

Analyzing regimes of mid-latitude atmosphere circulation by novel nonlinear data decomposition method

Dmitry Mukhin

Institute of Applied Physics of the Russian Academy of Sciences, Nizhny Novgorod, Russia

Abdel Hannachi

Stockholm University, Stockholm, Sweden

Circulation of the mid-latitude atmosphere:

high-dimensional and multiscale variability

Time scales:

synoptic (weather), intra- and inter-seasonal, interannual, decadal, etc.

Low-frequency variability (LFV)

LFV modulates synoptic activity:

weather regimes – recurrent and persistence patterns of LFV
(recurrent trajectories in the phase space)

See, e.g.,

Hannachi, A., et al. (2017). *Reviews of Geophysics*, 55(1), 199–234. <https://doi.org/10.1002/2015RG000509>

Ghil, M., et al. (2019). *Sub-Seasonal to Seasonal Prediction*. Elsevier, 119–142. <https://doi.org/10.1016/B978-0-12-811714-9.00006-1>

Problems related to long-term forecast of LFV:

- Revealing significant intra-seasonal LFV regimes (modes) from data
- Low-dimensional representation of their long-term (climatic) activity

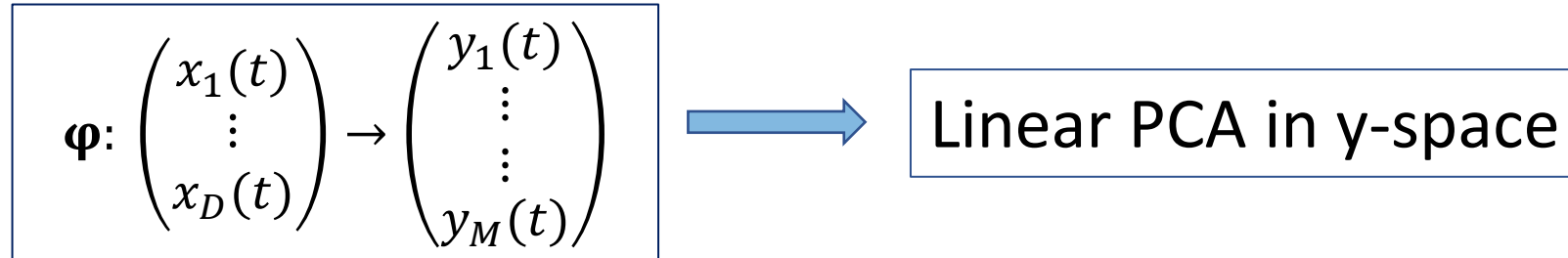
For analyzing the dynamics of the LFV modes, we need:

- **A method for detecting statistically significant modes**
- **An “embedding” for the LFV regimes:** a few dynamical variables for representing dynamics of the modes, e.g., transitions from one mode to another

Here we suggest an approach that combines the nonlinear principal component analysis (kernel PCA) with the recurrence network analysis

Nonlinear principal component analysis: Kernel PCA

Constructing a nonlinear embedding of the data space



Temporal covariances (**dot products**) in a new M-dimensional space:

$$W_{ij} = \sum_{k=1}^M \varphi_k(\mathbf{x}(t_i)) \varphi_k(\mathbf{x}(t_j)) = K(\mathbf{x}(t_i), \mathbf{x}(t_j)), \quad i, j = 1, \dots, N$$

Kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ – all we need to know for obtaining PCs (φ_k is unknown).

The kernels emphasizing similarity between two states (spatial patterns) $\mathbf{x}_i = \mathbf{x}(t_i)$, $\mathbf{x}_j = \mathbf{x}(t_j)$:

Gaussian kernels: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left\{ -\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{2a^2} \right\}$ $d(\mathbf{x}_i, \mathbf{x}_j)$ – distance between the states

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=0}^{\infty} \gamma_k \varphi_k(\mathbf{x}_i) \varphi_k(\mathbf{x}_j)$$

Formally, we do the linear PCA in an infinite dimensional space

$$\text{Distance matrix } d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$$

Kernel PCA

$$\mathbf{K}_{ij} = \exp \left\{ -\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{2a^2} \right\}$$

Decomposing $\mathbf{K}$ into orthogonal basis:

$$\mathbf{K} = \sum_l \mathbf{P}_l \mathbf{P}_l^T \sigma_l^2$$

$\mathbf{P}_l$ is l th principal component

Very different system's states should be well-separated in the space of a few leading component $\mathbf{P}$

Recurrence network

$$\mathbf{R}_{ij} = \begin{cases} 1, & \mathbf{K}_{ij} > \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

Partitioning the graph with the adjacency matrix $\mathbf{R}_{ij}$: finding communities in the set of system's states.

Spectral method for detecting significant communities:

M. E. J. Newman. Modularity and community structure in networks. (2006) PNAS, 103 (23) DOI: 10.1073/pnas.0601602103

This method allows to divide the states into communities which we associate with the modes. Also, it allows to estimate the centrality of each state in its community.

