CEFET/RJ

TIME SERIES PREDICTION

Eduardo Ogasawara
eogasawara@ieee.org
https://eic.cefet-rj.br/~eogasawara

Biography

- Doctorate in Systems and Computer Engineering (COPPE/UFRJ) in 2011
- Professor at EIC - CEFET/RJ
  - DEPIN
  - COINFO
- Permanent professor at
  - Postgraduate Program in Computer Science (PPCIC)
  - Postgraduate Program in Production and Systems Engineering (PPPRO)
- Member of IEEE, SBC, ACM, and INNS
- Institutional representative of SBC

eogasawara@ieee.org
https://eic.cefet-rj.br/~eogasawara

Notices

IX EIC Workshop

- Workshop has more than 200 participants
- Many interesting themes
- Confirmed talks:
  - Oct 19 – 4pm – Pesquisa e Extensão na EIC: De onde viemos? Quem somos? Para onde iremos? – Carmen de Queiroz, Jorge Soares, Joel Santos, Eduardo Ogasawara
  - Oct 20 – 2pm – Gabriela Ruberg – BCB
  - Oct 20 – 6pm – High Performance Data Science – Marta Mattoso, Alvaro Coutinho, Fabio Porto, Daniel Oliveira , Kary Ocana, Eduardo Ogasawara

October 18-22, 2021

Brazilian Symposium on Databases (SBBD)

Join!

SBBD 2021 ONLINE

SBBD 2021 – BRAZILIAN SYMPOSIUM ON DATABASES

The Premier Brazilian Conference on Data Science and Big Data

The annual **Brazilian Symposium on Databases (SBBD)** is the official event on databases of the Brazilian Computer Society (SBC). The symposium includes a technical program with research and industrial talks, tutorials, demos, and focused workshops. It also hosts invited talks by distinguished speakers from the international research community.

Due to COVID-19 and coronavirus pandemic, all activities of the 36th edition of the SBBD will happen **online** only, **October 4-8, 2021** – organized by CEFET/RJ (Rio de Janeiro, Brazil).

REALIZATION	EXECUTION	ACADEMIC SUPPORT
SBC Sociedade Brasileira de Computação	PPCIC PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO	CEFET/RJ

SPONSOR	SUPPORT
Loggi	PUC RIO

TRACKS COORDINATION

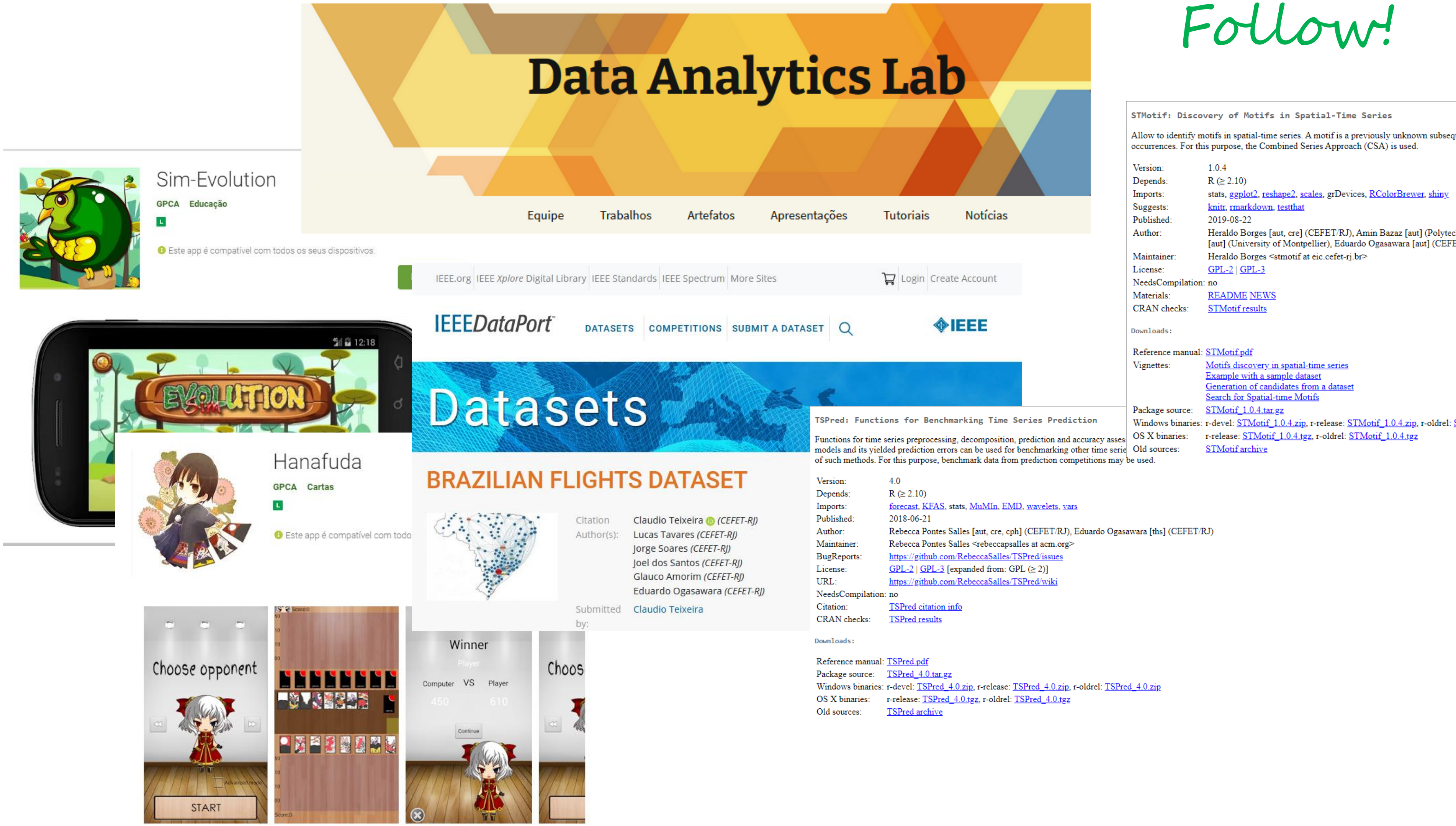
Program Chair: Ricardo Torres (NTNU, Norway)
Short Vision Industrial Chair: Damires Souza (IFPB, Brazil)
Steering Committee Chair: Fabio Porto (LNCC, Brazil)
Demos and Applications Chair: Leonardo Ribeiro (UFG, Brazil)
Thesis and Dissertation Workshop Chair (WTDBD): Julio Reis (Unicamp, Brazil)
CTDBD Chair: Cristina Ciferri (USP, Brazil)
Short courses Chair: Alessandréia M de Oliveira (UFJF, Brazil)
Tutorials Chair: Daniel de Oliveira (UFF, Brazil)
Workshop Chair: Eduardo Almeida (UFPR, Brazil)
WTAG Chair: André Carvalho (UFAM, Brazil)

ONLINE ORGANIZATION

SBBD General Chair:
Eduardo Ogawasara (CEFET/RJ) – eduardo.ogasawara@cefet-rj.br
Rafaelli Coutinho (CEFET/RJ) – rafaelli.coutinho@cefet-rj.br

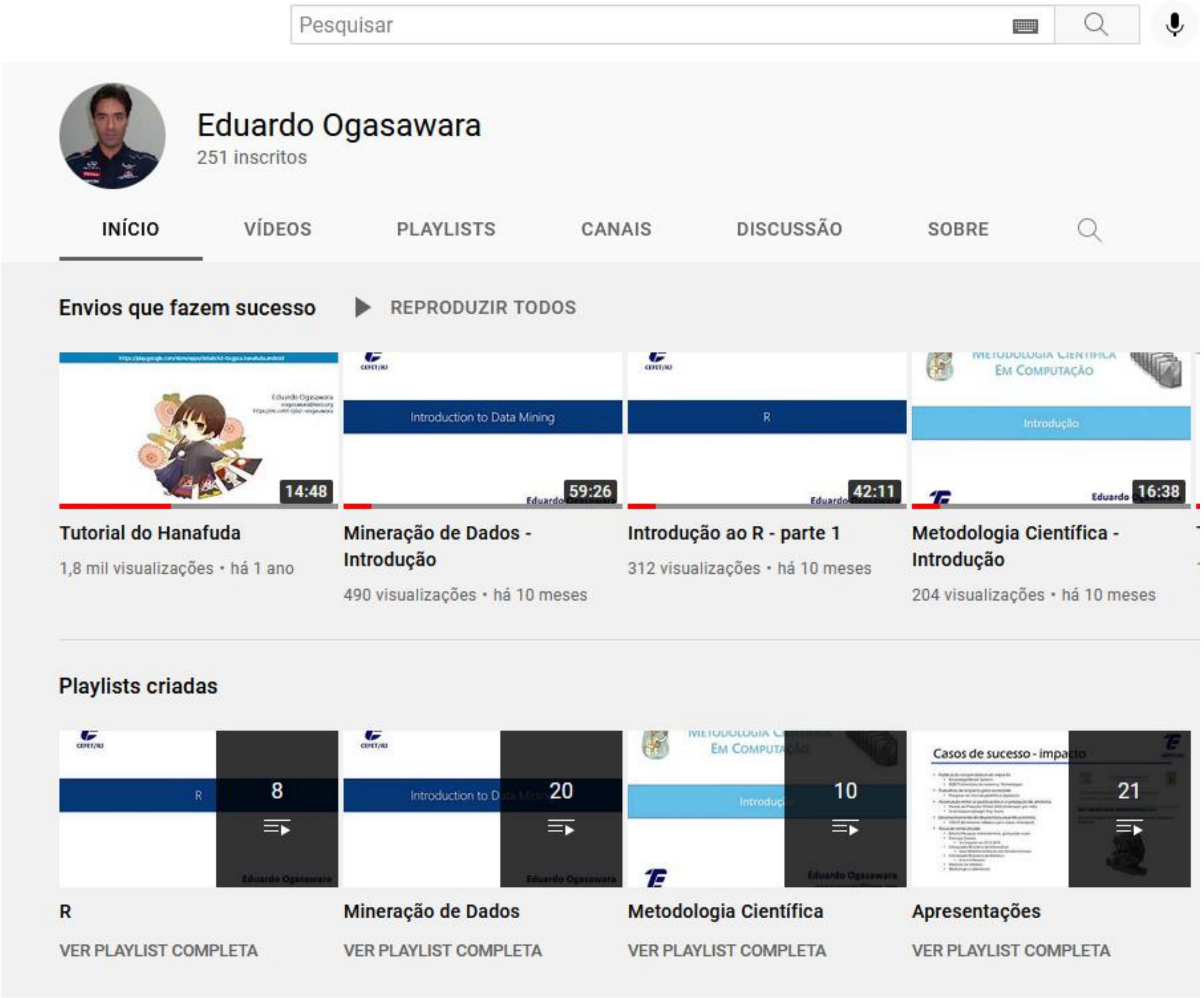
Data Analytics Lab

Follow!



