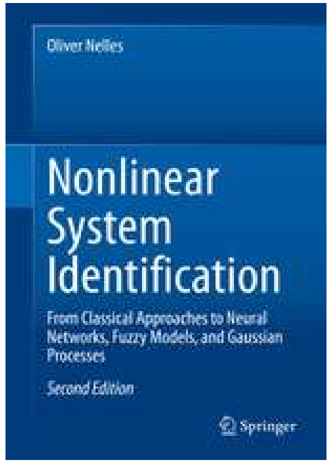






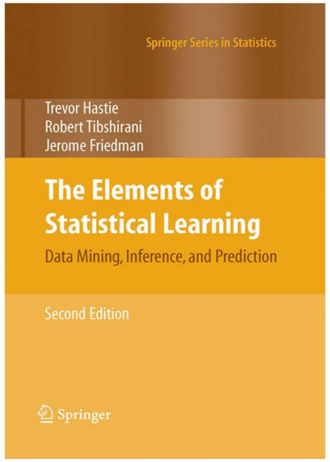


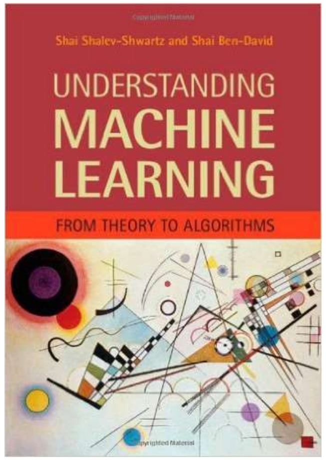


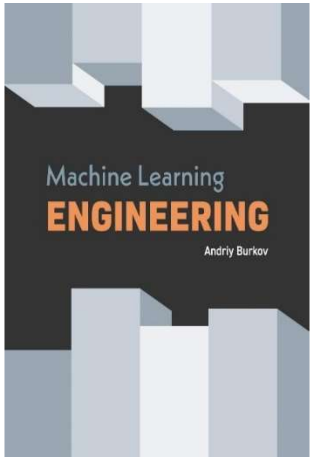


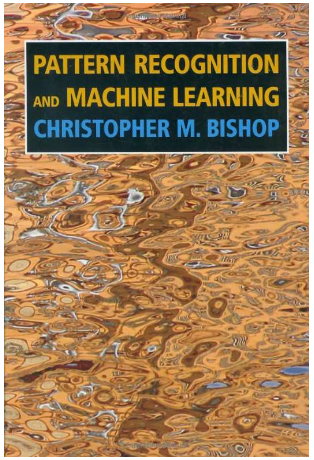
Books

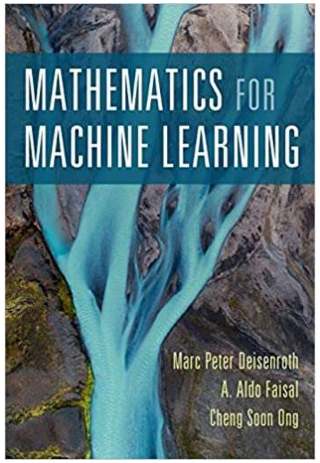
Machine learning

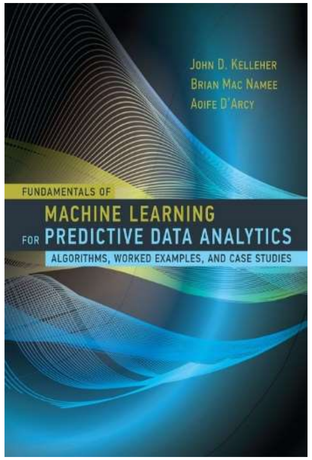


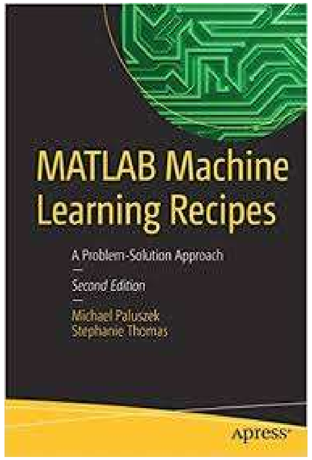


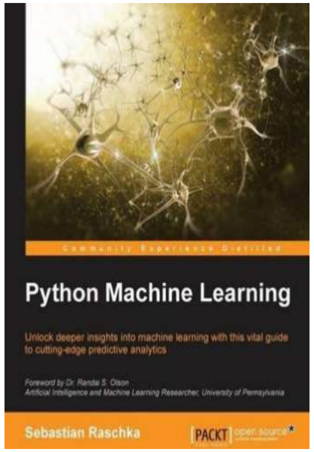


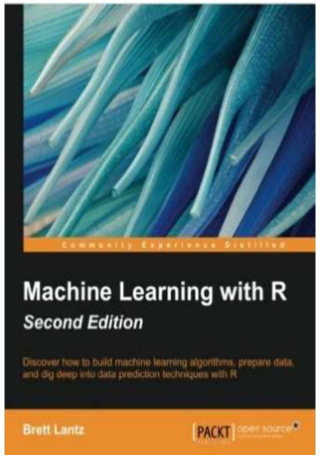




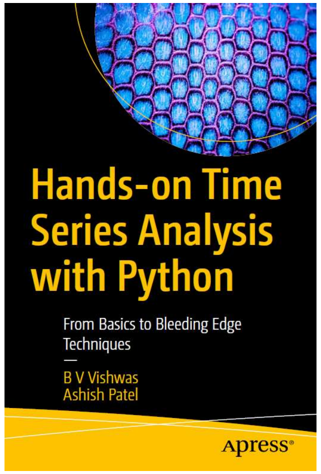


