



Department of
Meteorology, MISU

Common EOFs: A tool for multi-model comparison/evaluation

Abdel Hannachi, Kathrin Finke & Nickolay trendafilov

1. Background
2. Common EOFs
3. Application to Re-analysis
4. Application CMIP
5. Conclusion

1. Background

- Many reanalysis products are being produced
- Many climate model simulations are also produce – e.g. CMIP, etc.



Need for robust/efficient tool for data comparison and/or evaluation

Conventional tools: EOFs of, e.g. model mean as platform for comparison: (1) arbitrary, and (2) does not embed models variability

Here: we propose common EOF method

2. Common EOFs

Given m populations with $p \times p$ sample covariances

$S_1, \dots, S_m$, we seek an orthonormal basis $B = [b_1, \dots, b_p]$

s.t. $S_k \approx B\Lambda_k^2 B^T$, $\Lambda_k^2 = \text{diag}(\lambda_{k1}^2, \dots, \lambda_{kp}^2)$, $k = 1, \dots, m$,

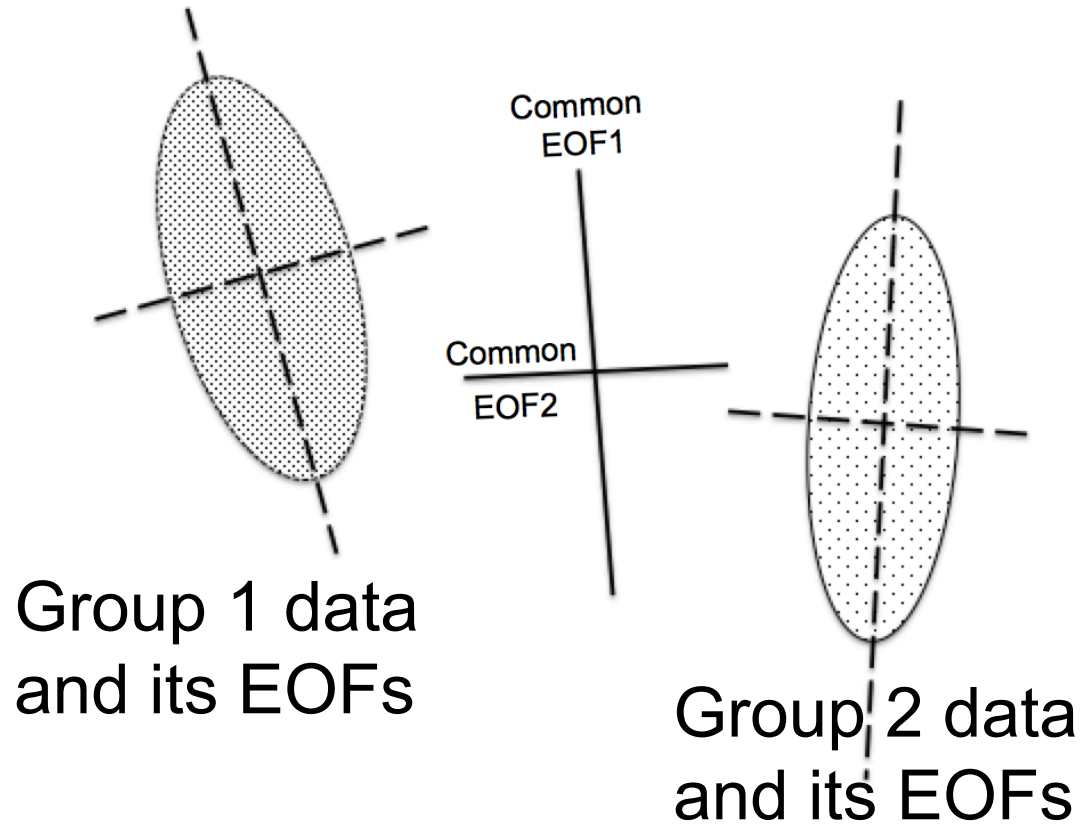


Simultaneous diagonalisation problem



Common PCA (Flury 1984)

Interpretation of common EOFs



2.1 Flury & Gautschi (1986) -FG86- algorithm

Maximum likelihood –MLE--

$$L(\Sigma_1, \dots, \Sigma_m) \propto \prod_{k=1}^m |\Sigma_k|^{-n_k/2} \exp \left[\text{tr} \left(-\frac{n_k}{2} \Sigma_k^{-1} \mathbf{S}_k \right) \right],$$