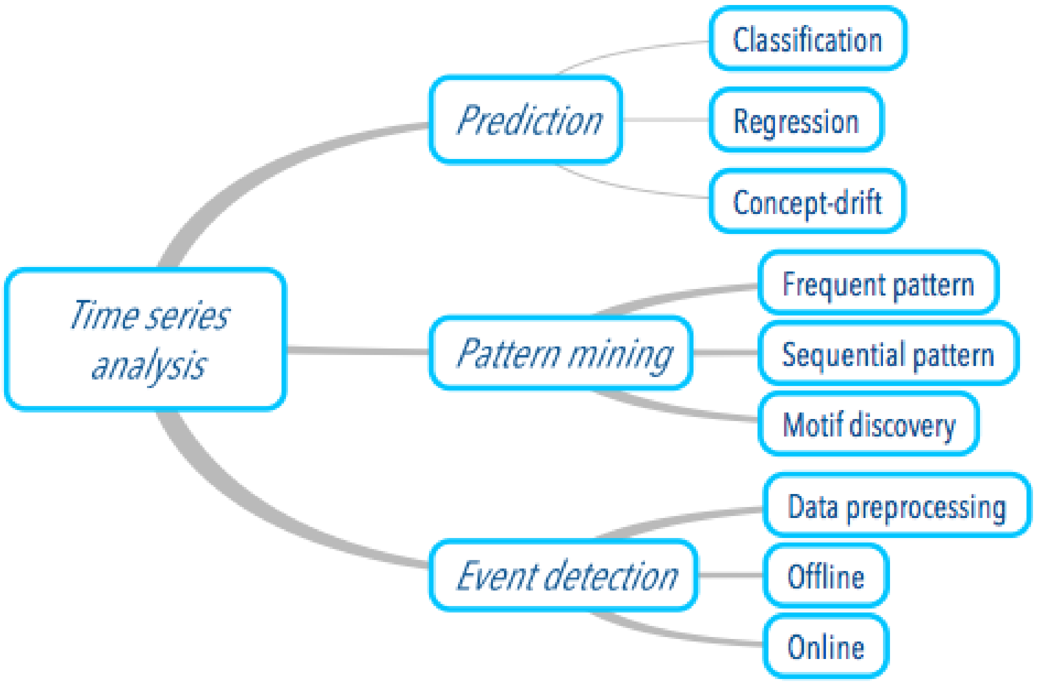
https://eic.cefet-rj.br/~dal/
http://dgp.cnpq.br/dgp/espelhogrupo/9806930220192669
https://www.youtube.com/channel/UCmn4Kh8fgl7VSM8X9t6qWmA

YouTube channel



like!

Research Themes



Let’s start

- A time series is a sequence of observations of a phenomenon of interest collected over time
  - $y = < y_1, y_2, \ldots, y_n >, |y| = n$


```
install.packages("astsa")
library(astsa)
plot(chicken)
```

[1] R.H. Shumway and D.S. Stoffer, 2017, Time Series Analysis and Its Applications: With R Examples. Springer.

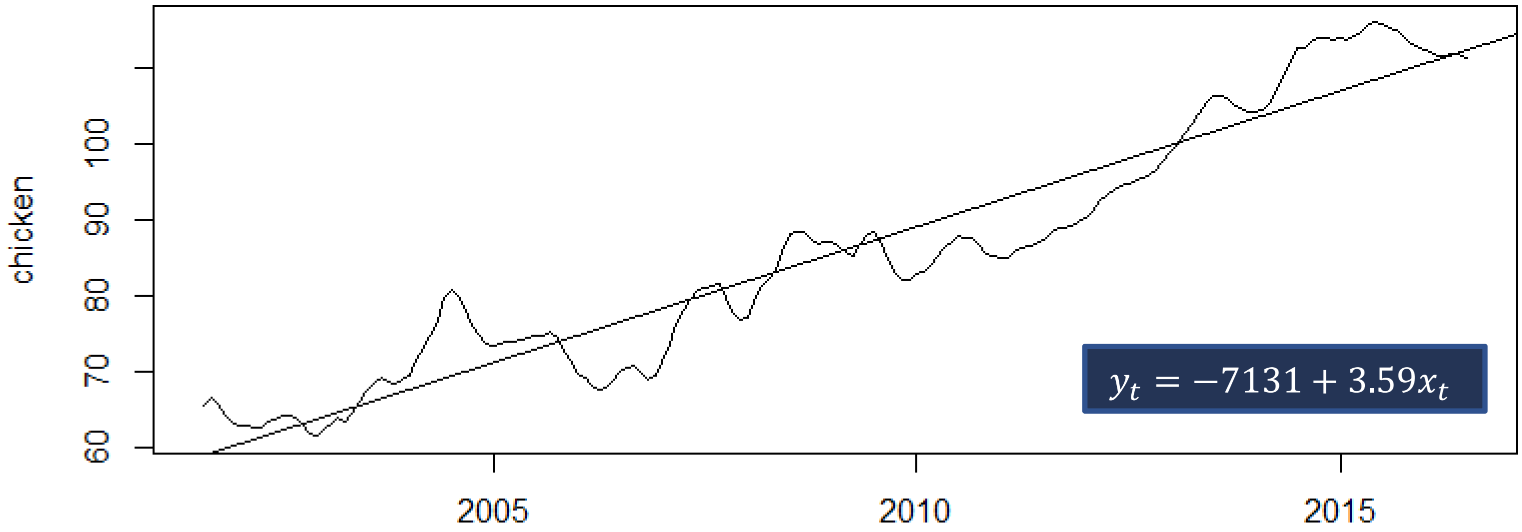
Linear regression

- $y_t = \beta_0 + \beta_1 x_t + \omega_t$
- Capture linear trend
- x_t can also be a time variable

```
summary(fit <- lm(chicken~time(chicken), na.action=NULL))

#                    Estimate  Std. Error  t.value
#    (Intercept)     -7131.02      162.41   -43.91
#  time(chicken)         3.59        0.08    44.43
#  ...
#  Residual standard error: 4.696 on 178 degrees of freedom

plot(chicken)
abline(fit)
```

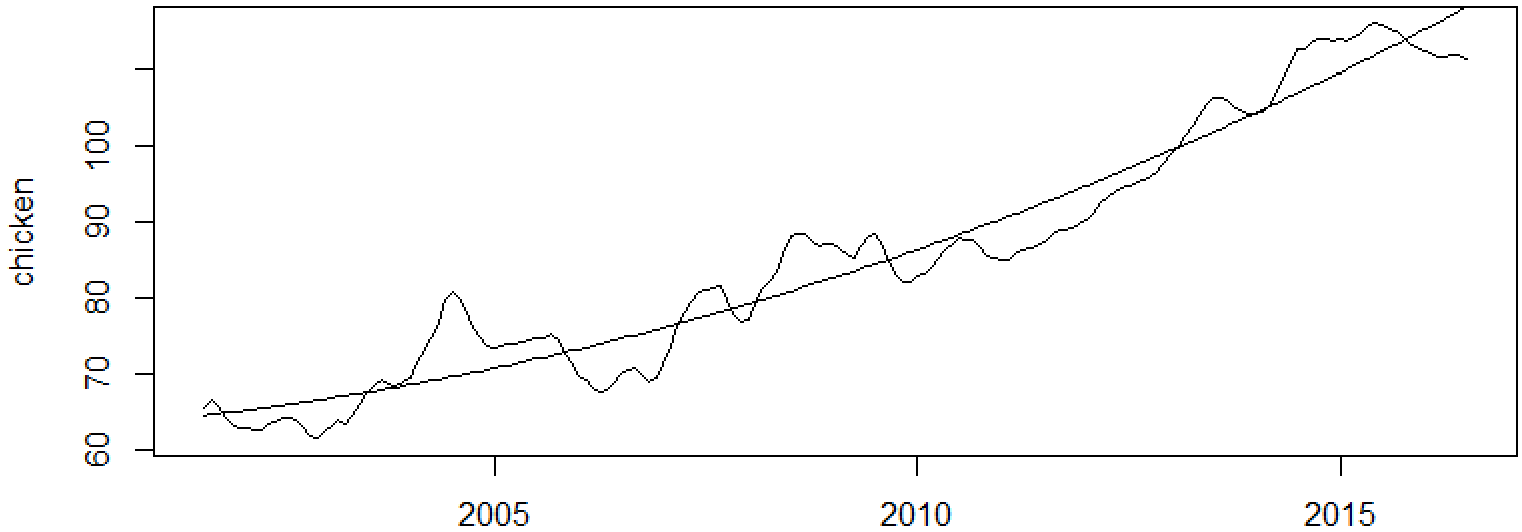


[1] R.H. Shumway and D.S. Stoffer, 2017, *Time Series Analysis and Its Applications: With R Examples*. Springer.
[2] R.J. Larsen and M.L. Marx, 2017, *An Introduction to Mathematical Statistics and Its Applications*. Pearson Education.
[3] D.N. Gujarati and D.C. Porter, 2008, *Basic Econometrics*. McGraw-Hill Publishing.

