The “Caterpillar”-SSA approach to time series analysis and its automatization

Th.Alexandrov, N.Golyandina

theo@pdmi.ras.ru, nina@ng1174.spb.edu

St.Petersburg State University

Outline

1. History
2. Possibilities and advantages
3. “Caterpillar”-SSA: basic algorithm
4. Decomposition feasibility
5. Identification of SVD components
6. Example: trend and seasonality extraction
7. Forecast
8. Example: signal forecast
9. AutoSSA:
    1) extraction of trend
    2) extraction of cyclic components
    3) model example
    4) conclusions
10. Future plans

History

Origins:

- **Singular system approach to the method of delays**. Dynamic Systems – analysis of attractors [middle of 80's] (*Broomhead*)
- **Singular Spectrum Analysis**. Geophysics/meteorology – signal/noise enhancing, distinguishing of a time series from the red noise realization (Monte Carlo SSA) [90's] (*Vautard, Ghil, Fraedrich*)
- **"Caterpillar"**. Principal Component Analysis evolution [end of 90's] (*Danilov, Zhigljavskij, Solntsev, Nekrutkin, Goljandina*)

Books:

- Elsner, Tsonis. *Singular Spectrum Analysis. A New Tool in Time Series Analysis*, 1996.
- Golyandina, Nekrutkin, and Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*, 2001.

Internet links and software:

- http://www.atmos.ucla.edu/tcd/ssa/
- http://www.gistatgroup.com/cat/
- http://www.pdmi.ras.ru/~theo/autossa/

Possibilities and advantages

Basic possibilities of the "Caterpillar"-SSA technique:

- Finding trend of different resolution
- Smoothing
- Extraction of seasonality components
  - Simultaneous extraction of cycles with small and large periods
  - Extraction periodicities with varying amplitudes
  - Simultaneous extraction of complex trends and periodicities
- Forecast
- Change-point detection

Advantages:

- Doesn't require the knowledge of parametric model of time series
- Works with wide spectrum of real-life time series
- Matches up for non-stationary time series
- Allows to find structure in short time series

"Caterpillar"-SSA: basic algorithm

- Decomposes time series into sum of additive components: $F_N = F_N^{(1)} + \ldots + F_N^{(m)}$
- Provides the information about each component

Algorithm:

1. Trajectory matrix construction:
   $F_N = (f_0, \ldots, f_{N-1}),\ \ F_N \to \mathbf{X} \in \mathbb{R}^{L \times K}$
   (L – window length, parameter)

   $$\mathbf{X} = \begin{bmatrix} f_0 & f_1 & \ldots & f_{N-L} \\ f_1 & f_2 & \ldots & f_{N-L+1} \\ \vdots & \ddots & \ddots & \vdots \\ f_{L-1} & f_L & \ldots & f_{N-1} \end{bmatrix}$$

2. Singular Value Decomposition (SVD):
   $\mathbf{X} = \sum \mathbf{X}_j$

   $\mathbf{X}_j = \sqrt{\lambda_j} U_j V_j^{\mathrm{T}}$
   λ_j – eigenvalue, U_j – e.vector of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$,
   V_j – e.vector of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $V_j = \mathbf{X}^{\mathrm{T}} U_j / \sqrt{\lambda_j}$

3. Grouping of SVD components:
   $\{1, \ldots, d\} = \bigoplus I_k$,

   $\mathbf{X}^{(k)} = \sum_{j \in I_k} \mathbf{X}_j$

4. Reconstruction by diagonal averaging:
   $\mathbf{X}^{(k)} \to \widetilde{F}_N^{(k)}$

   $F_N = \widetilde{F}_N^{(1)} + \ldots + \widetilde{F}_N^{(m)}$

Decomposition feasibility

$F_N = F_N^{(1)} + F_N^{(2)}, \quad \mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)}$

Separability: we can form the group I_1 of SVD components so that $I_1 \leftrightarrow \mathbf{X}^{(1)}$

- Separability is the necessary condition for "Caterpillar"-SSA application
- Exact separability impose strong constraints at the spectrum of time series which could be processed

Real-life: asymptotic separability

The case of finite N: $F_N = F_N^{(1)} + F_N^{(2)}, \quad I_1 \leftrightarrow \mathbf{X}^{(1)} \leftrightarrow \widetilde{F}_N^{(1)}$ –approximation of the $F_N^{(1)}$

Examples of asymptotic separability:

- A determinate signal is asympt. separable from a white noise
- A periodicity is asympt. separable from a trend

Fulfilment of (asymptotic) separability conditions places limitations on the value of window length L

Only time series which generate finite amount of SVD components – hard constraint?

Any linear combination of product of **exponents**, **harmonics** and **polynomials** generates finite amount of SVD components

Identification of SVD components

Identification – choosing of SVD components on the stage of grouping.

Important examples

Exponential trend: $f_n = Ae^{\alpha n}$

- it generates one SVD component,
- eigenvector:
  $U = (u_1, \ldots, u_L)^{\mathrm{T}}: \quad u_k = Ce^{\alpha k}$
  ("exponential" form with the same α)

Exponentially-modulated harmonic: $f_n = Ae^{\alpha n}\cos(2\pi\omega n)$

- it generates two SVD components,
- eigenvectors:

$$U_1 = (u_1^{(1)}, \ldots, u_L^{(1)})^{\mathrm{T}}: \quad u_k^{(1)} = C_1 e^{\alpha k}\cos(2\pi\omega k)$$
$$U_2 = (u_1^{(2)}, \ldots, u_L^{(2)})^{\mathrm{T}}: \quad u_k^{(2)} = C_2 e^{\alpha k}\sin(2\pi\omega k)$$

("exponentially-modulated" form with the same α and ω)

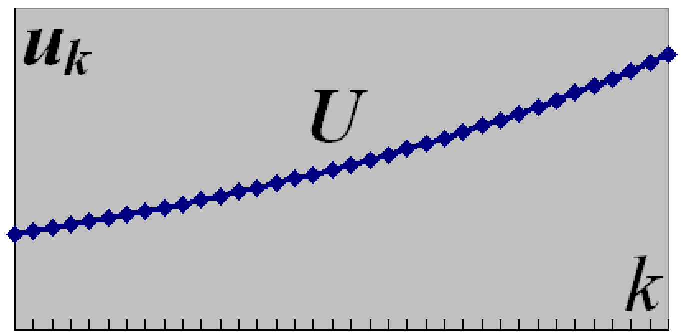
Identification of SVD components

Identification – choosing of SVD components on the stage of grouping.