With optimality conditions:

$$\mathbf{b}_i \left(\sum_{k=1}^m n_k \frac{\lambda_{ki}^2 - \lambda_{kj}^2}{\lambda_{ki}^2 \lambda_{kj}^2} \right) \mathbf{b}_j = 0, \quad 1 \leq i \leq j \leq p, \quad \text{and} \quad \mathbf{B}^T \mathbf{B} = \mathbf{I}_p$$

The optimization problem is based on maximizing the diagonality of: $(\mathbf{B}^T \mathbf{S}_k \mathbf{B})$, $k = 1, \dots, m$,

FG86 algorithm: based on matrix optimization, and is a two-step algorithm.

Two main weaknesses of FG86

- Two-step algorithm: It is extremely slow and not efficient for reasonably high dimensions (e.g. weather & climate)
- Main focus on diagonalisation – but not on dimensionality reduction – an important objective in weather & climate data analysis:

FG86 algorithm does not guarantee simultaneous decrease of eigenvalues in all groups

2.2 Step-wise algorithm

Transform the original (matrix) optimization problem:

$$\underset{\mathbf{B}}{\operatorname{argmin}} \sum_{k=1}^m n_k \log |diag(\mathbf{B}^T \mathbf{S}_k \mathbf{B})|, \quad \text{s.t. } \mathbf{B}^T \mathbf{B} = \mathbf{I}_p.$$

into a vectorized problem:

$$\begin{aligned} & \max_{\mathbf{b}} \sum_{k=1}^m n_k \log |diag(\mathbf{b}^T \mathbf{S}_k \mathbf{b})| \\ & \text{st } \mathbf{b}^T \mathbf{b} = 1, \text{ and } \mathbf{b}^T \mathbf{B}_{j-1} = \mathbf{0}^T \end{aligned}$$

where $\mathbf{B}_{j-1} = [\mathbf{b}_1, \dots, \mathbf{b}_{j-1}]$, $\mathbf{B}_0 = \mathbf{0}$,

algorithm is based on projecting the gradient of the objective function onto the orthogonal of the space spanned by common EOFs of the previous step:

$$\left(\sum_{k=1}^m \frac{n_k}{n} \frac{\mathbf{S}_k}{\mathbf{b}_1^T \mathbf{S}_k \mathbf{b}_1} - \mathbf{I}_p \right) \mathbf{b}_1 = \mathbf{0},$$

and for $j = 2, \dots, p$:

$$\left[(\mathbf{I}_p - \mathbf{B}_{j-1} \mathbf{B}_{j-1}^T) \left(\sum_{k=1}^m \frac{n_k}{n} \frac{\mathbf{S}_k}{\mathbf{b}_j^T \mathbf{S}_k \mathbf{b}_j} \right) - \mathbf{I}_p \right] \mathbf{b}_j = \mathbf{0},$$

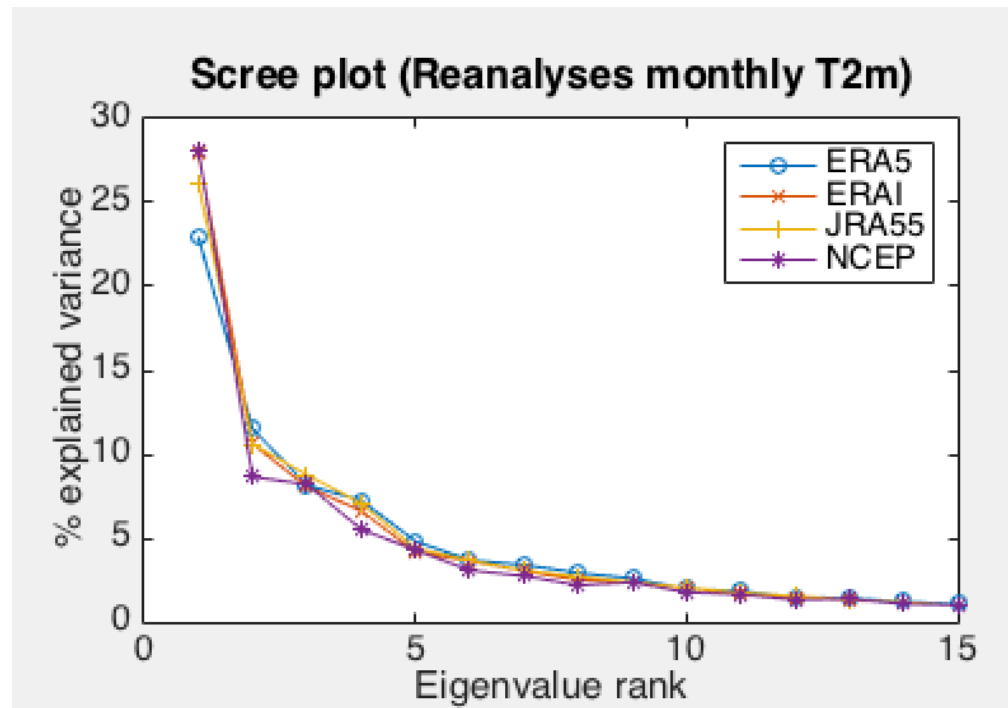
These equations are solved by standard power method: starting with the largest eigenvalue and then going in decreasing order.

