Spatiotemporal modes of climatic variability: building blocks of complex networks?

M. Vejmelka J. Hlinka D. Hartman M. Paluš

Institute of Computer Science, Academy of Sciences of the Czech Republic,
Pod vodárenskou věží 2, Prague 8, Czech Republic

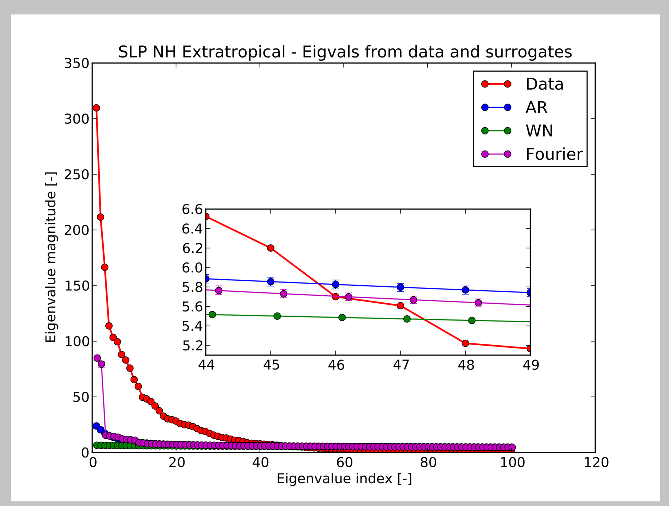
This study was supported by the Czech Science Foundation project No. P103/11/J068.

Complex network approach in Climatic data

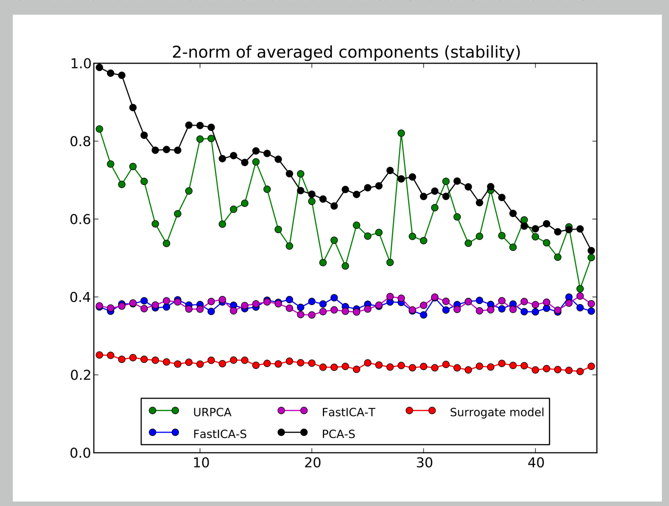
The theory of complex networks offers a rich set of tools aimed at understanding various aspects of high-dimensional spatiotemporal systems [1]. Recently, methods based on complex network theory have been applied to quantities characterizing climatic variability (e.g. [9]). However, complex network methods are typically applied directly to reanalysis data, which is available on a high-resolution planetary grid. However, reanalysis data products results from data assimilation of sparse measurement by means of a global Climate Data Assimilation System [5] and are thus spatially interpolated. Strong local correlations bias computations of complex network statistics, such as the clustering coefficient [8]. Additionally, care must be taken in constructing the network structure from climatic data, which manifest spatial inhomogeneity biasing the estimates of the statistical dependence between regions [7]. Our approach is to reduce the full complexity of the dataset using a suitable method to a much smaller number of spatiotemporal components which preserve allmost all of the variance. Network characteristics are then computed and analyzed on this reduced dataset.

Number of components

In this work we find that it is possible to identify 45 dimensions in the extratropical Northern Hemisphere ($20^\circ - 87.5^\circ$ N) in monthly Surface Level Pressure (SLP) anomalies which together account for 96.8% variance of the original dataset. The number of dimensions 45 was indicated consistently using two surrogate models. The rest of the principal components had eigenvalues in range of the corresponding null distributions and were thus indistinguishable from fluctuations of independent processes with similar autocorrelation structure. The permutation surrogate model does not reproduce the autocorrelation of the time series in the surrogate model and underestimates the covariance from random fluctuations.



(a) Comparison of eigenvalues of covariance matrix from data and from surrogate models.

