Singular System Analysis with missing values: theoretical performance and application to hydrological time series

H.Lange[1] and L. Gudmundsson[2]

[1] Norwegian Forest and Landscape Institute, Ås, Norway
[2] Department of Geosciences, University of Oslo, Norway

1. Singular System Analysis (gapfree original version)
2. The problem of gaps in data
3. SSA variants for data with gaps
4. Gap-filling with SSA
5. A fundamental problem: negative eigenvalues
6. Conclusions

Singular System Analysis (SSA)
„Principal Component Analysis for a single time series“

Basic procedure for a time series:

(Golyandina et al. 2001)

Embedding $X_i(t) = (x_i - \mu, x_{i+1} - \mu, \ldots, x_{i+M-1} - \mu)$

Trajectory matrix $Y = [X_1 X_2 \ldots X_{N-M+1}]$

Lag-Covariance matrix $C = YY^T / (N - M + 1)$

or use stationarity assumption to simplify (”Toeplitz modification”)

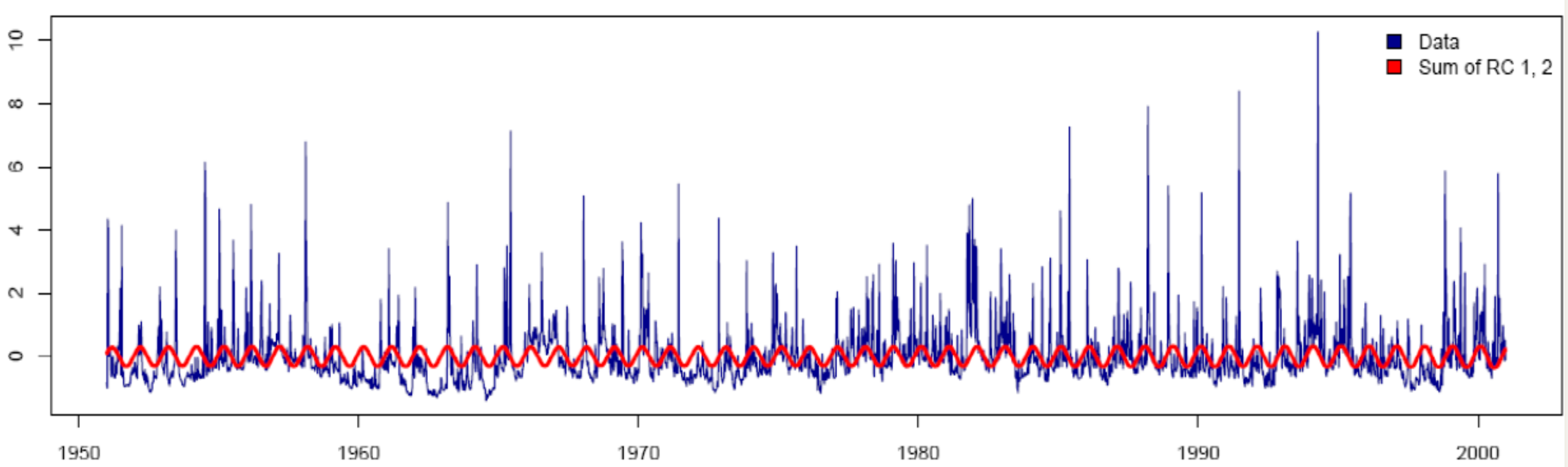
$C = E^T \Lambda E$ where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)$

SSA-PCs: $A_k(i) = \sum_{j=1}^{M} x_{i+j-1} E_k(j)$

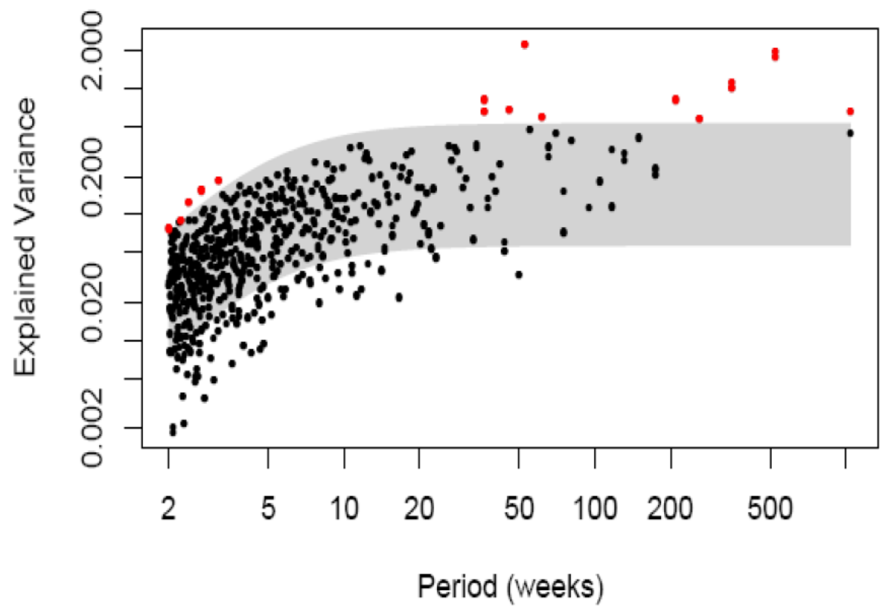
Reconstructed Component: $R_K(i) = \frac{1}{M_t} \sum_{k \in K} \sum_{j=1}^{M} A_k(i-j) E_k(j)$