Important examples

Exponential trend: $f_n = Ae^{\alpha n}$

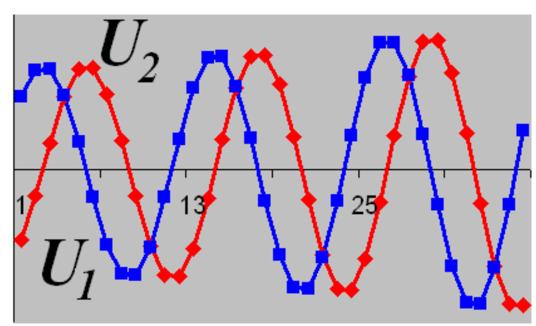
- it generates one SVD component,
- eigenvector:
  $U = (u_1, \ldots, u_L)^{\mathrm{T}}: \quad u_k = Ce^{\alpha k}$
  ("exponential" form with the same α)



Exponentially-modulated harmonic: $f_n = Ae^{\alpha n}\cos(2\pi\omega n)$

- it generates two SVD components,
- eigenvectors:

$U_1 = (u_1^{(1)}, \ldots, u_L^{(1)})^{\mathrm{T}}: \quad u_k^{(1)} = C_1 e^{\alpha k}\cos(2\pi\omega k)$
$U_2 = (u_1^{(2)}, \ldots, u_L^{(2)})^{\mathrm{T}}: \quad u_k^{(2)} = C_2 e^{\alpha k}\sin(2\pi\omega k)$
("exponentially-modulated" form with the same α and ω)



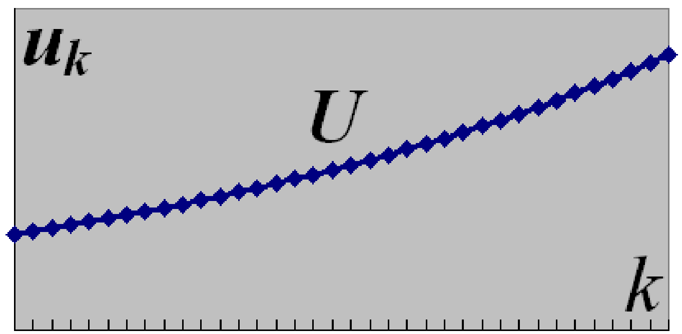
Identification of SVD components

Identification – choosing of SVD components on the stage of grouping.

Important examples

Exponential trend: $f_n = Ae^{\alpha n}$

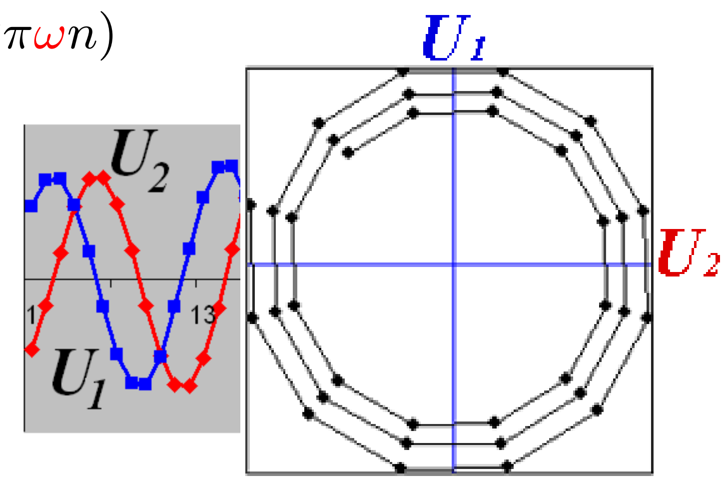
- it generates one SVD component,
- eigenvector:
  $U = (u_1, \ldots, u_L)^{\mathrm{T}}: \quad u_k = Ce^{\alpha k}$
  ("exponential" form with the same α)



Exponentially-modulated harmonic: $f_n = Ae^{\alpha n}\cos(2\pi\omega n)$

- it generates two SVD components,
- eigenvectors:

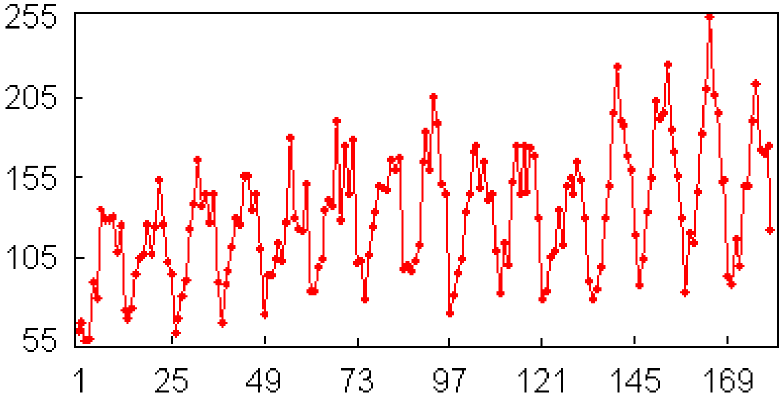
$U_1 = (u_1^{(1)}, \ldots, u_L^{(1)})^{\mathrm{T}}: \quad u_k^{(1)} = C_1 e^{\alpha k}\cos(2\pi\omega k)$
$U_2 = (u_1^{(2)}, \ldots, u_L^{(2)})^{\mathrm{T}}: \quad u_k^{(2)} = C_2 e^{\alpha k}\sin(2\pi\omega k)$
("exponentially-modulated" form with the same α and ω)



