Quantifying Differences in Weather Types on the basis of PDF dissimilarity measures

Henning Rust\textsuperscript{1}, Mathieu Vrac\textsuperscript{1}, Matthieu Lengaigne\textsuperscript{2}, Benjamin Sultan\textsuperscript{2}

\textsuperscript{1} Laboratoire des Sciences du Climat et de l’Environnement (LSCE/IPSL), Gif-sur-Yvette, France
\textsuperscript{2} Laboratoire d’Océanographie et du Climat: Experimentation et Approches Numériques (LOCEAN/IPSL), Paris, France

Weather Types
Weather Types, Circulation Patterns and Regimes

Definition, an attempt ...

Classes of large scale atmospheric circulation (characterised by, e.g., pressure fields) with

- recurrent states (circulation patterns), regions of increased probability density in state space
- persistent states (circulation regimes), or
- states related to local weather (weather types).

Approaches

- subjectively: e.g., Lamb WTs
- objectively: using clustering algorithms k-means, hierarchical clustering, Gaussian mixture models
Weather Types, Circulation Patterns and Regimes

**Definition, an attempt ...**

Classes of large scale atmospheric circulation (characterised by, e.g., pressure fields) with

- recurrent states (circulation patterns), regions of increased probability density in state space
- persistent states (circulation regimes), or
- states related to local weather (weather types).

**Approaches**

- subjectively: e.g., Lamb WTs
- objectively: using clustering algorithms k-means, hierarchical clustering, Gaussian mixture models
Weather Types, Circulation Patterns and Regimes

Definition, an attempt ...

Classes of large scale atmospheric circulation (characterised by, e.g., pressure fields) with

- recurrent states (circulation patterns), regions of increased probability density in state space
- persistent states (circulation regimes), or
- states related to local weather (weather types).

Approaches

- subjectively: e.g., Lamb WTs
- objectively: using clustering algorithms, k-means, hierarchical clustering, Gaussian mixture models
Weather Types, Circulation Patterns and Regimes

Definition, an attempt ...

Classes of large scale atmospheric circulation (characterised by, e.g., pressure fields) with

- recurrent states (circulation patterns), regions of increased probability density in state space
- persistent states (circulation regimes), or
- states related to local weather (weather types).

Approaches

- subjectively: e.g., Lamb WTs
- objectively: using clustering algorithms k-means, hierarchical clustering, Gaussian mixture models
Why?

- downscaling/WT related studies:
  - do GCMs reproduce (certain) WTs?
  - do WTs change in time?
- general understanding:
  - what WTs are not reproduced
  - do WTs change for different forcings?

How?

- Compare mean states
  - visually
  - Euclidean distance
  - pattern correlation

Problems

- mean states sufficient representatives?
- spread(variance)/spatial extension??

Compare WTs including shape and size
Comparing Weather Types

Why?

- downscaling/WT related studies:
  - do GCMs reproduce (certain) WTs?
  - do WTs change in time?
- general understanding:
  - what WTs are not reproduced
  - do WTs change for different forcings?

How?

- Compare mean states
  - visually
  - Euclidean distance
  - pattern correlation

Problems

- mean states sufficient representatives?
- spread(variance)/spatial extension??

Compare WTs including shape and size

5/21
Comparing Weather Types

**Why?**

- downscaling/WT related studies:
  - do GCMs reproduce (certain) WTs?
  - do WTs change in time?
- general understanding:
  - what WTs are not reproduced
  - do WTs change for different forcings?

**How?**

- Compare mean states
  - visually
  - Euclidean distance
  - pattern correlation

**Problems**

- mean states sufficient representatives?
- spread(variance)/spatial extension??

**Summary**

Comparing WTs including shape and size
Comparing Weather Types

Why?
- downscaling/WT related studies:
  - do GCMs reproduce (certain) WTs?
  - do WTs change in time?
- general understanding:
  - what WTs are not reproduced
  - do WTs change for different forcings?

How?
- Compare mean states
  - visually
  - Euclidean distance
  - pattern correlation

Problems
- mean states sufficient representatives?
- spread(variance)/spatial extension??

Compare WTs including shape and size
Comparing Weather Types

Why?

- downscaling/WT related studies:
  - do GCMs reproduce (certain) WTs?
  - do WTs change in time?
- general understanding:
  - what WTs are not reproduced
  - do WTs change for different forcings?

How?

Compare mean states
- visually
- Euclidean distance
- pattern correlation

Problems
- mean states sufficient representatives?
- spread(variance)/spatial extension??

