

# Modified segmentation methods of quasi-stationary time series

**Irina Roslyakova**

Department of Scale Bridging Thermodynamic and Kinetic Simulation  
ICAMS (Interdisciplinary Centre for Advanced Materials Simulation)

Ruhr-Universität Bochum, UHW 10/1022  
Stiepeler Str. 129, 44801 Bochum

Tel.: +49 234 32 22449, Mobil: +49 (163) 196 3327  
E-mail: [irina.roslyakova@rub.de](mailto:irina.roslyakova@rub.de)

ICAMS



# Contents

- Problem importance
- Method of Pedro Bernaola-Galván et al.
- Modified segmentation algorithm
- Comparison: Modified segmentation algorithm vs. R-Package “*strucchange*”
- Conclusion

This work is part of my master thesis performed in BASF SE, Germany



# Problem importance

- Many time series are a sequence of stationary intervals
- The statistical analysis of such quasi-stationary processes requires a division of the measurements into different stationary time segments.
- The segmentation of quasi-stationary time series is a tedious computational problem
- For large samples statistical methods will be too time consuming. Heuristics have to be applied.
- Possible application
  - Medicine
  - Chemical production processes
  - Internet traffic fluctuations

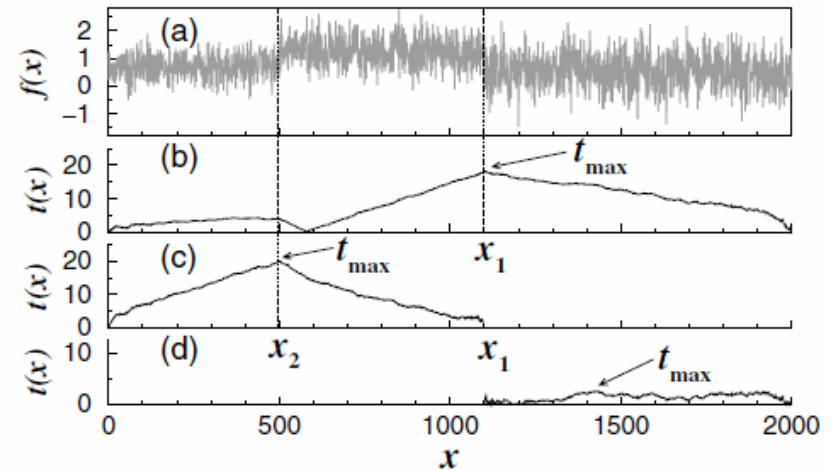
# Method of Pedro Bernaola-Galván et al.

Statistic  $t$  for comparison mean value of two samples is calculated for every position of sliding point  $p$

from start of interval

$$t_i(p) = \left| \frac{m_x^{left} - m_x^{right}}{S_D} \right|, \quad S_D \text{ is the pooled variance}$$

till the end of interval



DOI: 10.1103/PhysRevLett.87.168105

Point  $p$  with maximal difference in mean value is checked by using modified T-Test

$$P(t_{\max}) \approx \left\{ 1 - I_{[v/(v+t_{\max}^2)]}(\delta v, \delta) \right\}^\eta$$

where  $\nu = N - 2$  is the number of degrees of freedom,  $I_x(a, b)$  is the incomplete beta function

Significance of cutting point  $p$  is checked under condition that minimal length of subintervals have to be not less than some used defined value

$$P(t_{\max}) \geq P_0 = 0.95 \quad \text{and} \quad \begin{cases} l^{left} \geq l_0 \\ l^{right} \geq l_0 \end{cases} \quad \text{we cut the series at point } \boxed{p_c = p(t_{\max})}$$