Example: trend and seasonality extraction

Traffic fatalities. Ontario, monthly, 1960-1974

Abraham, Redolter. *Stat. Methods for Forecasting*, 1983

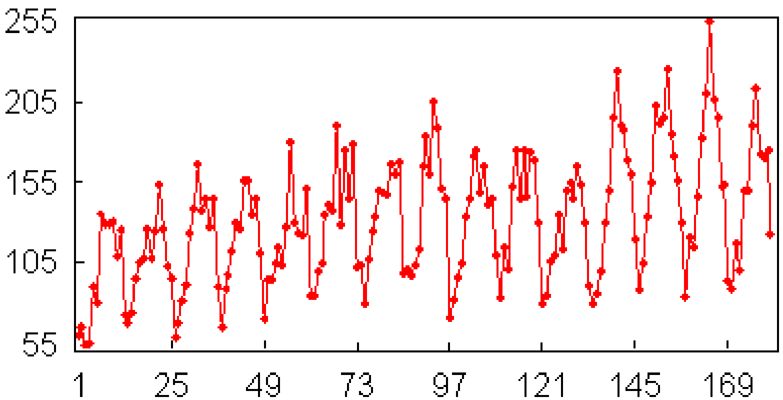


N=180, L=60

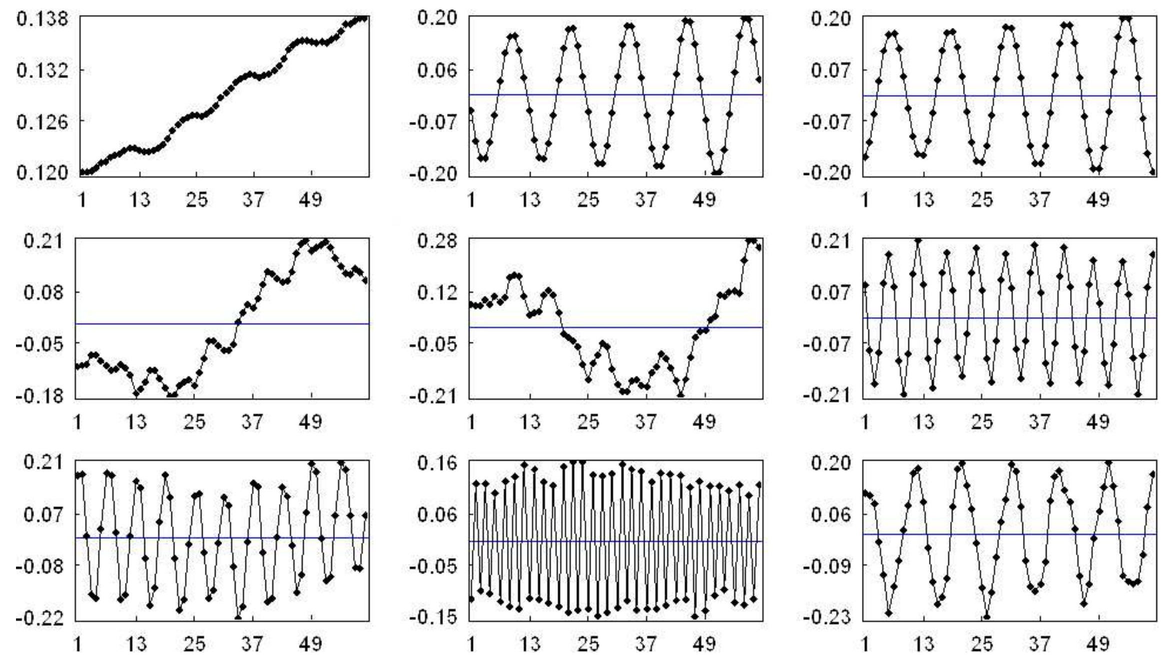
Example: trend and seasonality extraction