Kernel PCA

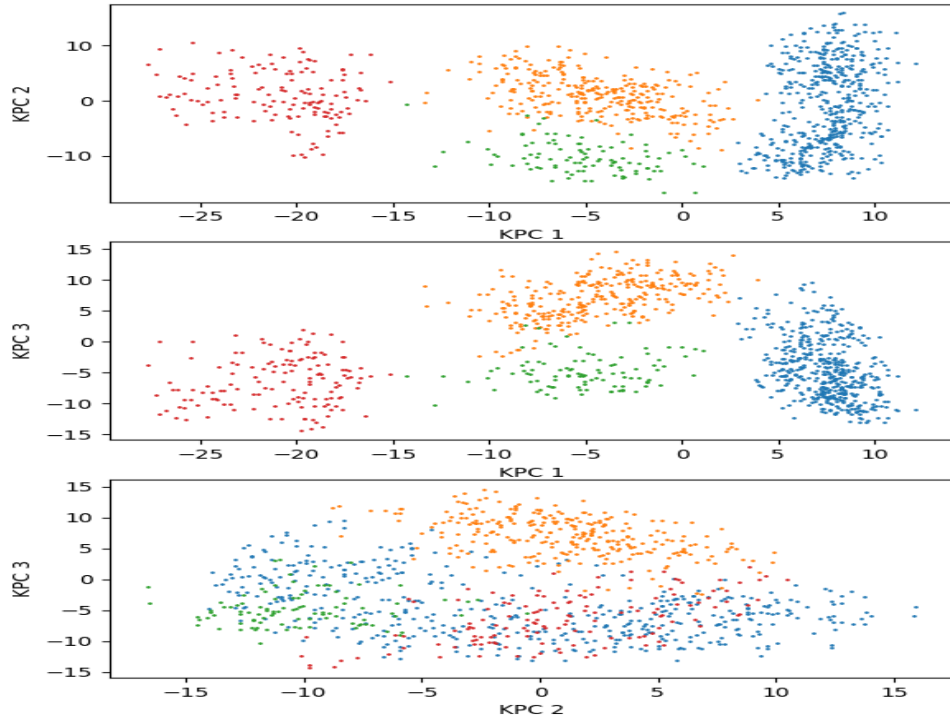
$$\mathbf{K}_{ij} = \exp \left\{ -\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{2a^2} \right\}$$



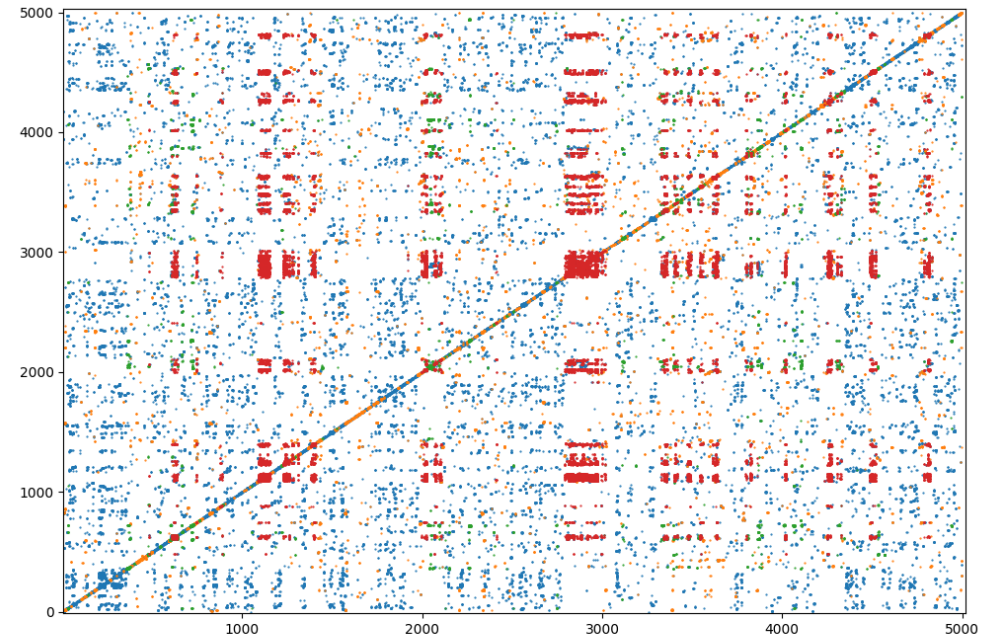
Recurrence network

$$\mathbf{R}_{ij} = \begin{cases} 1, & \mathbf{K}_{ij} > \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

Here we use Euclidian metric $d(\mathbf{x}_i, \mathbf{x}_j)^2 = |\mathbf{x}_i - \mathbf{x}_j|^2$