Polynomial regression

- $y_t = \beta_0 + \beta_1 x_t + \beta_2 {x_t}^2 + \cdots + \beta_n {x_t}^n + \omega_t$
- Capture other degree components

```
x <- time(chicken)
x2 <- x^2
summary(fit2 <- lm(chicken~x+x2, na.action=NULL))
#                  Estimate  Std. Error  t.value
#    (Intercept)   -611100       70560    8.661
#              x    -611.9       70.25   -8.711
#             x2    0.1532        0.02    8.762
#  ...
#  Residual standard error: 3.933 on 177 degrees of freedom
plot(chicken)
lines(fit2$fitted.values)
```



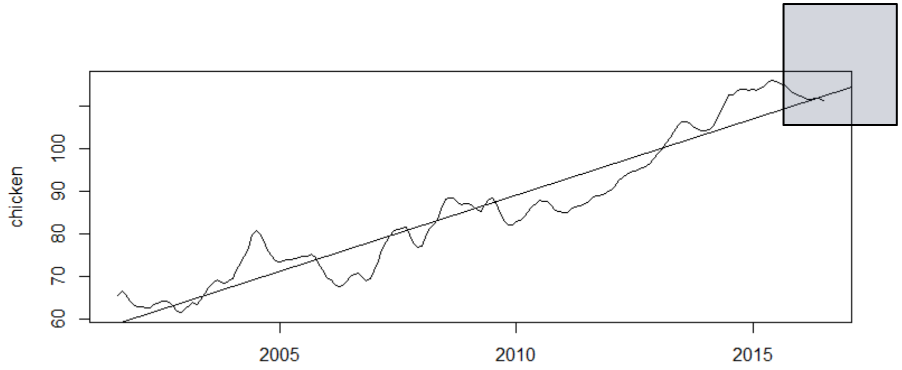
[1] R.H. Shumway and D.S. Stoffer, 2017, *Time Series Analysis and Its Applications: With R Examples*. Springer.
[2] R.J. Larsen and M.L. Marx, 2017, *An Introduction to Mathematical Statistics and Its Applications*. Pearson Education.
[3] D.N. Gujarati and D.C. Porter, 2008, *Basic Econometrics*. McGraw-Hill Publishing.

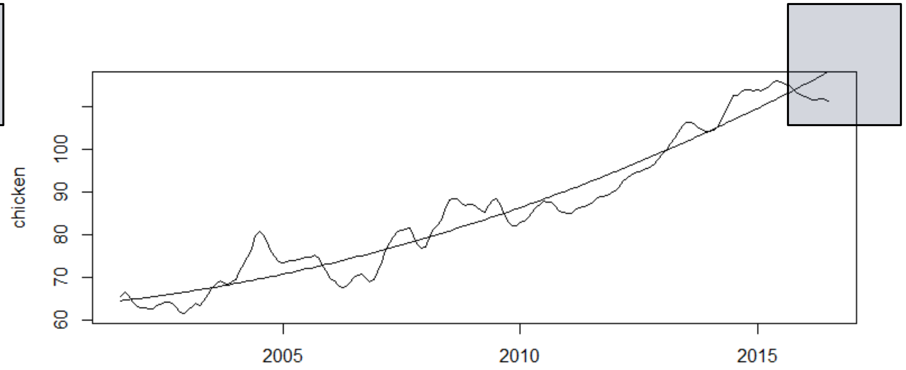
Increasing the complexity

- Theory is important to support other degrees

```
> anova(fit, fit2)
Analysis of Variance Table

Model 1: chicken ~ x
Model 2: chicken ~ x + x2
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    178 3925.9
2    177 2738.2  1    1187.7 76.773 1.527e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```





A detailed video explaining how to choose models is coming soon

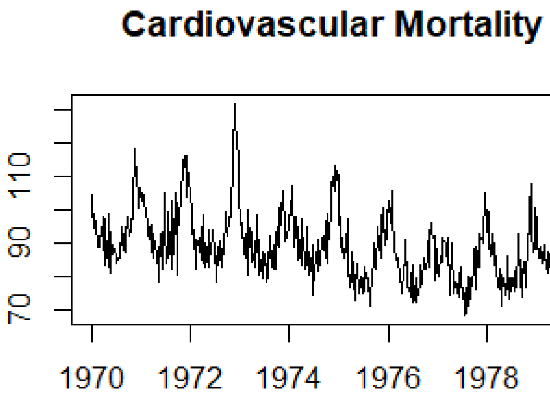
[1] R.H. Shumway and D.S. Stoffer, 2017, *Time Series Analysis and Its Applications: With R Examples*. Springer.

Multiple regression problem

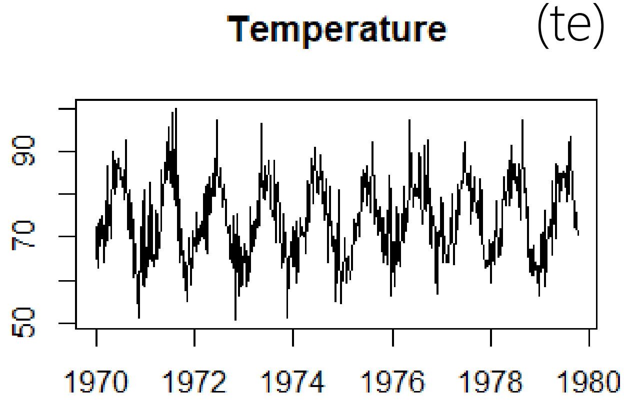
- $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_n x_{tn} + \omega_t$
- $x_{t1}, x_{t2}, \ldots, x_{tn}$ are independent variables
  - They were commonly theoretically established
- ω_t is an intrinsic error, noise variable

[1] R.H. Shumway and D.S. Stoffer, 2017, *Time Series Analysis and Its Applications: With R Examples*. Springer.
[2] R.J. Larsen and M.L. Marx, 2017, *An Introduction to Mathematical Statistics and Its Applications*. Pearson Education.

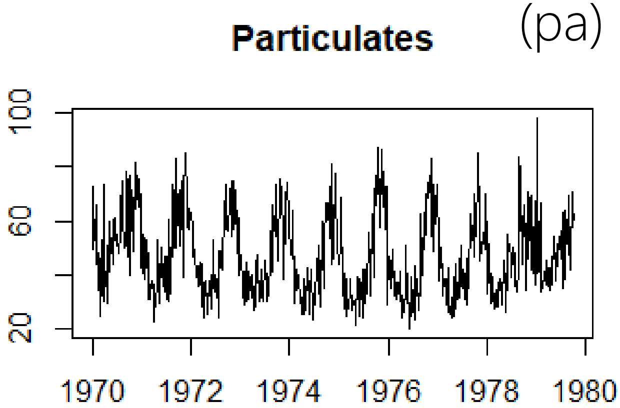
Cardiovascular mortality in Los Angeles



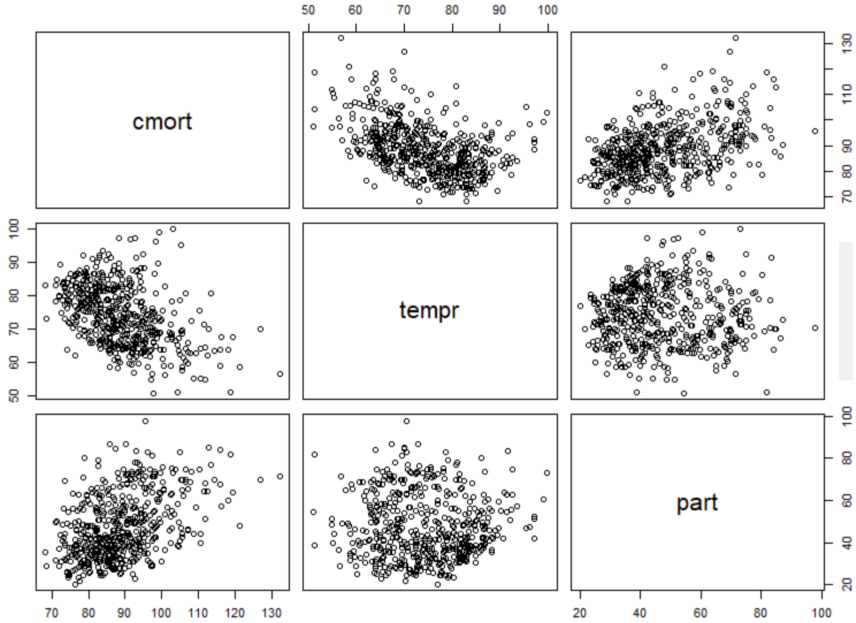
(c)



(te)



(pa)




```
plot(cmort, main="Cardiovascular Mortality", xlab="", ylab="")
plot(tempr, main="Temperature", xlab="", ylab="")
plot(part, main="Particulates", xlab="", ylab="")
pairs(data.frame(cmort,tempr,part))
```

[1] R.H. Shumway, A.S. Azari, and Y. Pawitan, 1988, Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects, *Environmental Research*, v. 45, n. 2 (Apr.), p. 224–241.

Model building

- Model 1: $c_t = \beta_0 + \beta_1 t + \omega_t$
- Model 2: $c_t = \beta_0 + \beta_1 t + \beta_2 (te_t - \overline{te}) + \beta_3 (te_t - \overline{te})^2 + \beta_4 pa_t + \omega_t$

```
temp = tempr-mean(tempr) # center temperature
temp2 = temp^2
trend = time(cmort) # time

fit = lm(cmort~ trend, na.action=NULL)
summary(fit)
summary(aov(fit))

fit2 = lm(cmort~ trend + temp + temp2 + part, na.action=NULL)
summary(fit2)
summary(aov(fit2))

anova(fit, fit2)
```

```
> anova(fit, fit2)
Analysis of Variance Table

Model 1: cmort ~ trend
Model 2: cmort ~ trend + temp + temp2 + part
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    506 40020
2    503 20508  3     19511 159.52 < 2.2e-16 ***
```