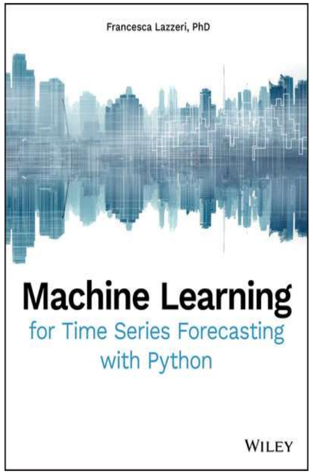


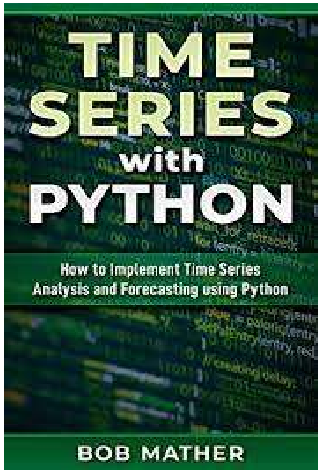


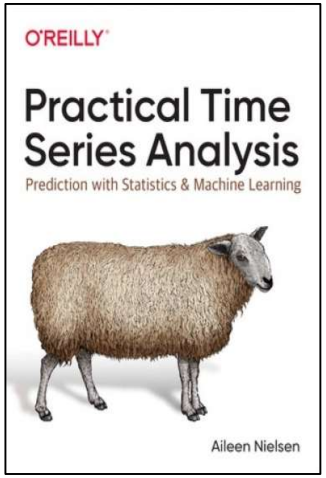


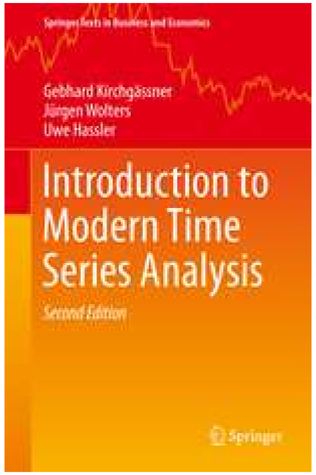
Books

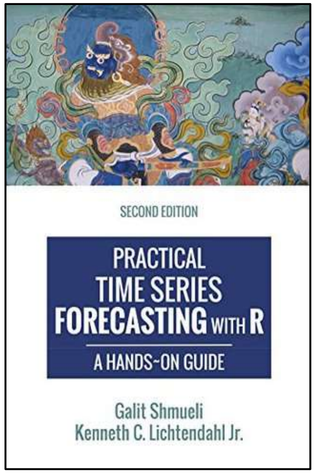

Hands-on Time Series Analysis with Python
From Basics to Bleeding Edge Techniques
B V Vishwas
Ashish Patel
Apress

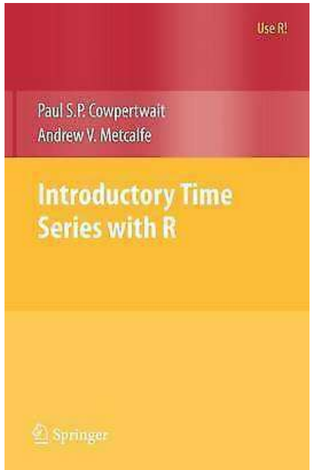

Francesca Lazzeri, PhD
Machine Learning
for Time Series Forecasting with Python
WILEY