Traffic fatalities. Ontario, monthly, 1960-1974

Abraham, Redolter. *Stat. Methods for Forecasting*, 1983



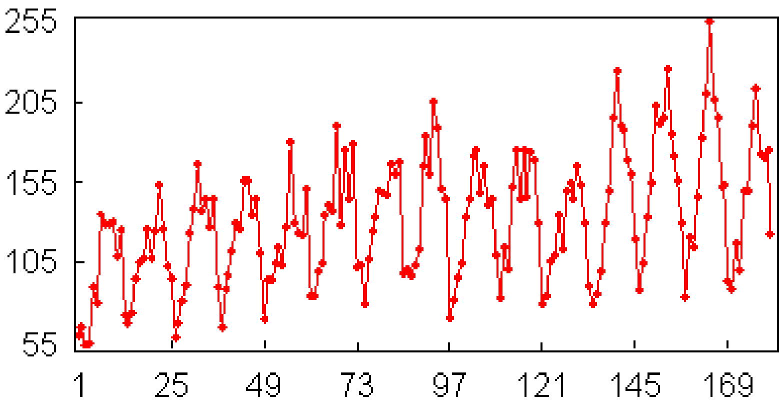
N=180, L=60



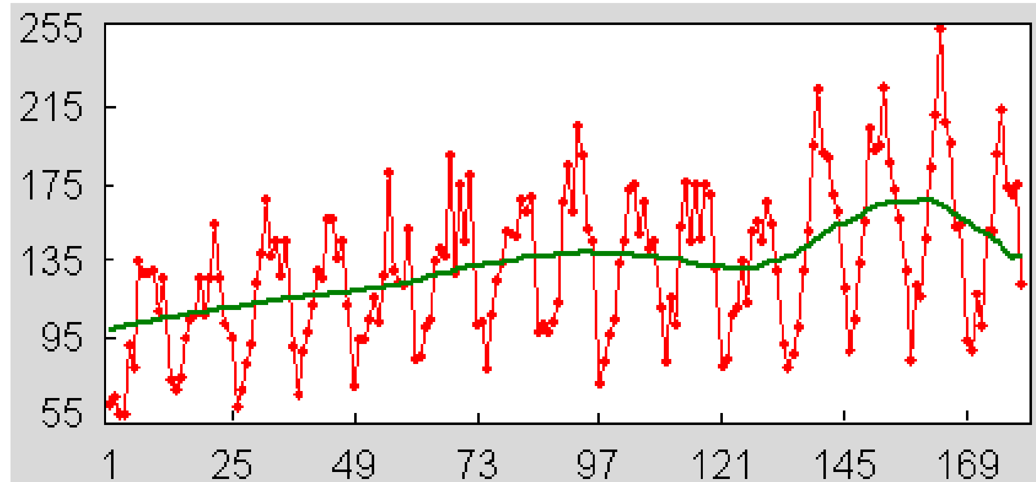
Example: trend and seasonality extraction

Traffic fatalities. Ontario, monthly, 1960-1974

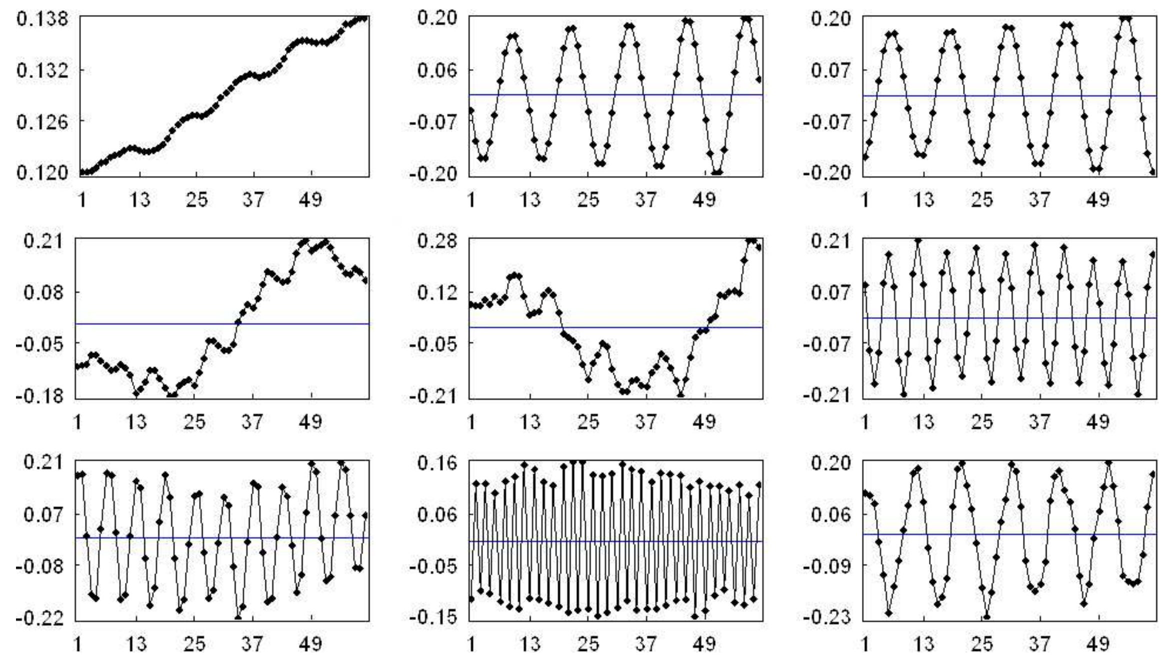
Abraham, Redolter. *Stat. Methods for Forecasting*, 1983



N=180, L=60



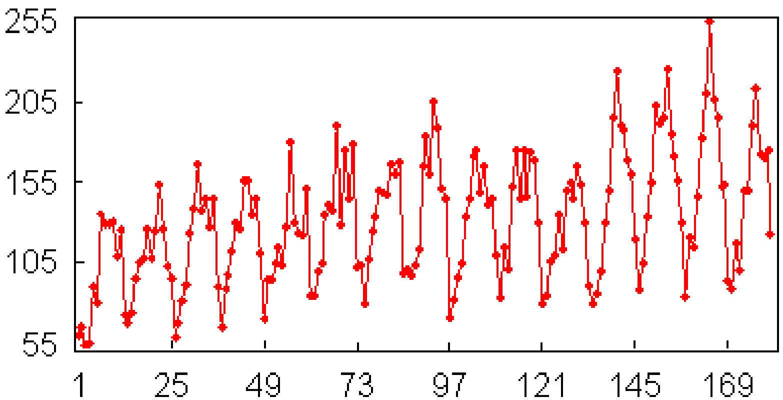
SVD components: 1, 4, 5



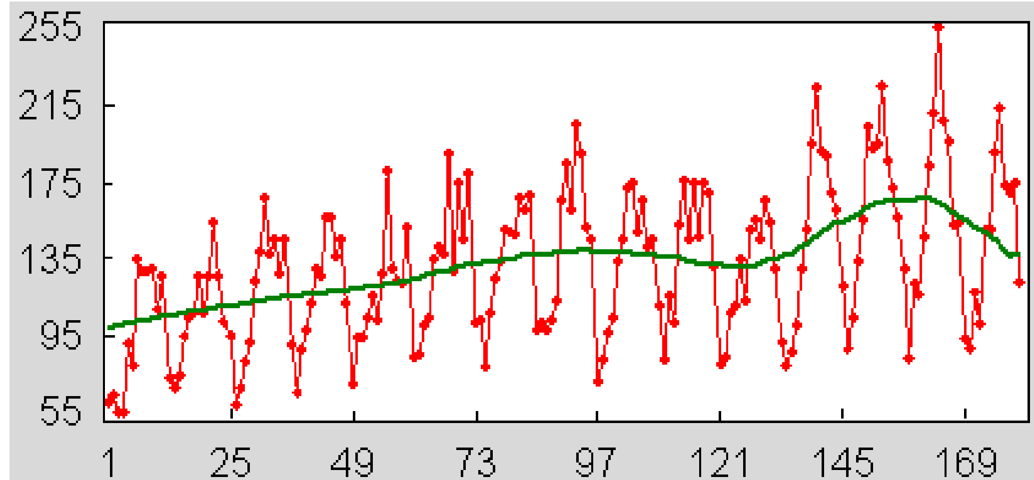
Example: trend and seasonality extraction