```
> num = length(cmort) # sample size
> AIC(fit)/num - log(2*pi) # AIC
[1] 5.37846
> BIC(fit)/num - log(2*pi) # BIC
[1] 5.403443
> (AICc = log(sum(resid(fit)^2)/num)
+    + (num+5)/(num-5-2)) # AICc
[1] 5.390601
>
> num = length(cmort) # sample size
> AIC(fit2)/num - log(2*pi) # AIC
[1] 4.721732
> BIC(fit2)/num - log(2*pi) # BIC
[1] 4.771699
> (AICc = log(sum(resid(fit2)^2)/num)
+    + (num+5)/(num-5-2)) # AICc
[1] 4.722062
```

A detailed video explaining how to choose models is coming soon

Regression with lagged values

- Independent variables can be lagged versions of y_t
  - $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_n y_{t-n} + \omega_t$
  - ω_t is an intrinsic error, noise variable
- Open room for data-driven models

[1] A. Ronald, R. Salles, K. Belloze, D. Pastore, and E. Ogasawara, 2019, Modelo Autorregressivo de Integração Adaptativa, In: *Anais do Simpósio Brasileiro de Banco de Dados (SBBD)*, p. 175–180
[2] R.S. Tsay, 2010, *Analysis of Financial Time Series*. Wiley.

Autoregressive Integrated Moving Average

- ARIMA(p, d, q)
  - AR(p)
    - $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \omega_t$
  - MA (q)
    - $y_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \cdots + \theta_n \omega_{t-q}$
  - Differentiation (d)

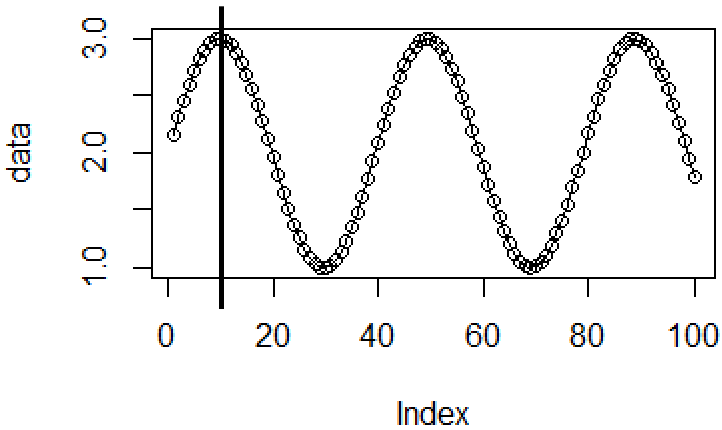
A detailed video explaining AR and MA is coming soon

[1] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung, 2015, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Subsequences and sliding windows

- Subsequence is a continuous sample of a time series
  - $seq_{p,i}(y) = < y_i, y_{i+1}, \dots, y_{i+p-1} >$
    - $|seq_{p,i}(y)| = p$
    - $1 \leq i \leq |y| - p$
- Sliding window explores all subsequences of a time series
  - $sw_p(y) = A$
    - $\forall a_i \in A, a_i = seq_{p,i}(y)$

	t0
1	2.158484
2	2.312962
3	2.459529
4	2.594481
5	2.714406
6	2.816273
7	2.897507
8	2.956056
9	2.990438
10	2.999785



	t4	t3	t2	t1	t0
1	2.158484	2.312962	2.459529	2.594481	2.714406
2	2.312962	2.459529	2.594481	2.714406	2.816273
3	2.459529	2.594481	2.714406	2.816273	2.897507
4	2.594481	2.714406	2.816273	2.897507	2.956056
5	2.714406	2.816273	2.897507	2.956056	2.990438
6	2.816273	2.897507	2.956056	2.990438	2.999785

