The “Caterpillar”-SSA approach: automatic trend extraction and other applications

Theodore Alexandrov
theo@pdmi.ras.ru

St. Petersburg State University, Russia

Bremen University, 28 Feb 2006
History – Present

Origins of the “Caterpillar”-SSA approach

- **Singular System Analysis** *(Broomhead)*
  Dynamic Systems, method of delays for analysis of attractors [middle of 80’s],

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

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

Present

- **Automation**: papers are published, see [http://www.pdmi.ras.ru/~theo/autossa/](http://www.pdmi.ras.ru/~theo/autossa/) *(Alexandrov, Golyandina)*

- **Change-point detection** *(Golyandina, Nekrutkin, Zhigljavsky)*

- **Missed observations**: a paper is published, a software is on [www.gistatgroup.com](http://www.gistatgroup.com) *(Golyandina, Osipov)*

- **2-channel SSA**: a paper is published, see [www.gistatgroup.com](http://www.gistatgroup.com) *(Golyandina, Stepanov)*

- **Some generalizations**

Future

- **2D, online “Caterpillar”-SSA...**

Tasks and advantages

“Caterpillar”-SSA kernel

Theoretical framework, the most important concepts are

- **Time series of finite rank** (=order of Linear Recurrent Formula)
- **Separability** (possibility to separate/extract additive components)

General tasks

- **Additive components extraction** (for example trend, harmonics, exp.modulated harmonics)
- **Smoothing** (self-adaptive linear filter with small \( L \))
- **Automatic calculation of LRF for t.s. of finite rank** => prolongation of an extracted additive component => forecast of an extracted additive component
- **Change-point detection**

Advantages

- **Model-free**
- **Works with non-stationary time series** (constrains will be described)
- **Suits for short t.s., robust to noise model etc.**

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
“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:
   \( F_N = (f_0, \ldots, f_{N-1}), \ F_N \rightarrow X \in \mathbb{R}^{L \times K} \)
   \( (L – \text{window length, parameter}) \)

2. Singular Value Decomposition (SVD):
   \( X = \sum X_j \)

3. Grouping of SVD components:
   \( \{1, \ldots, d\} = \bigoplus I_k \),
   \( X^{(k)} = \sum_{j \in I_k} X_j \)

4. Reconstruction by diagonal averaging:
   \( X^{(k)} \rightarrow \tilde{F}_N^{(k)} \)

Does exist an SVD such that it forms necessary additive component & how to group SVD components?

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Example: trend and seasonality extraction


N=180, L=60

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Example: trend and seasonality extraction


N=180, L=60

AutoSSA: http://www.pdmu.ru/~theo/autossa/
Example: trend and seasonality extraction


N=180, L=60

SVD components: 1, 4, 5

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Example: trend and seasonality extraction


N=180, L=60

SVD components: 1, 4, 5

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Example: trend and seasonality extraction


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)

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
We said model-free, but the area of action is constrained to: \( \text{span}(\exp \ast \cos \ast Pn) \).

**Important concepts**

- \( \mathcal{L}^{(L)} = \mathcal{L}^{(L)} = \text{span}(X_1, \ldots, X_K) \) – the trajectory space for \( F_N \), \( X_i = (f_{i-1}, \ldots, f_{i+L-2})^T \).
- Time series \( F_N \) is a time series of (finite) rank \( d \) (\( \text{rank}(F_N) = d \)), if \( \forall L \) \( \dim \mathcal{L}^{(L)} = d \).