Traffic fatalities. Ontario, monthly, 1960-1974

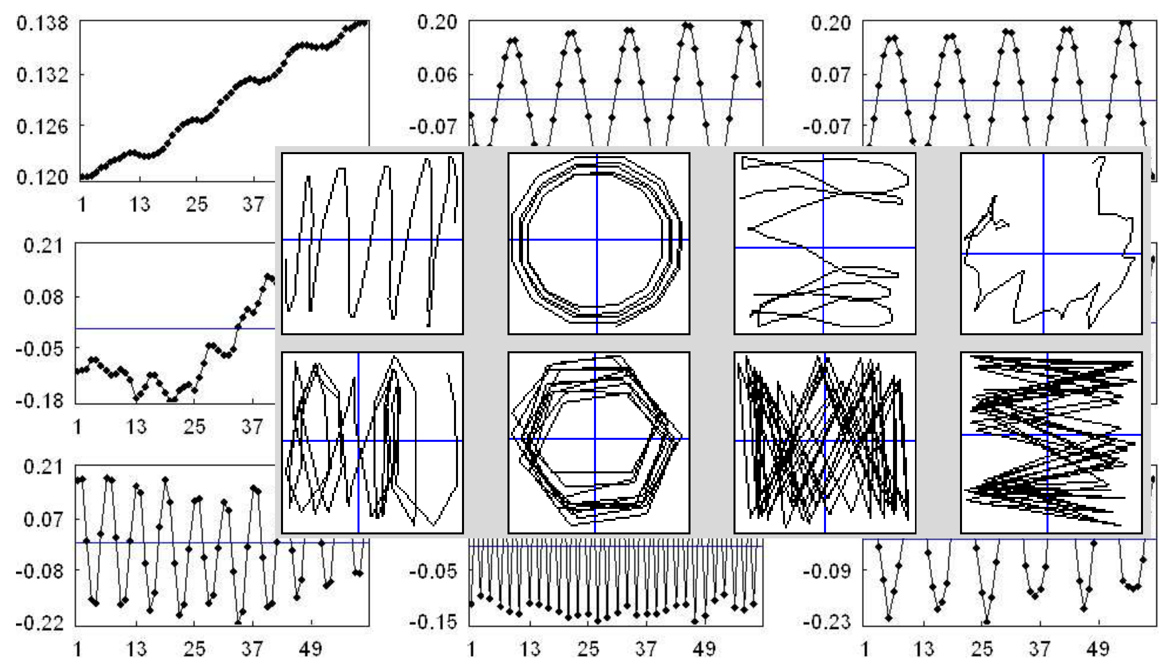
Abraham, Redolter. *Stat. Methods for Forecasting*, 1983



N=180, L=60



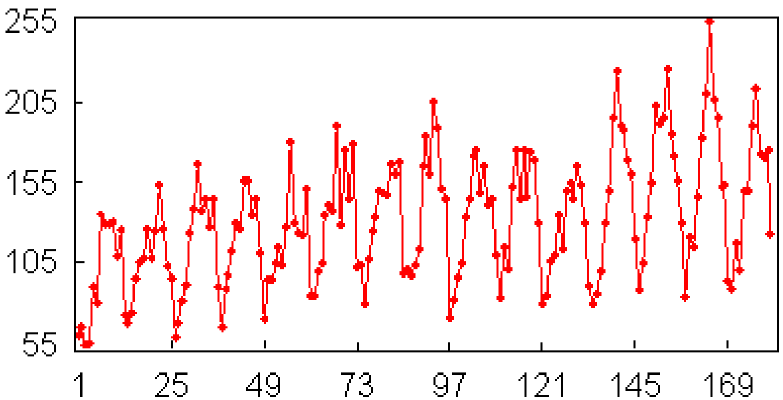
SVD components: 1, 4, 5



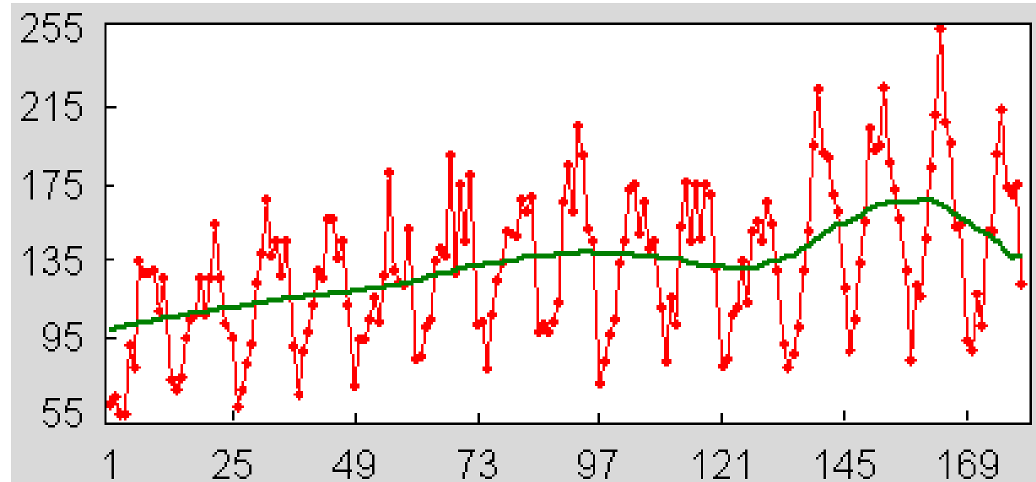
Example: trend and seasonality extraction