where K is an arbitrary user-selected index set

Singular System Analysis – example result

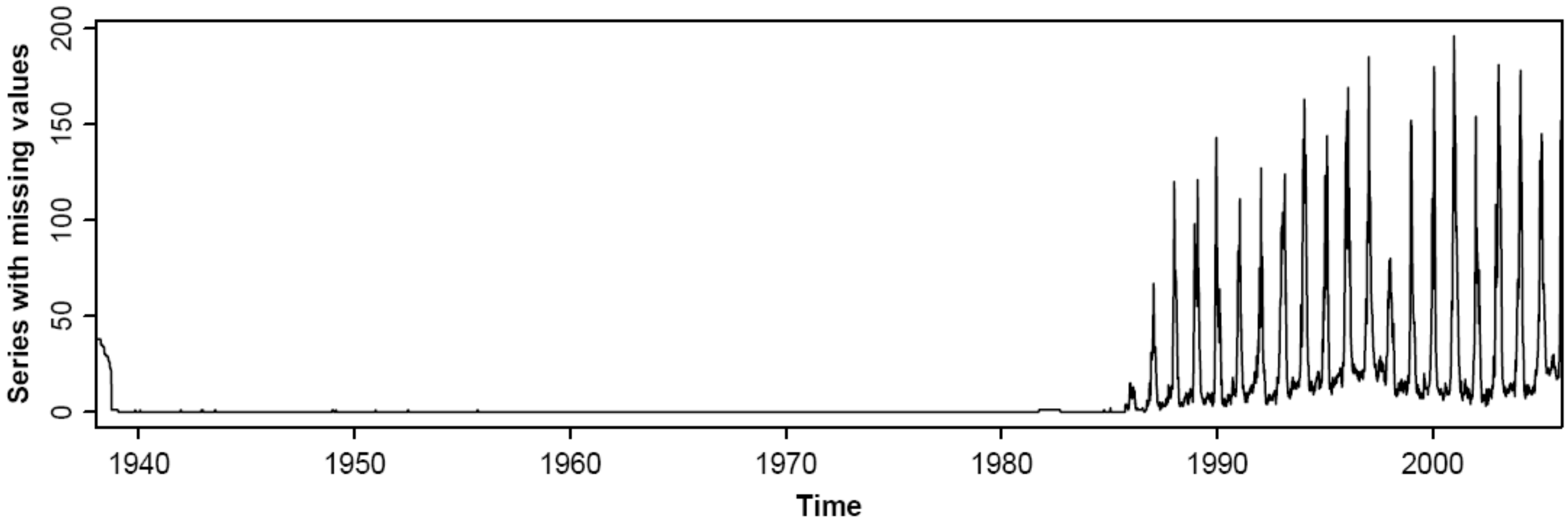


Spectral analysis
and significance tests:



The problem of gaps in data

- Many analysis methods require gapfree data
- Here, the lag-covariance matrix is only calculable / diagonalizable when all data are present
- Measured time series ***do*** contain missing values:



USGS Hydro-Climatic Data Network (HCDN) runoff time series

Gudmundsson 2007

Variants for dealing with gaps in SSA

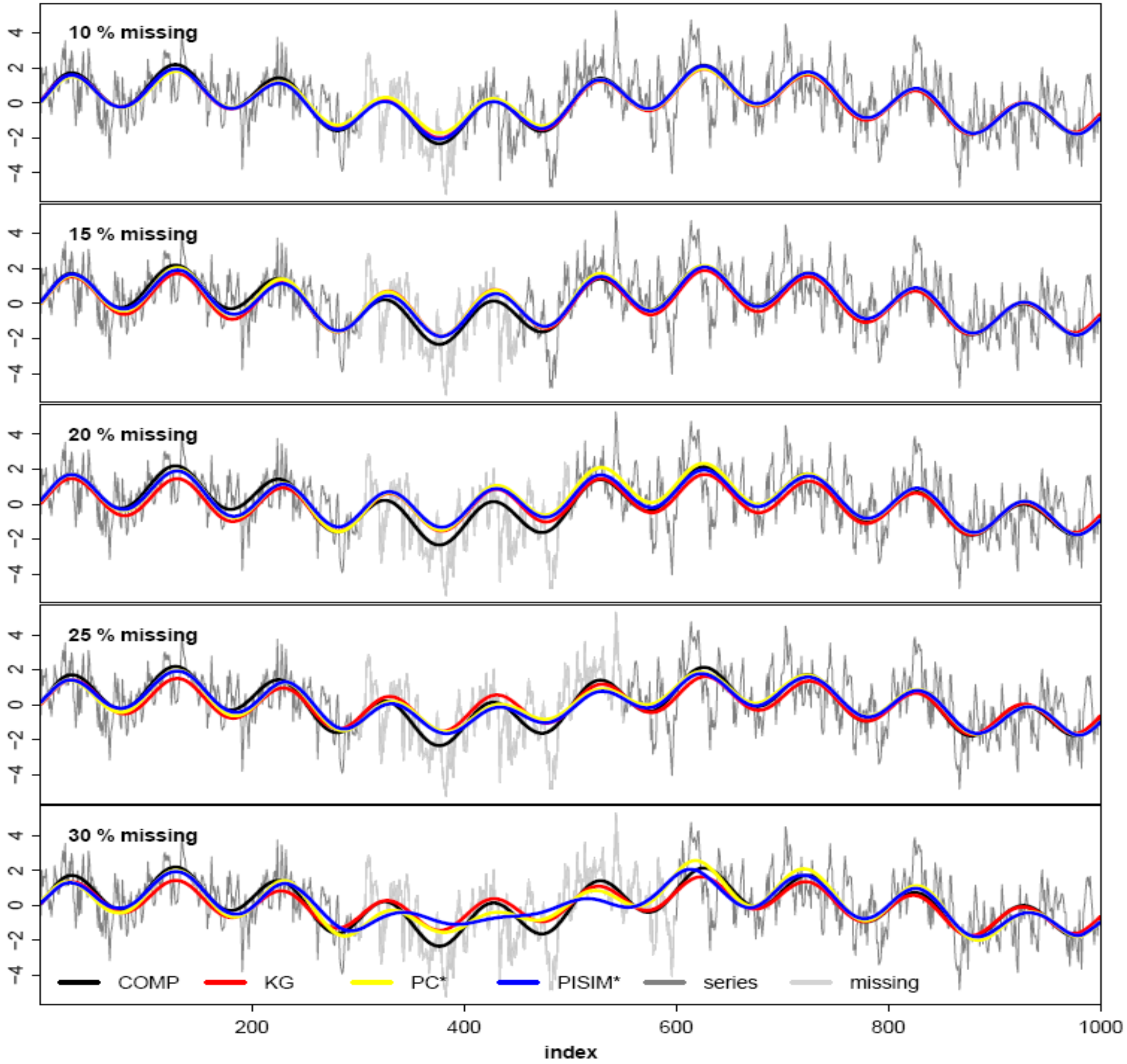
- Kondrashov and Ghil (2006): iterative algorithm
  - Substituting gaps with zeros
  - perform usual SSA, replacing zeros with values from first RC
  - Repeat until convergence (inner loop)
  - Increase number of RCs (outer loop)
- Schoellhamer (2001) and Golyandina and Osipov (2007): direct method(s)
  - generalize inner product (star product '*')

$$\vec{A} * \vec{B} = \frac{n}{n - |A \cup B|} \sum_{k \notin A \cup B} a_k b_k$$

