Possibilities of Automation of the “Caterpillar”-SSA Method for Time Series Analysis and Forecast

Th.Alexandrov, N.Golyandina

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

St.Petersburg State University, Russia

Workshop on Nonlinear and Nonstationary time series, 20.-21.09.2005, Kaiserslautern

Origins of “Caterpillar”-SSA approach

- Singular System Analysis

  Dynamic Systems, method of delays for analysis of attractors [middle of 80’s],
  (Broomhead)

- Singular Spectrum Analysis

  Geophysics/meteorology – signal/noise enhancing, signal detection in red noise (Monte Carlo SSA) [90’s],
  (Vautard, Ghil, Fraedrich)

- “Caterpillar”

  Principal Component Analysis for time series [end of 90’s],
  (Danilov, Zhigljavsky, Solntsev, Nekrutkin, Golyandina)

Main sources of information about “Caterpillar”-SSA and AutoSSA

- “Caterpillar”-SSA:
  - [GNZ] Golyandina, Nekrutkin, Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques, 2001*
  - http://www.gistatgroup.com/cat/
- AutoSSA: http://www.pdmi.ras.ru/˜theo/autossa/

Basic possibilities of the “Caterpillar”-SSA technique

- Finding trends 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 non-stationary time series
- Allows one 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:
   $F_N = (f_0, \ldots, f_{N-1}),\quad F_N \to \mathbf{X} \in \mathbb{R}^{L\times K}$
   (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):
   $\mathbf{X} = \sum \mathbf{X}_j$

$$\mathbf{X}_j = \sqrt{\lambda_j} U_j V_j^{\mathsf{T}}$$

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

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:
   $\mathbf{X}^{(k)} \to \widetilde{F}_N^{(k)}$

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

Does exist an SVD such that it forms trend/periodicity & how to group components?

Identification of SVD components

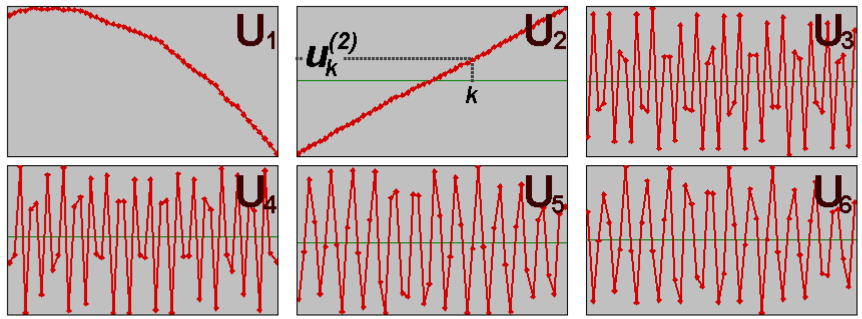
Identification – choosing of SVD components on the stage of grouping.

Trend and periodicity (sum of harmonics)

Trend

SVD components corr. to a trend have slowly-varying eigenvectors

Figure depicts eigenvectors (sequences of their elements), abscissa axis: indices of vector elements.



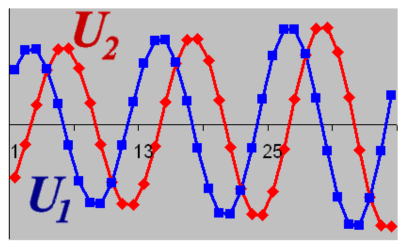
Exponentially-modulated harmonic: $f_n = Ae^{\alpha n}\cos(2\pi\omega n)$

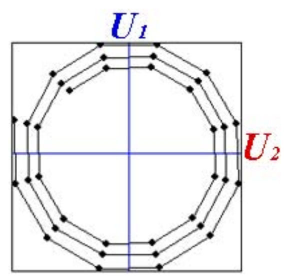
- it generates two SVD components,
- eigenvectors:

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

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

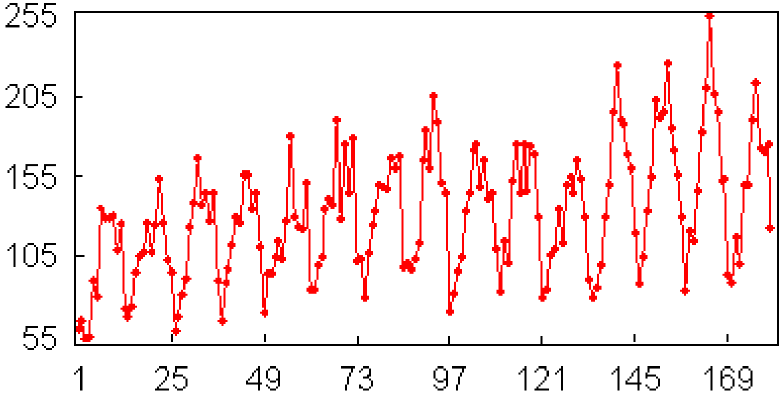
("e-m harmonical" form with the same α and ω)





