Multi-channel Singular Spectrum Analysis (MSSA)

Teija Seitola
FMI
16.11.2015

Contents

- What is Multi-channel singular spectrum analysis (MSSA)?
- Experiments with temperature data set
- Monte-Carlo MSSA in identifying the significant oscillations

Multi-channel singular spectrum analysis

- MSSA provides an efficient method to identify oscillatory behavior in a high-dimensional multivariate data set
- The idea is to identify spatially and temporally coherent patterns that maximize the lagged covariance of the data set
- MSSA has similarities to traditional PCA → main difference is that MSSA also takes into account the lagged correlations
- Non-parametric method → In contrast with e.g. Fourier analysis with fixed basis of sine and cosine functions, MSSA uses an adaptive basis generated by the time series itself

MSSA

$$\begin{array}{cc} L & \\ \boldsymbol{X}=\begin{bmatrix} \boxed{\begin{matrix} x_{1,1} \\ x_{2,1} \\ x_{3,1} \\ \vdots \\ x_{N,1} \end{matrix}} & \begin{matrix} x_{1,2} \\ x_{2,2} \\ x_{3,2} \\ \vdots \\ x_{N,2} \end{matrix} & \begin{matrix} x_{1,3} \\ x_{2,3} \\ x_{3,3} \\ \vdots \\ x_{N,3} \end{matrix} & \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} x_{1,L} \\ x_{2,L} \\ x_{3,L} \\ \vdots \\ x_{N,L} \end{matrix} \end{bmatrix} & N \end{array}$$

Original data ***X:***

- *N* = number of time steps
- *L* = number of channels, e.g. grid points
- Possible preprocessing steps, eg. centering, scaling

Lag 0

Lag 1

Lag 2

Lag *M*

Next step is to construct an augmented data matrix:

- Choose *M* (lag window) → *M* lagged copies of each channel in ***X***
- *N'* = *N* – *M* + 1

$$\boldsymbol{Y}_{\boldsymbol{i}}=\begin{bmatrix} x_{1,i} & x_{2,i} & x_{3,i} & \cdots & x_{M,i} \\ x_{2,i} & x_{3,i} & x_{4,i} & \cdots & x_{M+1,i} \\ x_{3,i} & x_{4,i} & x_{5,i} & \cdots & x_{M+2,i} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{N',i} & x_{N'+1,i} & x_{N'+2,i} & \cdots & x_{N,i} \end{bmatrix}, i=1\ldots L$$

$$\boldsymbol{A}=[\boldsymbol{Y}_{\boldsymbol{1}} \;\; \boldsymbol{Y}_{\boldsymbol{2}} \;\; \cdots \;\; \boldsymbol{Y}_{\boldsymbol{L}}]$$