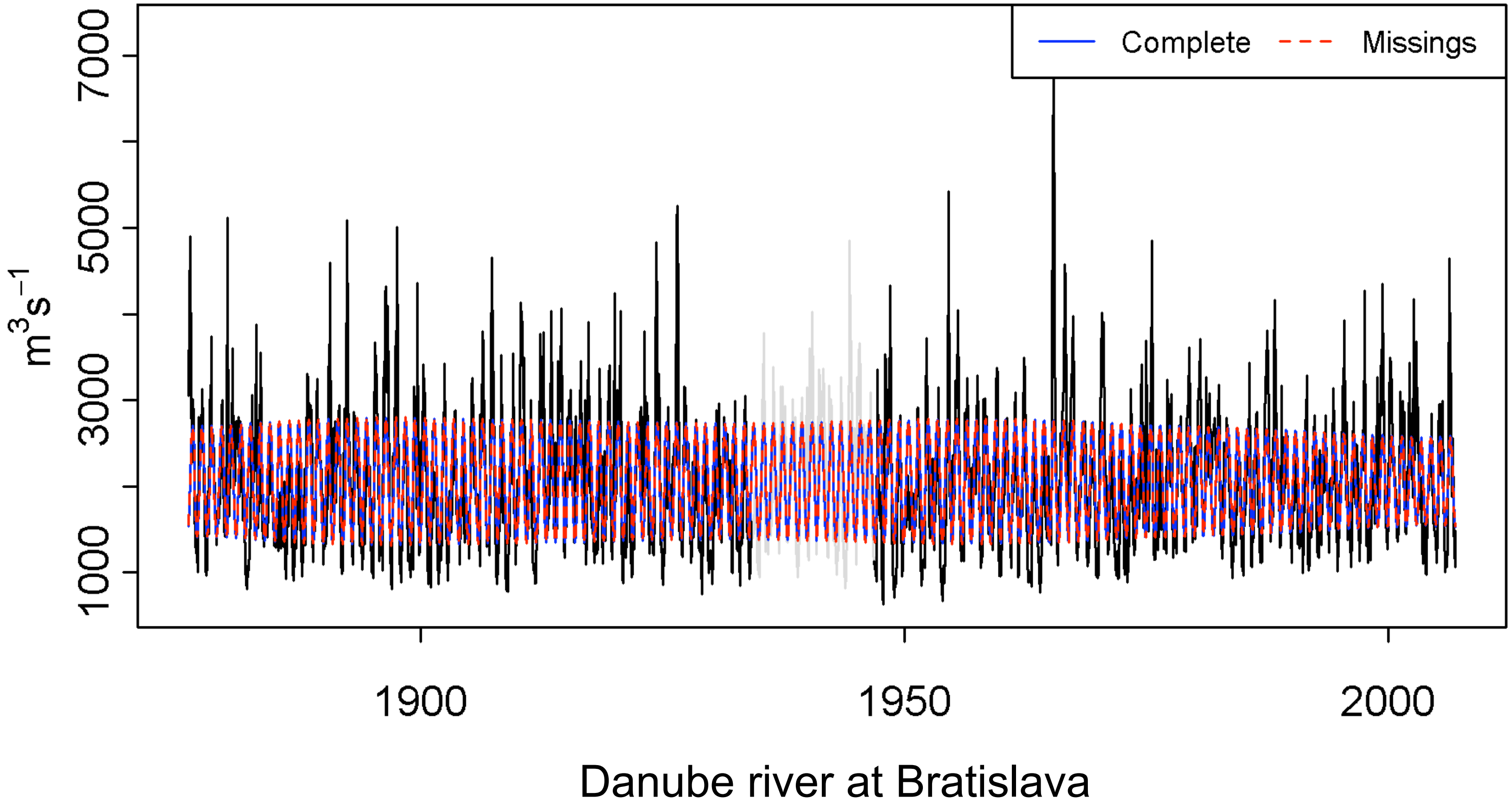
  - recovery by means of principal components
  - (or): projection and simultaneous filling in
  - several extensions/variants…("PC", "PISIM", …)

Gapfilling with SSA: artificial example

Two sine waves
plus autoregressive noise;
first four RCs used;
single large gap present;
uses Toeplitz modification



Gapfilling with SSA: real-world example



- 12-year gap artificially introduced
- first two RCs displayed
- perfect agreement