Example: trend and seasonality extraction

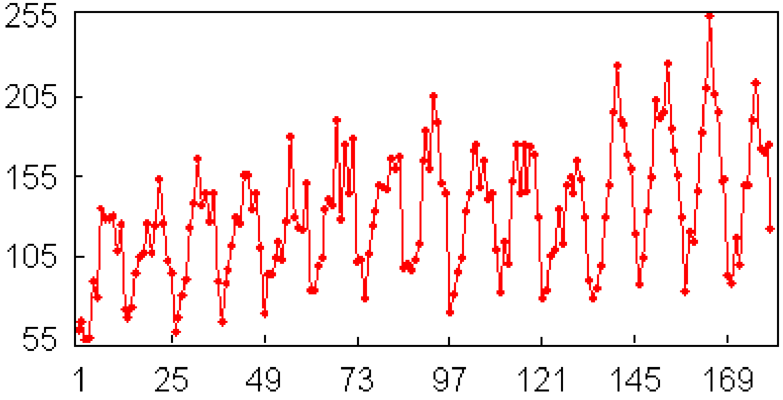
Traffic fatalities. Ontario, monthly, 1960-1974 (Abraham, Redolter. *Stat. Methods for Forecasting*, 1983)



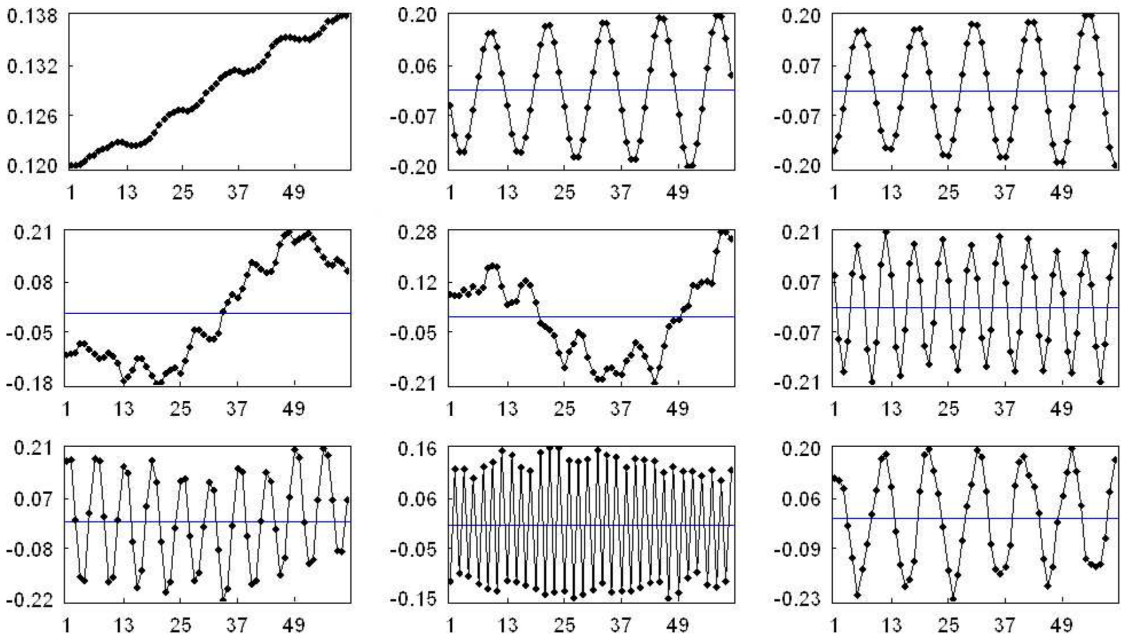
N=180, L=60

Example: trend and seasonality extraction

Traffic fatalities. Ontario, monthly, 1960-1974 (Abraham, Redolter. *Stat. Methods for Forecasting*, 1983)