- Dimensions of ***A***
  - *Rows: N'*
  - *Columns: M*L*

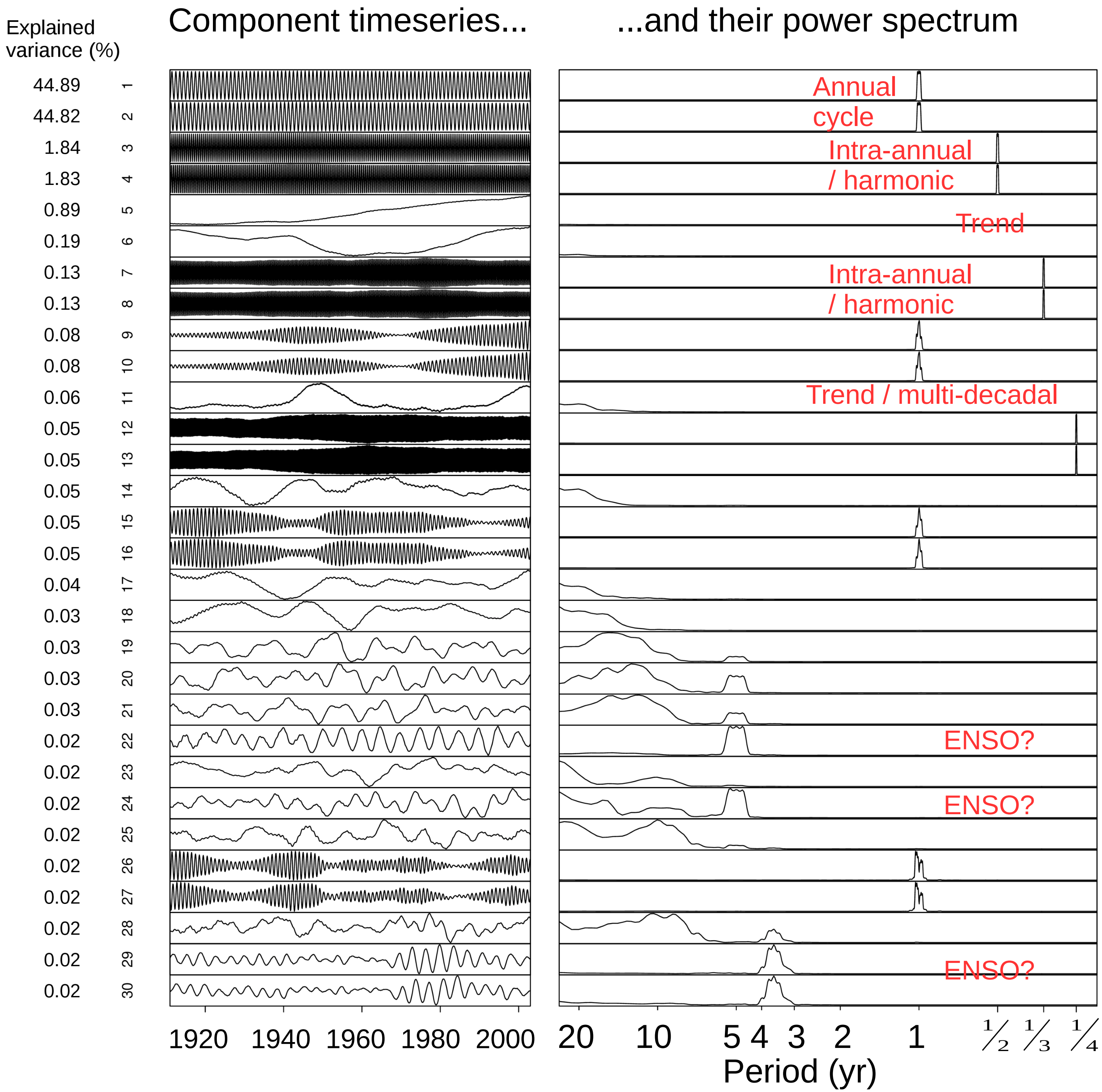
Calculate decomposition of A
→ use e.g. SVD (Singular Value Decomposition)

$$\boldsymbol{A}=\underbrace{\boldsymbol{U}_A}_{\text{ST-PCs}}\boldsymbol{D}_A^{1/2}\underbrace{\boldsymbol{V}_A^T}_{\text{ST-EOFs}}$$

Data set

- Monthly surface temperature field from the 20th Century Reanalysis V2 data
- 1344 time steps (1901/01—2012/12)
- ~2.0 degree latitude x 1.75 degree longitude global grid (192 x 94 = 18 048 grid points)
- $\boldsymbol{X}_{N \times L}$, where N = 1344 and L = 18048

Initial experiments with the data set



- M = 240 months = 20 yr
- ST-PCs/EOFs often come in pairs explaining approx. the same variance and are $\pi/2$ out of phase
- Modes with period ≤ *M* can be only presented by such pairs
- BUT: such pairs can also be generated by non-oscillatory processes, such as first-order autoregressive noise → Monte-Carlo test for MSSA results

Monte-Carlo MSSA

(Allen and Robertson, 1996)

- Components are tested against a null-hypothesis of the data being generated by independent AR(1) processes (i.e. red noise)
- The red noise model:

$$u_{t+1,s} = \gamma_s u_{t,s} + \alpha_s w_{t,s}$$

- γ_s is the lag-1 autocorrelation of channel *s* (in the original data set)
- $\alpha_s = \sqrt{c_s(1-\gamma_s^2)}$, c_s is the variance of channel *s*
- $w_{t,s}$ is gaussian white noise