Compare WTs including shape and size
Comparing Weather Types

same centre, different orientation
Comparing Weather Types

same centre, different orientation
Comparing Weather Types

same centre, different extension
Comparing Weather Types

same centre, different extension
Gaussian Mixture Models
Gaussian Mixture Models

Modelling

- “truth”
- is unknown
- model selection
- estimation
- pdf, classification
- uncertainty

\[ p(x \mid \theta) = \sum_{k=1}^{3} a_k f(x; \mu_k, \Sigma_k), \]
Gaussian Mixture Models

Modelling

- “truth”
- is unknown
- model selection
- estimation
- pdf, classification
- uncertainty
Gaussian Mixture Models

Modelling

- “truth”
- is unknown
- model selection
  - estimation
  - pdf, classification
  - uncertainty

\[ \text{BIC} = -2 \log L(\theta; x) + m \log N, \]

\( \theta \): parameter vector, \( m \): number of parameters
Gaussian Mixture Models

Estimated parameters (EM):

\[ \hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3, \hat{a}_1, \hat{a}_2, \hat{a}_3) \]
Gaussian Mixture Models

Modelling

- “truth”
- is unknown
- model selection
- estimation
- pdf, classification
- uncertainty

\[
p(x | \hat{\theta}) = \sum_{k=1}^{3} \hat{a}_k f(x; \hat{\mu}_k, \hat{\Sigma}_k), \quad Cl_{x_i} \in \{1, 2, 3\}
\]
Similarity Measures

[Diagram showing different similarity measures]
Similarity Measures for (Gaussian) pdfs

1. **Euclidian distance:**
   \[ d_{\text{Eucl}}^2(P, Q) = \| \mu_p - \mu_q \|^2 = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q) \]

2. **Mahalanobis distance:**
   \[ d_{\text{Mah}}(P, Q) = \| \mu_p - \mu_q \|^2 = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q) \]

3. **Kullback-Leibler discrimination (KL):**
   \[ d_{\text{KL}}(P, Q) = I(P \mid Q) = \int_{\mathbb{R}} \log \left( \frac{q(x)}{p(x)} \right) q(z) dx \]

4. **J-coefficient:**
   \[ d_J(P, Q) := \frac{I(P \mid Q) + I(Q \mid P)}{2} \]

5. **Hellinger coefficient \((s=1/2)\):**
   \[ d_H(P, Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)} dx, \quad d_H \in [0, 1] \]
1. Euclidian distance: 
\[ d_{\text{Eucl}}^2(P, Q) = \|\mu_p - \mu_q\|^2 = (\mu_p - \mu_q)^T \mathbf{1} (\mu_p - \mu_q) \]

2. Mahalanobis distance: 
\[ d_{\text{Mah}}^2(P, Q) = \|\mu_p - \mu_q\|^2_{\Sigma_p^{-1}} = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q) \]

3. Kullback-Leibler discrimination (KL): 
\[ d_{\text{KL}}(P, Q) = I(P \mid Q) = \int_{\mathbb{R}} \log \left( \frac{q(x)}{p(x)} \right) q(x) dx \]

4. J-coefficient: 
\[ d_J(P, Q) := \frac{I(P \mid Q) + I(Q \mid P)}{2} \]

5. Hellinger coefficient (s=1/2): 
\[ d_H(P, Q) = \int_{\mathbb{R}} q(x)^{s} p(x)^{(1-s)} dx, \quad d_H \in [0, 1] \]
Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
   \[ d_{\text{Eucl}}^2(P, Q) = \| \mu_p - \mu_q \|^2 = (\mu_p - \mu_q)^T \mathbb{1}(\mu_p - \mu_q) \]

2. Mahalanobis distance:
   \[ d_{\text{Mah}}(P, Q) = \| \mu_p - \mu_q \|^2_{\Sigma_p^{-1}} = (\mu_p - \mu_q)^T \Sigma_p^{-1}(\mu_p - \mu_q) \]

3. Kullback-Leibler discrimination (KL):
   \[ d_{\text{KL}}(P, Q) = I(P \mid Q) = \int_\mathbb{R} \log \left( \frac{q(x)}{p(x)} \right) q(z) dx \]

4. J-coefficient:
   \[ d_J(P, Q) := \left( I(P \mid Q) + I(Q \mid P) \right) / 2 \]

5. Hellinger coefficient (s=1/2):
   \[ d_H(P, Q) = \int_\mathbb{R} q(x)^s p(x)^{(1-s)} dx, \quad d_H \in [0, 1] \]
Similarity Measures for (Gaussian) pdfs