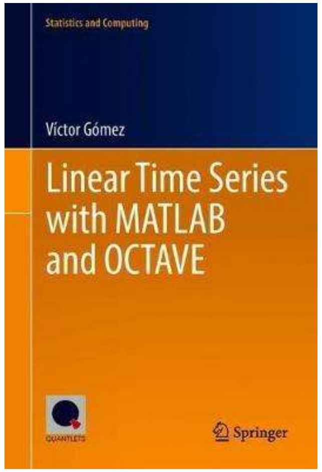

TIME SERIES with PYTHON
How to Implement Time Series Analysis and Forecasting using Python
BOB MATHER

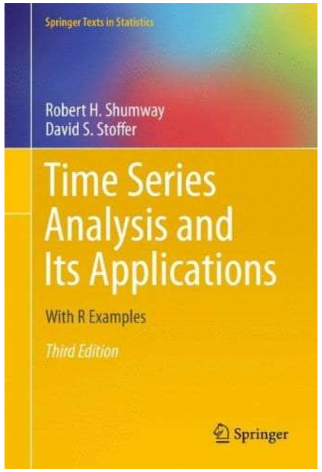

O'REILLY
Practical Time Series Analysis
Prediction with Statistics & Machine Learning
Aileen Nielsen

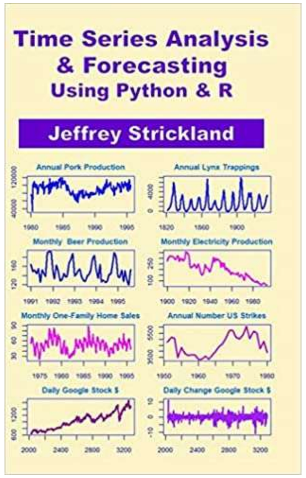

Gebhard Kirchgässner
Jürgen Wolters
Uwe Hassler
Introduction to Modern Time Series Analysis
Second Edition
Springer


SECOND EDITION
PRACTICAL TIME SERIES FORECASTING WITH R
A HANDS-ON GUIDE
Galit Shmueli
Kenneth C. Lichtendahl Jr.


Use R!
Paul S.P. Cowpertwait
Andrew V. Metcalfe
Introductory Time Series with R
Springer


Statistics and Computing
Víctor Gómez
Linear Time Series with MATLAB and OCTAVE
Springer


Springer Texts in Statistics
Robert H. Shumway
David S. Stoffer
Time Series Analysis and Its Applications
With R Examples
Third Edition
Springer


Time Series Analysis & Forecasting Using Python & R
Jeffrey Strickland

Quote

If time travel is possible, where are the tourists from the future?

Stephen William Hawking (1942-2018)

English theoretical physicist, cosmologist, and author who was director of research at the Centre for Theoretical Cosmology at the University of Cambridge at the time of his death.

It is a curious fact that Stephen William Hawking was born on 8th January 1942, exactly 300 years after the death of the Italian astronomer, Galileo Galilei.

sociedade brasileira de AUTOMÁTICA

III Webinar 2021

26/08/2021 , 17h30

CT de Identificação de Sistemas e Ciência de Dados

https://www.youtube.com/watch?v=5bB6hOcFSJQ

Análise e previsão de séries temporais

Time series analysis and forecasting

Leandro dos Santos Coelho

Pontifícia Universidade Católica do Paraná (PUCPR), Escola Politécnica
Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS), Graduação em Engenharia de Controle e Automação (Mecatrônica)
Rua Imaculada Conceição, 1155, CEP 80215-901, Curitiba, PR, Brasil

Universidade Federal do Paraná (UFPR), Graduação e Pós-Graduação em Engenharia Elétrica, Campus Centro Politécnico
Av. Cel. Francisco H. dos Santos, 100, CEP 81530-000, Curitiba, PR, Brasil

Lattes (CNPq): *http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4792095Y4*

Google Scholar: *https://scholar.google.com/citations?user=0X7VkC4AAAAJ&hl=pt-PT*

Linkedin: *https://www.linkedin.com/in/leandro-dos-santos-coelho-07a08893*

Top H-Index for Scientists in Brazil: *http://www.guide2research.com/scientists/BR*

Email: lscoelho2009@gmail.com