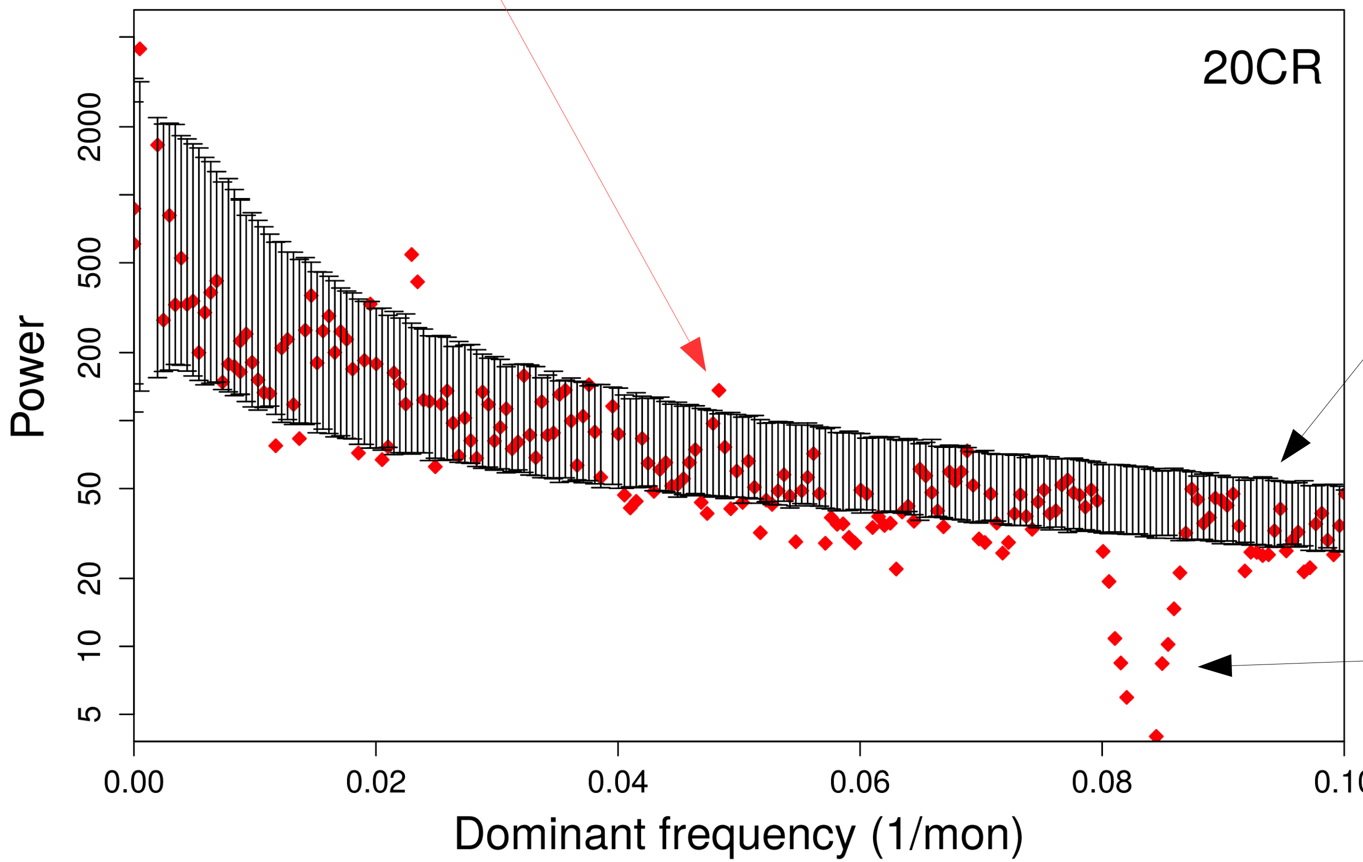
Application of Monte-Carlo MSSA

- Some preprocessing:
  - The original data set $\boldsymbol{X}_{N \times L}$ was standardized (0-mean, unit variance)
  - The annual cycle was dominating → estimated by STL (Loess based Seasonal-Trend Decomposition, Cleveland et al. (1990)) and removed
- In the test the input channels should be uncorrelated at zero-lag → SVD of the original data set was calculated and 50 first PCs retained (~70 % of the variance)
- 50 PCs were used as input channels in MC-MSSA
- 1000 realizations of red-noise surrogates were generated → these were analyzed in the same way as the 'real' data set

Example of MC-MSSA result (M=20 yr)

The 'real' data eigenvalues (plotted against the dominant periodicity of the ST-PC corresponding to each eigenvalue)

The periodicities that correspond to eigenvalues rising above the 97.5th percentiles are considered significant at 95% level.



