



Department of
Meteorology, MISU

Archetypal Analysis: Towards Analysing multivariate weather/climate extremes

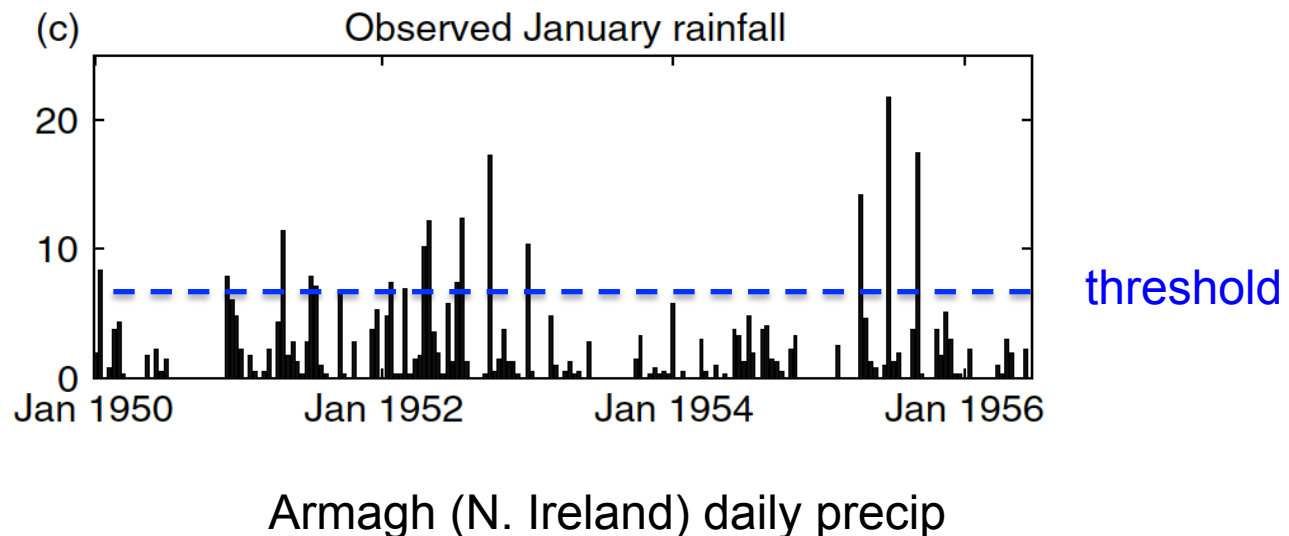
Abdel Hannachi & Nickolay trendafilov

1. Background
2. Challenges and strategies
3. Archetypal analysis
4. Numerical solution
5. Application to SST
6. Summary and conclusion

1. Background

Weather/climate $\equiv$ high-dimensional and nonlinear system:

- Seek dimension reduction
- Investigate extremes and simultaneous extremes



Extremes

- 1D $\rightarrow$ EVT through GEV

$$G(x) = \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right]$$

$\rightarrow$ threshold models $Y_i = X_i - u | X_i > u$

$$H(Y_i < y) = \Pr(X_i - u < y | X_i > u) = 1 - \left(1 + \xi \frac{y}{\sigma_u} \right)^{-1/\xi}$$

- 2D $\rightarrow$ Possible use of copula
- Higher dimension $\rightarrow$ the above become intractable

2. Challenges & strategies

Challenges

High-dimensionality

→ Possible dimension reduction: EOFs, RP, nonlinear mapping, etc.

A number of drawbacks, e.g. mixing, scaling

→ limitation, BUT EOFs are flexible (eg SVD)

→ Features: can be obtained via clustering (e.g. k-means) $\equiv$ like the data

Strategy

Methods that

- 1 Address extremes in high dimensions
- 2 Unlike EOFs, find modes (or patterns) looking like the (physical) variables through a proper combination of those original variables.