1. **Euclidian distance:**
   \[ d_{\text{Eucl}}^2(P, Q) = \| \mu_p - \mu_q \|^2 = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q) \]

2. **Mahalanobis distance:**
   \[ d_{\text{Mah}}(P, Q) = \| \mu_p - \mu_q \|^2 \Sigma_p^{-1} = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q) \]

3. **Kullback-Leibler discrimination (KL):**
   \[ d_{\text{KL}}(P, Q) = I(P \mid Q) = \int_{\mathbb{R}} \log \left( \frac{q(x)}{p(x)} \right) q(z) dx \]

4. **J-coefficient:**
   \[ d_J(P, Q) := \frac{I(P \mid Q) + I(Q \mid P)}{2} \]

5. **Hellinger coefficient \((s=1/2):\)**
   \[ d_H(P, Q) = \int_{\mathbb{R}} q(x)^s p(x)^{1-s} dx, \quad d_H \in [0, 1] \]
Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
   \[ d_{\text{Eucl}}^2(P, Q) = \| \mu_p - \mu_q \|^2 = (\mu_p - \mu_q)^T \Sigma^{-1} (\mu_p - \mu_q) \]

2. Mahalanobis distance:
   \[ d_{\text{Mah}}(P, Q) = \| \mu_p - \mu_q \|^2_{\Sigma^{-1}_p} = (\mu_p - \mu_q)^T \Sigma^{-1}_p (\mu_p - \mu_q) \]

3. Kullback-Leibler discrimination (KL):
   \[ d_{\text{KL}}(P, Q) = I(P \mid Q) = \int_{\mathbb{R}} \log \left( \frac{q(x)}{p(x)} \right) q(x) dx \]

4. J-coefficient:
   \[ d_J(P, Q) := \frac{1}{2} (I(P \mid Q) + I(Q \mid P)) \]

5. Hellinger coefficient (s=1/2):
   \[ d_H(P, Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)} dx, \quad d_H \in [0, 1] \]
**Similarity Measures for (Gaussian) pdfs**

1. **Euclidian distance:**
   
   $d_{\text{Eucl}}^2(P, Q) = \| \mu_p - \mu_q \|^2 = (\mu_p - \mu_q)^T \mathbf{1}(\mu_p - \mu_q)$

2. **Mahalanobis distance:**

   $d_{\text{Mah}}(P, Q) = \| \mu_p - \mu_q \|^2_{\Sigma_p^{-1}} = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q)$

3. **Kullback-Leibler discrimination (KL):**

   $d_{\text{KL}}(P, Q) = I(P \mid Q) = \int_{\mathbb{R}} \log \left( \frac{q(x)}{p(x)} \right) q(z) \, dx$

4. **J-coefficient:**

   $d_J(P, Q) := \left( I(P \mid Q) + I(Q \mid P) \right) / 2$

5. **Hellinger coefficient (s=1/2):**

   $d_H(P, Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)} \, dx, \quad d_H \in [0, 1]$
Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
   \[ d_{\text{Eucl}}^2(P, Q) = \| \mu_p - \mu_q \|^2 = (\mu_p - \mu_q)^T \mathbf{1} (\mu_p - \mu_q) \]

2. Mahalanobis distance:
   \[ d_{\text{Mah}}(P, Q) = \| \mu_p - \mu_q \|^2_{\Sigma_p^{-1}} = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q) \]

3. Kullback-Leibler discrimination (KL):
   \[ d_{\text{KL}}(P, Q) = I(P \mid Q) = \int_{\mathbb{R}} \log \left( \frac{q(x)}{p(x)} \right) q(z) dx \]

4. J-coefficient:
   \[ d_J(P, Q) := (I(P \mid Q) + I(Q \mid P))/2 \]

5. Hellinger coefficient (s=1/2):
   \[ d_H(P, Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)} dx, \quad d_H \in [0, 1] \]
Similarity Measures, Example

- Euclidean
- Mahalanobis
- Kullback-Leibler
- J-coef
- Hellinger
Similarity Measures, Example

- Euclidian
- Mahalanobis
- Kullback-Leibler
- J-coef
- Hellinger
Similarity Measures, Example

WTs
Mixtures
Definition
Example
Case Study
Summary

WT differences
Henning Rust

Euclidean
Mahalanobis
Kullback-Leibler
J-coef
Hellinger

0
5
10
15
20
25
0 0.25 0.5 0.75 1
Similarity Measures, Example

Euclidean
Mahalanobis
Kullback-Leibler
J-coef
Hellinger

Case Study
Similarity Measures, Example

WT differences
Henning Rust
WTs
Mixtures
Similarity
Definition
Example
Case Study
Summary