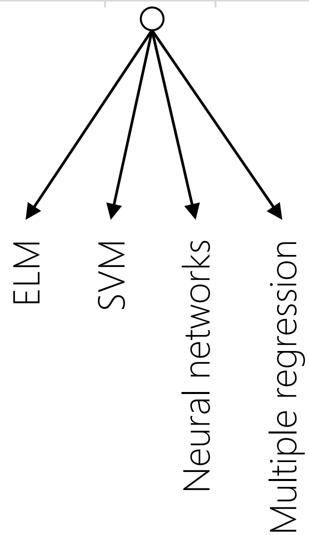
Sliding window of size 5

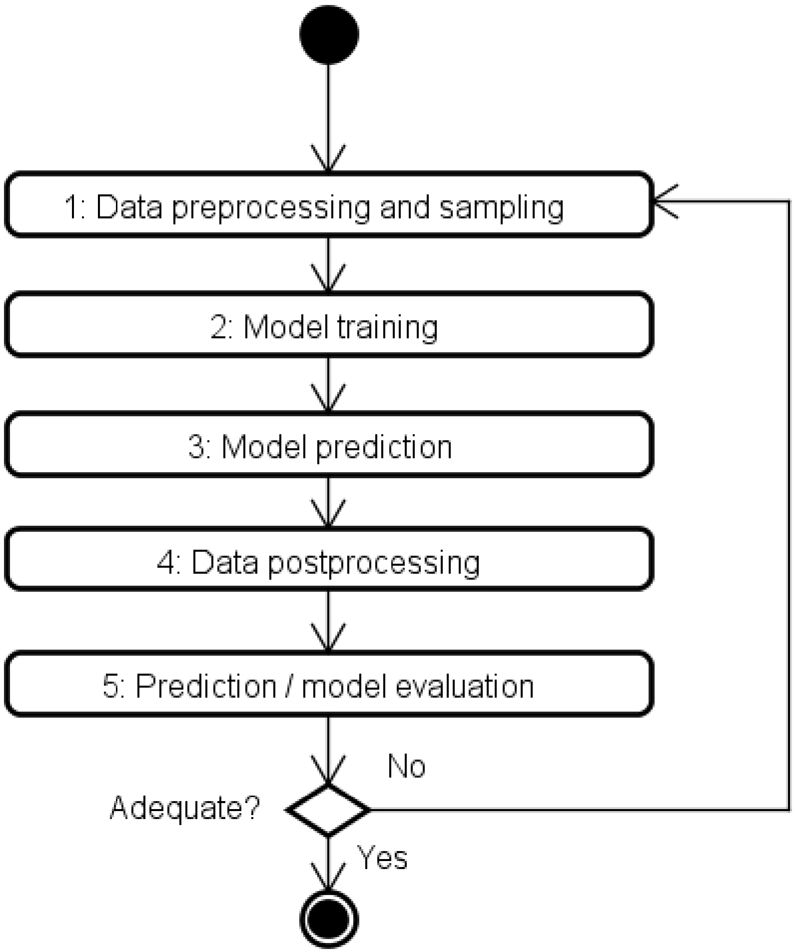
[1] H. Borges, M. Dutra, A. Bazaz, R. Coutinho, F. Perosi, F. Porto, F. Masseglia, E. Pacitti, and E. Ogasawara, 2020, Spatial-time motifs discovery, *Intelligent Data Analysis*, v. 24, n. 5, p. 1121–1140.

Prediction using sliding windows (lagged terms)
Mining process

input output

	t4	t3	t2	t1	t0
1	2.158484	2.312962	2.459529	2.594481	2.714406
2	2.312962	2.459529	2.594481	2.714406	2.816273
3	2.459529	2.594481	2.714406	2.816273	2.897507
4	2.594481	2.714406	2.816273	2.897507	2.956056
5	2.714406	2.816273	2.897507	2.956056	2.990438
6	2.816273	2.897507	2.956056	2.990438	2.999785



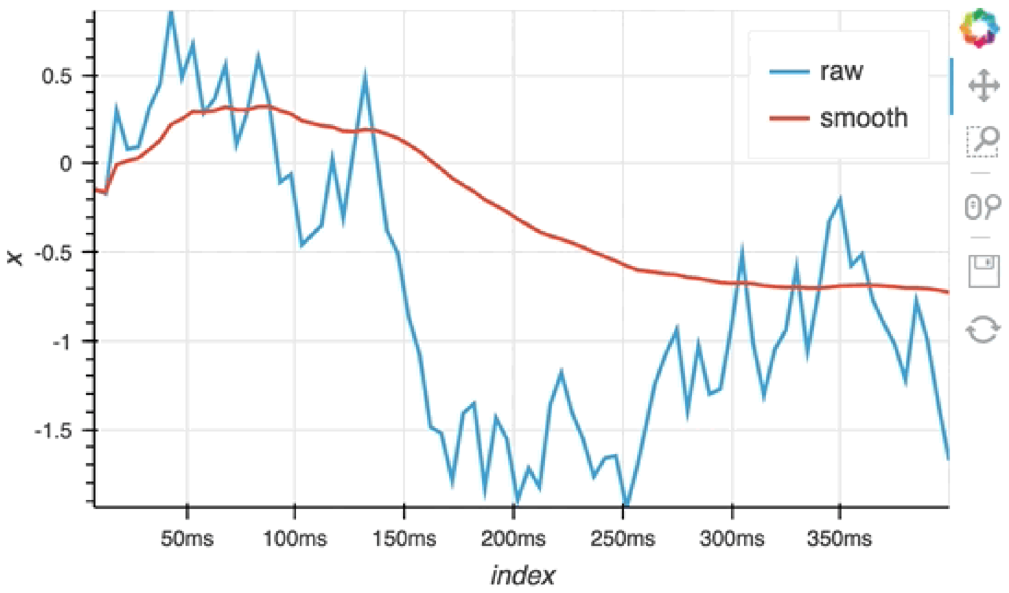


A detailed video explaining mining process is coming soon

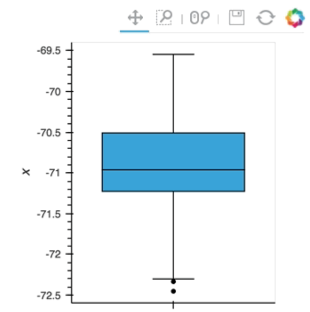
[1] R. Salles, L. Assis, G. Guedes, E. Bezerra, F. Porto, and E. Ogasawara, 2017, A framework for benchmarking machine learning methods using linear models for univariate time series prediction, In: *Proceedings of the International Joint Conference on Neural Networks*, p. 2338–2345

Nonstationarity

- Statistical properties may vary over time
  - $\chi(\hat{y}_s) \neq \chi(\hat{y}_t)$



#	index	x
0	2017-10-27 19:43:04	-70.87262710547351
1	2017-10-27 19:43:04	-71.00788295730518
2	2017-10-27 19:43:04	-71.01297392873352
3	2017-10-27 19:43:04	-71.4148637783796
4	2017-10-27 19:43:04	-71.60890969520968
5	2017-10-27 19:43:04	-71.32610485802545
6	2017-10-27 19:43:04	-71.1935680343768
7	2017-10-27 19:43:04	-70.76579342173753
8	2017-10-27 19:43:04	-70.59743524950701
9	2017-10-27 19:43:04	-70.99300214112863



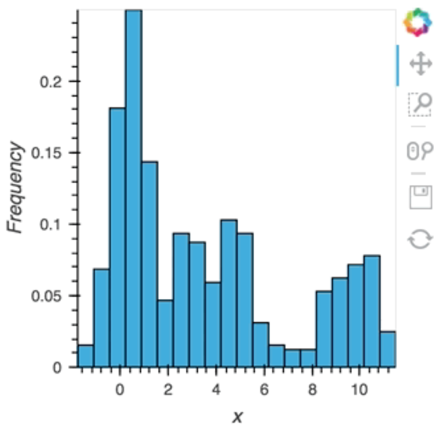


Image source: http://holoviews.org/user_guide/Streaming_Data.html

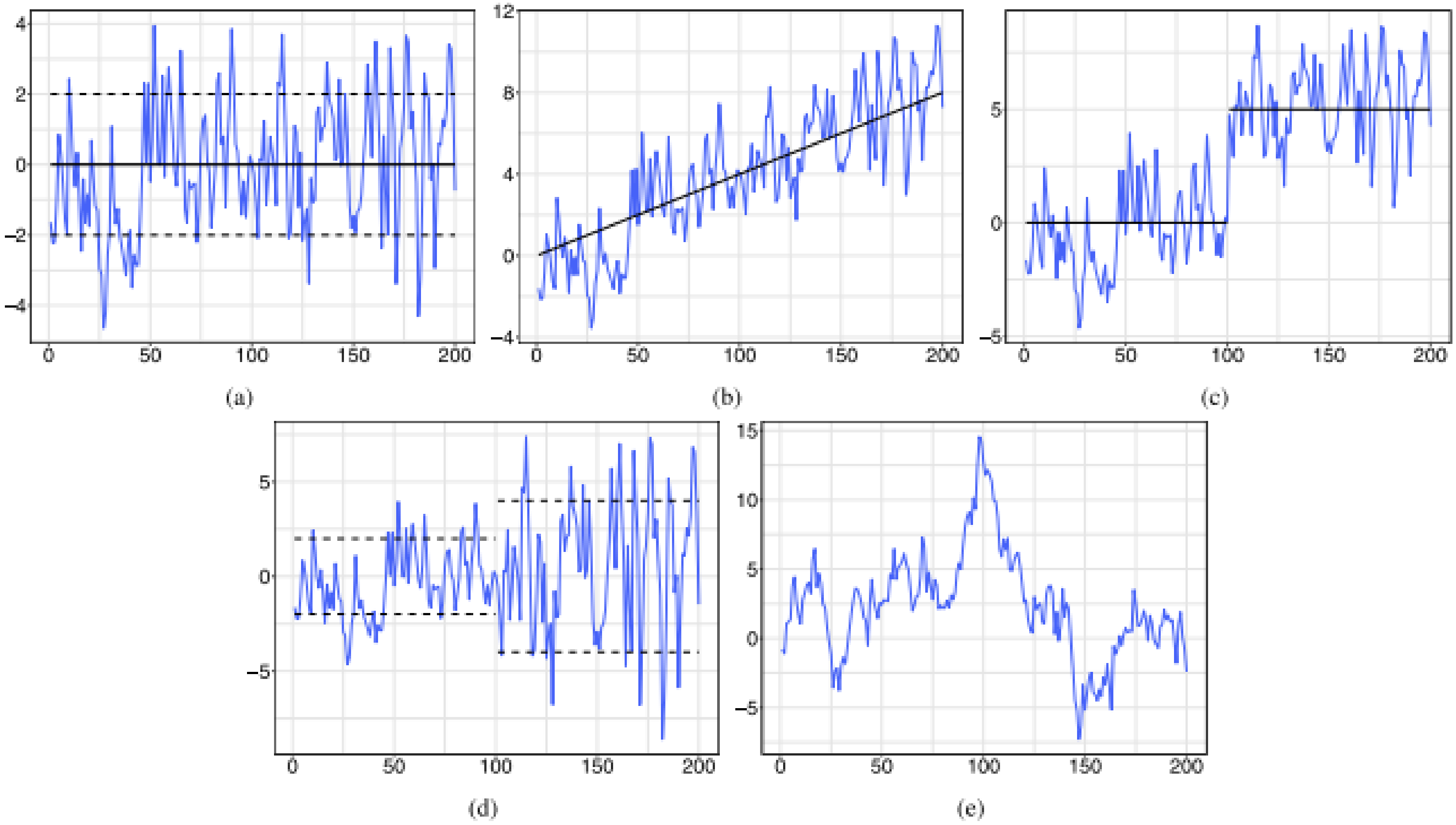
[1] G.I. Webb, R. Hyde, H. Cao, H.L. Nguyen, e F. Petitjean, 2016, Characterizing concept drift, *Data Mining and Knowledge Discovery*, v. 30, n. 4, p. 964–994.

Stationarity

- Stationarity
  - Time series y
  - Samples $\hat{y}_s$ from y
  - Statistical properties in $\hat{y}_s$ do not vary over time
    - Mean $\mu(\hat{y}_s) \cong \mu(\hat{y}_t)$
    - Variance:$\sigma^2(\hat{y}_s) \cong \sigma^2(\hat{y}_t)$
    - Covariance: $cov(\hat{y}_s, \hat{y}_{s+d}) \cong cov(\hat{y}_t, \hat{y}_{t+d})$