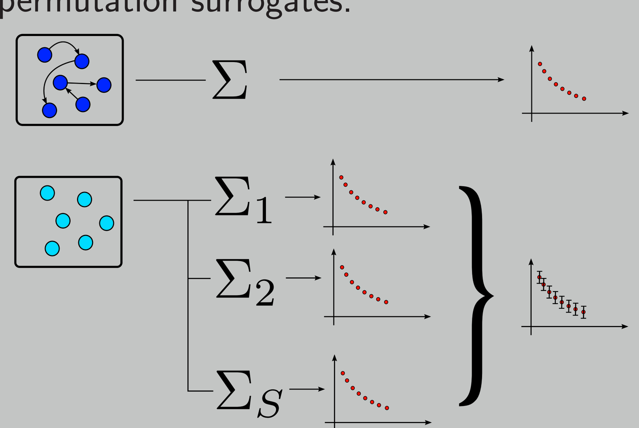


(b) Component stability for computed components (see block Component maps).

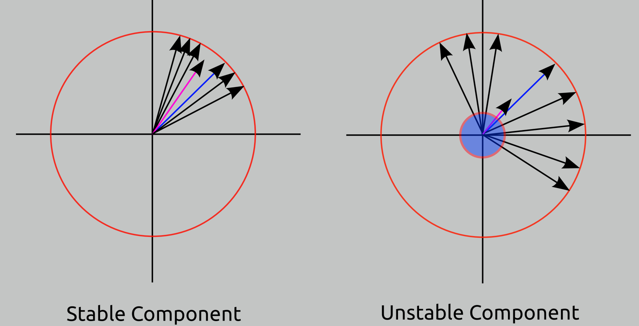
Methods and Materials

Preprocessing The monthly NCEP/NCAR Surface Level Pressure (SLP) data between Jan 1950 and Jan 2012 is transformed to anomalies and the variance of the time series is normalized to 1 for each month separately. The SLP data is available on a grid with grid points closer together as one approaches the North pole, resulting in a smaller area covered by each grid point. To compensate for the unequal area covered by the grid, we scale the time series by the cosine of the lattitude [4]. Hence the sample covariance matrix can be estimated (however this is not done explicitly in this work).

Estimating the number of components We estimate the number of principal components which contain useful information. Many more or less formal criteria for selecting the number of principal components exist, such as the scree plot, Kaiser's criterion, or recently [2]. Some of these are not suitable for analysing large sets of data originating from complex systems while recent ones work not with time series but independent samples. We have devised a method that compares the covariance structure of a surrogate model and finds principal components the eigenvalues of which are outside the distribution of the corresponding eigenvalue in the surrogate model. Then the False Discovery Rate multiple testing strategy is applied to control the occurrence of false positives in the result. Three surrogate models are compared: autoregressive process fitted using the Bayesian criterion to each time series, fourier transform surrogates [**?**] and simple permutation surrogates.



(c) Obtaining null distribution of eigenvalues of covariance matrix



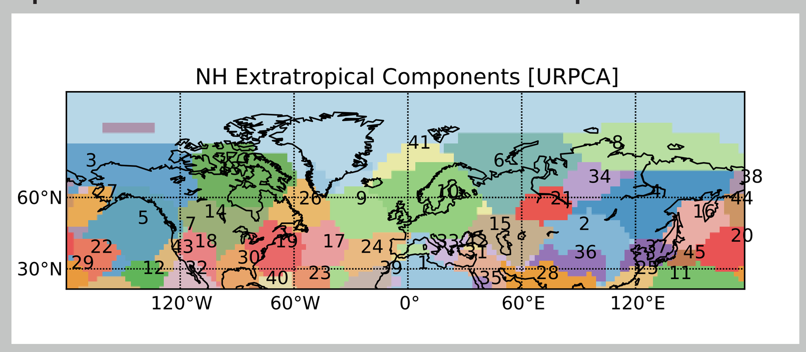
(d) Computing the mean of boostrapped vectors, toy example.