Similarity Measures, Example

- Euclidian
- Mahalanobis
- Kullback-Leibler
- J-coef
- Hellinger

Graph showing similarity measures with data points and comparison.
Similarity Measures, Example

- Euclidian
- Mahalanobis
- Kullback-Leibler
- J-coef
- Hellinger

WT differences
Henning Rust
WTs
Mixtures
Similarity
Definition
Example
Case Study
Summary

11/21
Similarity Measures, Example

- Euclidian
- Mahalanobis
- Kullback-Leibler
- J coef
- Hellinger

The diagram illustrates the similarity measures with respective values on the x-axis and y-axis.
Case Study

Comparing North-Atlantic WTs
Data Preparation

Description

- North-Atlantic region
- daily SLP anomalies
- 1975 – 2000, NDJFM
- datasets, interpolated to NCEP/NCAR grid
- common PCA 95%
## Data Preparation

### Description

- North-Atlantic region
- Daily SLP anomalies
- 1975 – 2000, NDJFM
- Datasets, interpolated to NCEP/NCAR grid
- Common PCA 95%

### Reanalyses

- NCEP/NCAR
- ERA-40

### 14 IPCC Models

- CCCMA CGCM3.1, T47
- CNRM CM3.0
- CSIRO MK3.0 and MK3.5
- GFDL CM2.0 and CM2.1
- INGV ECHAM4
- INM CM3.0
- IPSL CM4
- MIROC 3.2 high/medium resolution
- MIUB ECHO.G
- MPI ECHAM5
- MRI CGCM 2.3.2a
Data Preparation

Description

- North-Atlantic region
- daily SLP anomalies
- 1975 – 2000, NDJFM
- datasets, interpolated to NCEP/NCAR grid
- common PCA 95%

Summary

North-Atlantic region

daily SLP anomalies

1975 – 2000, NDJFM
datasets, interpolated to NCEP/NCAR grid

common PCA 95%
Defining Weather Types with Gaussian Mixtures

Reproduce $k$-Means Result (Plaut & Simmonet, 2001)

1. Force 5 spherical clusters ($\Sigma_k = \sigma_k I$)
2. Define WTs on NCEP/NCAR
3. Associate NCEP/NCAR means to reference
4. Define WTs on ERA-40/GCMs
5. Associate GCM WTs to NCEP/NCAR WTs
Defining Weather Types with Gaussian Mixtures

Reproduce $k$-Means Result (Plaut&Simmonet, 2001)

1. force 5 spherical clusters ($\Sigma_k = \sigma_k \mathbb{1}$)
2. define WT$\bar{s}$ on NCEP/NCAR
3. associate NCEP/NCAR means to reference
4. define WT$\bar{s}$ on ERA-40/GCMs
5. associate GCM WT$\bar{s}$ to NCEP/NCAR WT$\bar{s}$
Defining Weather Types with Gaussian Mixtures

Reproduce \( k \)-Means Result (Plaut & Simmonet, 2001)

1. force 5 spherical clusters \( \Sigma_k = \sigma_k I \)
2. define WTs on NCEP/NCAR
3. associate NCEP/NCAR means to reference
4. define WTs on ERA-40/GCMs
5. associate GCM WTs to NCEP/NCAR WTs
Defining Weather Types with Gaussian Mixtures

Reproduce $k$-Means Result (Plaut & Simonnet, 2001)

1. Force 5 spherical clusters ($\Sigma_k = \sigma_k I$)
2. Define WTs on NCEP/NCAR
3. Associate NCEP/NCAR means to reference
4. Define WTs on ERA-40/GCMs
5. Associate GCM WTs to NCEP/NCAR WTs

from Plaut & Simonnet, (2001)
Defining Weather Types with Gaussian Mixtures

from Plaut\&Simonnet, (2001)
Defining Weather Types with Gaussian Mixtures

Reproduce $k$-Means Result (Plaut & Simmonet, 2001)

1. force 5 spherical clusters ($\Sigma_k = \sigma_k I$)
2. define WTs on NCEP/NCAR
3. associate NCEP/NCAR means to reference
4. define WTs on ERA-40/GCMs
5. associate GCM WTs to NCEP/NCAR WTs
Weather Type Mean Values

WT differences
Henning Rust

WTs
Mixtures
Similarity
Case Study
Data Preparation
Defining Weather Types
Visual
Diff. to NCEP/NCAR
Summary
Weather Type Mean Values