Traffic fatalities. Ontario, monthly, 1960-1974

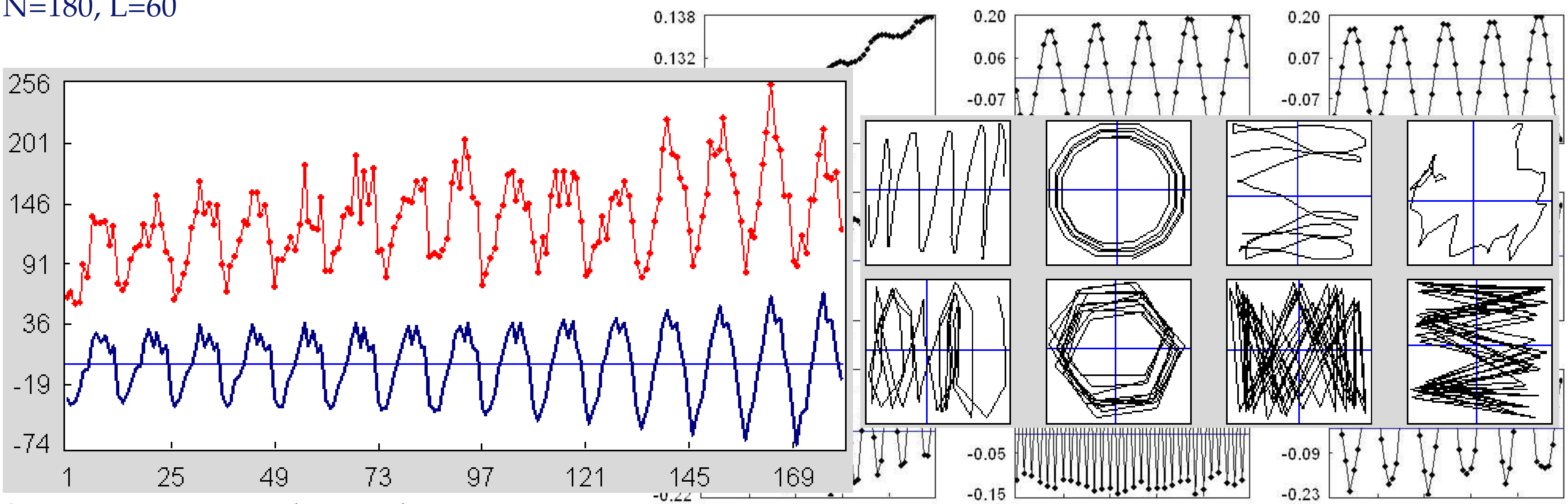
Abraham, Redolter. *Stat. Methods for Forecasting*, 1983



N=180, L=60



SVD components: 1, 4, 5



SVD components with periods estimations:

2-3 (12), 4-5, 6-7(6), 8(2), 9-10(10), 11-12(4), 13-14(2.4)

Forecast

Recurrent forecast in the framework of the "Caterpillar"-SSA.

Linear Recurrent Formula (LRF):

$$F_N = (f_0, \ldots, f_{N-1}) : \quad f_n = \sum_{k=1}^{d} a_k f_{n-k}$$

Theory:

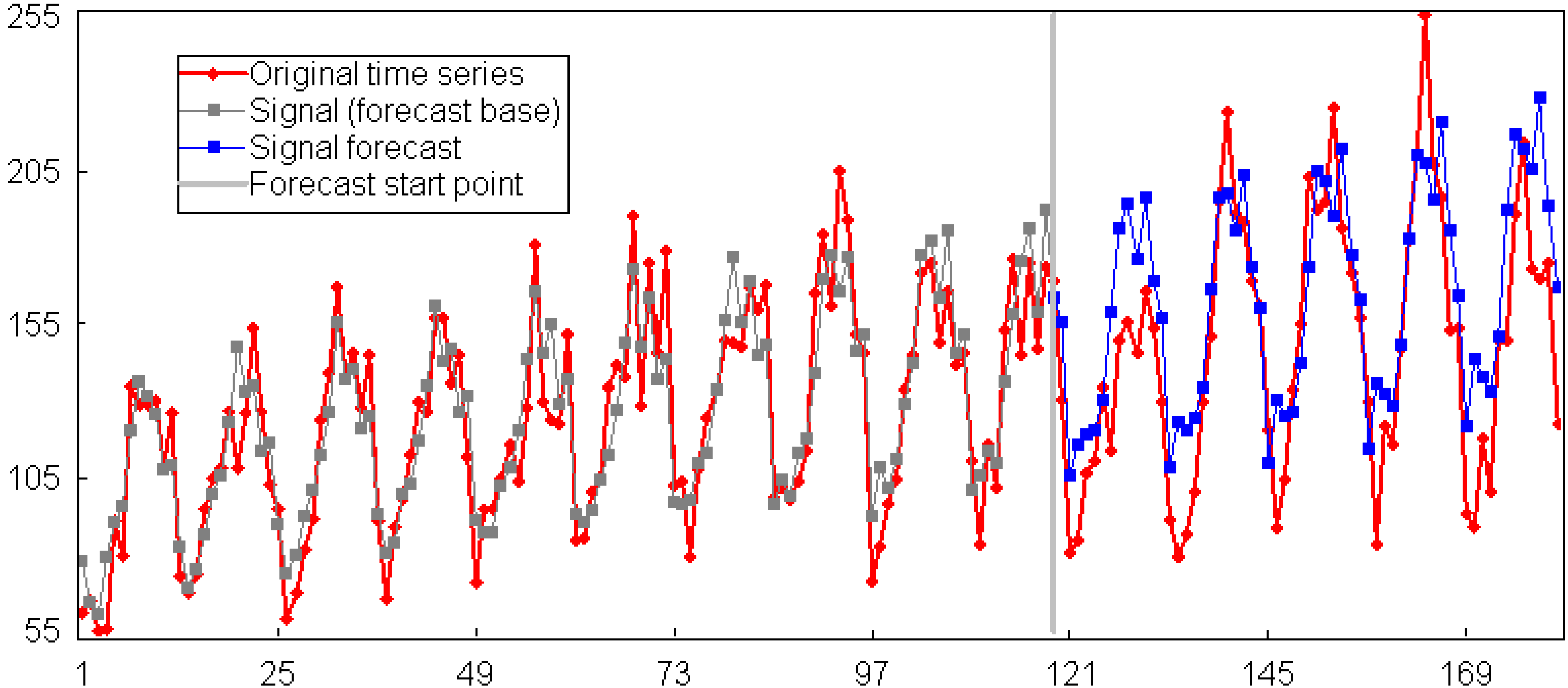
- All time series with finite number of SVD components generates finite order LRFs
- $F_N = F_N^{(1)} + F_N^{(2)}$, if $\{U_j\}_{j \in I_1} \leftrightarrow F_N^{(1)}$ then we can find a LRF governing the $F_N^{(1)}$