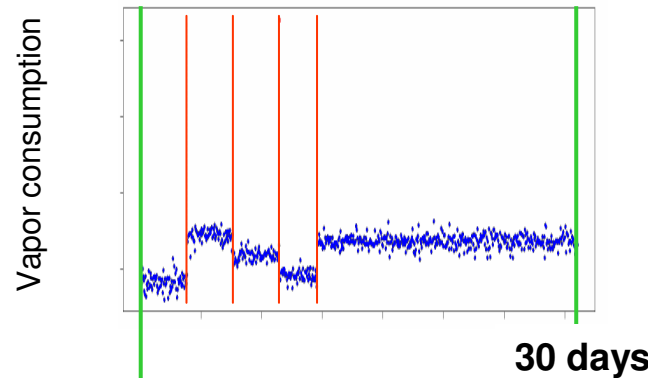


# Method of Pedro Bernaola-Galván et al.

## Analysis of the vapor consumption in a chemical production

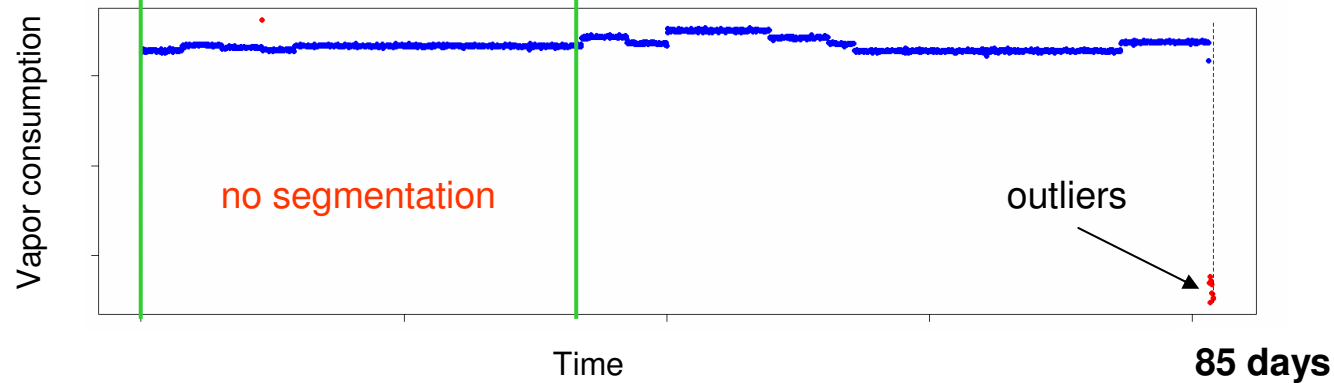
Observation window:

30 days



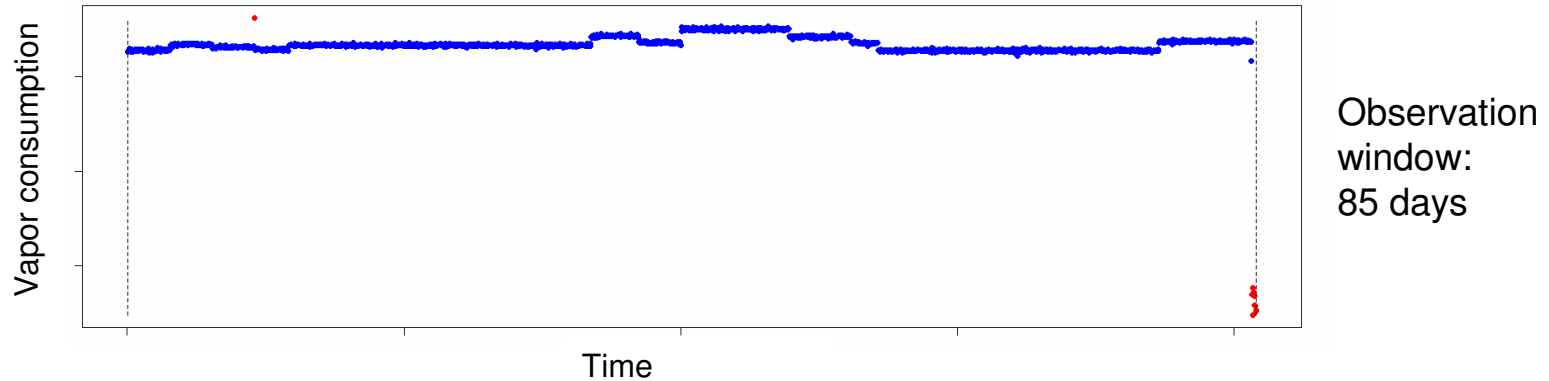
Method is sensitive but not robust against significant outliers near boundaries