Test cases for gappy data

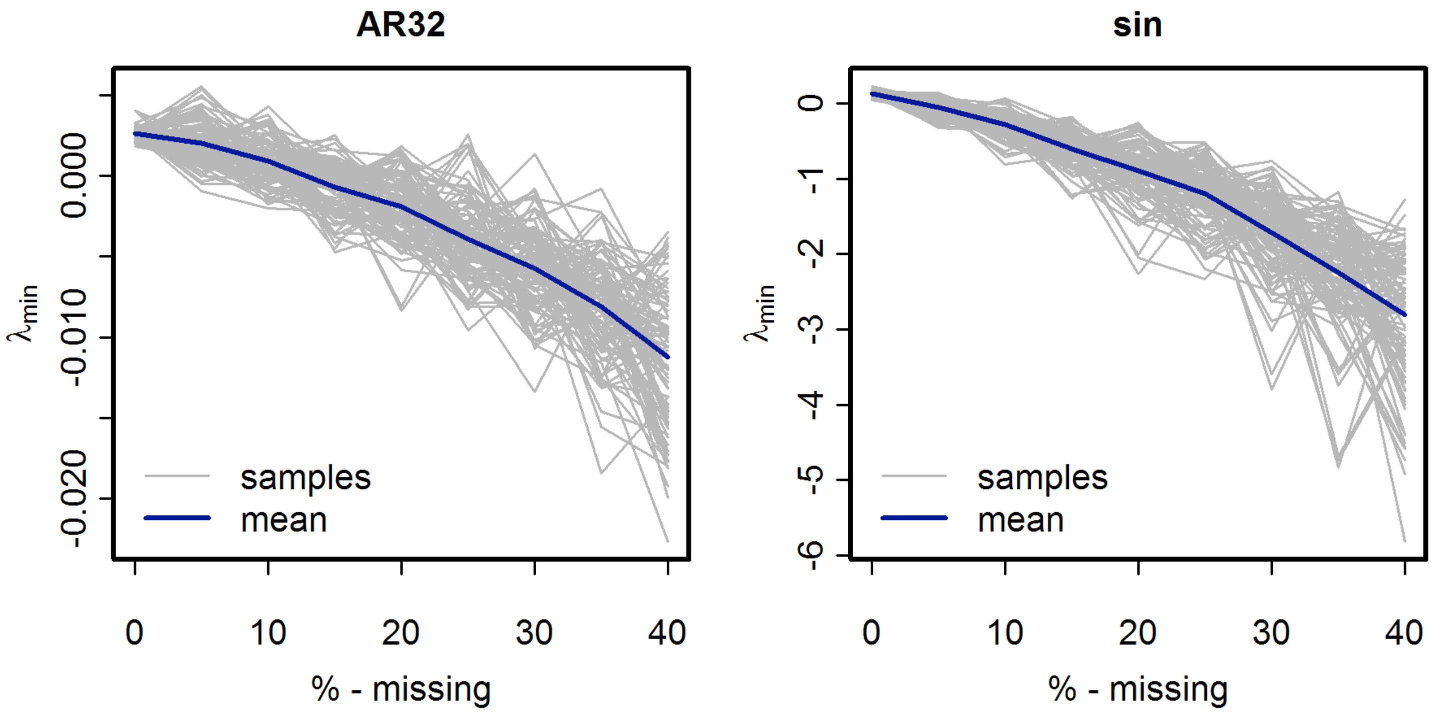
Two synthetic series:

- ARMA(3,2) process
- sine wave (period 25)
- series length N=200, window length M=50
- many realizations

Two gap variants:

- periodic gap occurrence, period length=10,
  number of missing values from 0 to 4
- Poisson-distributed gaps, gapfraction from 0 to 40%

A fundamental problem



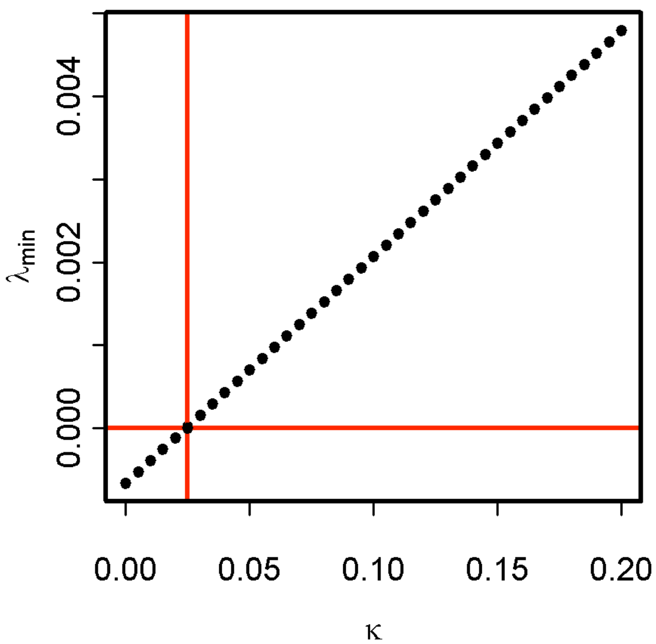
- smallest eigenvalue turns negative
- however, covariance matrix C is a positive definite matrix!
- normalized eigenvalues give the percentage of explained variance ??
- corresponding components unusable

Solution to the fundamental problem: shrinkage

- Introduce a parameter κ, the shrinkage intensity
- Drive C into a direction towards ”positive definite”:

$$C^* = (1-\kappa)C + \kappa I$$

- Analytical approach (Ledoit-Wolf theorem, minimizing quadratic loss) fails in our case
- Heuristically, the optimal shrinkage intensity can be approximated