Algorithm of $F_N^{(1)}$ one-step forecast:

1. Choose window length L
2. Identify SVD components corresponding to the $F_N^{(1)}$: $\quad \{U_j\}_{j \in I_1} \leftrightarrow F_N^{(1)}$
3. Reconstruct via diagonal averaging: $\quad \{U_j\}_{j \in I_1} \Rightarrow \widetilde{F}_N^{(1)}$
4. Find a LRF governing $\widetilde{F}_N^{(1)}$ using the theorem
5. Apply the LRF to last values of the $\widetilde{F}_N^{(1)}$: $\quad f_N$

Example: signal forecast

First 119 points were given as the base for the signal reconstruction and forecast
Remaining part of the time series is figured to estimate the forecast quality



N=119, L=60, forecast of points 120-180

SVD components: 1 (trend); 2-3, 5-6, 9-10 (harmonics with periods 12, 4, 2.4); 4 (harmonic with period 2)

AutoSSA: trend extraction

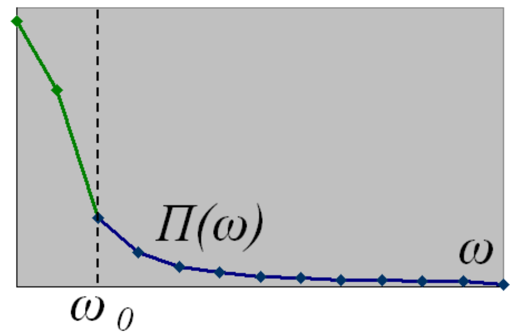
Let us investigate every eigenvector U_j and take $U = (u_1, \ldots, u_L)^{\mathrm{T}}$.

Low Frequencies method

- $$u_n = c_0 + \sum_{1 \leqslant k \leqslant \frac{L-1}{2}} \big(c_k \cos(2\pi nk/L) + s_k \sin(2\pi nk/L)\big) + (-1)^n c_{L/2},$$

- Periodogram:

$$\Pi_U^L(k/L) = \frac{L}{2} \begin{cases} 2{c_0}^2, & k = 0, \\ {c_k}^2 + {s_k}^2, & 1 \leqslant k \leqslant \frac{L-1}{2}, \\ 2{c_{L/2}}^2, & L \text{ is even and } k = L/2. \end{cases}$$



$\Pi_U^L(\omega)$, $\omega \in \{k/L\}$, **reflects the contribution of a harmonic with the frequency** ω **into the Fourier decomposition of** U**.**

- Parameter: ω_0 – upper boundary for the “low frequencies” interval

$$\mathcal{C}(U) = \frac{\sum_{0 \leqslant k \leqslant L\omega_0} \Pi_U^L(k/L)}{\sum_{0 \leqslant k \leqslant L/2} \Pi_U^L(k/L)}$$ – contribution of LF frequencies.

$\mathcal{C}(U) \geqslant \mathcal{C}_0 \Rightarrow$ **eigenvector** U **corresponds to a trend,** where $\mathcal{C}_0 \in (0, 1)$ – threshold

“Usually” $\mathcal{C}_0 = 0.1$

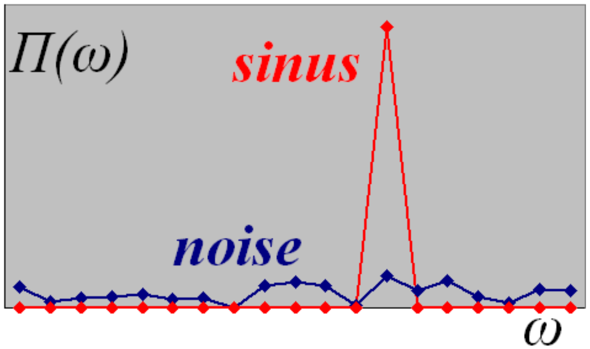
AutoSSA: periodicity extraction

We consider every pair of neighbor eigenvectors U_j,U_{j+1}.

Fourier method

- **Stage 1.** Check "maximal" frequencies: $\theta_j = \arg\min_k \Pi^L_{U_j}(k/L)$,
  $L|\theta_j - \theta_{j+1}| \leqslant s_0 \Rightarrow$ the pair $(j, j+1)$ is a "harmonical" pair.

- **Stage 2.** Check the form of periodogram:
  $\rho_{(j,j+1)} = \frac{1}{2}\max_k \left(\Pi^L_{U_j}(k/L) + \Pi^L_{U_{j+1}}(k/L)\right)$ for a harm. pair $\rho_{(j,j+1)} = 1$.