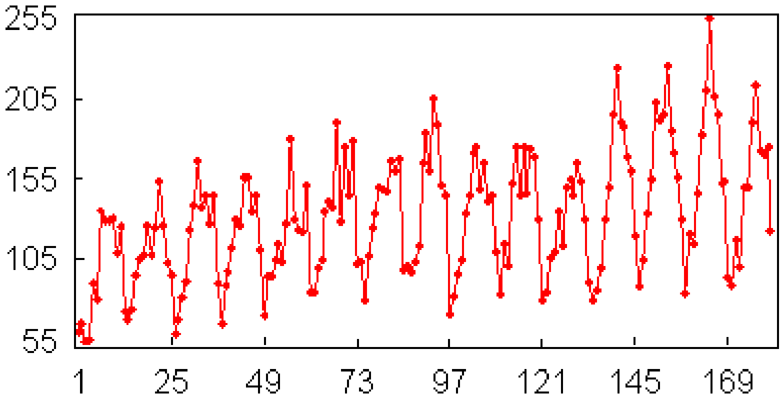


N=180, L=60

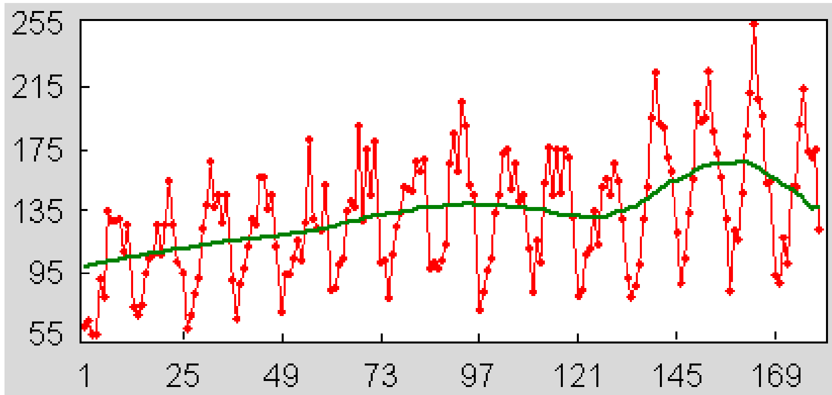


Example: trend and seasonality extraction

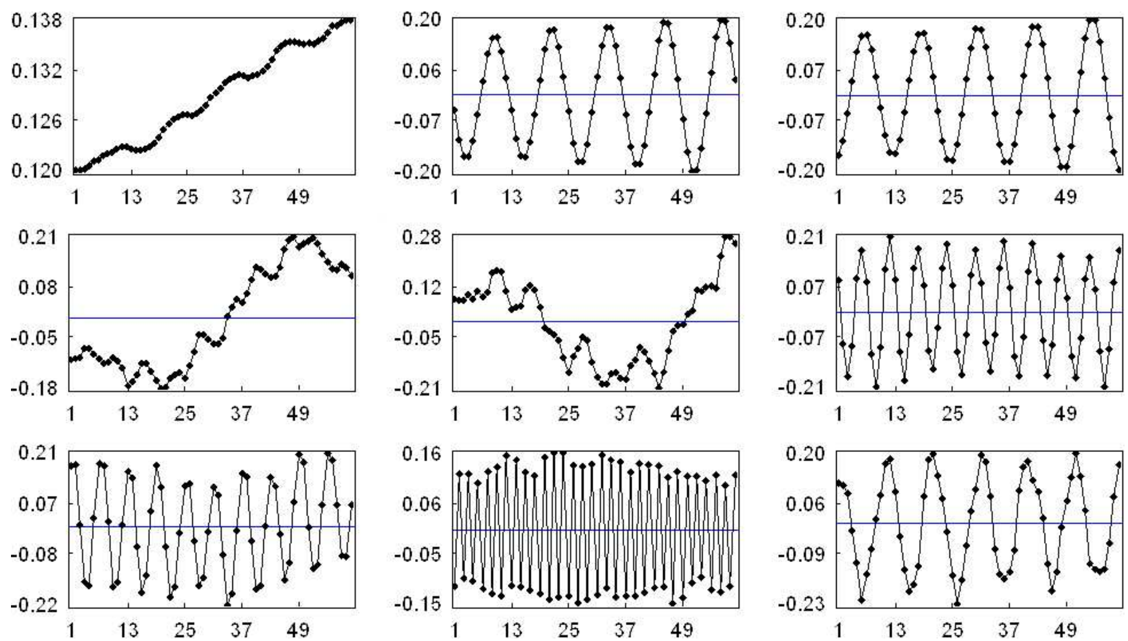
Traffic fatalities. Ontario, monthly, 1960-1974 (Abraham, Redolter. *Stat. Methods for Forecasting*, 1983)