85 days

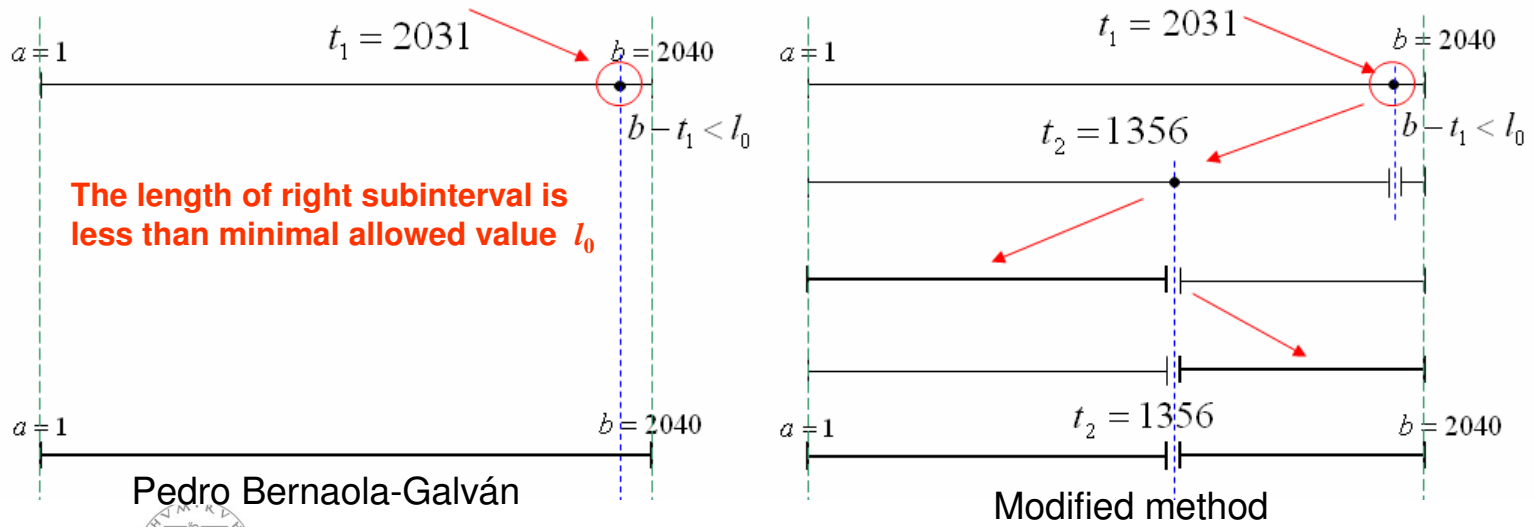


# Modified segmentation algorithm

- Bernaola-Galván's method is sensitive but not robust against significant outliers near boundaries



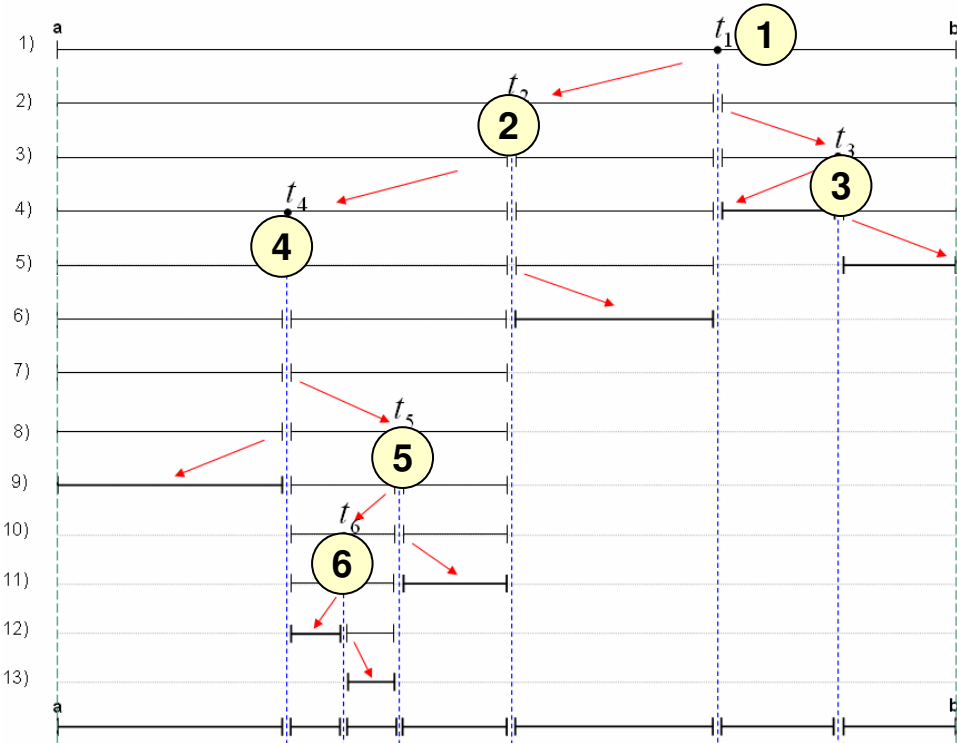
- Modification to make algorithm robust against outliers near boundaries