- From now on, we work with *shrinked* covariance matrices

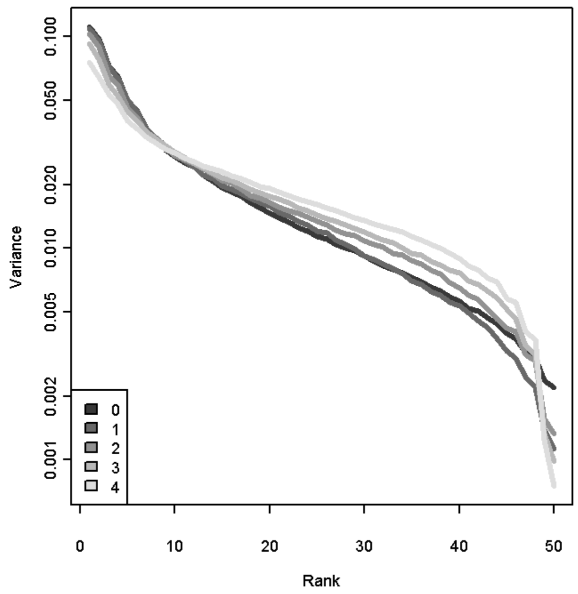
- Loss quantification: Stein’s loss equation

$$L_S = tr(C^* C^{-1}) - \log\det(C^* C^{-1}) - rank(C)$$

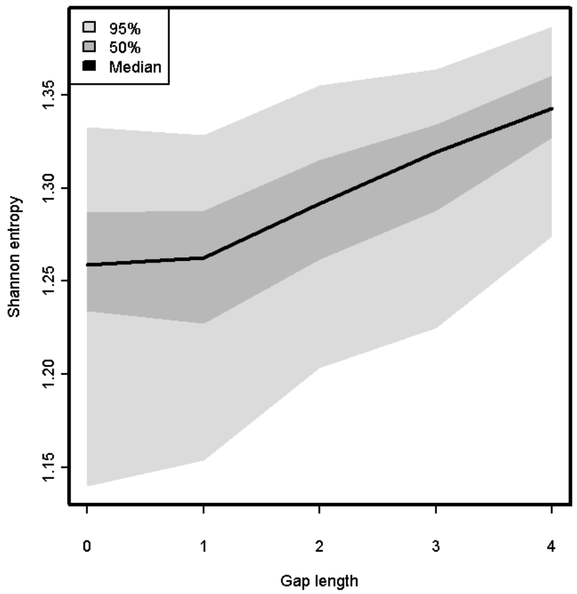
Performance of gapfilling

Example: ARMA(3,2) with periodic gaps

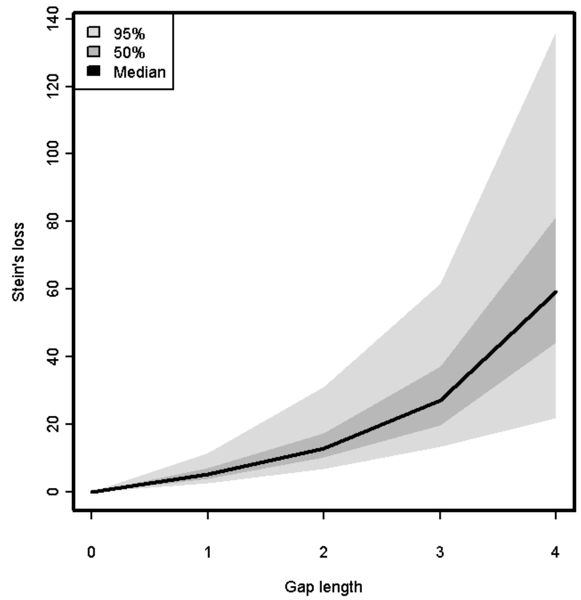