**Rank ↔ amount of SVD components ↔ order of LRF**

- \( \text{rank}_L(F_N) = \text{rank} X \Rightarrow \) amount of SVD components with \( \lambda_j \neq 0 \) is equal to the rank.
- \( F = (\ldots, f_{-1}, f_0, f_1, \ldots) \) – infinite time series, then
  \[
  f_{i+d} = \sum_{k=1}^d a_k f_{i+d-k}, \ a_d \neq 0 \quad \Leftrightarrow \quad \text{rank}(F) = d.
  \]

**Examples of finite rank time series**

- **Exponentially modulated (e-m) harmonic** \( F_N \): \( f_n = Ae^{\alpha n} \cos(2\pi\omega n + \phi) \).
  - e-m harmonic (\( 0 < \omega < 1/2 \)): \( \text{rank} = 2 \)
  - e-m saw (\( \omega = 1/2 \)): \( \text{rank} = 1 \)
  - exponential time series (\( \omega = 0 \)): \( \text{rank} = 1 \)
  - harmonic (\( \alpha = 1 \)): \( \text{rank} = 2 \)
- **Polynomial** \( F_N \): \( f_n = \sum_{k=0}^m a_k n^k, \ a_m \neq 0 \): \( \text{rank} = m + 1 \)

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Separability

\[ F_N = F^{(1)}_N + F^{(2)}_N, \] window length \( L \), traj.matrices \( X = X^{(1)} + X^{(1)} \), traj.spaces \( \mathcal{L}^{(L,1)}, \mathcal{L}^{(L,2)} \).

\( F^{(1)}_N \) and \( F^{(2)}_N \) are the \( L \)-separable if \( \mathcal{L}^{(L,1)} \perp \mathcal{L}^{(L,2)} \) and \( \mathcal{L}^{(K,1)} \perp \mathcal{L}^{(K,2)} \).

If \( F^{(1)}_N \) and \( F^{(2)}_N \) are separable then the SVD components of \( X \) can be grouped so that the first group corresponds to \( X^{(1)} \) and the second to \( X^{(2)} \).

i.e. separability (separation of trajectory spaces) ⇔ separation of additive components

Reality:
- Approximate separability (approximate orthogonality of trajectory spaces)
- Asymptotic separability (with \( L, N \to \infty \))

Examples

| Separability (strict, asymptotic) on some conditions |
|-----------------|-----------------|----------------|-----------------|-----------------|-----------------|
|                | const | cos | exp | exp*cos | Pn |
| const          | -    | +   | -   | +       | -   | -   |
| cos            | +    | +   | -   | -       | -   | +   |
| exp            | -    | -   | -   | +       | +   | -   |
| exp*cos        | -    | -   | +   | +       | +   | -   |
| Pn             | -    | -   | +   | -       | +   | -   |

- signal is asymptotically separated from noise
- periodicity is asymptotically separated from trend

Separability conditions (and the rate of convergence) ⇒ rules for \( L \) setting (this problem had no solution before)

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
General trend extraction

Trend – slow varying deterministic additive component.
Examples of parametric trends: exp, Pn, harmonic with large \( T (T > N/2) \).

How to identify trend SVD components

- **Eigenvalues**
  \( \lambda_j \) – contribution of \( F^{(j)} \) to the form of \( F_N \) \( (F^{(j)} \) is reconstructed by \( \sqrt{\lambda_j} U_j V_j^T \). 
  Trend is large \( \Rightarrow \) its SVD components are the first.

- **Eigenvectors**
  \( U_j = (u_1^{(j)}, \ldots, u_L^{(j)})^T \)

Form of eigenvectors for some slow-varying time series

<table>
<thead>
<tr>
<th>( f_n )</th>
<th>( u_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e^{\alpha n} )</td>
<td>( e^{\alpha k} )</td>
</tr>
<tr>
<td>( \sum_m a_m n^m )</td>
<td>( \sum_m b_m k^m )</td>
</tr>
<tr>
<td>( e^{\alpha n} \cos(2\pi \omega n + \phi) )</td>
<td>( e^{\alpha k} \cos(2\pi \omega k + \psi) )</td>
</tr>
</tbody>
</table>

Trend SVD components have slow-varying eigenvectors.