Archetypal analysis - AA

3. Archetypal analysis

nxm data matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$ (n obs., m variables)

EOFs $\rightarrow$ linear combination of variables

AA $\rightarrow$ typical patterns:

+ convex combination of obs.

$$\mathbf{z}_k = \sum_{t=1}^n b_{tk} \mathbf{x}_t, \quad b_{tk} \geq 0 \quad \text{and} \quad \sum_{t=1}^n b_{tk} = 1.$$

++ $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_p]$: approximate the data:

$$\text{Min Res} = \sum_{t=1}^n \left\| \mathbf{x}_t - \sum_{k=1}^p a_{kt} \mathbf{z}_k \right\|^2, \quad a_{jt} \geq 0, \quad \sum_{k=1}^p a_{kt} = 1,$$

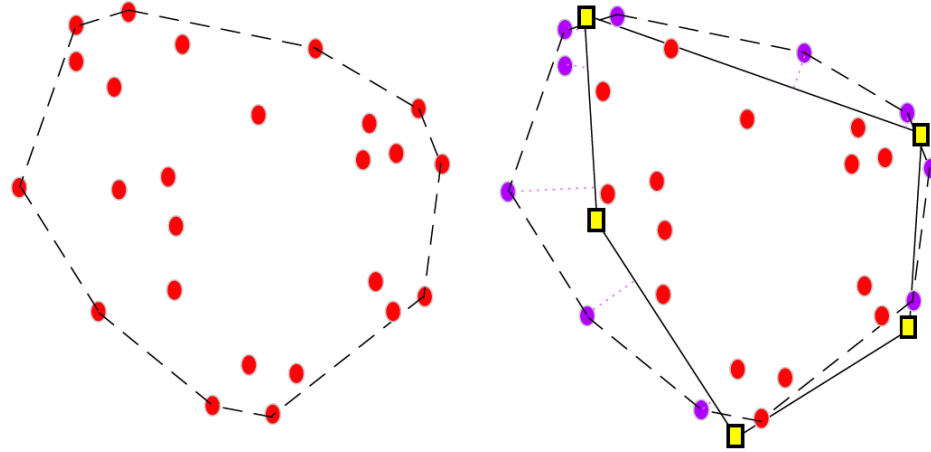
AA problem

Seek $\mathbf{A} = (a_{ij})$, and $\mathbf{B} = (b_{lk})$, $\{\mathbf{A}, \mathbf{B}\} = \underset{\mathbf{A}, \mathbf{B}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{A}^T \mathbf{B}^T \mathbf{X}\|_F^2$.

Subject to $\mathbf{A}, \mathbf{B} \geq 0$, $\mathbf{A}^T \mathbf{1}_p = \mathbf{1}_n$, and $\mathbf{B}^T \mathbf{1}_n = \mathbf{1}_p$,

➡ $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_p] = \mathbf{X}^T \mathbf{B}$. archetypes

Interpretation



archetypes:

- (1) *approximate the data convex hull (envelope)*
- (2) *represent the corners or extremes of the data in state space.*

4. AA solution and algorithm

4.1 Manifold-based algorithm

+ conventional: alternating algorithm,

Find $\mathbf{A}$ for fixed $\mathbf{Z}$, then find $\mathbf{Z}$ for fixed $\mathbf{A}$, etc.

+ present algorithm: non-alternating and manifold-based,

$$\begin{aligned}\text{Minimize } \mathcal{R} &= \|\mathbf{X} - (\mathbf{A} \circ \mathbf{A})^T (\mathbf{B} \circ \mathbf{B})^T \mathbf{X}\|_F^2 \\ &= \text{tr}(\mathcal{Z}) - 2\text{tr}(\mathcal{Z}\mathbf{W}) + \text{tr}(\mathcal{Z}\mathbf{W}^T\mathbf{W}),\end{aligned}$$