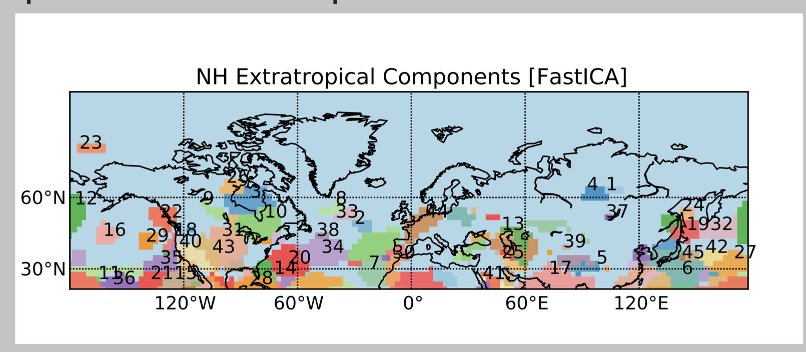
Estimating the components Spatiotemporal components can be found even in purely random data. To estimate the stability of the computed spatiotemporal components, we resample the time series using the bootstrap method (in time) and compute new sets of components. Since most methods return components in an undefined order and polarity, the components must be matched. Our method maximizes the sum of the congruence coefficients between each bootstrap sample and components of the original SLP data. The matched components can then be averaged. If a component is stable, it will be present in many bootstrap samples and the average of this component across bootstrap samples will have a high magnitude. If however, the matched bootstrap components differ widely, then their average will have a magnitude close to zero.

Applied methods temporal/spatial ICA, temporal/spatial PCA, VARIMAX rotated PCA (multiple variants).

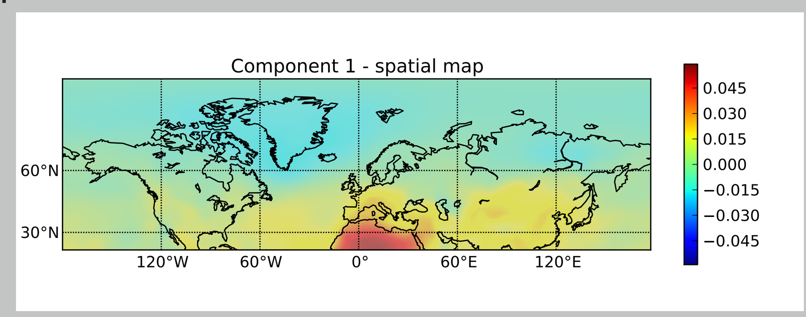
Component maps

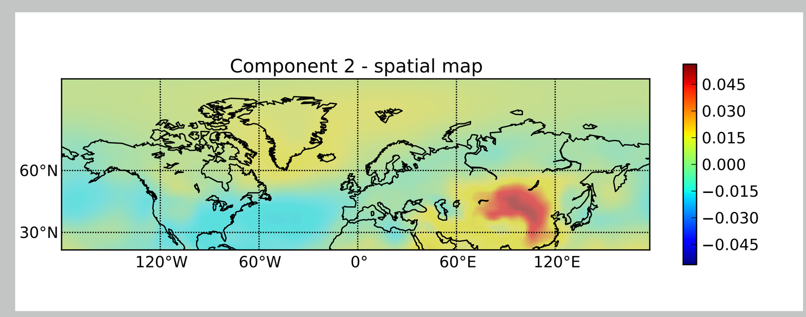
We have applied the machinery detailed in the previous section to several variants of methods of identifying spatiotemporal components and we present results for two of them: VARIMAX rotation of unit size PCs (URPCA) and Independent Component Analysis (ICA). Of the tested methods, URPCA was most stable (except simple PCA) and yielded the best coverage of grid points. Here we show the resulting "clustering", obtained by assigning each grid point to the component with the maximal loading, of URPCA and Spatial FastICA. URPCA attempts to build components with simple structure.