- Stability of eigenspectrum



Eigenvalues



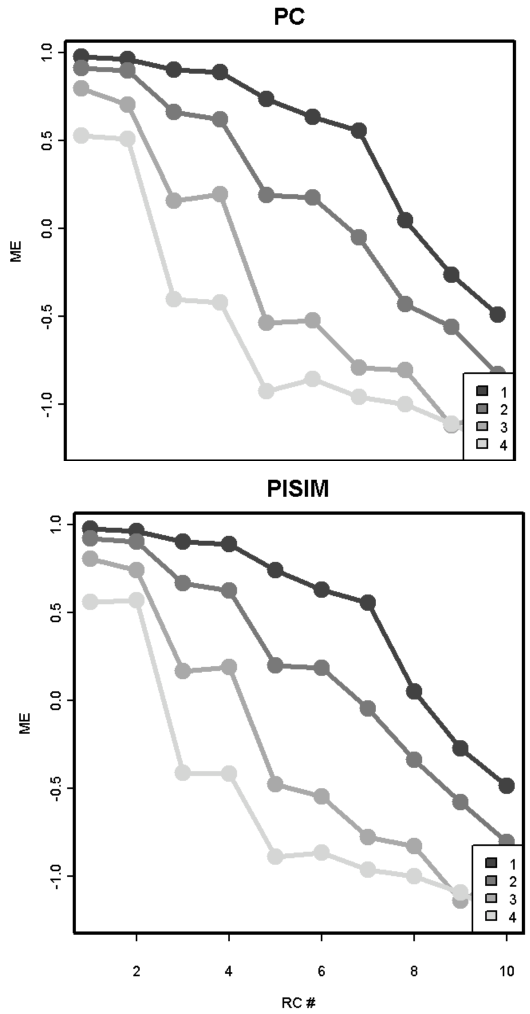
Entropy of eigenvalues



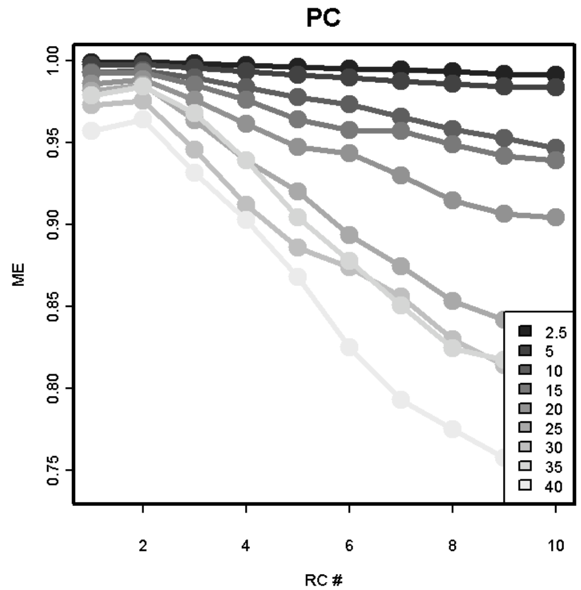
Loss function

Performance of gapfilling

ARMA(3,2) with periodic gaps



Model efficiency or RCs (separately)

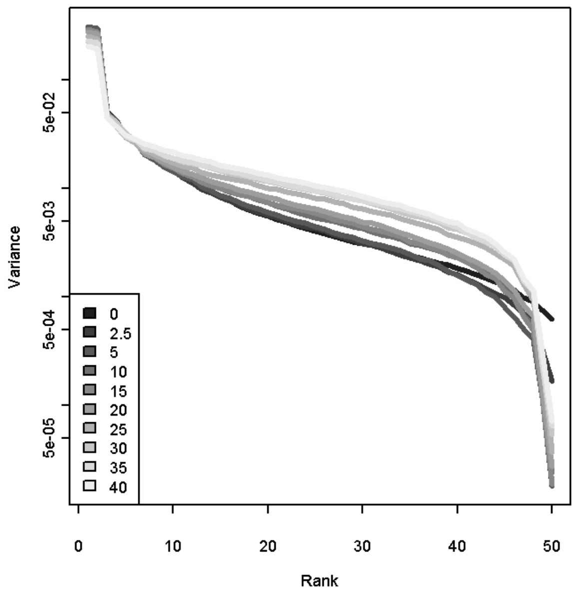


Model efficiency or RCs (accumulated)