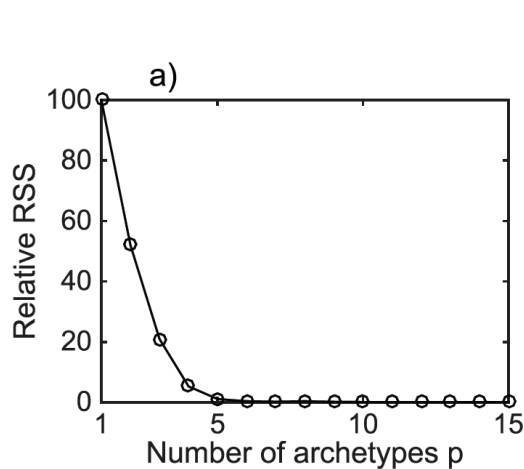
where $\mathcal{Z} = \mathbf{X}\mathbf{X}^T$ and $\mathbf{W} = (\mathbf{A} \circ \mathbf{A})^T (\mathbf{B} \circ \mathbf{B})^T$, and

$\mathbf{A} \circ \mathbf{C}$: the Hadamard matrix product, i.e. $\mathbf{A} \circ \mathbf{C} = (a_{ij}c_{ij})$

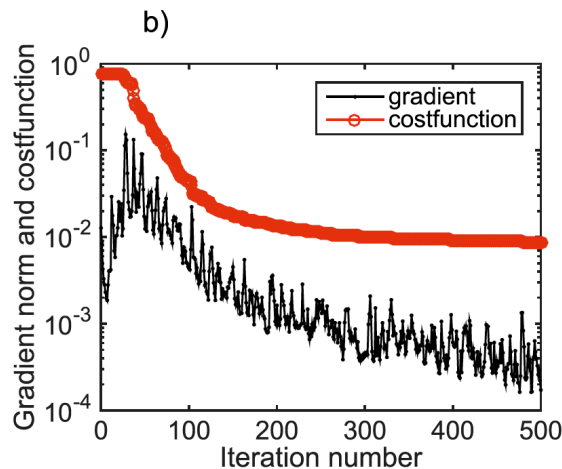
Gradient of $\mathcal{R}$: computed using manifold theory + conjugate gradient, for a given number p of archetypes (H&T 2017)

4.2 Illustration

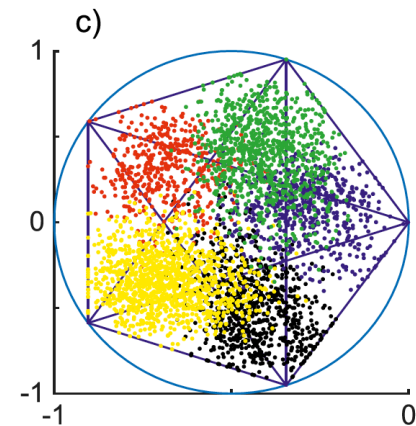
3D Gaussian clusters centered at 5 vertices of a polytope



Scree plot (optimal value: $p=5$)



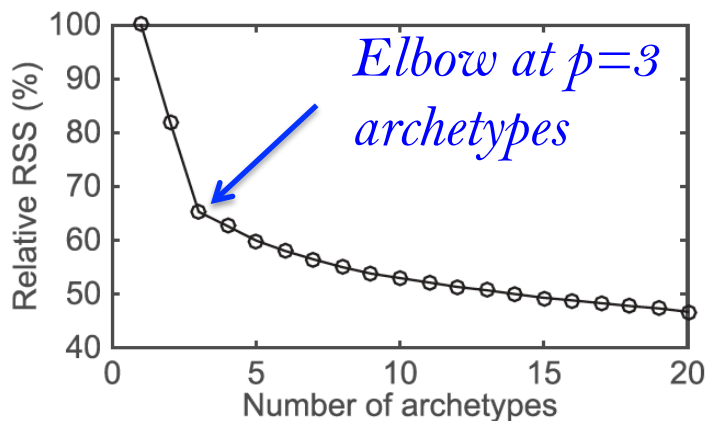
Costfunction and its gradient



Skew projection of the elements of $\mathbf{A} \circ \mathbf{A}$.