$\rho_{(j,j+1)} \geqslant \rho_0 \Rightarrow$ **the pair** $(j, j+1)$ **corresponds to a harmonic**, where $\rho_0 \in (0,1)$ – threshold

"Usually" $\rho_0 = 0.7$

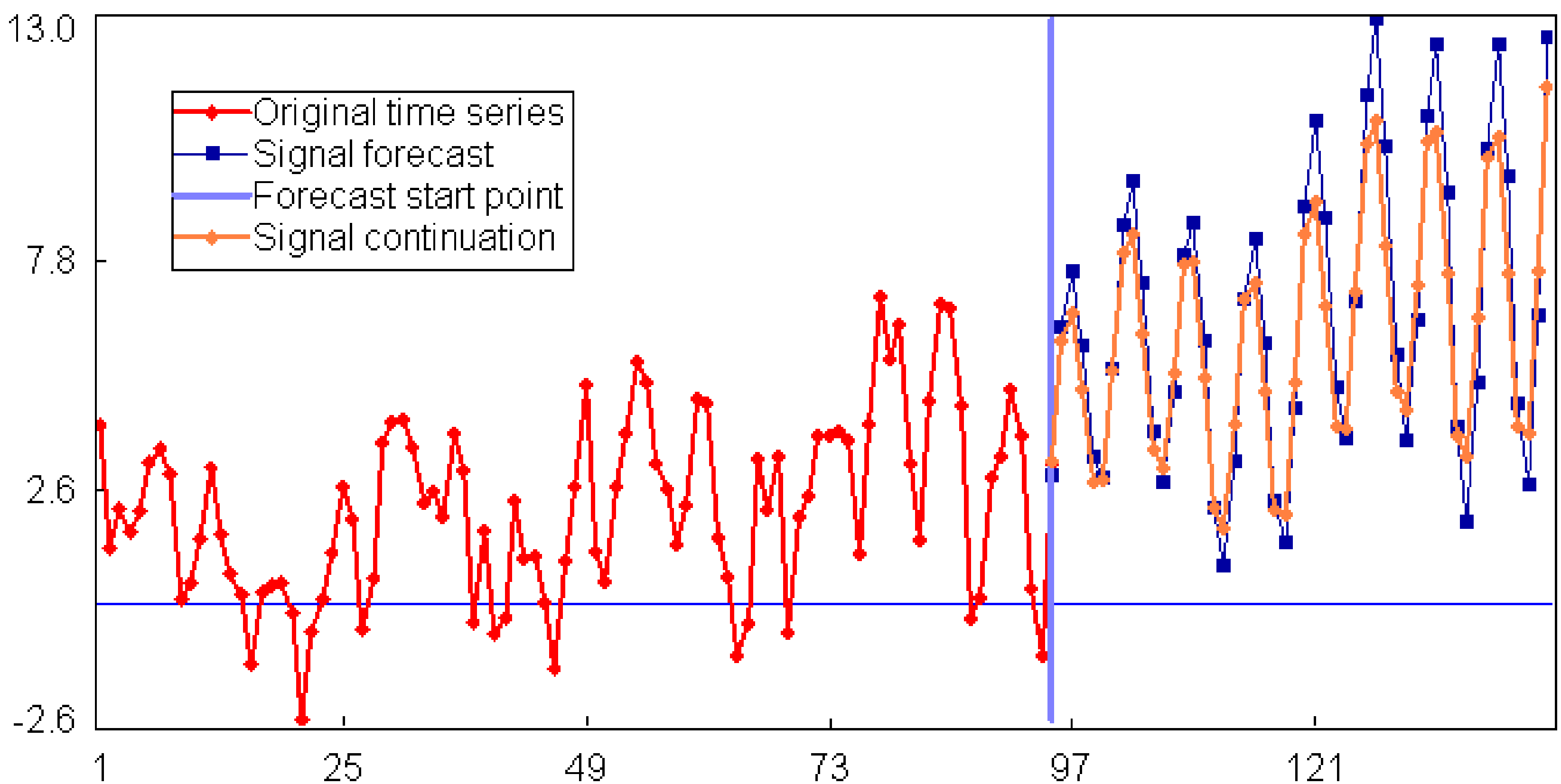
AutoSSA: model example

Signal forecast

$F_N: \ f_n = s_n + \varepsilon_n, \quad s_n = e^{0.015n} + 2e^{-0.005n}\sin(2\pi n/24) + e^{0.01n}\sin(2\pi n/6+2), \quad \varepsilon_n \sim N(0,1)$

N=95, L=48, forecast of points 96-144

Trend: 1st SVD component ($\mathcal{C}(U_1) = 0.997 > 0.1$)
Periodicity, T=6: 2-3 ($\rho_{(2,3)} = 0.95 > 0.7$); T=24: 4-5 ($\rho_{(2,3)} = 0.96 > 0.7$)



AutoSSA: conclusions

Main questions: **does automatization make sense?** and **how to choose thresholds?**

There is no general answer.

One application of AutoSSA: processing of time series from the given class (e.g. trend extraction if we know that trend has exponential form with $\alpha < \alpha_0$)

It's reasonable for batch processing of time series data

We invested applicability of AutoSSA for extraction and forecast exponential trend/e-m harmonic. For this task

- AutoSSA provides good results and its quality is comparable with interactive identification
- We realized that thresholds for extraction and forecast are the same
- It's possible to work out instructions of thresholds setting for a class of time series

Future plans

Open problems:

- SSA: complex time series models. For instance, with modulation of parameters (e.g. harmonic frequency) in time

Future plans:

- AutoSSA: methodology
- SSA: comparison of the “Caterpillar”-SSA with other techniques:
  - standard techniques
  - wavelets
  - neural networks
- SSA: application of the “Caterpillar”-SSA in other areas