WT differences
Henning Rust
WTs
Mixtures
Similarity
Case Study
Data Preparation
Defining Weather Types
Visual
Diff. to NCEP/NCAR
Summary
Weather Type Mean Values

WT differences

Henning Rust

WTs

Mixtures

Similarity

Case Study

Data Preparation

Defining Weather Types

Visual

Diff. to NCEP/NCAR

Summary
Difference to NCEP/NCAR WT vs
Quantifying Differences to NCEP/NCAR

WT differences
Henning Rust
WTs
Mixtures
Similarity
Case Study
Data Preparation
Defining Weather Types
Visual
DIFF To NCEP/NCAR
Summary

Euclidian/Mahalanobis/KL/J-coef.

Hellinger
AR
Euclidian
Mahalanobis
KullbackLeibler
Jcoef
Hellinger

0 5 10 ... 1/3 2/3 1

 Hellinger

0 1/3 2/3 1

 ncep era40
 miub_echo_g
csiro_mk3_0
csiro3_2_medres
ingv_echam4
gfdl_cm2_0
mpi_echam5
inmcm3_0
mri_cgcm2_3_2a
mri_cgcm2_3_2a
ipsl_cm4
ipsl_cm4
ccma_cgcm3_1
ccma_cgcm3_1
Quantifying Differences to NCEP/NCAR

WT differences
Henning Rust
WTs
Mixtures
Similarity
Case Study
Data Preparation
Defining Weather Types
Visual Differences to NCEP/NCAR
Summary
### Hellinger Coefficient \((1 - d_H)\)

<table>
<thead>
<tr>
<th>Weather Type</th>
<th>AR</th>
<th>BL</th>
<th>GA</th>
<th>WBL</th>
<th>ZO</th>
<th>Mean(Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reanalysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCEP</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000(0.000)</td>
</tr>
<tr>
<td>ERA-40</td>
<td>0.004</td>
<td>0.009</td>
<td>0.011</td>
<td>0.022</td>
<td>0.004</td>
<td>0.010(0.007)</td>
</tr>
<tr>
<td><strong>GCM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSIRO MK3.5</td>
<td>0.175</td>
<td>0.182</td>
<td>0.433</td>
<td>0.418</td>
<td>0.294</td>
<td>0.301(0.124)</td>
</tr>
<tr>
<td>MIUB ECHO G</td>
<td>0.105</td>
<td>0.430</td>
<td>0.287</td>
<td>0.564</td>
<td>0.174</td>
<td>0.312(0.187)</td>
</tr>
<tr>
<td>MRI CGCM2.3</td>
<td>0.350</td>
<td>0.455</td>
<td>0.183</td>
<td>0.482</td>
<td>0.279</td>
<td>0.350(0.124)</td>
</tr>
<tr>
<td>MPI ECHAM5</td>
<td>0.279</td>
<td>0.105</td>
<td>0.543</td>
<td>0.492</td>
<td>0.635</td>
<td>0.411(0.215)</td>
</tr>
</tbody>
</table>

4 GCMs remain the top 4 across all measures.
**Summary**

**Quantifying WT Differences**

- describe pdf in state space with Gaussian mixtures (5 spherical clusters for comparison)
- complement Euclidean distance by
  - Mahalanobis distance
  - Kullback-Leibler discrimination
  - J-coefficient
  - Hellinger coefficient
- compare 14 GCMs to NCEP/NCAR by WT in NA region
- best GCMs on average reproducing NA-WTs: CSIRO MK3.5, MIUB ECHO G, MPI ECHAM5, MRI CGCM2.3
- quality varies with WTs
Quantifying WT Differences

- describe pdf in state space with Gaussian mixtures (5 spherical clusters for comparison)
- complement Euclidean distance by
  - Mahalanobis distance
  - Kullback-Leibler discrimination
  - J-coefficient
  - Hellinger coefficient
- compare 14 GCMs to NCEP/NCAR by WT in NA region
- best GCMs on average reproducing NA-WTs: CSIRO MK3.5, MIUB ECHO G, MPI ECHAM5, MRI CGCM2.3
- quality varies with WT
Summary

Quantifying WT Differences

- describe pdf in state space with Gaussian mixtures (5 spherical clusters for comparison)
- complement Euclidean distance by
  - Mahalanobis distance
  - Kullback-Leibler discrimination
  - $J$-coefficient
  - Hellinger coefficient
- compare 14 GCMs to NCEP/NCAR by WT in NA region
  - best GCMs on average reproducing NA-WTs: CSIRO MK3.5, MIUB ECHO G, MPI ECHAM5, MRI CGCM2.3
  - quality varies with WTs
Summary