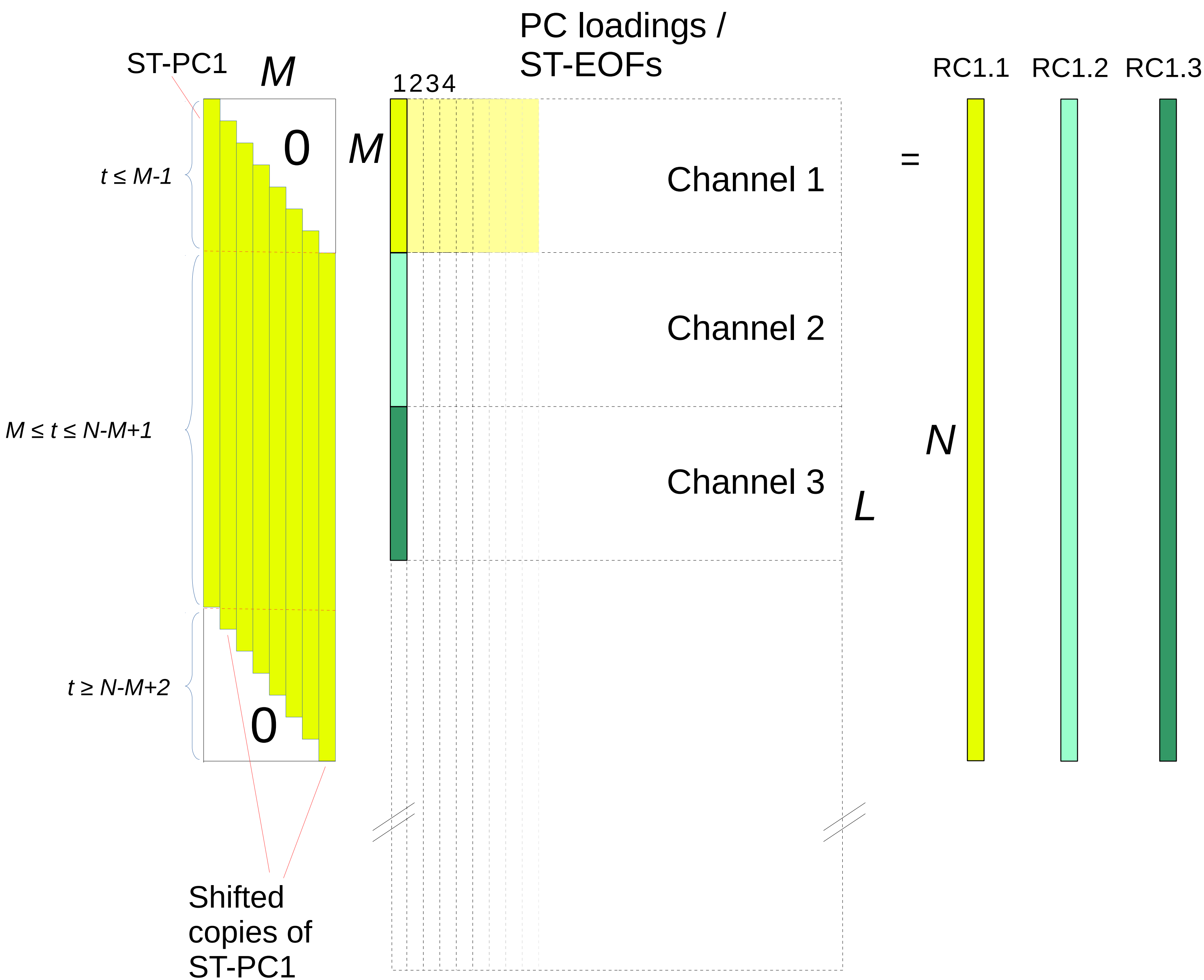
The 2.5th and 97.5th percentiles of the eigenvalue distribution calculated from 1000 realizations of the red-noise surrogates.

Missing power at ~1 yr due to the removal of the annual cycle.