N=180, L=60

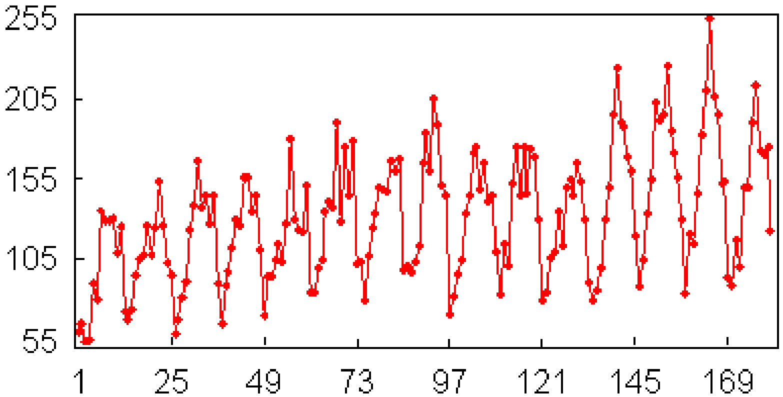


SVD components: 1, 4, 5

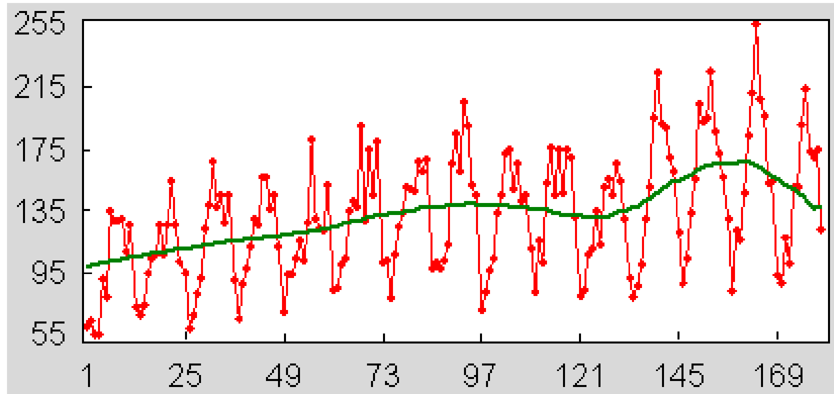


Example: trend and seasonality extraction

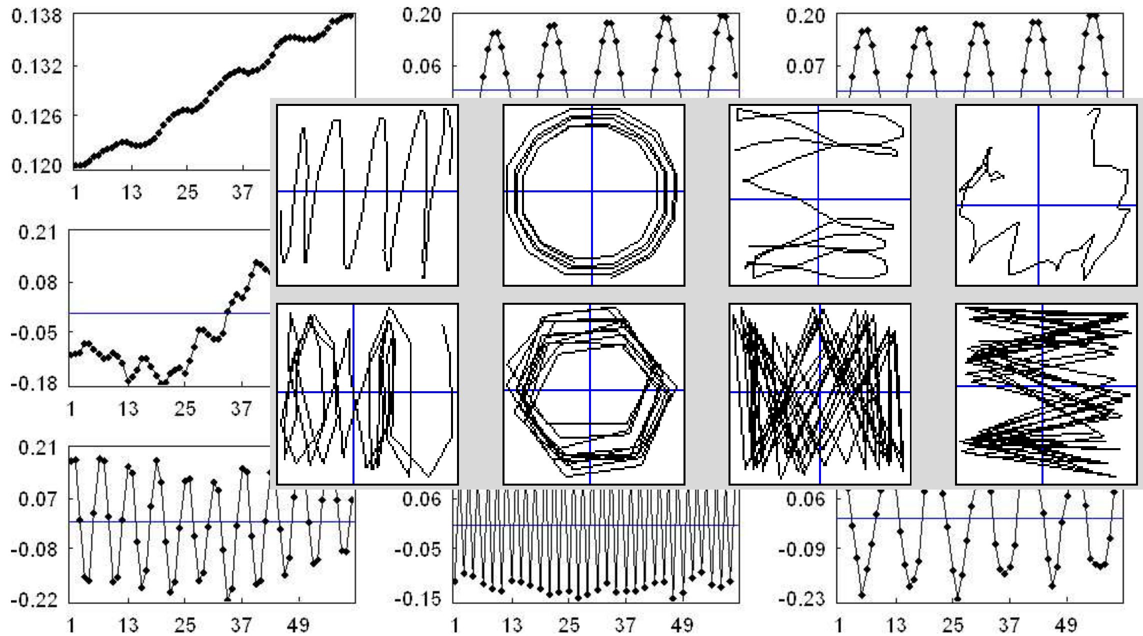
Traffic fatalities. Ontario, monthly, 1960-1974 (Abraham, Redolter. *Stat. Methods for Forecasting*, 1983)