Quantifying WT Differences

- describe pdf in state space with Gaussian mixtures (5 spherical clusters for comparison)
- complement Euclidean distance by
  - Mahalanobis distance
  - Kullback-Leibler discrimination
  - J-coefficient
  - Hellinger coefficient
- compare 14 GCMs to NCEP/NCAR by WT in NA region
- best GCMs on average reproducing NA-WTs: CSIRO MK3.5, MIUB ECHO G, MPI ECHAM5, MRI CGCM2.3
- quality varies with WTs
Quantifying WT Differences

- describe pdf in state space with Gaussian mixtures (5 spherical clusters for comparison)
- complement Euclidean distance by
  - Mahalanobis distance
  - Kullback-Leibler discrimination
  - J-coefficient
  - Hellinger coefficient
- compare 14 GCMs to NCEP/NCAR by WT in NA region
- best GCMs on average reproducing NA-WTs: CSIRO MK3.5, MIUB ECHO G, MPI ECHAM5, MRI CGCM2.3
- quality varies with WTs
Downscaling of Precipitation (not NA)

- quantify temporal change in WTs
- select GCMs according to relevant WTs

In General

- quantify separation of WTs within models
- understand why certain WTs are not reproduced
- do WTs change for changing GCM forcings
- . . .
Applications/Outlook

Downscaling of Precipitation (not NA)

- quantify temporal change in WTs
- select GCMs according to relevant WTs

In General

- quantify separation of WTs within models
- understand why certain WTs are not reproduced
- do WTs change for changing GCM forcings
- ...
I do not want to advocate 5 spherical clusters for the NA region!

- Gaussian mixtures + BIC
Appendix

6. Common PCA

7. WT Association

8. Projection onto NCEP

9. Summary Table

10. Differences to NCEP

11. Ranking using other measures

12. BIC
Common PCA

PCA

- 250 grid points, highly correlated
- PCA | all models (250x161330)
- 20 PCs ≈ 95% total variance
- models use PCs differently
- > 95% (indv.) included
Common PCA

## PCA

- 250 grid points, highly correlated
- PCA | all models (250x161330)
- 20 PCs ≈ 95% total variance
- models use PCs differently
- > 95% (indv.) included

<table>
<thead>
<tr>
<th>model</th>
<th>date</th>
<th>grid1</th>
<th>grid2</th>
<th>...</th>
<th>grid250</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA-40</td>
<td>01/01/1975</td>
<td>0.376</td>
<td>0.435</td>
<td>...</td>
<td>1.344</td>
</tr>
<tr>
<td>ERA-40</td>
<td>01/02/1975</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
<tr>
<td>ERA-40</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>ERA-40</td>
<td>12/31/2000</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
<tr>
<td>NCEP</td>
<td>01/01/1975</td>
<td>0.376</td>
<td>0.435</td>
<td>...</td>
<td>1.344</td>
</tr>
<tr>
<td>NCEP</td>
<td>01/02/1975</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
<tr>
<td>NCEP</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>NCEP</td>
<td>12/31/2000</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
<tr>
<td>CCCMA</td>
<td>01/01/1975</td>
<td>0.376</td>
<td>0.435</td>
<td>...</td>
<td>1.344</td>
</tr>
<tr>
<td>CCCMA</td>
<td>01/02/1975</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
<tr>
<td>CCCMA</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>CCCMA</td>
<td>12/31/2000</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
<tr>
<td>MRI</td>
<td>01/01/1975</td>
<td>0.376</td>
<td>0.435</td>
<td>...</td>
<td>1.344</td>
</tr>
<tr>
<td>MRI</td>
<td>01/02/1975</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
<tr>
<td>MRI</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>MRI</td>
<td>12/31/2000</td>
<td>0.276</td>
<td>0.335</td>
<td>...</td>
<td>1.244</td>
</tr>
</tbody>
</table>
Common PCA

**PCA**

- 250 grid points, highly correlated
- PCA | all models (250x161330)
- 20 PCs ≈ 95% total variance
- models use PCs differently
- > 95% (indv.) included

![Graph showing percentage of cumulative variance vs PC for different models.](image)
Association of WT

What GCM WT is Associated to NCEP WT?