- Non-stationarity
  - When stationary does not hold

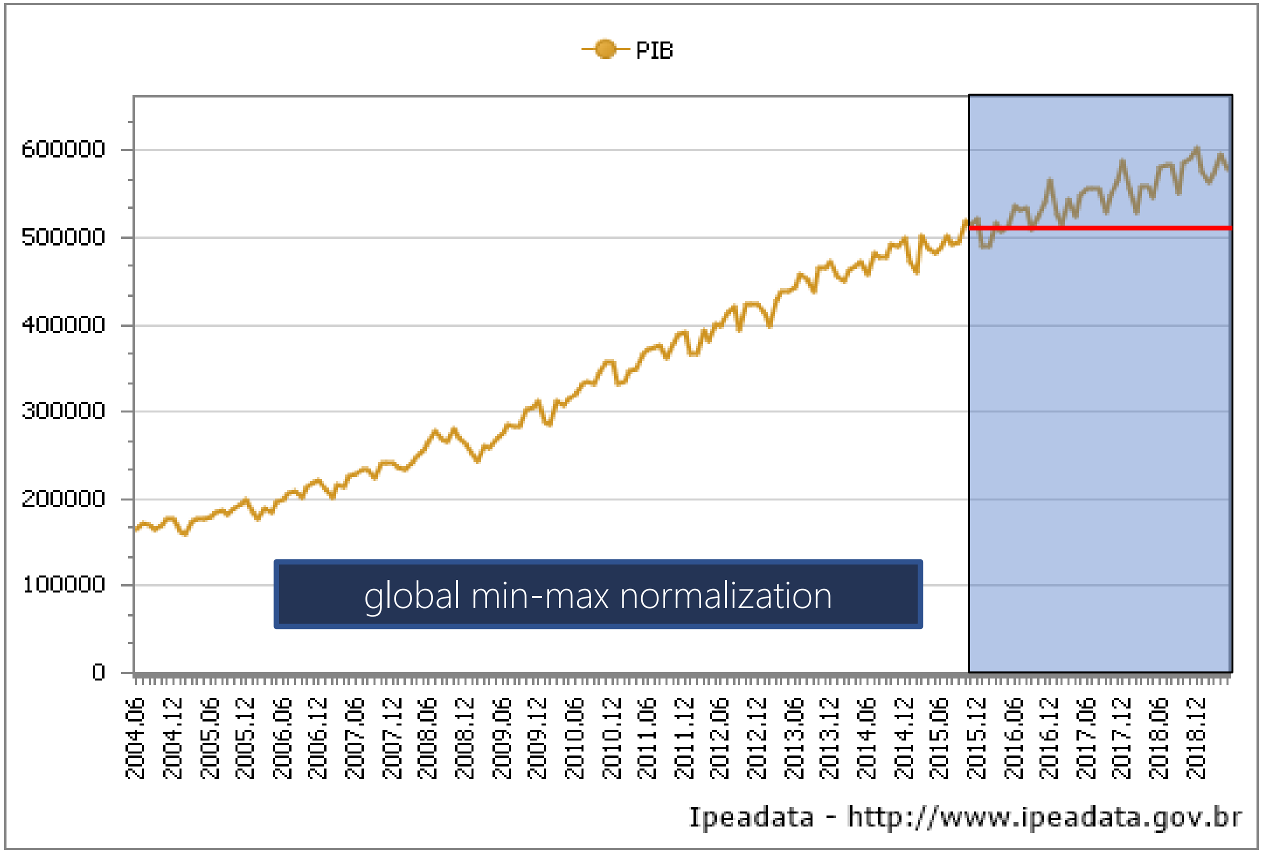
[1] R.H. Shumway e D.S. Stoffer, 2017, Time Series Analysis and Its Applications: With R Examples. 4 ed. New York, NY, Springer.

Stationarity and non-stationary time series



Drawbacks of non-stationarity

- Most data analytics methods implicitly assume stationarity



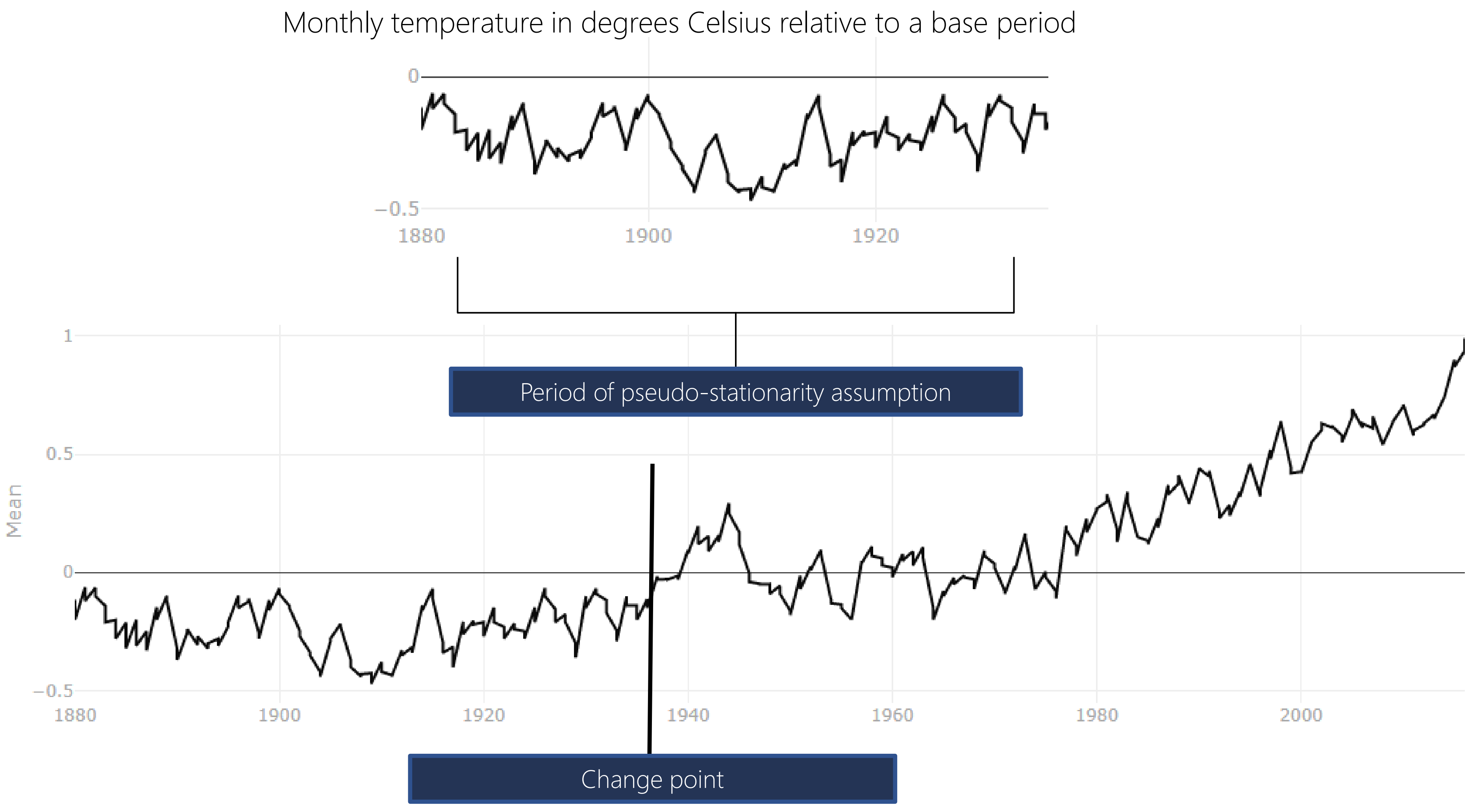
[1] E. Ogasawara, L.C. Martinez, D. De Oliveira, G. Zimbrão, G.L. Pappa, e M. Mattoso, 2010, Adaptive Normalization: A novel data normalization approach for non-stationary time series, In: Proceedings of the International Joint Conference on Neural Networks

Possible solutions

- Assumption of stationarity
- Adaptability
  - Drift detection
  - Memory management
- Transformation methods

[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4
[2] G. Ditzler, M. Roveri, C. Alippi, e R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, v. 10, n. 4, p. 12–25.
[3] R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, e E. Ogasawara, "Nonstationary time series transformation methods: An experimental review", *Knowledge-Based Systems*, nov. 2018.

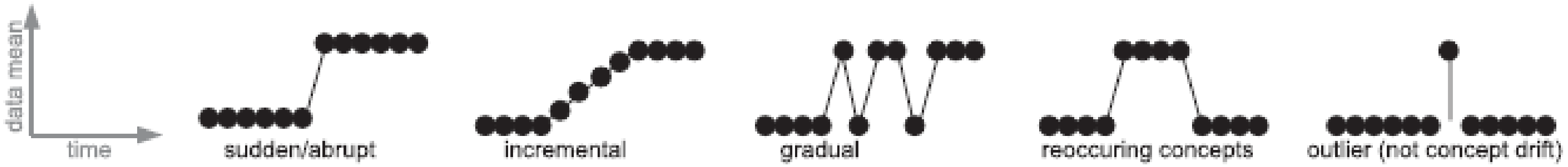
Assumption of stationarity



Adaptability

- Some machine learning methods (e.g., neural networks) are known for adaptability
  - Ability to update model due to changes in the environment
  - Incremental training
- Adaptive systems aim to address non-stationarity
  - Seeking robustness, adaptability is adopted
  - Greater adaptability, more susceptible to spurious situations, less robust
  - Dilemma: finding the right time to adapt

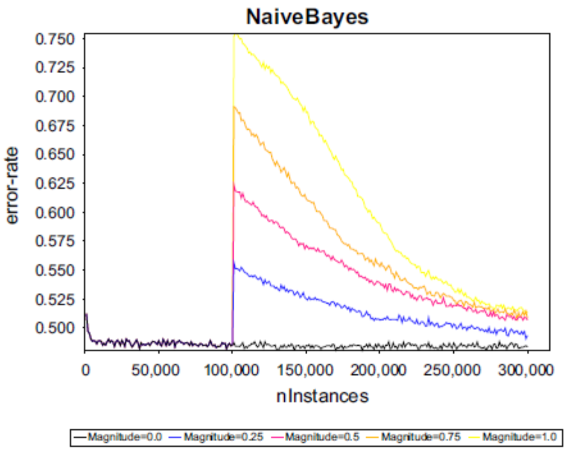
Plasticity-stability dilemma

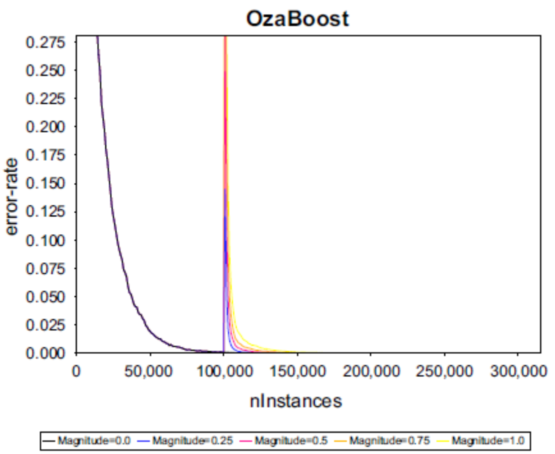


[1] S.O. Haykin, 2008, *Neural Networks and Learning Machines*. 3 ed. New York, Prentice Hall.
[2] Grossberg, S., 1988. Neural Networks and Natural Intelligence, Cambridge, MA: MIT Press.
[3] G. Ditzler, M. Roveri, C. Alippi, e R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, v. 10, n. 4, p. 12–25.

Drift detection

- Drift detection
  - Active
  - Passive
- Learning
  - Incremental
  - Non-incremental
- Models
  - Single
  - Ensemble (Boosting)