3. Application to reanalyses Background

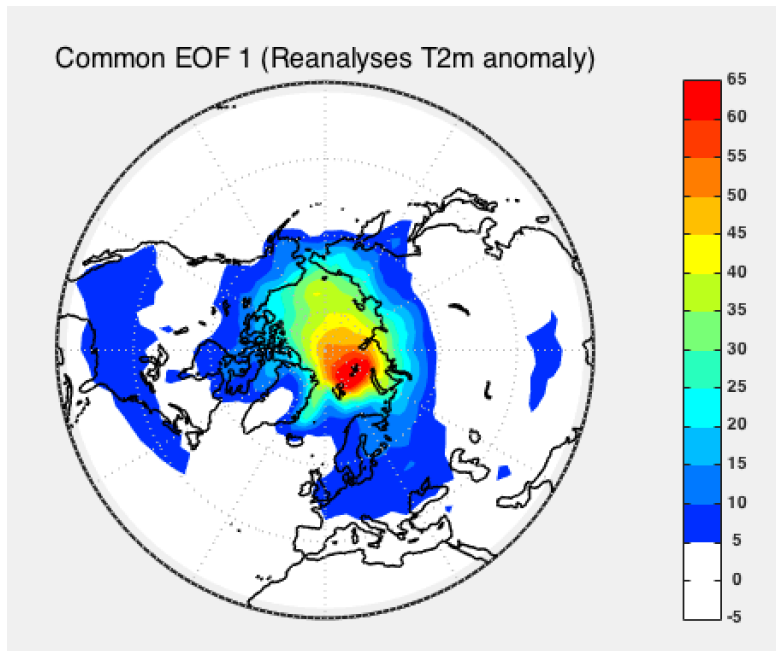
- Reanalysis products: ERA5, ERAI, JRA55 & NCEP
- Data: 1979-2008, 3.75x3.7 grid over 20°N-90°N
- Fields: T2m, SLP, Z500, and Z10, monthly and DJF daily time series

Monthly T2m

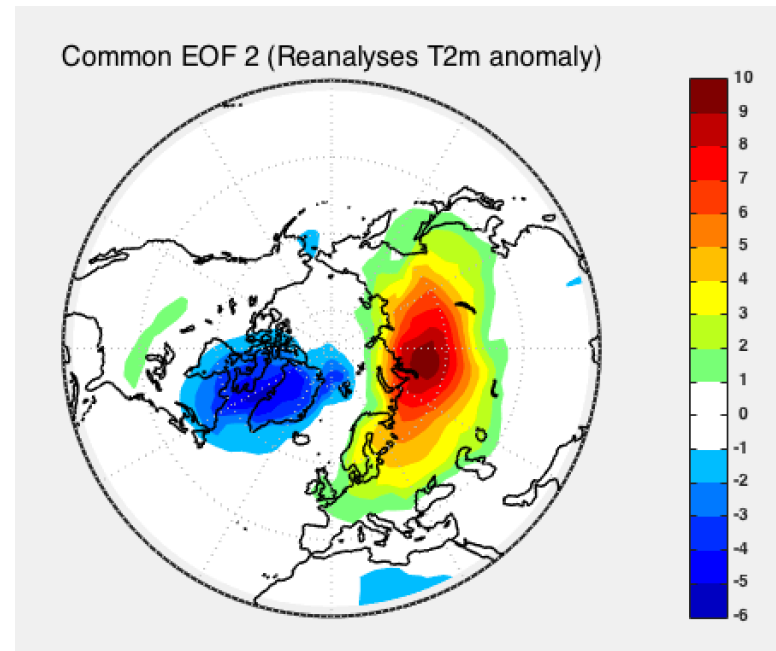
- Variation within
a range of 5% of
explained variance