<table>
<thead>
<tr>
<th>model</th>
<th>NCEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Π₁</td>
<td>1</td>
</tr>
<tr>
<td>Π₂</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Π₅</td>
<td>5</td>
</tr>
</tbody>
</table>

Mapping:

Use Mapping Minimising Sum of Distances

\[ \Pi_0 = \arg \min_{\Pi \in \Pi \{1,2,3\}} \sum_{i=1}^{5} d(WT_{\text{NCEP}}^i, C_l^\Pi_{\text{GCM}}), \]

\(\Pi\): permutation out of all possible permutations of \{1, 2, 3\}. 

Comparison 24/21
Association of WTs

What GCM WT is Associated to NCEP WTs?

<table>
<thead>
<tr>
<th>model</th>
<th>NCEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Π₁</td>
<td>1</td>
</tr>
<tr>
<td>Π₂</td>
<td>2</td>
</tr>
<tr>
<td>⋮</td>
<td>⋮</td>
</tr>
<tr>
<td>Π₅</td>
<td>5</td>
</tr>
</tbody>
</table>

Mapping:

Use Mapping Minimising Sum of Distances

\[ \Pi_0 = \arg \min_{\Pi \in \Pi \{1,2,3\}} \sum_{i=1}^{5} d(WT_{NCEP}^i, C_{GCM}^i) \]

\( \Pi \): permutation out of all possible permutations of \( \{1, 2, 3\} \).
Centroids of Projection onto NCEP/NCAR WT

WT differences
Henning Rust
Appendix
Common PCA
WT Association
Projection onto NCEP
Summary Table
Differences to NCEP
Ranking using other measures
BIC

NCEP/NCAR  – ERA-40  – MPI ECHAM5 – CSIRO MK3.0
### Characteristics of Measures

#### Summary Table

<table>
<thead>
<tr>
<th>Measure</th>
<th>Symm.</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean</td>
<td>yes</td>
<td>distance in means</td>
</tr>
<tr>
<td>Mahalanobis</td>
<td>no</td>
<td>distance in means, metric depending on one covariance matrix</td>
</tr>
<tr>
<td>Kullback-Leibler</td>
<td>no</td>
<td>metric depends on both covariance matrices</td>
</tr>
<tr>
<td>$J$-coefficient</td>
<td>yes</td>
<td>symmetrised KL</td>
</tr>
<tr>
<td>Hellinger(s=1/2)</td>
<td>yes</td>
<td>measures “overlap”</td>
</tr>
</tbody>
</table>
Quantifying Differences to NCEP/NCAR

WT differences
Henning Rust
Appendix
Common PCA
WT Association
Projection onto NCEP
Summary Table
Differences to NCEP
Ranking using other measures
BIC
Quantifying Differences to NCEP/NCAR

WT differences
Henning Rust
Appendix
Common PCA
WT Association
Projection onto NCEP
Summary Table
Differences to NCEP
Ranking using other measures
BIC
Quantifying Differences to NCEP/NCAR

Differences to NCEP

Ranking using other measures

BIC
Quantifying Differences to NCEP/NCAR

- **WT differences**
- **Henning Rust**
- **Appendix**
- **Common PCA**
- **WT Association**
- **Projection onto NCEP**
- **Summary Table**
- **Differences to NCEP**
- **Ranking using other measures**
- **BIC**
Quantifying Differences to NCEP/NCAR

WT differences
Henning Rust

Appendix
Common PCA
WT Association
Projection onto NCEP
Summary Table
Differences to NCEP
Ranking using other measures
BIC
### Euclidean Distance

<table>
<thead>
<tr>
<th>Weather Type</th>
<th>AR</th>
<th>BL</th>
<th>GA</th>
<th>WBL</th>
<th>ZO</th>
<th>Mean(Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reanalysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCEP</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000(0.000)</td>
</tr>
<tr>
<td>ERA-40</td>
<td>0.848</td>
<td>1.188</td>
<td>1.493</td>
<td>1.922</td>
<td>0.954</td>
<td>1.281(0.436)</td>
</tr>
<tr>
<td><strong>GCM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSIRO MK3.5</td>
<td>6.071</td>
<td>5.756</td>
<td>11.099</td>
<td>9.136</td>
<td>8.486</td>
<td>8.110(2.226)</td>
</tr>
<tr>
<td>MRI CGCM2.3</td>
<td>7.575</td>
<td>10.199</td>
<td>6.231</td>
<td>10.148</td>
<td>7.662</td>
<td>8.363(1.748)</td>
</tr>
<tr>
<td>MPI ECHAM5</td>
<td>7.917</td>
<td>4.270</td>
<td>13.188</td>
<td>10.360</td>
<td>14.056</td>
<td>9.958(3.997)</td>
</tr>
</tbody>
</table>
### Mahalanobis Distance