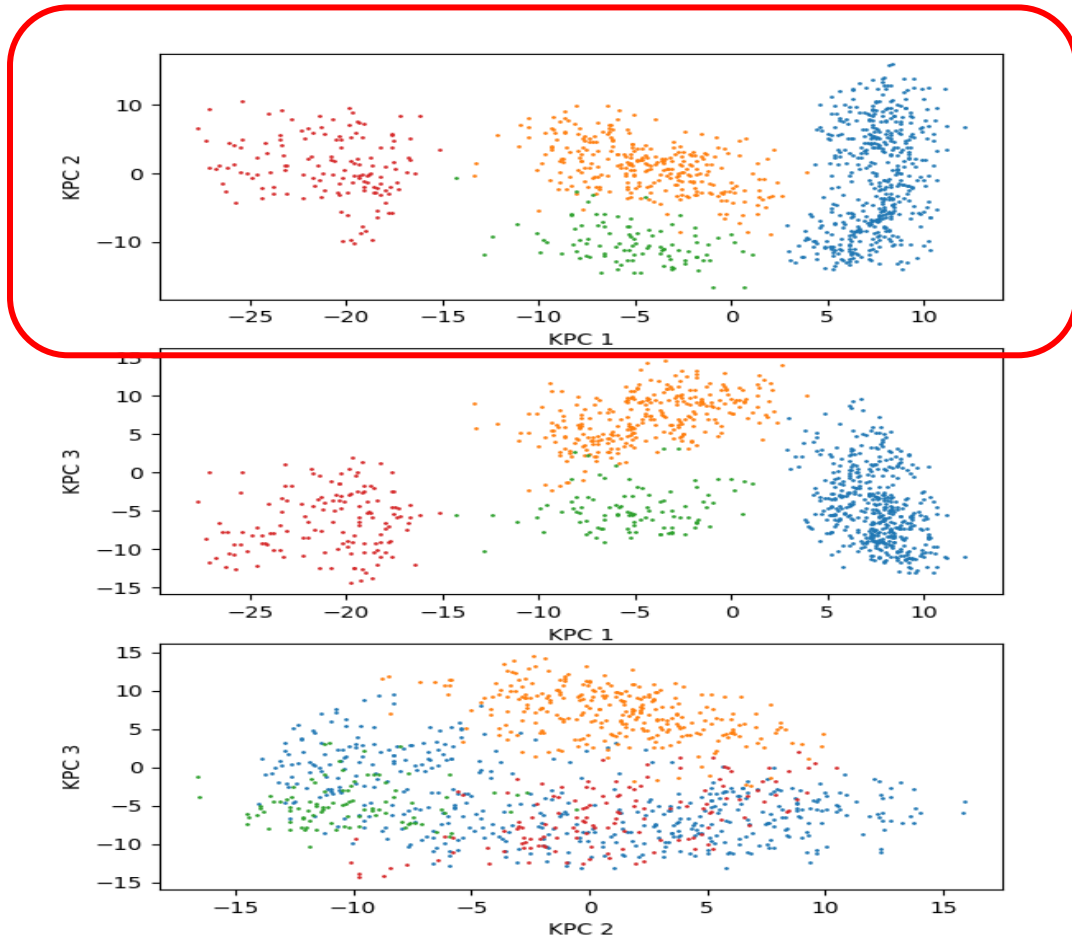
4 modes
detected



Quasigeostrophic 3-layer (QG3) model of the winter North Hemisphere atmosphere. Data is the stream function field anomalies from the middle layer (500 hPa) of 5000 days duration.

QG3 model, daily data

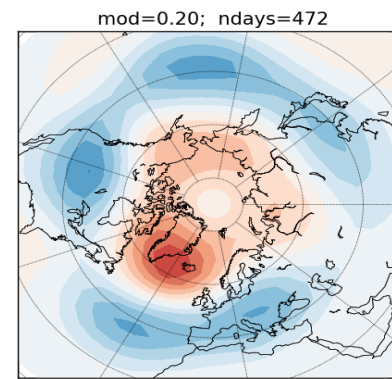
The modes are embedded in PC1-PC2 plain:



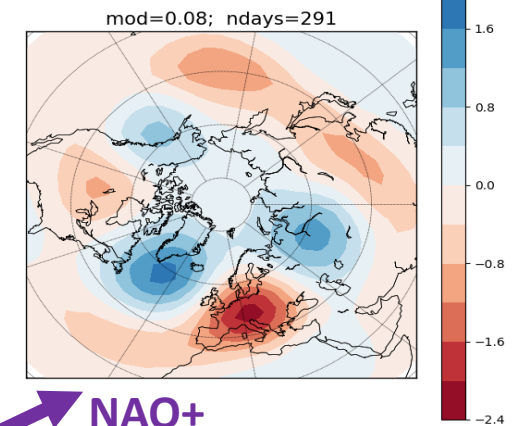
Here we use Euclidian metric $d(\mathbf{x}_i, \mathbf{x}_j)^2 = |\mathbf{x}_i - \mathbf{x}_j|^2$

Composite patterns
(means over 20% most central points for each mode)