5. Application to SST anomalies

Data: HadISST, 1deg x 1deg, 1870-2014, 45.5S-45.5N



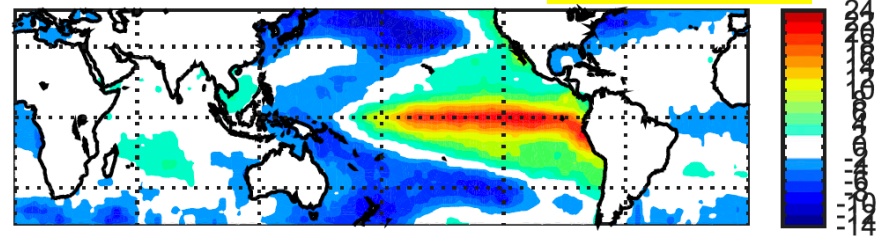
Scree plot for SSTA

La-Nina

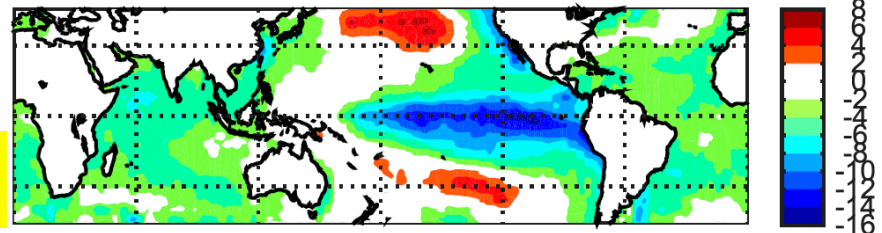
WBC: Western
boundary current

WBC

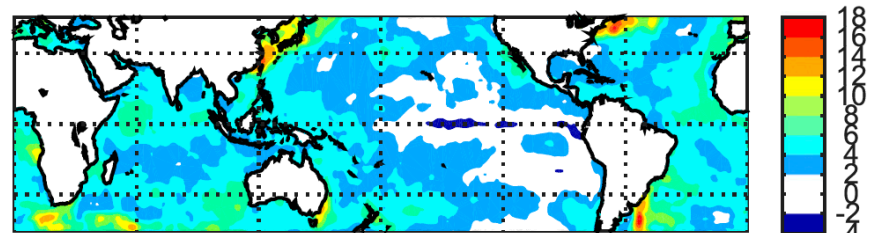
a) Archetype 1 **El-Nino**



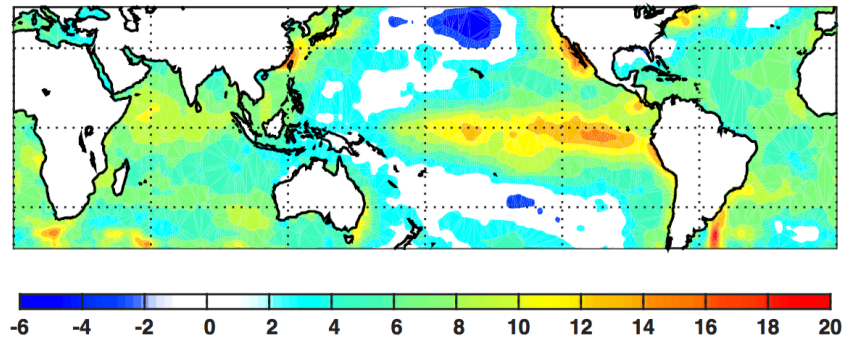
b) Archetype 2 **0.1degC**



c) Archetype 3

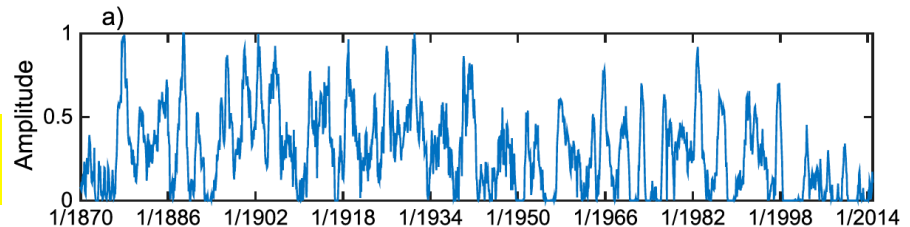


SSTA EOF1: mixing
Of ENSO and WBC

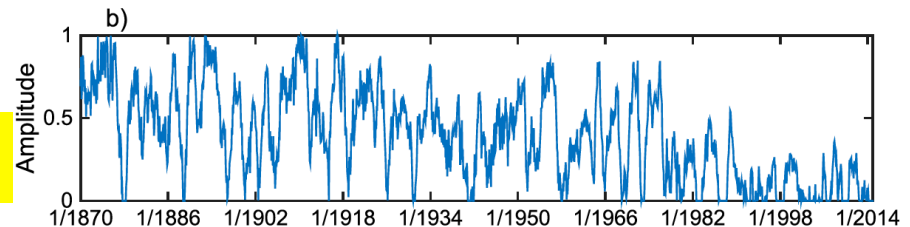


Time series of arhetypes amplitudes

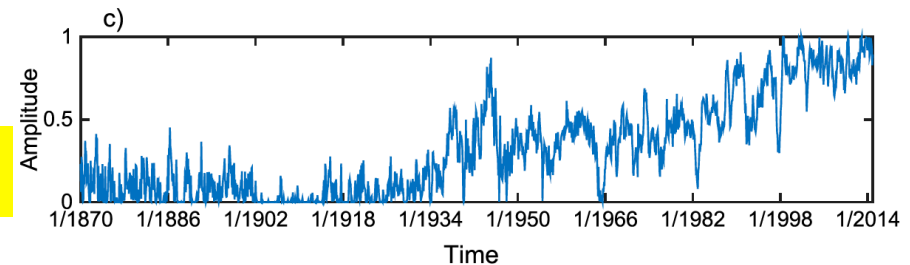
El-Nino



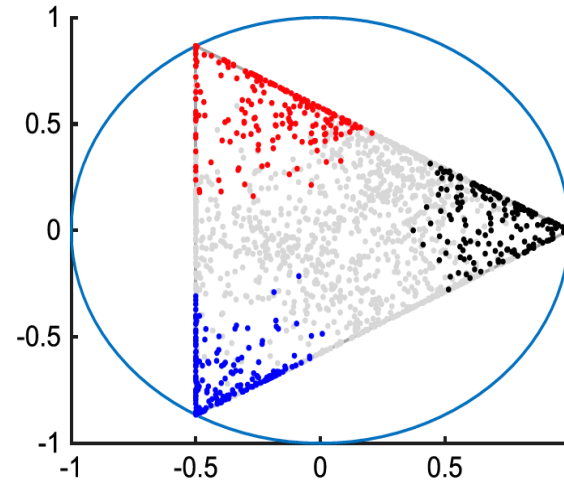
La-Nina



WBC



*2D simplex projection of