N=180, L=60

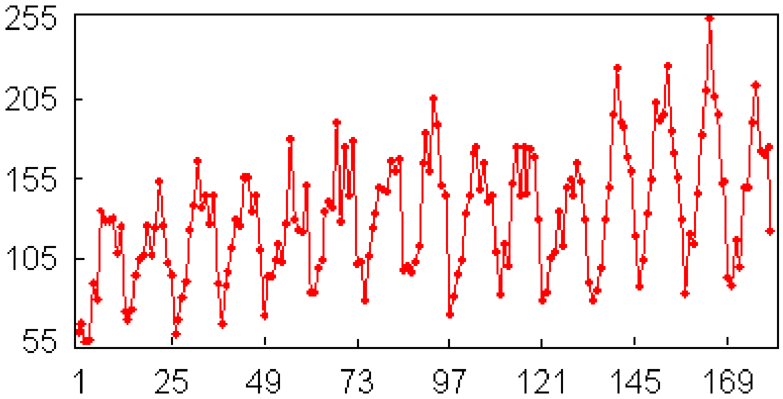


SVD components: 1, 4, 5

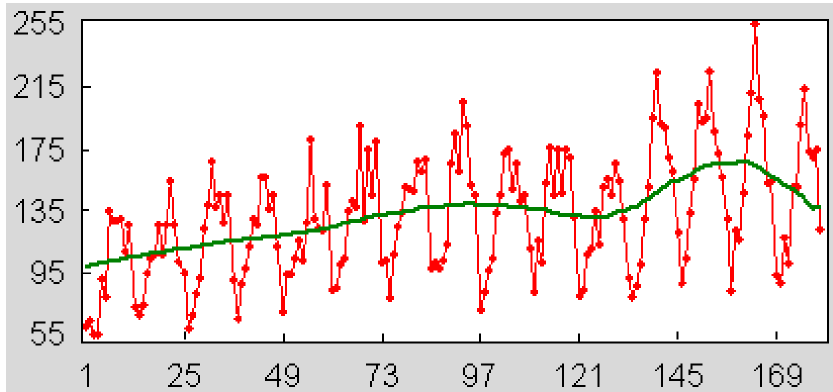


Example: trend and seasonality extraction

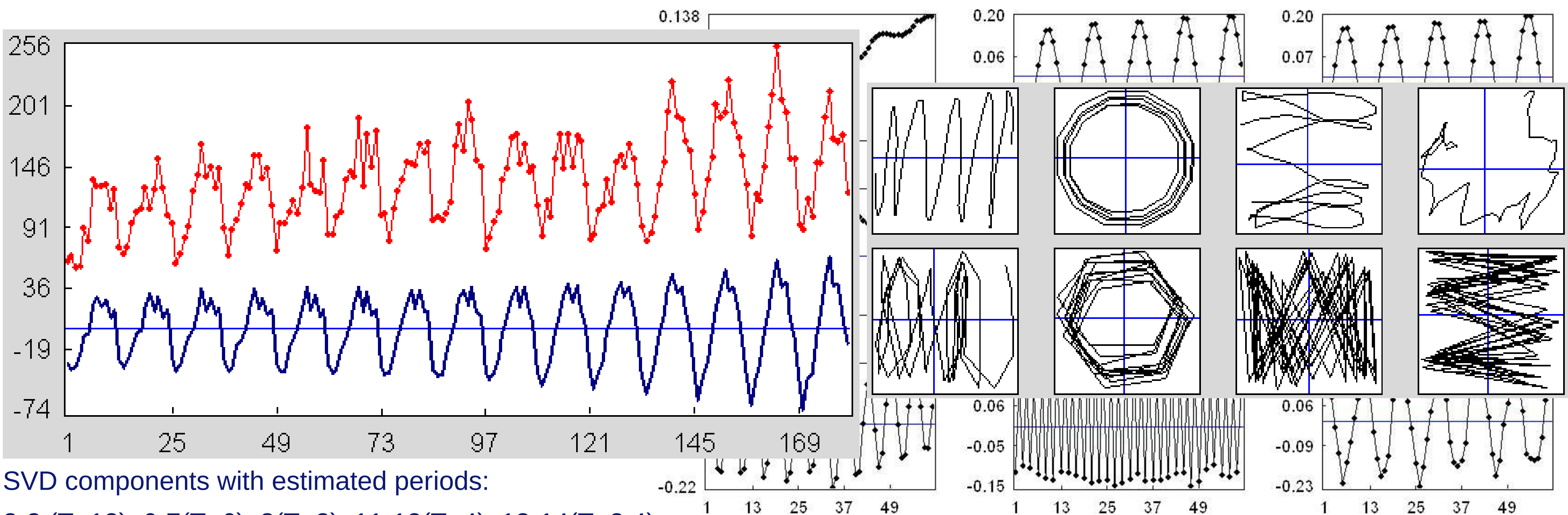
Traffic fatalities. Ontario, monthly, 1960-1974 (Abraham, Redolter. *Stat. Methods for Forecasting*, 1983)