Mode 1



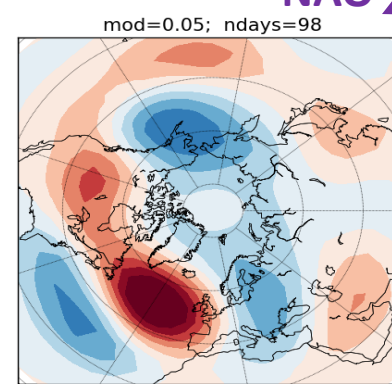
Mode 2



AO-

NAO+

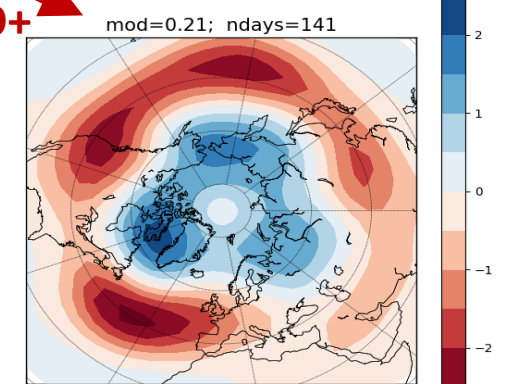
Mode 3



NAO-

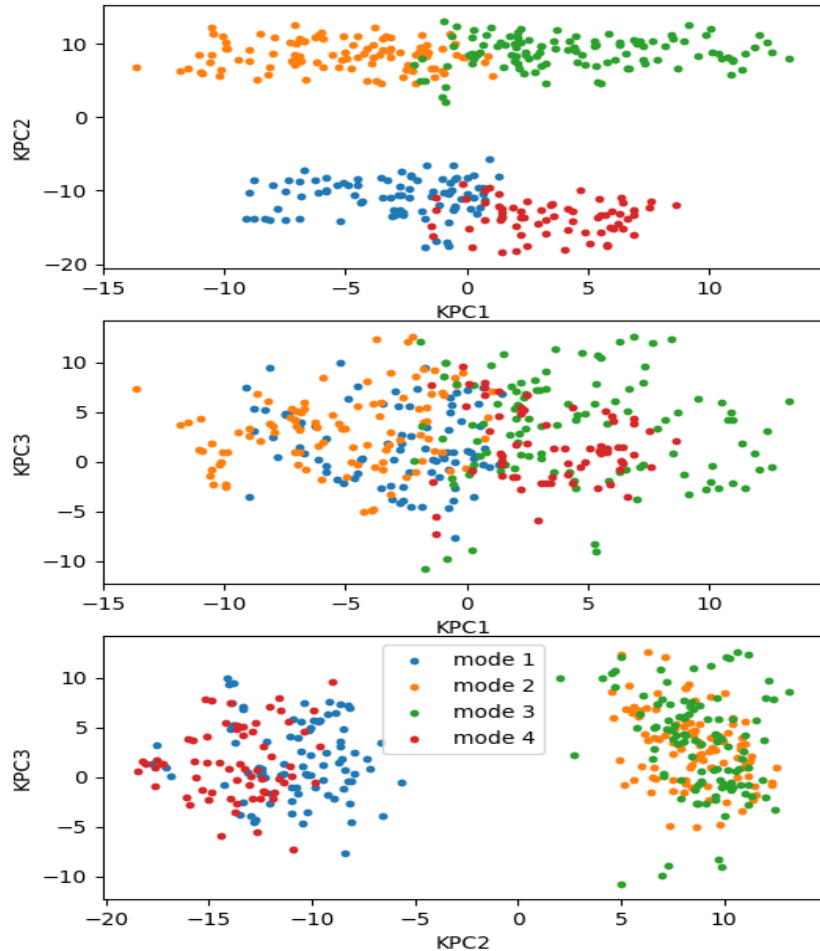
AO+

Mode 4

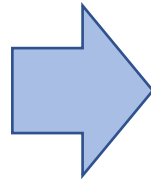


Reanalysis daily data: HGT-500, 1981-2019, winters

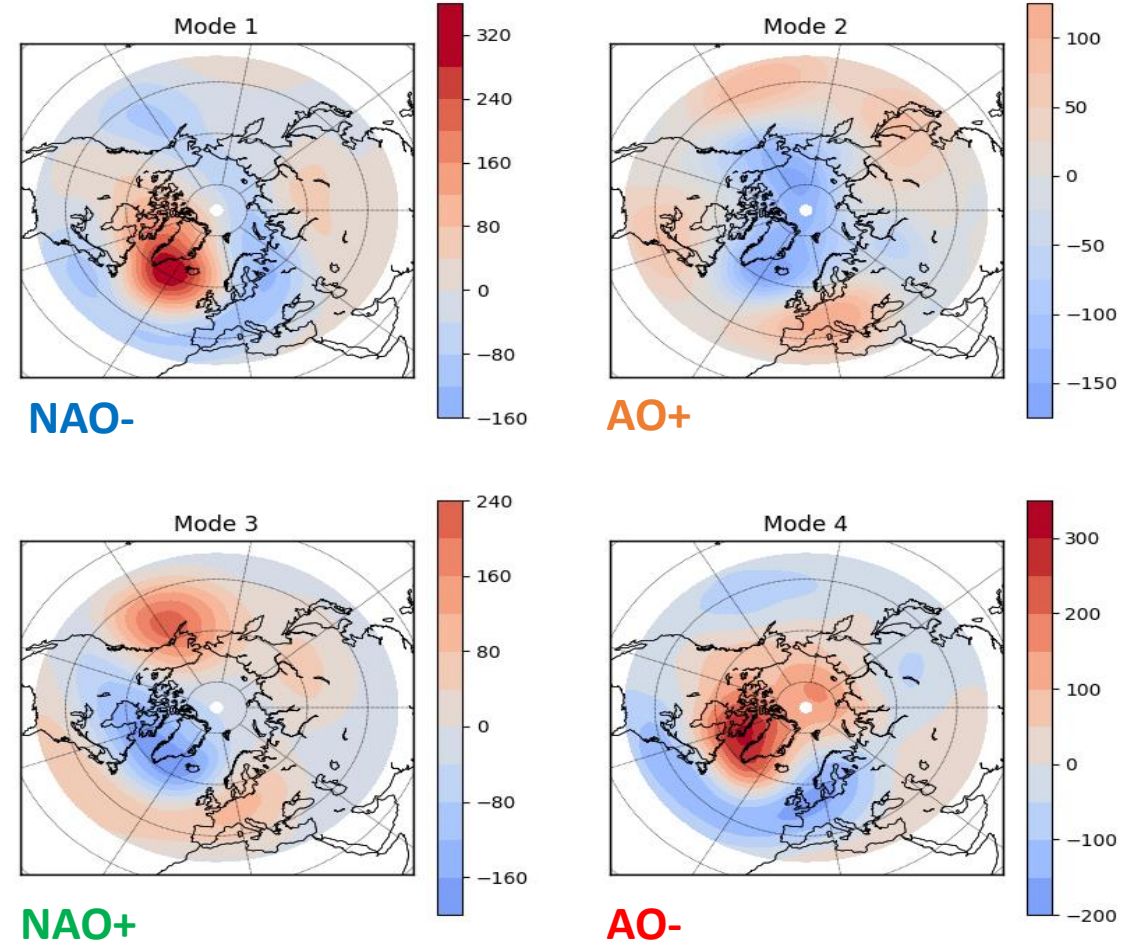
Modes in the kernel PC space
20% most central points are shown