[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4
[2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1
[3] G.I. Webb, R. Hyde, H. Cao, H.L. Nguyen, e F. Petitjean, 2016, Characterizing concept drift, *Data Mining and Knowledge Discovery*, v. 30, n. 4, p. 964–994.

Lucas critique

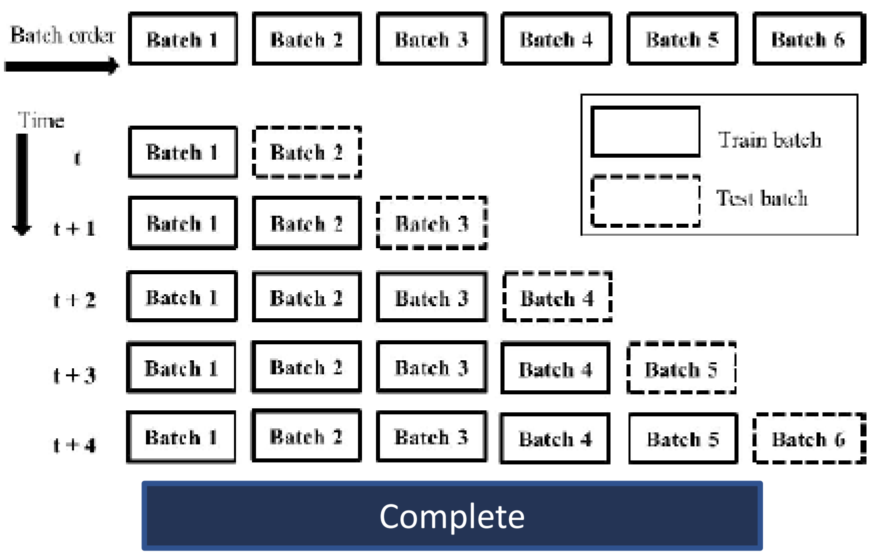
- "Given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models."

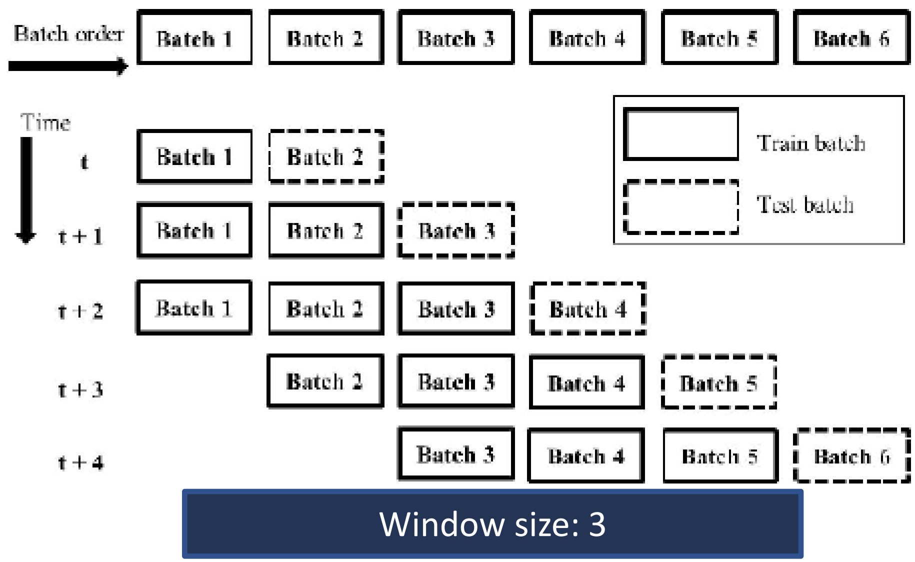
Goes toward memory management

[1] D. Gujarati, 2002, *Basic Econometrics*. 4 ed. Boston; Montreal, McGraw-Hill/Irwin.

Memory management

- Process
  - It is tested in the last batch (forecast)
  - The last batch is incorporated in the training
- Memory
  - complete
  - Without memory
  - sliding windows

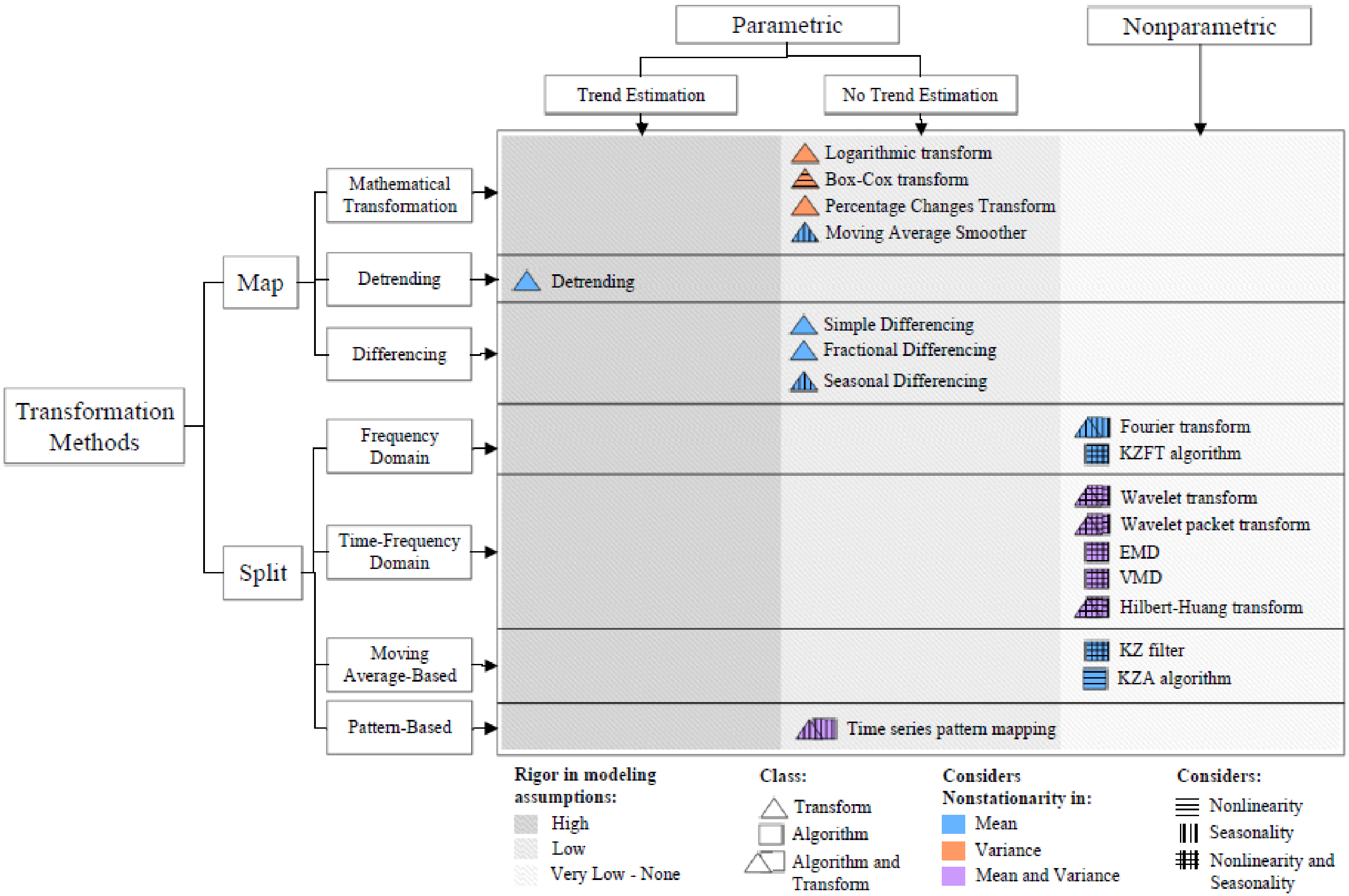




[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4

[2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1

Transformation methods

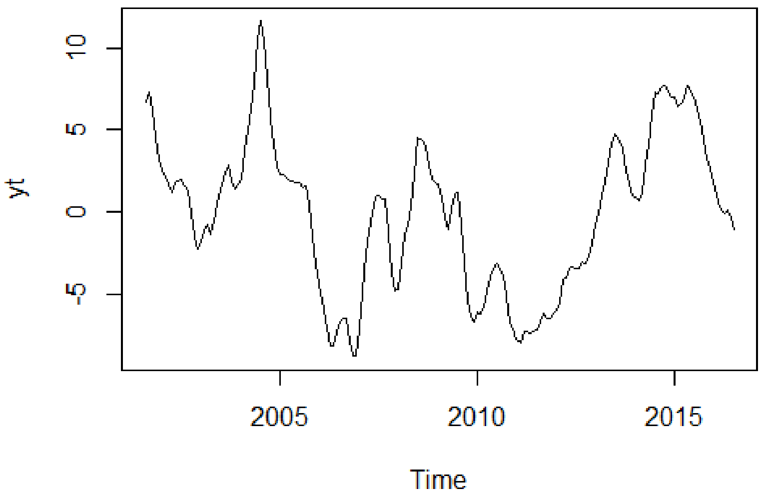


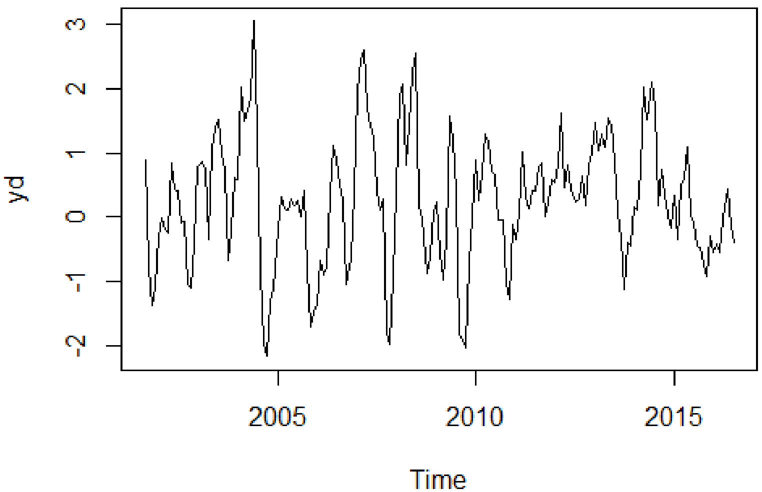
A detailed video explaining transformations is coming soon

R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, e E. Ogasawara, “Nonstationary time series transformation methods: An experimental review”, *Knowledge-Based Systems*, nov. 2018.

Detrending and differentiation

- Detrending
  - $\hat{y}_t = y_t - (\beta_0 + \beta_1 x_t)$
- Differentiation (d)
  - First order differentiation ($\mathbf{d} = 1$)
    - $\hat{y}_t = \nabla y_t = y_t - y_{t-1}$
  - General order differentiation ($\mathbf{d} > 1$)
    - $\nabla^d = (1 - B)^d, B^k y_t = y_{t-k}$

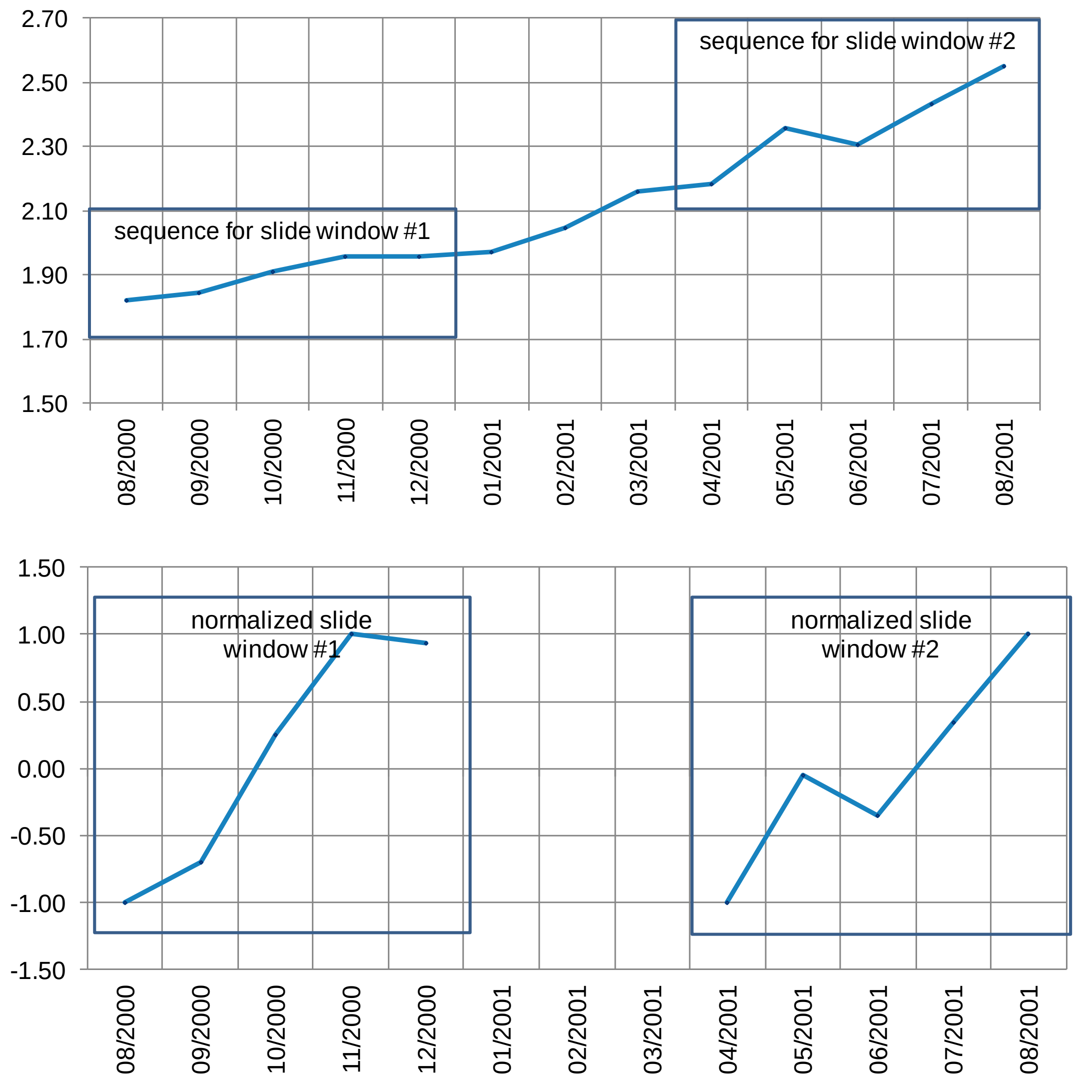




A detailed video explaining detrending and differentiation is coming soon

R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, e E. Ogasawara, “Nonstationary time series transformation methods: An experimental review”, Knowledge-Based Systems, nov. 2018.

Normalization issues using sliding windows



Monthly average exchange rate of U.S. Dollar to Brazilian Real
normalized by sliding window technique from aug/2000 to dec/2000 and from apr/2001 to aug/2001

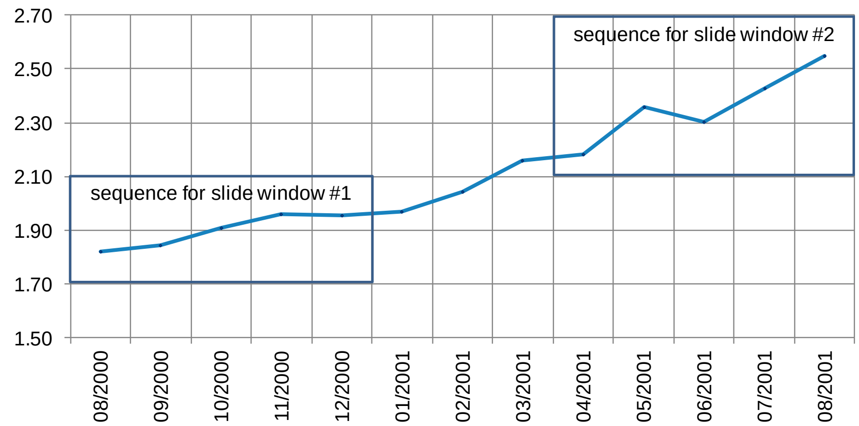
Adaptive Normalization

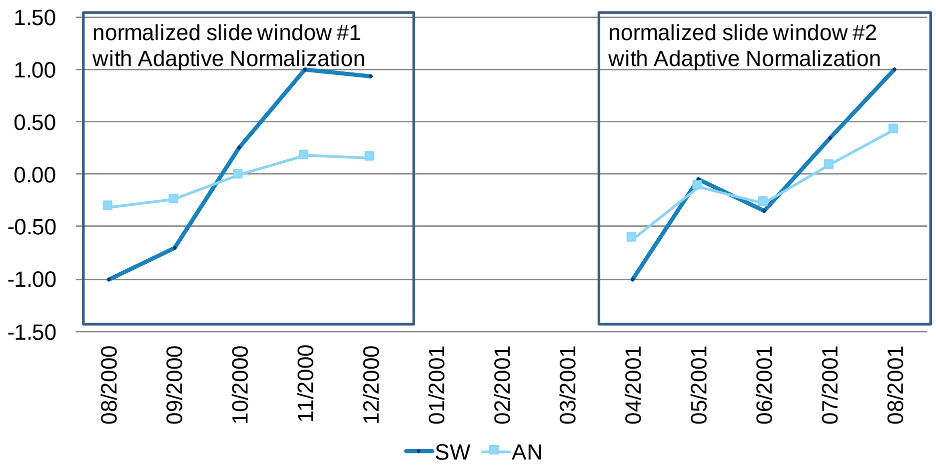
- Transformation
  - Using sliding windows