N=180, L=60



SVD components: 1, 4, 5



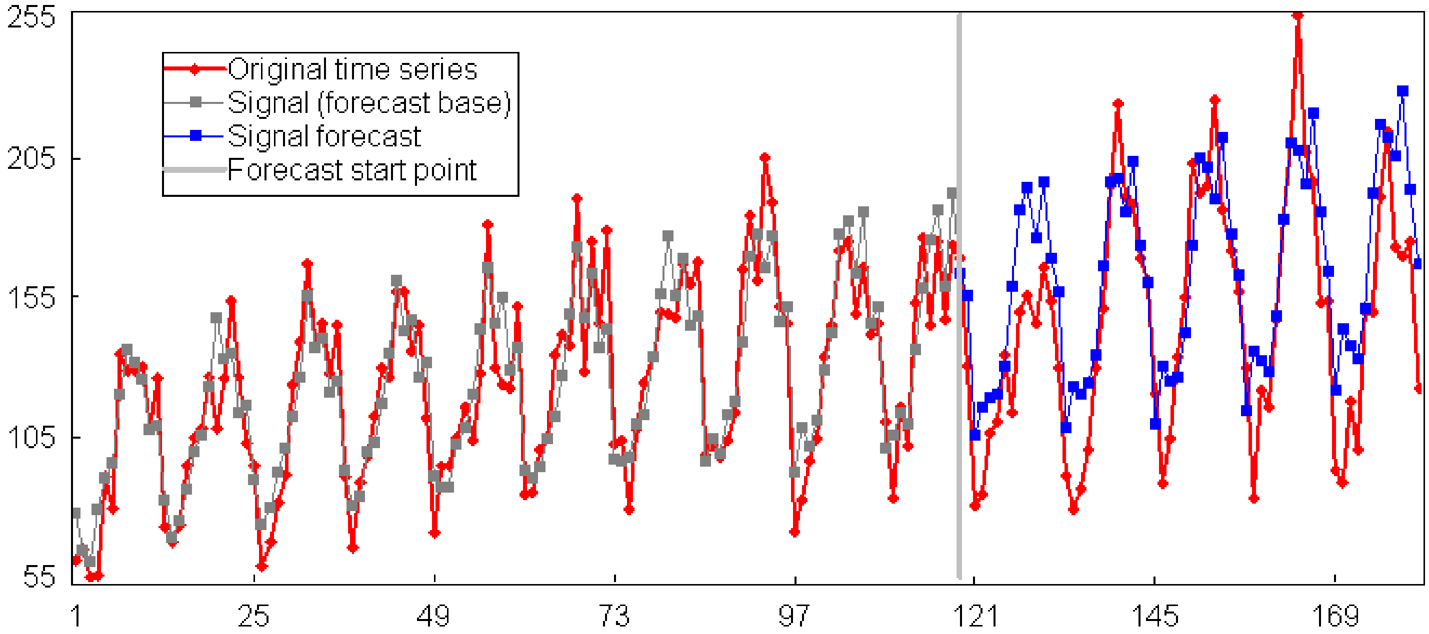
SVD components with estimated periods:

2-3 (T=12), 6-7(T=6), 8(T=2), 11-12(T=4), 13-14(T=2.4)

Example: signal forecast

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)



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

Forecast – using Linear Recurrent Formula (see [GNZ])

AutoSSA: motivation and problems statement

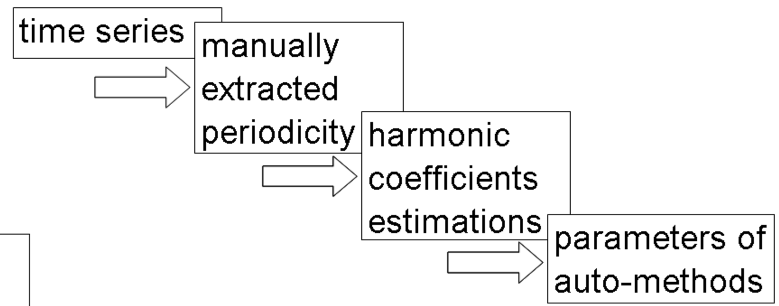
Main motive behind AutoSSA: batch processing of data, mostly **families of similar time series**.