4 modes
detected



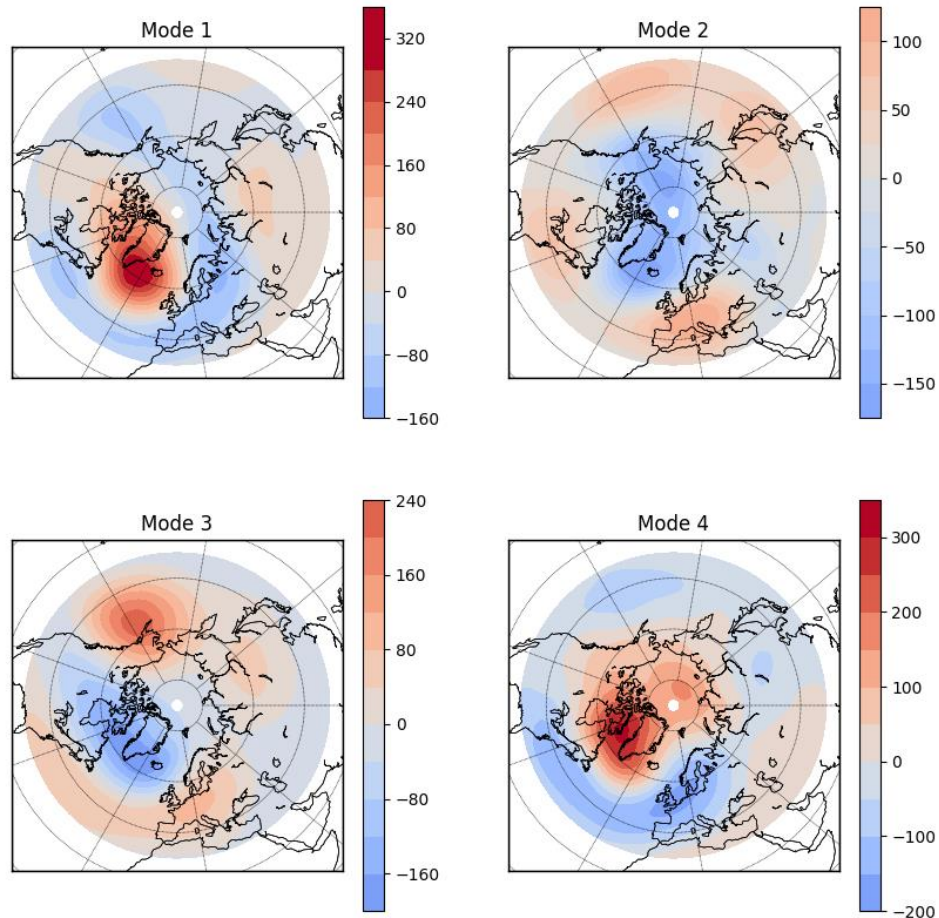
Composite patterns
(means over 20% most central points for each mode)



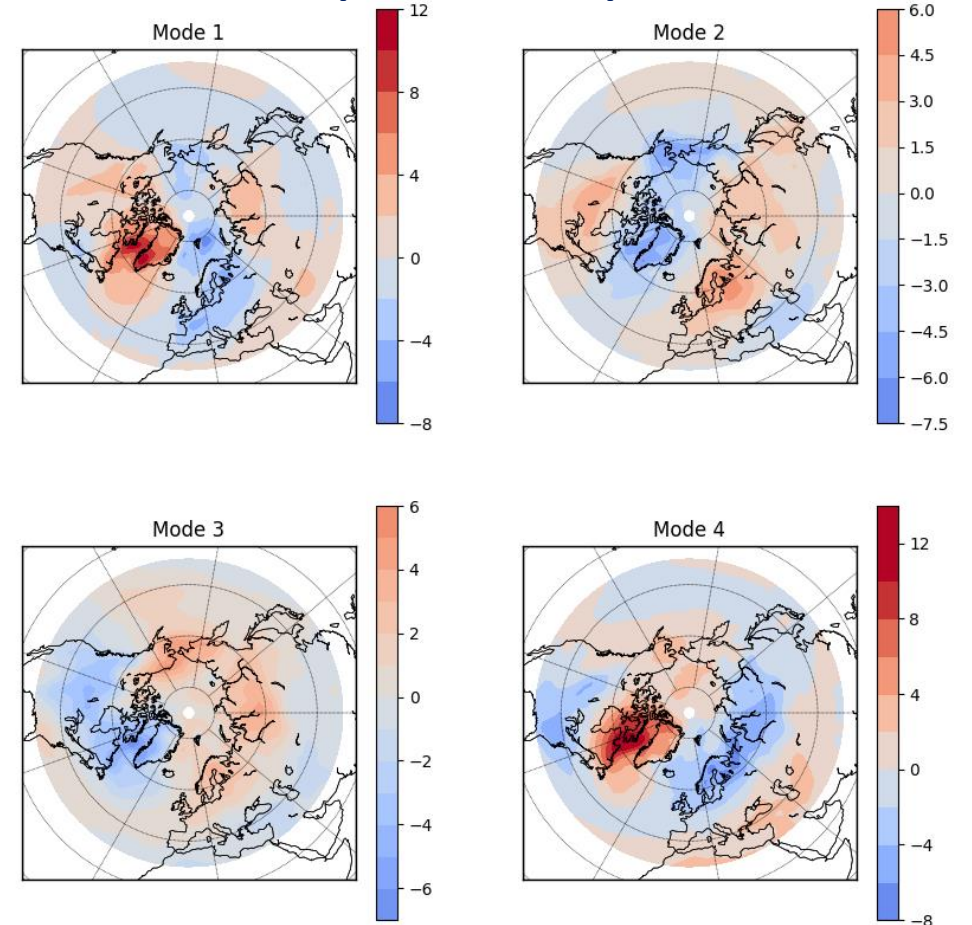
Reanalysis daily data: HGT-500, 1981-2019, winters