  - Compute moving average (inertia)
  - Remove inertia
  - Outlier removal
  - Sliding window min-max
- Inverse transform
  - Prediction
  - Denormalization
  - Add inertia

A detailed video explaining AN is coming soon

[1] E. Ogasawara, L.C. Martinez, D. De Oliveira, G. Zimbrão, G.L. Pappa, and M. Mattoso, 2010, Adaptive Normalization: A novel data normalization approach for non-stationary time series, In: *Proceedings of the International Joint Conference on Neural Networks*

Intuition

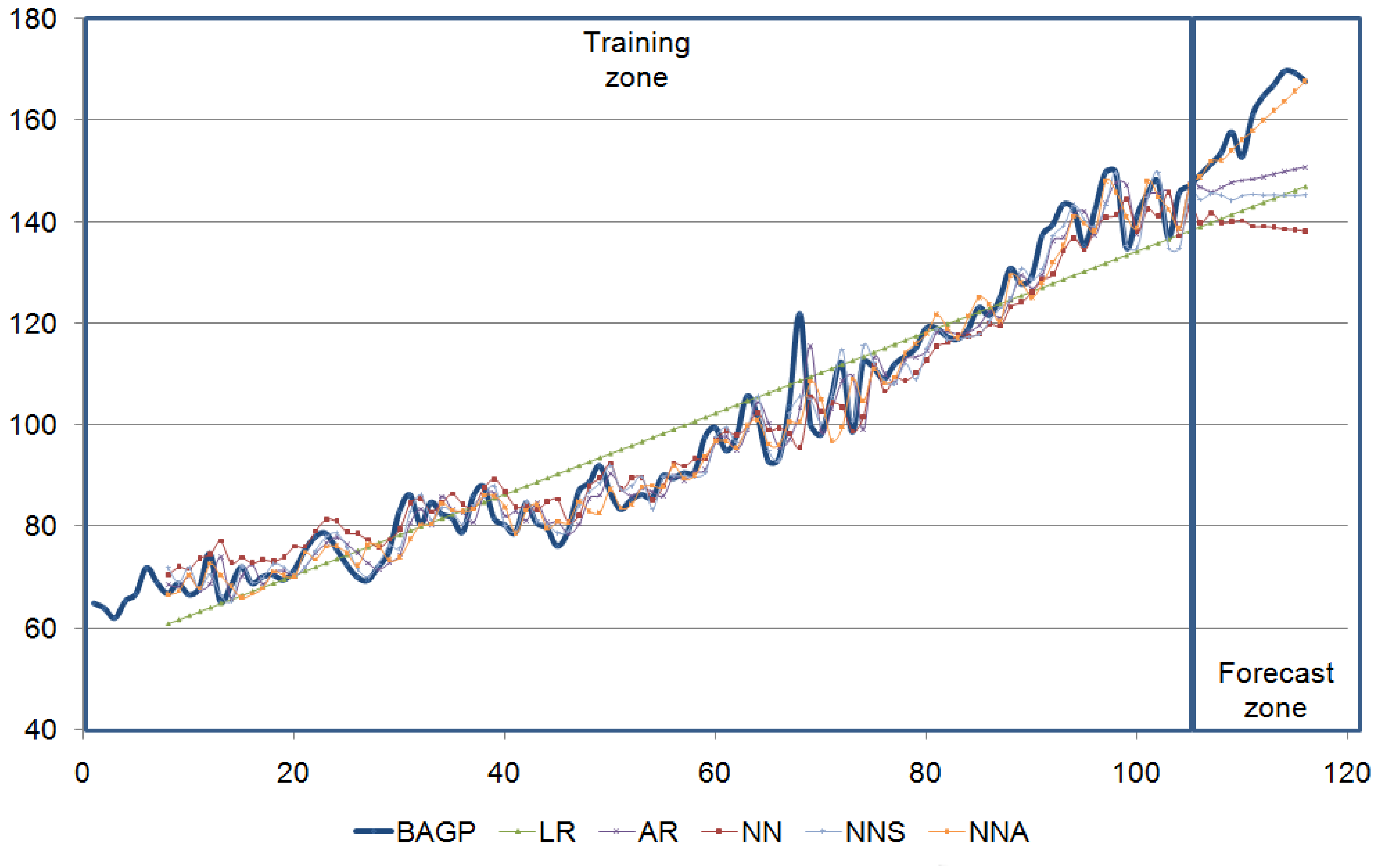




A detailed video explaining AN is coming soon

[1] E. Ogasawara, L.C. Martinez, D. De Oliveira, G. Zimbrão, G.L. Pappa, and M. Mattoso, 2010, Adaptive Normalization: A novel data normalization approach for non-stationary time series, In: *Proceedings of the International Joint Conference on Neural Networks*

Different preprocessing and prediction methods



[1] E. Ogasawara, L.C. Martinez, D. De Oliveira, G. Zimbrão, G.L. Pappa, and M. Mattoso, 2010, Adaptive Normalization: A novel data normalization approach for non-stationary time series, In: *Proceedings of the International Joint Conference on Neural Networks*

Road map

Next videos

- Linear model fitting
- Linear model selection
- Trends and Differentiation
- Seasonal Adjustment
- Spectral Analysis
- Smoothing and Filtering
- Autocorrelation
- ARIMA
- GARCH
- State Space Models
- Sliding windows and normalization
- Adaptive normalization
- Machine learning models
- Data sampling
- Mining process
- Performance evaluation
  - evaluation on a rolling forecasting origin (time series cross validation)

https://www.youtube.com/channel/UCAm1hAXWEqYJfXz4EzzBhVg

[1] R.J. Hyndman and G. Athanasopoulos, 2018, *Forecasting: principles and practice*. OTexts.
[2] R.H. Shumway and D.S. Stoffer, 2017, *Time Series Analysis and Its Applications: With R Examples*. Springer.

Students

D.Sc.

Juan Fabian (LNCC)

Lais Baroni (CEFET/RJ)

Leonardo Carvalho (CEFET/RJ)

Paulo Elias (UFF)

Rebecca Salles (CEFET/RJ)

M.Sc.

Cristiane Gea (CEFET/RJ)

Janio Lima (CEFET/RJ)

Lucas Giusti (CEFET/RJ)

Tacito Braga (CEFET/RJ)

Arthur Severiano
Diego Sá
Flavio Marques

CEFET/RJ

TIME SERIES PREDICTION

Eduardo Ogasawara
eogasawara@ieee.org
https://eic.cefet-rj.br/~eogasawara