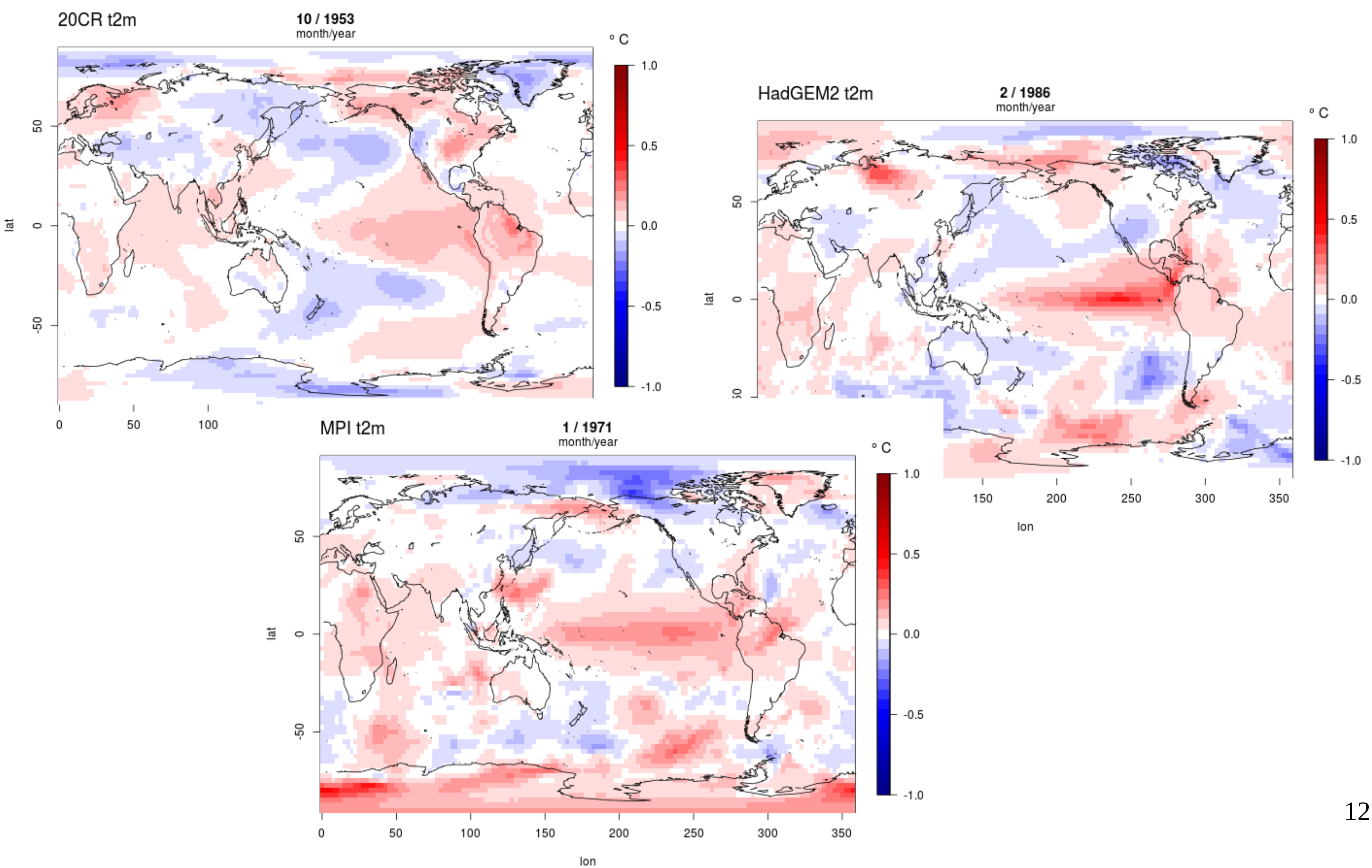
Reconstructed components

- MSSA modes (ST-PCs) are represented in the original index space by their reconstructed components (RC)
- Each RC is a kind of filtered version of the original time series
- RCs are not mutually orthogonal, but their sum across all MSSA modes is identical to the original time series

How to calculate reconstructed components (RCs)



Animations of ~5 yr cycle of monthly near-surface temperature



Randomized algorithm for MSSA

1) construct the original data matrix $\boldsymbol{X}_{N\,x\,L}$
2) Pre-processing, if needed
3) generate k L-dimensional vectors of gaussian distributed random numbers → matrix $\boldsymbol{R}_{L\,x\,k}$ (optional orthogonalization)
4) project $\boldsymbol{X}_{N\,x\,L}$ onto $\boldsymbol{R}_{L\,x\,k}$: $\boldsymbol{P}_{N\times k}=\frac{1}{\sqrt{k}}\boldsymbol{X}_{N\times L}\boldsymbol{R}_{L\times k}$
5) construct augmented matrix $\boldsymbol{A}$ of $\boldsymbol{P}$
6) Calculate SVD of $\boldsymbol{A}$

Summary

- MSSA an efficient method to identify oscillatory behavior in a high-dimensional multivariate data set
- Applied in different areas: climatology, marine science, geophysics, engineering, image processing, medicine, econometrics ...
- Applications: e.g. trend extraction, periodicity detection, seasonal adjustment, smoothing, noise reduction ...

References

- Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., ... & Yiou, P. (2002). Advanced spectral methods for climatic time series. Reviews of geophysics, 40(1), 3-1.
- Allen, M. R., & Robertson, A. W. (1996). Distinguishing modulated oscillations from coloured noise in multivariate datasets. Climate Dynamics, 12(11), 775-784.