➤ Pattern



Common EOF1 –
explained most by
NCEP and least by
ERA5



Common EOF2 –
Explained most by
ERA% and least by
NCEP

Monthly mass field

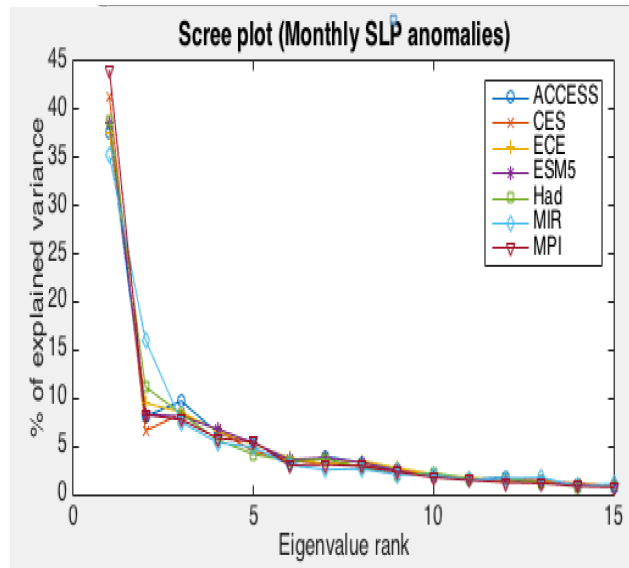
- Leading common mode: Arctic Oscillation AO
- SLP: Excellent coherence between the 4 products

Daily DJF fields

- Similar variation range from T2m
- Leading common EOF1: dipole over Siberia and North pole, and common EOF2: 3 centres over Siberia, Canada and the Aleutian.
- Mass field and Z500: coherence between the 4 products

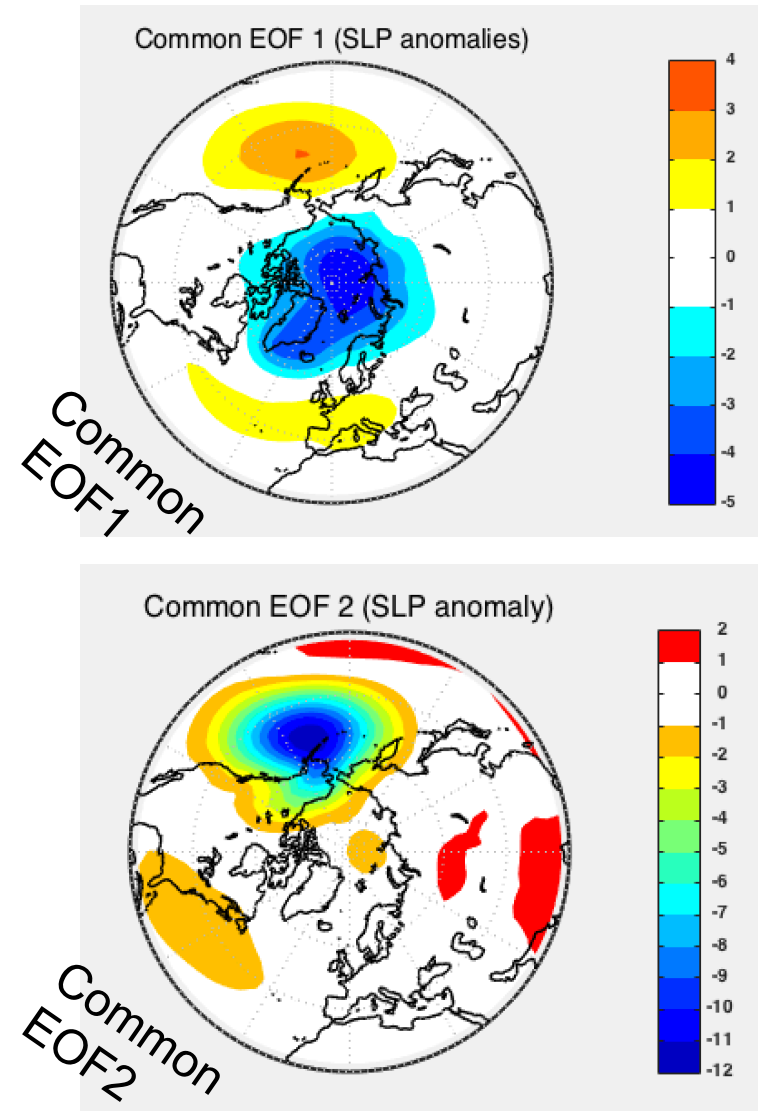
4. Application to CMIP

4.1 Monthly timescale



Scree plot

- ✓ Variation within a range of 10% of explained variance (common EOF1)