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Continuation and forecast

Continuation

\[ F_N = (f_0, \ldots, f_{N-1}), \quad \text{rank}(F_N) = d < L, \text{ then typically, } F_N \text{ is governed by LRF of order } d. \]

Main variant of the continuation: recurrent continuation using LRF.

- There are the unique minimal LRF (order \(d\)) and many LRFs of order \(> d\)
- The “Caterpillar”-SSA:
  - \(\mathcal{L}(L)\): an orthogonal basis (e.g. eigenvectors) \(\rightarrow\) the LRF of order \(L - 1\) (automatically)
  - deflation of LRF (considering characteristic polynomial of LRF) (can be automated)

Continuation of an additive component

\[ F_N = F_N^{(1)} + F_N^{(2)}, \quad \text{rank}(F_N^{(1)}) = d_1 < L, \quad \text{rank}(F_N^{(1)}) = d_1 < L. \]

\(F_N^{(1)}\) and \(F_N^{(2)}\) are separable \(\Rightarrow\) we can continue them separately.

Forecast (approximate)

\[ F_N = F_N^{(1)} + F_N^{(2)}, \text{ approximate separability (for example, signal+noise or slightly distorted time series)} \Rightarrow \]

approximation of \(F_N^{(1)}\) by \(d\)-dimension trajectory space \(\Rightarrow\) approximation by LRF.
Change-point detection in brief

Problem statement

\( F_N \) is homogeneous if it is governed by some LRF with order \( \ll N \).

Assume that at one time it stops following the original LRF and after a certain time period it again becomes governed by an LRF.

We want to a posteriori detect such heterogeneities (change-points) and compare somehow homogeneous partes before and after change.

Solution

- \( F_N \) is governed by the LRF \( \Rightarrow \) for suff. large \( N \), \( L: \mathcal{L}^{(L)} = \text{span}(X_1, \ldots, X_K) \) is independent of \( N \)

- Minimal LRF \( \leftrightarrow \mathcal{L}^{(L)} \)

\( \Rightarrow \) LRF heterogeneities can be described in terms of corresponding lagged vectors: the perturbations force the lagged vectors to leave the space \( \mathcal{L}^{(L)} \)

We can measure this distance and detect a change-point

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Automation of time series processing

Problem statement
Automation of manual processing of large set of similar time series (family)

Manual processing is the ideal ⇒ quality in comparison with manual processing
⇒ we can use the stated theory

Why family?
Family processing ⇒ several randomly taken time series can be used for:
■ testing the auto-procedure, whether it works in general (necessary condition)
■ finding proper parameters of the auto-procedure for the whole family (performance optimization)

Bootstrap test of the procedure
We must know noise (residual) model (or its approximation)

■ Extract trend $\tilde{F}_N^{(T)}$ manually
■ Consider $F_N - \tilde{F}_N^{(T)}$ and estimate parameters of noise
■ Simulate noise using these parameters and generate surrogate data: $G_N = \tilde{F}_N^{(T)} + noise$
■ Extract trend from surrogate data $\tilde{G}_N^{(T)}$
■ MSE($\tilde{F}_N^{(T)}$, $\tilde{G}_N^{(T)}$) – a measure of procedure quality

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Auto-method of trend extraction

Eigenvectors of trend SVD components have slow-varying form

Search all eigenvectors, let us assume we process an $U = (u_1, \ldots, u_L)^T$.

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

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

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

Low Frequencies method

Parameter – $\omega_0$, upper boundary for the “low frequencies” interval

Define $C(U) = \sum_{0 \leq \omega \leq \omega_0} \Pi(\omega)/\Pi(0)$, $\omega \in k/L, k \in \mathbb{Z}$ – contribution of LF frequencies

$C(U) \geq C_0 \Rightarrow$ eigenvector $U$ corresponds to a trend, where $C_0 \in (0, 1)$ – the threshold