SSTA with 3 archetypes –
Colors: 200 points closest
to the archetypes.*



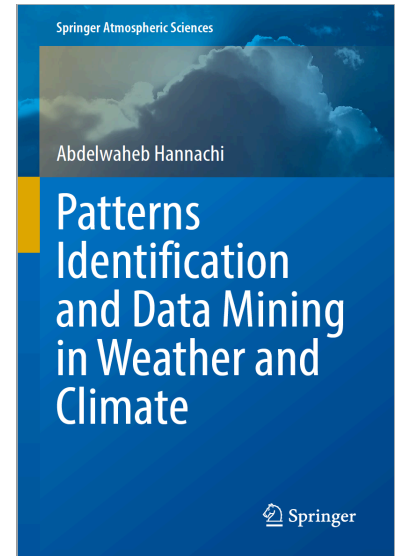
AA → Decrease of Lanina in recent decades
→ increase of WBC amplitudes: clear trend starting
1930's

AA of detrended SSTA → 2 representative archetypes:
El-Nino and La-Nina

Trend of extremes in SSTA: mainly explained by WBC not
ENSO

6. Summary and conclusion

- AA: very useful tool to analyse multivariate extremes
- Efficient manifold-based algorithm
- Application to SST anomalies:
 - 3 Representative archetypes – El Niño, La Niña, and western boundary currents
- Trend in extremes of SSTA seems to be explained mainly by WBC



In press, May 2021