<table>
<thead>
<tr>
<th></th>
<th>AR</th>
<th>BL</th>
<th>GA</th>
<th>WBL</th>
<th>ZO</th>
<th>Mean(Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reanalysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCEP</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000(0.000)</td>
</tr>
<tr>
<td>ERA-40</td>
<td>0.015</td>
<td>0.033</td>
<td>0.042</td>
<td>0.090</td>
<td>0.017</td>
<td>0.039(0.030)</td>
</tr>
<tr>
<td><strong>GCM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSIRO MK3.5</td>
<td>0.770</td>
<td>0.781</td>
<td>2.308</td>
<td>2.035</td>
<td>1.366</td>
<td>1.452(0.706)</td>
</tr>
<tr>
<td>MIUB ECHO G</td>
<td>0.318</td>
<td>2.209</td>
<td>1.332</td>
<td>3.289</td>
<td>0.763</td>
<td>1.582(1.187)</td>
</tr>
<tr>
<td>MRI CGCM2.3</td>
<td>1.198</td>
<td>2.453</td>
<td>0.727</td>
<td>2.511</td>
<td>1.113</td>
<td>1.600(0.824)</td>
</tr>
<tr>
<td>MPI ECHAM5</td>
<td>1.308</td>
<td>0.430</td>
<td>3.258</td>
<td>2.617</td>
<td>3.747</td>
<td>2.272(1.377)</td>
</tr>
<tr>
<td>Weather Type</td>
<td>AR</td>
<td>BL</td>
<td>GA</td>
<td>WBL</td>
<td>ZO</td>
<td>Mean (Std)</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td><strong>Reanalysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCEP</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td>ERA-40</td>
<td>0.015</td>
<td>0.037</td>
<td>0.043</td>
<td>0.091</td>
<td>0.018</td>
<td>0.041 (0.030)</td>
</tr>
<tr>
<td><strong>GCM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSIRO MK3.5</td>
<td>0.770</td>
<td>0.853</td>
<td>2.314</td>
<td>2.511</td>
<td>1.371</td>
<td>1.564 (0.811)</td>
</tr>
<tr>
<td>MIUB ECHO G</td>
<td>0.418</td>
<td>2.214</td>
<td>1.335</td>
<td>3.290</td>
<td>0.792</td>
<td>1.610 (1.157)</td>
</tr>
<tr>
<td>MRI CGCM2.3</td>
<td>1.534</td>
<td>2.456</td>
<td>0.771</td>
<td>2.539</td>
<td>1.219</td>
<td>1.704 (0.774)</td>
</tr>
<tr>
<td>MPI ECHAM5</td>
<td>1.396</td>
<td>0.459</td>
<td>3.347</td>
<td>3.184</td>
<td>3.813</td>
<td>2.440 (1.438)</td>
</tr>
</tbody>
</table>
### J-Coefficient

<table>
<thead>
<tr>
<th></th>
<th>Weather Type</th>
<th></th>
<th></th>
<th></th>
<th>Mean(Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR</td>
<td>BL</td>
<td>GA</td>
<td>WBL</td>
<td>ZO</td>
</tr>
<tr>
<td><strong>Reanalysis</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>NCEP</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>ERA-40</strong></td>
<td>0.030</td>
<td>0.073</td>
<td>0.086</td>
<td>0.181</td>
<td>0.036</td>
</tr>
<tr>
<td><strong>GCM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSIRO MK3.5</td>
<td>1.540</td>
<td>1.614</td>
<td>4.546</td>
<td>4.419</td>
<td>2.787</td>
</tr>
<tr>
<td>MIUB ECHO G</td>
<td>0.896</td>
<td>4.501</td>
<td>2.706</td>
<td>6.634</td>
<td>1.527</td>
</tr>
<tr>
<td>MRI CGCM2.3</td>
<td>3.506</td>
<td>4.850</td>
<td>1.617</td>
<td>5.277</td>
<td>2.628</td>
</tr>
<tr>
<td>MPI ECHAM5</td>
<td>2.626</td>
<td>0.886</td>
<td>6.288</td>
<td>5.545</td>
<td>8.096</td>
</tr>
</tbody>
</table>

**Note:** The table shows the differences in J-coefficient for various weather types (AR, BL, GA, WBL, ZO) between reanalysis data and models (NCEP, ERA-40, GCM) and the mean and standard deviation (Mean(Std)) for the models.