Auto-methods are managed by parameters $\Rightarrow$ **how to set parameters?**

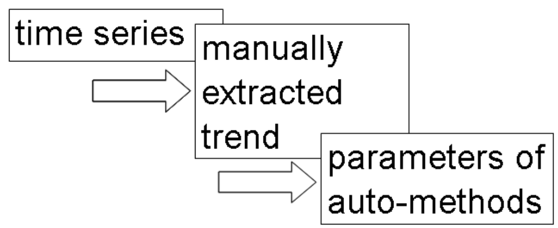
Main idea: to find parameters **examining only some time series** of a family.

What information can we obtain from a selected specimen?

- Periodicity extraction:
  parameters of its harmonic components,
  (T, modulation, periodicity/noise relation)



- Trend extraction:
  only form of a trend
  (it has no parametric form)



We will use **frequency approach to trend definition**, i.e. slowly-varying trend character in terms of Fourier decomposition $=$ harmonics with low freqs have large contribution.

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

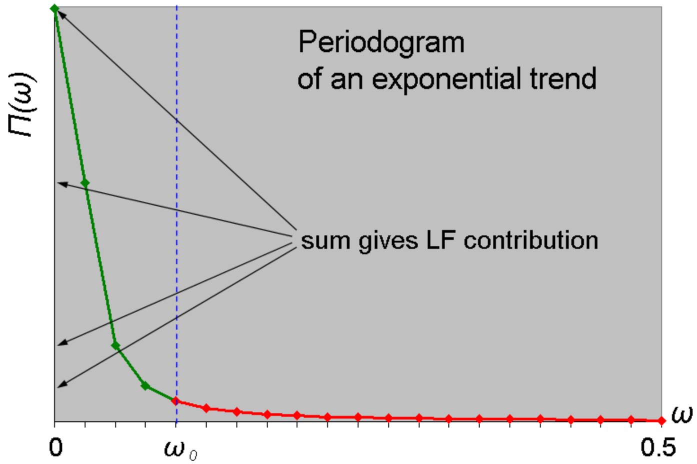
- Fourier decomposition of U:

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

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

Low Frequencies method

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



Define $\mathcal{C}(U) = \dfrac{\sum_{\mathbf{0} \leqslant \omega \leqslant \omega_\mathbf{0}} \mathbf{\Pi}(\omega)}{\sum_{\mathbf{0} \leqslant \omega \leqslant \mathbf{0.5}} \mathbf{\Pi}(\omega)}$ – contribution of LF frequencies ($\omega \in k/L, k \in \mathbb{Z}$).

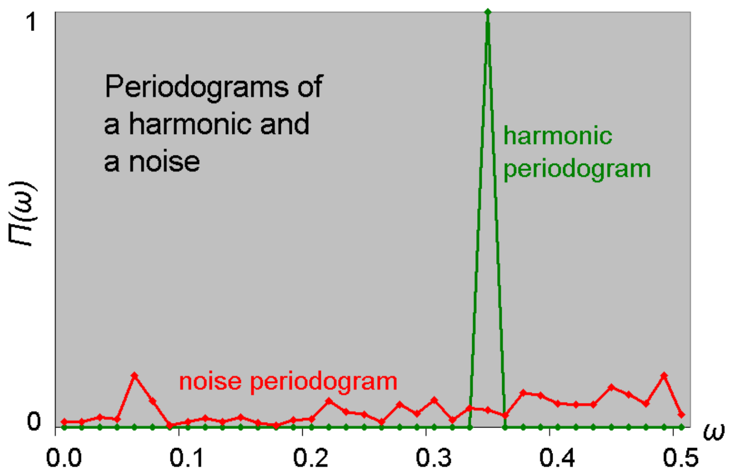
$\mathcal{C}(U) \geqslant \mathcal{C}_0 \Rightarrow$ eigenvector U corresponds to a trend,
where $\mathcal{C}_0 \in (0, 1)$ – the threshold

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

Fourier method

- Stage 1. Check if U_j, U_{j+1} have maximums of periodograms at the same frequency
- Stage 2. Check if their periodograms have "harmonic" forms

$$\rho_{(j,j+1)} = \frac{1}{2}\max_\omega\left(\Pi_{U_j}(\omega) + \Pi_{U_{j+1}}(\omega)\right),\ \omega \in k/L,$$ for a harm. pair $\rho_{(j,j+1)} = 1$.