- ✓ Largest % explained variance: MPI (44%) and CES (42%), and smallest: MIR (35%)
- ✓ Largest skewness of common CEOF1: Had and MPI
→ -AO
- ✓ Peaked distributions (positive excess kurtosis)

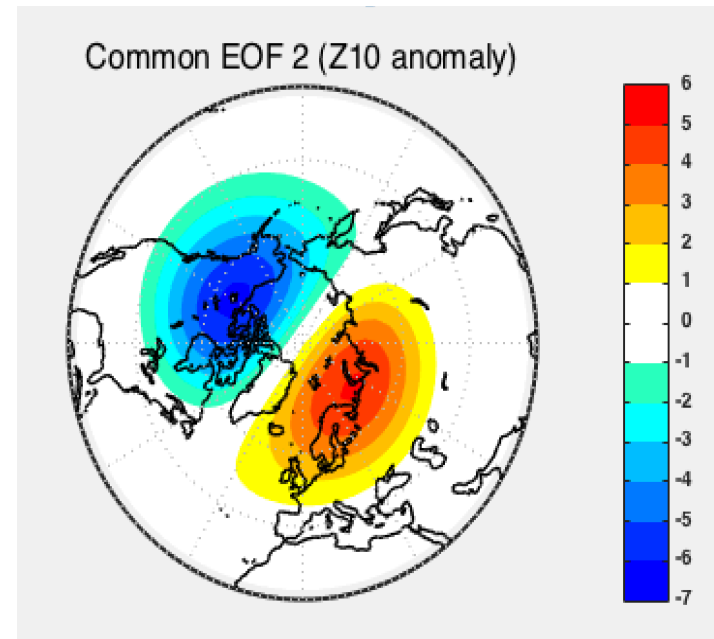
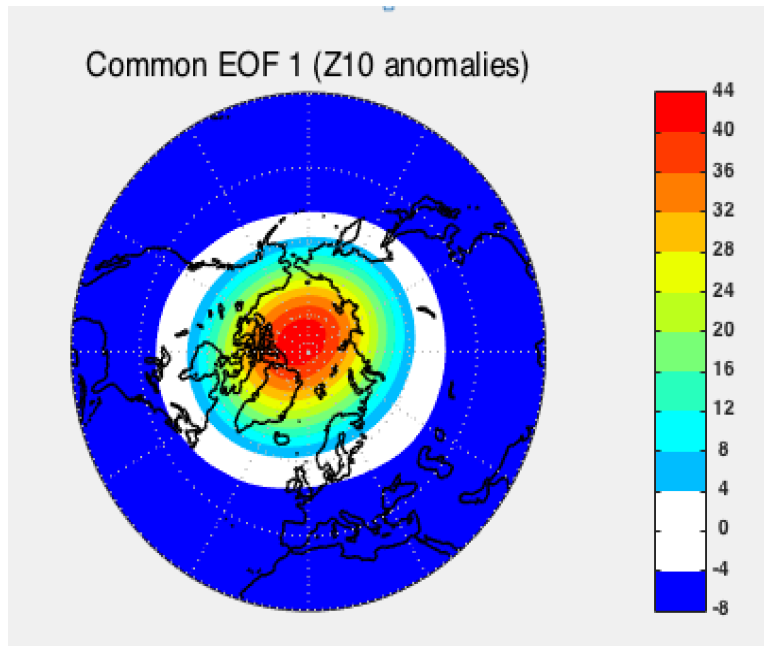
Stratosphere: monthly Z10

- ✓ Largest difference in % of explained variance: 20%: from 52% (GFDL) to 71% (ESM) for common EOF1.
- ✓ All models show large skewness (-AO) except MPI (+AO)

- ✓ Largest skewness (.93/-.61 for common EOF1/EOF2)