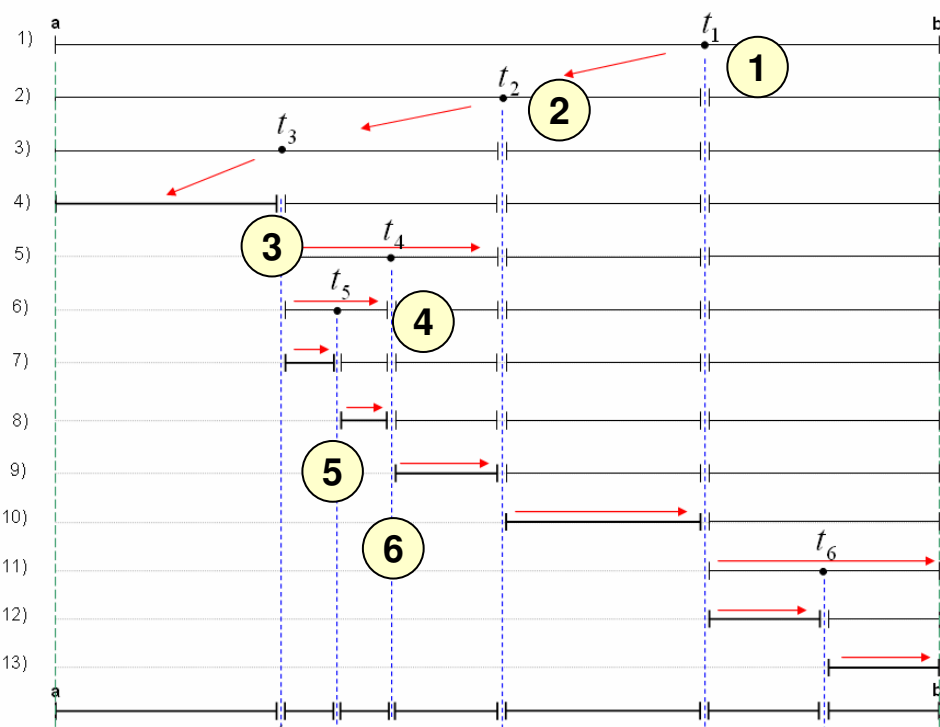


# Segmentation Steps in Comparison

Pedro Bernaola-Galván



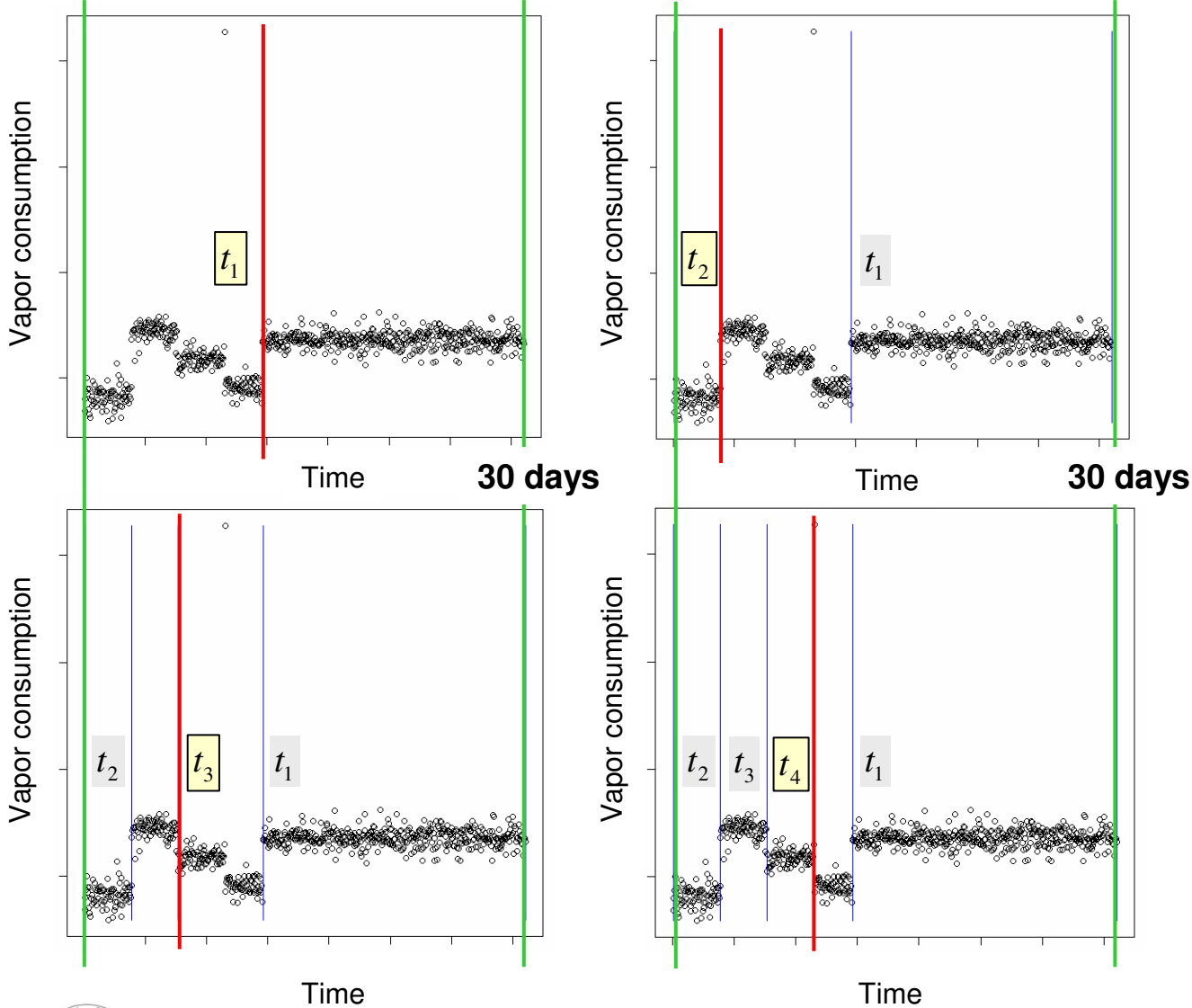
Modified method





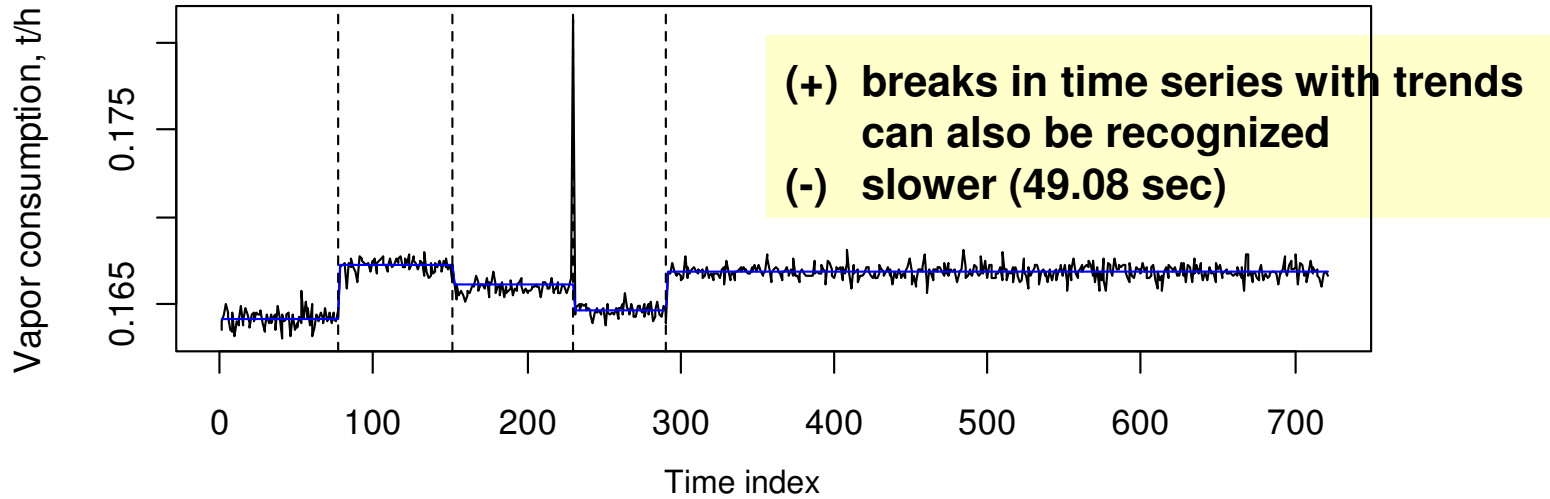
# Modified Segmentation: Steps

Observation window:  
**30 days**

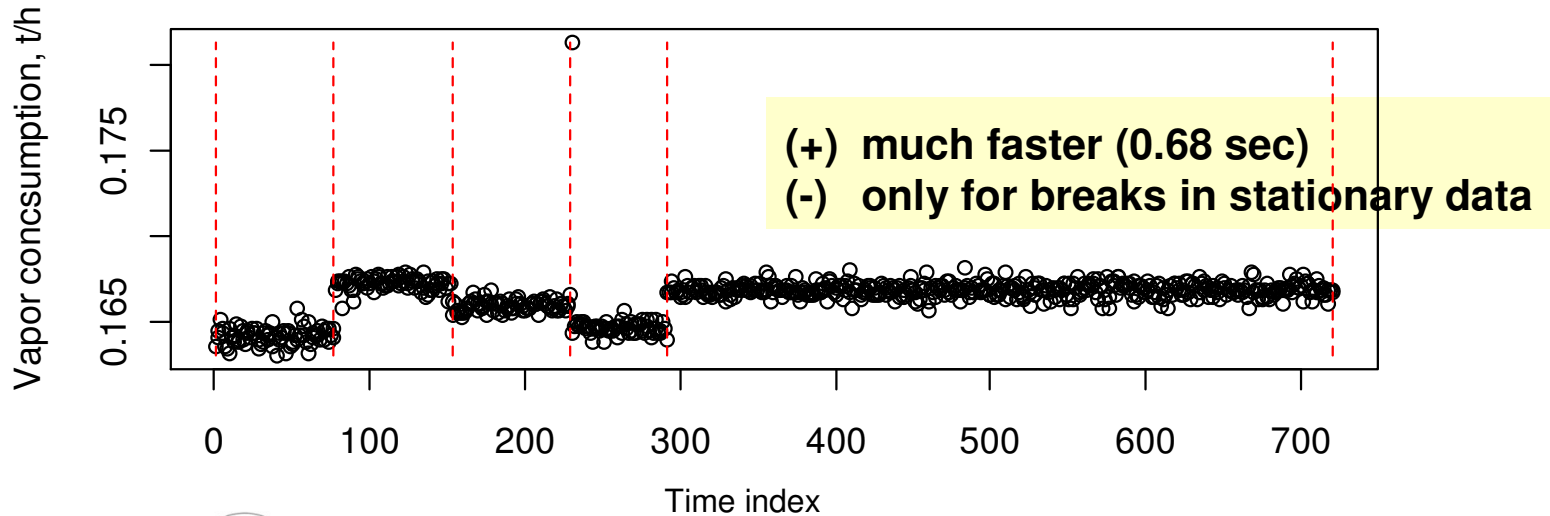


# Modified segmentation vs. "strucchange"

R-Package strucchange



Modified segmentation



# Conclusion

- Heuristic method proposed by Pedro Bernaola-Galván et al. was implemented in R and analyzed
- Analysis of original segmentation algorithms showed that this method is not robust to outliers near to bounds