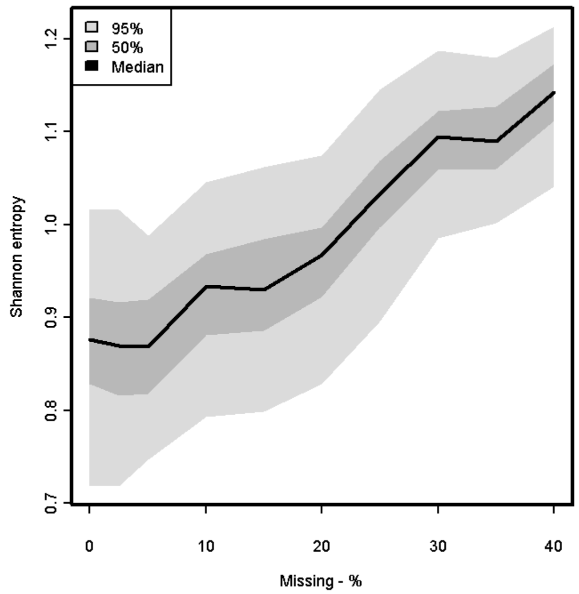
Performance of gapfilling

Example: Sine wave with Poisson gaps

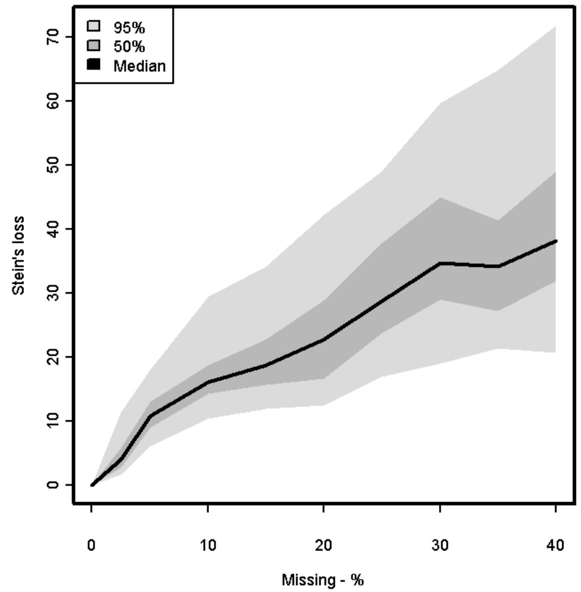
- Stability of eigenspectrum



Eigenvalues



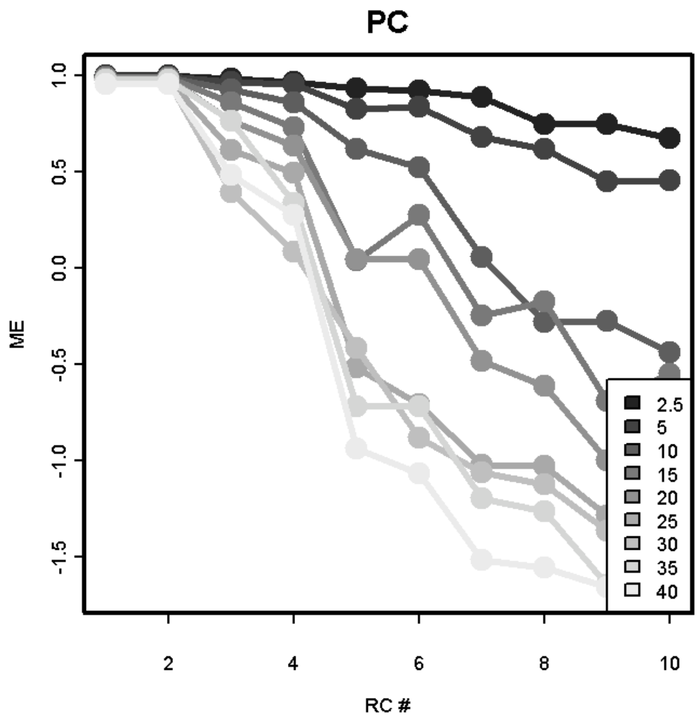
Entropy of eigenvalues



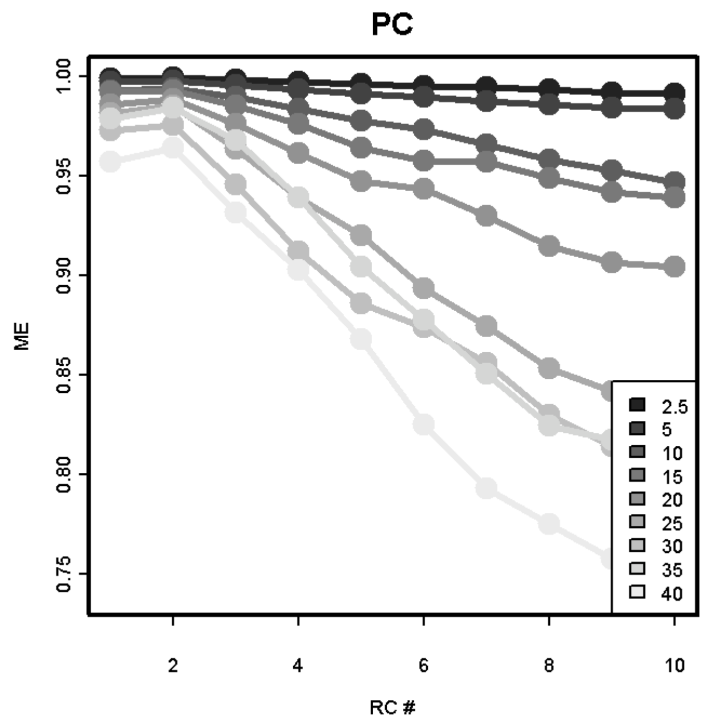
Loss function

Performance of gapfilling

Sine wave with Poisson gaps



Model efficiency or RCs (separately)



Model efficiency or RCs (accumulated)

Conclusions

- SSA extensible to gappy data
- Gapfilling with SSA possible
- Caution: covariance matrices have to be shrinked
- Eigenspectra robust at realistic gap ratios (depending on process and gap structure, however)
- Eigenfunctions (RCs) more sensitive to gap presence
- Real-world examples: successful gap filling, performance dependent on gap structure and timescale