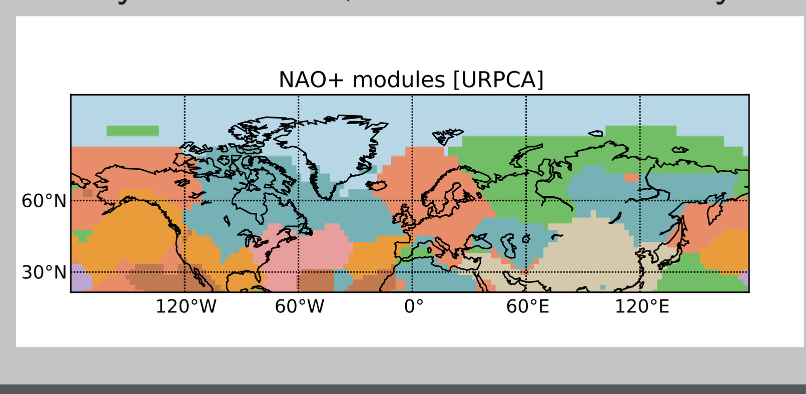


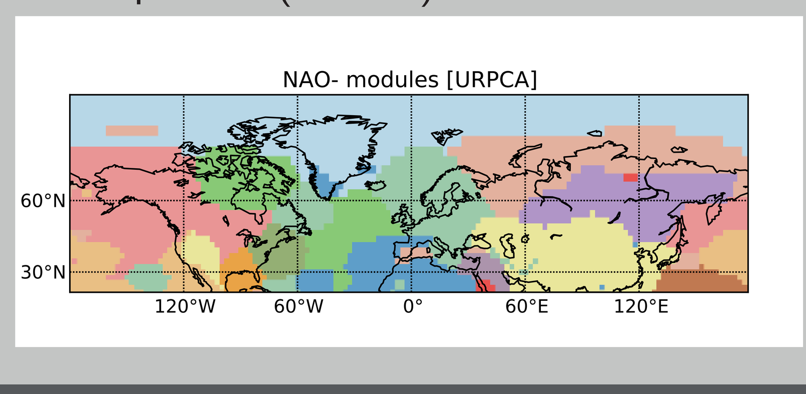
The URPCA decomposition contains two components related to the NAO phenomenon, which together explain about 50% of the variance of the Station-based NAO index and of the CPC NOAA NAO index. Both components are depicted below: one is coupled to pressure fluctuations in China, the second in Northern Africa.





The time series of the URPCA components have been split into NAO+ months (NAO index value in its top third) and NAO- months (NAO value in its bottom third). Network analysis (not shown here) indicates that the degree of the NAO-related components is higher NAO+ years than in NAO- years, while that of component 3 (Alaskan-Pacific) decreases (most). The opposite is true for clustering coefficients. This suggests that in NAO+ years, the NAO components connect more, but not locally, while in NAO- years, the NAO has a limited reach. These observations are supported by the modules computed from the component time series using the algorithm [6] based on a thresholded graph with the top 10% edges included. The optimal number of modules in NAO+ years is 7, while in NAO- years it is 13, while the modularity value is comparable (≈ 0.63).





Summary

- URPCA provides more compact and stable components than other tested methods on the analyzed SLP data, with the exception of PCA methods, which however provide very unevenly sized components, which are typically widely dispersed.
- Both temporal and spatial ICA seem less suitable to the construction of succinct interpretable components. One explanation may be that SLP data are close to multivariate Gaussian (some exceptions of known origins notwithstanding) [3], while ICA requires non-gaussian source to be able to identify them.
- The URPCA decomposition identifies two main NAO-related components which explain 50% variance of the Station NAO time series, for comparison the correlation between Station NAO and CPC/NOAA NAO on the examined segment is 0.67 (45% explained variance).
- Modularity analysis indicates differences in climatic structure: in the NAO+ phase, the NAO-related components are widely integrated, while in the NAO- phase, the large modules break up and NAO-related components are more local.

References

[1] **S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.U. Hwang**. Complex Networks: Structure and Dynamics, Phys. Rep., 424(4–5), 175–308 (2006).

[2] **S. Dray**. On the number of principal components: A test of dimensionality based on measurements of similarity between matrices, Comp. stat & Data Analysis, 52, pp 2228–2237 (2008).

[3] **J. Hlinka, D. Hartman, M. Vejmelka, M. Paluš**. Non-linear contributions to interactions in climate networks: sources, relevance, nonstationarity. EGU Talk: April 25th, Room 28, 14:45-15:00.

[4] **R. Huth**. The effect of various methodological options on the detection of leading modes of sea level pressure variability. Tellus, 58A, pp. 121–130 (2006).

[5] **R. Kistler et al**. The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, 82:247–268, 2001.

[6] **M.E.J. Newman** Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E, 74:036104 (2006).

[7] **M. Paluš, D. Hartman, J. Hlinka and M. Vejmelka**. Discerning connectivity from dynamics in climate networks. Nonlin. Processes Geophysics, 18, pp. 751–763 (2011).

[8] **A. Tsonis, K.L. Swanson and G. Wang**. Estimating the clustering coefficient in scale-free networks on lattices with local spatial correlation structure, Physica A, 387, pp. 5287–5294 (2008).

[9] **A. Tsonis and K.L. Swanson**. Topology and Predictability of El Nino and La Nina Networks, Phys. Rev. Lett., 100, 228502 (2008).

http://ndw.cs.cas.cz/ {vejmelka,hlinka,hartman,palus}@cs.cas.cz