- Original algorithm was modified and shows good resistance to influence of outliers near bounds
- Algorithm is limited to stationary segments.
- The modified segmentation method was compared with functions *breakpoints* from R-package *strucchange* and its computational efficiency was shown

# Literature

- Bernaola-Galván, Pedro; Ivanov, Plamen Ch.; Amaral, Luís A. Nunes; Stanley, H. Eugene: Scale Invariance in the Nonstationarity of Human Heart Rate. In: Physical review letters (2001), Volume 87, number 16 (abgerufen am 13. August 2009). <http://polymer.bu.edu/hes/articles/bias01.pdf>
- Fukuda, Kensuke; Stanley, H. Eugene; Amaral, Luís A. Nunes: Heuristic segmentation of a nonstationary time series. In: Physical review letters 69 (2004).
- Zeileis, Achim; Leisch, Friedrich; Hansen, Bruce; Hornik, Kurt; Kleiber, Christian: Package „strucchange“, 2009 (Version 1.3-7 )
- Zeileis, Achim; Leisch, Friedrich; Hornik, Kurt; Kleiber: strucchange: An R Package for Testing for Structural Change in Linear Regression Models (Version 1.3-7 ) <http://cran.r-project.org/web/packages/strucchange/vignettes/strucchange-intro.pdf>