Relation of the modes to weather anomalies

HGT500 composites



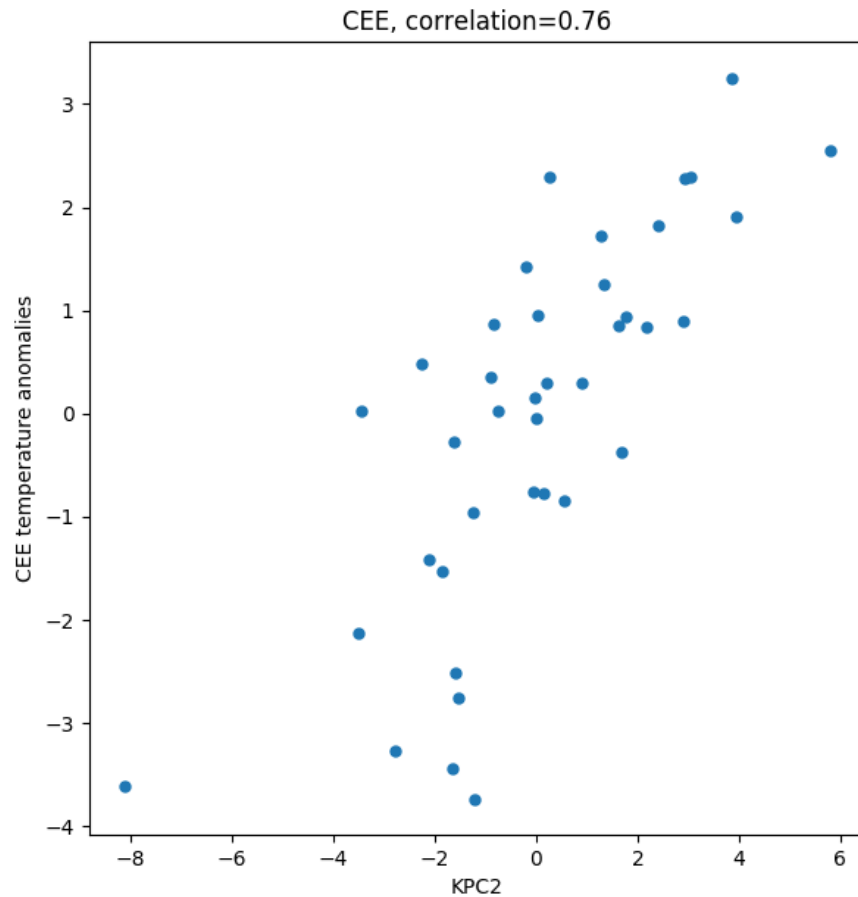
Surface air temperature composites



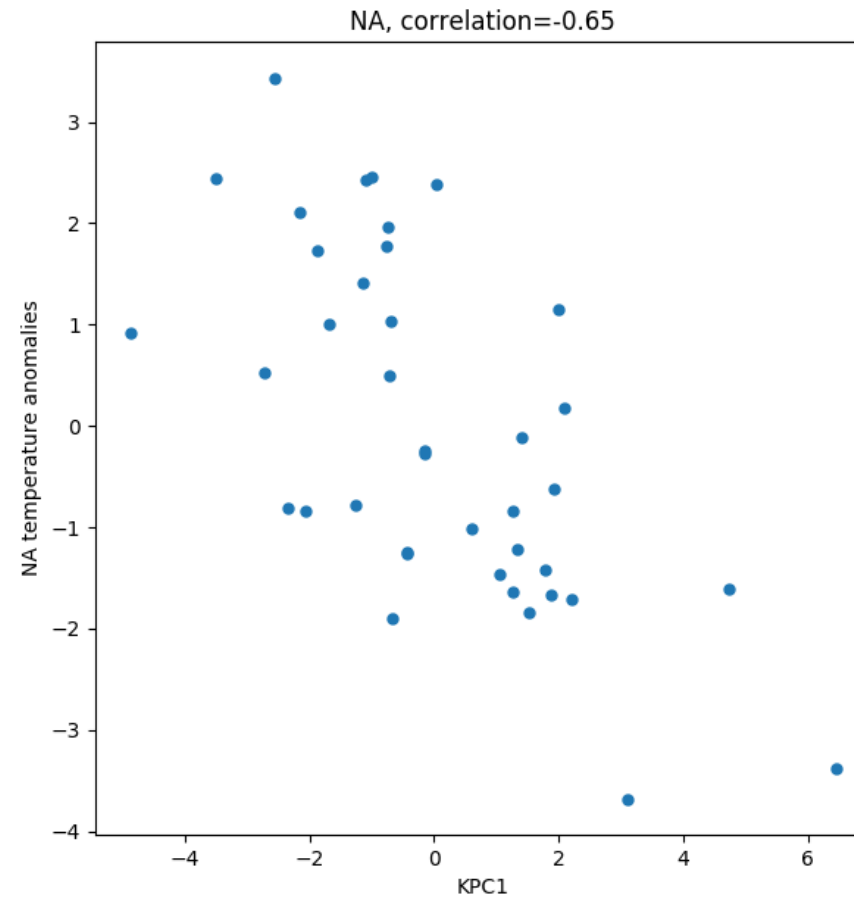
Reanalysis daily data: HGT-500, 1981-2019, winters