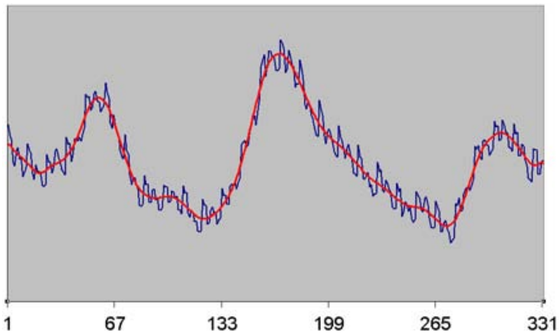


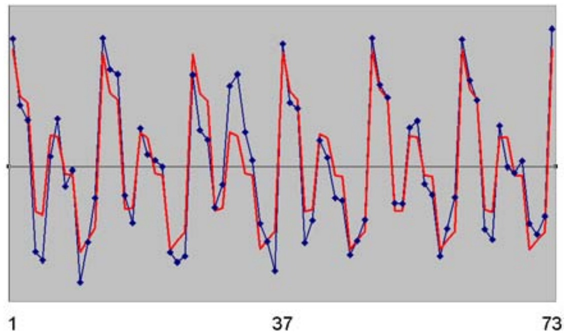
$\rho_{(j,j+1)} \geqslant \rho_0 \Rightarrow$ the pair $(j, j+1)$ corresponds to a harmonic,
where $\rho_0 \in (0,1)$ – the threshold parameter

AutoSSA: example of a family processing

Unemployment Level in different states of the USA, monthly, 1978-2005

Manually extracted trend and seasonality of a specimen time series:





Calculated methods parameters:

$\omega_0 = 0.07, \mathcal{C}_0 = 0.82$

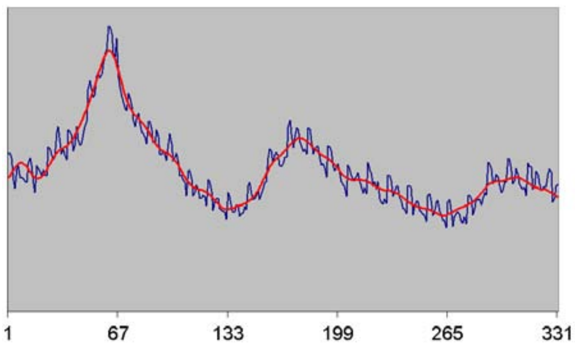
(applying to a trend, LowFreq method
with such parameters follows manual results)

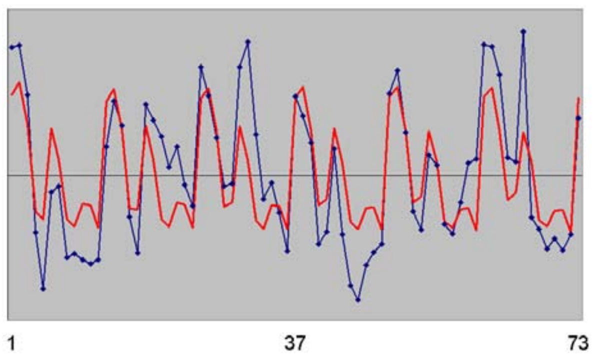
$\rho_0 = 0.89$

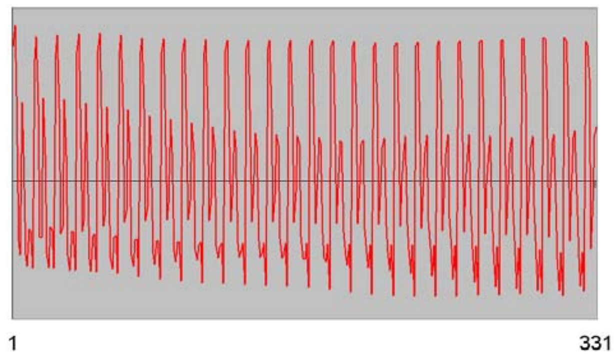
(method Fourier perfectly extracts
the seasonality harmonics with the same
properties as in the specimen
in the presence of the same noise)

Application of AutoSSA with these parameters gives such results for some two time series from the family:

1)







2)