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Choice of $\omega_0$

Examining the periodogram of an original time series

- Periodicity with period $T$ exists $\Rightarrow \omega_0 < 1/T$

Examples

Exp+cos and its periodogram,

$$f_n = e^{0.01n} + \cos(2\pi n/12)$$

$0.3 < \omega_0 < 0.8$

Pn+cos and its periodogram,

$$f_n = (x - 10)(x - 40)(x - 60)(x - 95)\cos(2\pi n/12)$$

$\omega_0 \approx 0.7 < 0.8$

Traffat (left), its periodogram (center) and periodogram of normalized time series (right)

$0.3 < \omega_0 < 0.8$

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/
Choice of $C_0$, measure of quality of trend extraction

If we have a measure $\mathcal{R}$ of quality of trend extraction \[ \Rightarrow \ C_{\text{opt}} = \arg\min_{C_0 \in [0,1]} \mathcal{R} \]

Measure
The natural measure of quality is $\text{MSE}(F(T), \tilde{F}_0(T))$, where $F(T)$ is the real trend and $\tilde{F}_0(T)$ is the extracted trend (with $C_0$), but it requires unknown $F(T)$.

We propose \[ \mathcal{R}(C_0) = \frac{C(F - \tilde{F}_0(T))}{C(F)}, \quad \tilde{F}_0(T) \text{ is extracted with } C_0 \]

$\mathcal{R}(C_0)$ is consistent with $\text{MSE}(F(T), \tilde{F}_0(T))$ in such a way:
- it behaves like MSE
- by means of $\mathcal{R}(C_0)$ we can define $C_{\text{opt}} = \arg\min_{C_0} \text{MSE}(F(T), \tilde{F}_0(T))$
Examples of $C_{\text{opt}}$ estimation

**Model example, Pn+noise**

\[ f_n = (n - 10)(n - 70)(n - 160)^2 \cdot \frac{(n - 290)^2}{1e11} + N(0, 25), \]
\[ N = 329, L = N/2 = 160, \]
\[ \omega_0 = 0.07 \]
\[ C_0 = 1 \ldots 0.9 : \text{graphics reflect stepwise identification of trend SVD components } \Rightarrow \text{considerable changes of } \]
\[ \text{MSE}(F(T), \tilde{F}_0(T)) \]
\[ C_{\text{opt}} < 0.9 (\approx 0.9) \]

**Real-life example, Massachusetts unemployment**

Massachusetts unemployment (thousands, monthly), from economagic.com
\[ N = 331, L = N/2 = 156, \]
\[ \omega_0 = 0.05 < 1/12 = 0.08(3) \]
\[ C_0 = 1 \ldots 0.75 : \text{graphics reflect stepwise identification of trend SVD components } \Rightarrow \text{considerable changes of } \]
\[ \text{MSE}(F(T), \tilde{F}_0(T)) \]
\[ C_{\text{opt}} < 0.75 (\approx 0.75) \]
Final slide: to sum up

I. We have a family of similar time series \( \mathcal{F} = \{ F_N \} \).

II. Take randomly (or somehow otherwise) a test subset \( \mathcal{T} \) of several characteristic time series.

On these time series perform:

1. Extract trends manually
2. Examine periodograms of the time series and choose \( \omega_0 \)
3. Check if the proposed auto-procedure works in general on such time series:
   - Bootstrap comparison: how trends automatically extracted from surrogate data are close to manually extracted trends
     If they are sufficiently close then auto-procedure is accepted
   - It requires knowledge of noise model but we can take a simple one as a first approximation
4. Estimate \( C_{opt}^{(F)} \) for each time series and take a minimum from them

\[
C_{opt}^{(\mathcal{F})} = \min_{F \in \mathcal{T}} C_{opt}^{(F)}
\]

as the optimal \( C_0 \) for the family \( \mathcal{F} \)

(this optimization step can be skipped, then during processing of \( \mathcal{F} \) we have to estimate \( C_{opt} \) for each time series)

III. Process all time series from the family \( \mathcal{F} \)

AutoSSA: http://www.pdmi.ras.ru/~theo/autossa/