- ✓ weaker vortex with preferred wave breaking over Canada

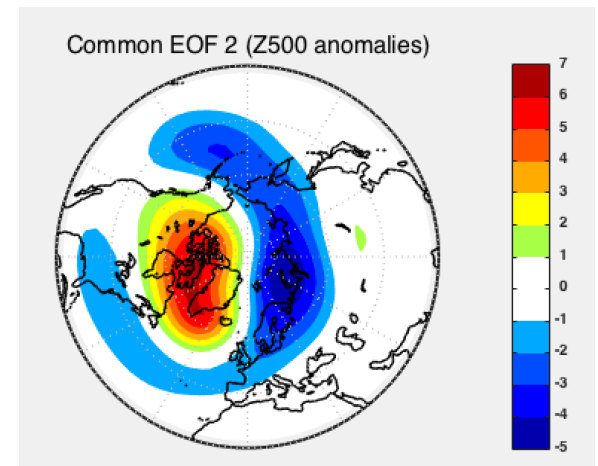
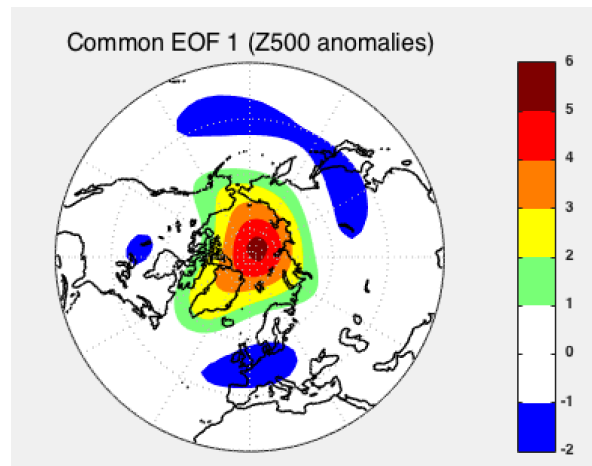
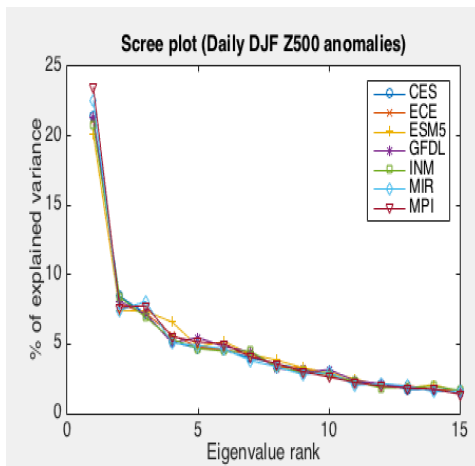


- ✓ Positive excess kurtosis  peaked distribution

4.2 DJF daily timescale

Troposphere (SLP and Z500)

- Variability: within a range of 5% explained variance
- SLP leading modes: AO/Siberia-NA dipole (modes 1,2)
- Mostly skewed, and –ve (MPI) and +ve (GFDL,INM) excess kurtosis
- Z500 leading modes: AO/PNA:



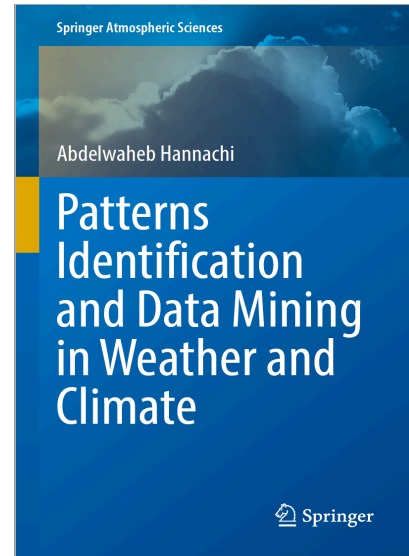
- Mostly skewed ($\rightarrow$ -AO, +PNA)
- -ve (MIR, MPI) and +ve (CES, INM) excess kurtosis

Stratosphere (Z10)

- Variability: within a range of 24% explained variance: from 35% (GFDL) to 59%(INM)
- Strongly skewed
- Mostly strong –ve kurtosis except MIR, GFDL

6. Summary and conclusion

- Common EOFs: appropriate tool for models' comparison/evaluation
- Stepwise algorithm: efficient and desirable for dimensionality reduction
- Application to reanalyses: difference in T2m, consistency in upper air & mass field
- CMIP fields: difference in troposphere: 5%-10% range in variability
- In the stratosphere: difference within range of 24% of percentage of explained variance



In press, May 2021