Relation of the modes to weather anomalies

Central & Eastern Europe temperature vs. KPC2
(winter means)



North of America temperature vs. KPC1
(winter means)



Reanalysis daily data: HGT-500, 1981-2019, winters

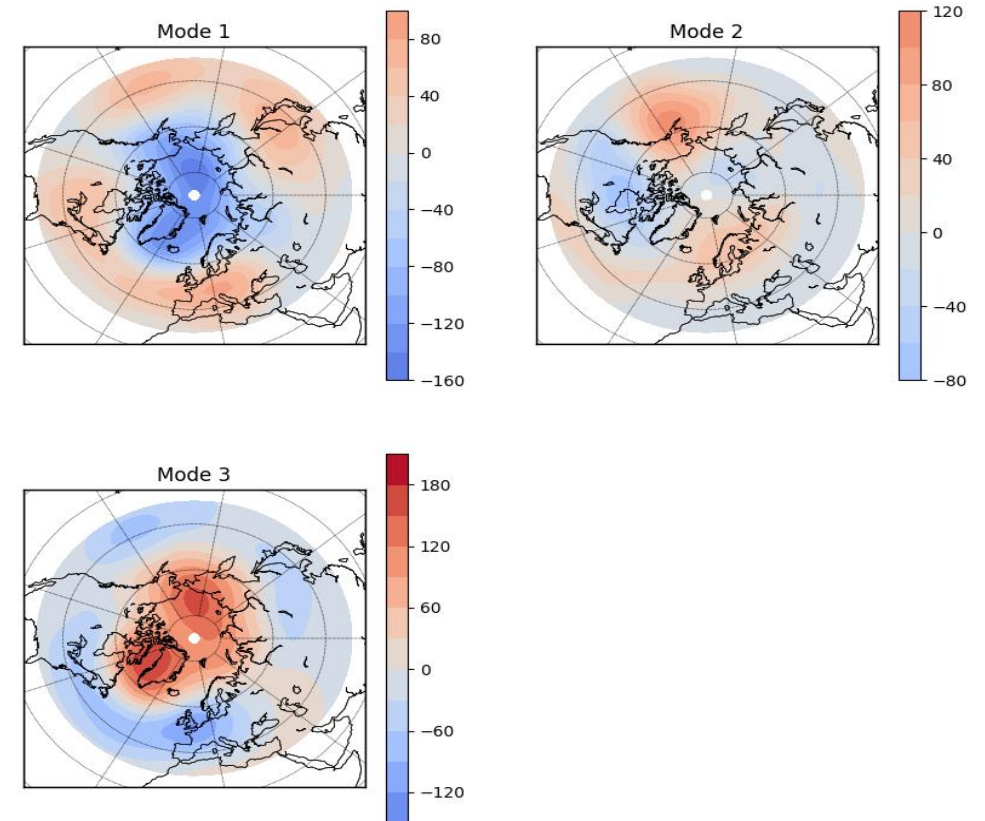
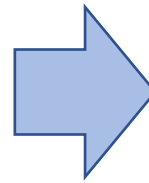
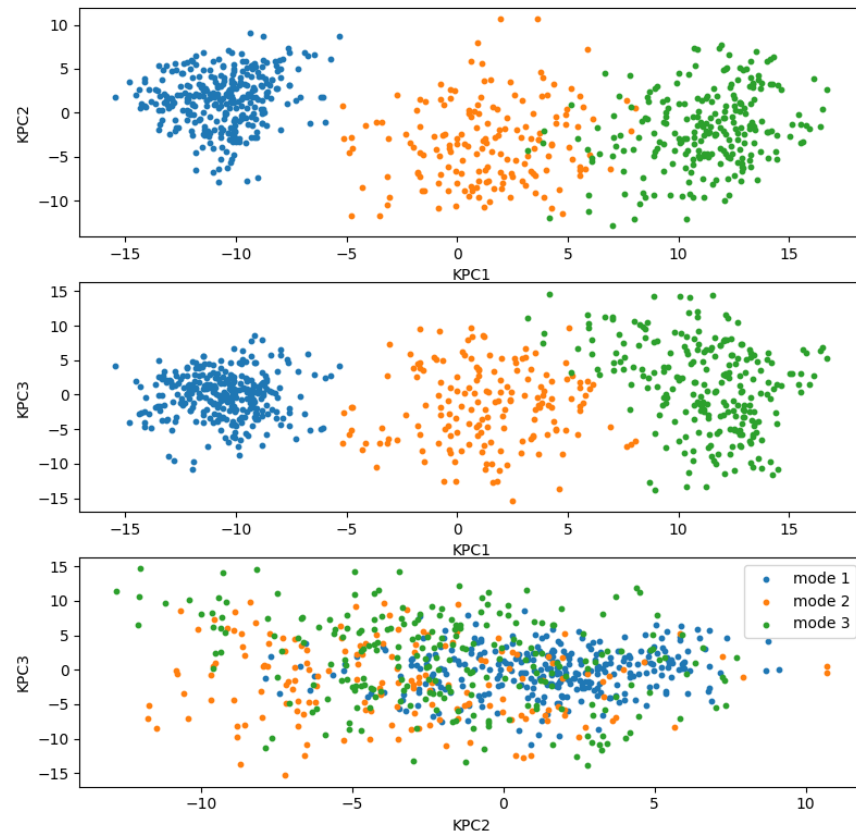
Redefining distance measure to capture zonal shifts of similar wave structures:

$$d_{xy} = \min_S |X - S \cdot Y|, \text{ } S \text{ is a zonal shifts}$$

The modes are separated along the 1st leading KPC

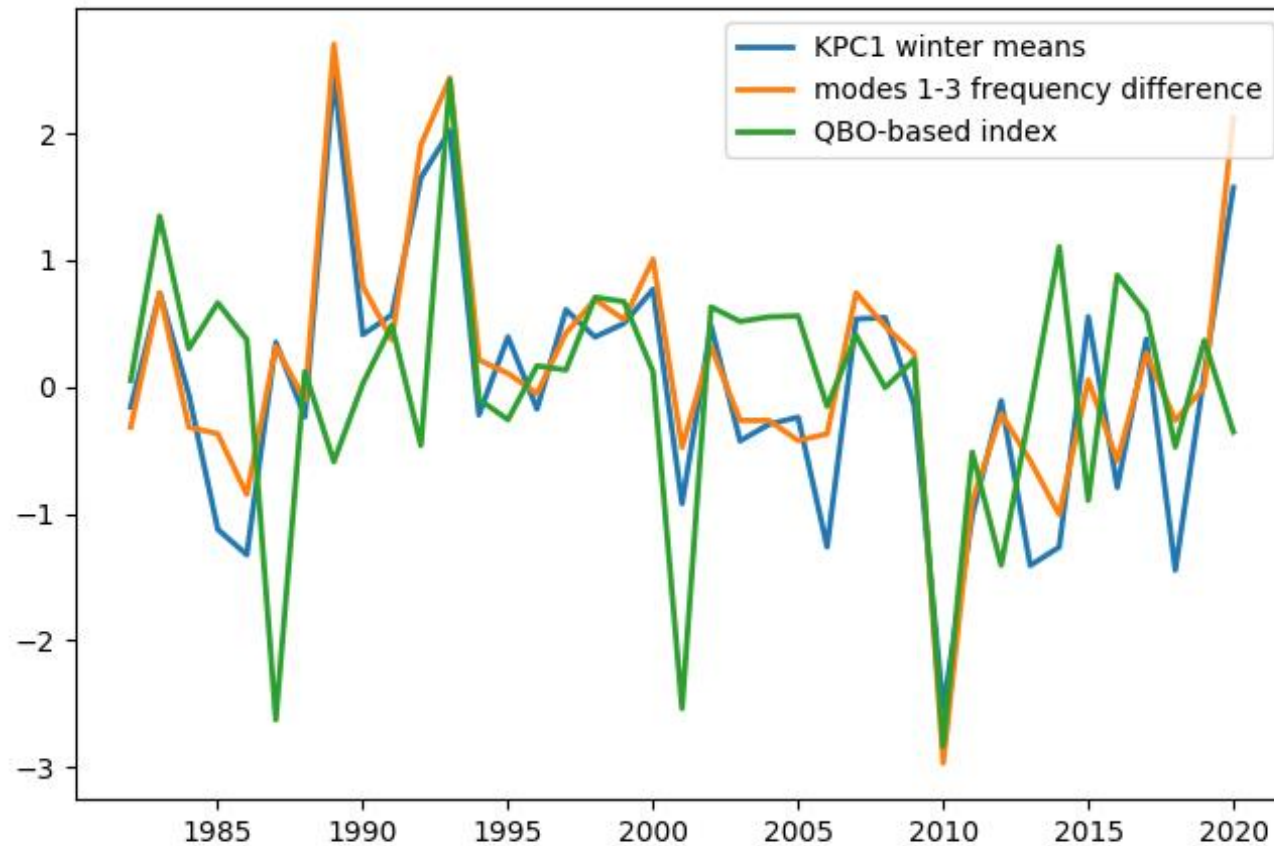
Composite patterns

(means over 20% most central points for each mode)



Reanalysis daily data: HGT-500, 1981-2019, winters

Relation of the Quasi-Biennial Oscillation (QBO) of the tropical stratosphere



QBO-based index is obtained by averaging of QBO signal over the previous winter and current fall

KPC1 averaged over each winter

Difference between frequencies of the mode 1 (strengthened zonal flow) **and the mode 3** (weakened zonal flow)

Conclusions

- A new approach, based on nonlinear PCA and recurrence networks, to identifying and analyzing the dominant dynamical regimes underlying the observed mid-latitude atmosphere circulation is suggested
- It allows both to detect significant regimes of LFV, and to obtain a dynamical variables serving as an embedding for the regimes detected
- The extracted LFV regimes of the atmosphere circulation correspond to large-scale weather patterns over the Northern Hemisphere mid-latitudes
- Inter-annual dynamics of the obtained regimes is related to the two-year state of the Quasi-Biennial oscillation of zonal winds in the tropical stratosphere