WT differences

Henning Rust

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

Henning Rust[1], Mathieu Vrac[1], Matthieu Lengaigne[2], Benjamin Sultan[2]

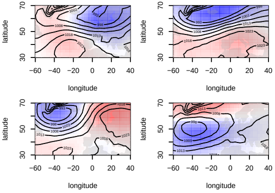
[1] Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL), Gif-sur-Yvette, France

[2] Laboratoire d'Océanographie et du Climat: Experimentation et Approches Numériques (LOCEAN/IPSL), Paris, France

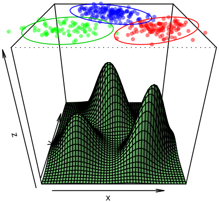
Dresden, 30. July, 2009

Outline

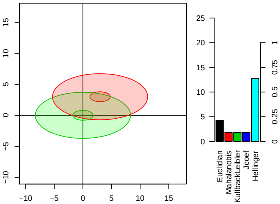
Weather Types



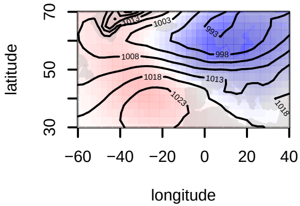
Mixture Models



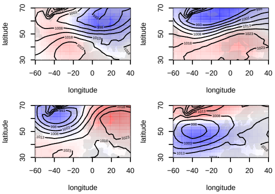
Similarity Measures



Case Study: North-Atlantic



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

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

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

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

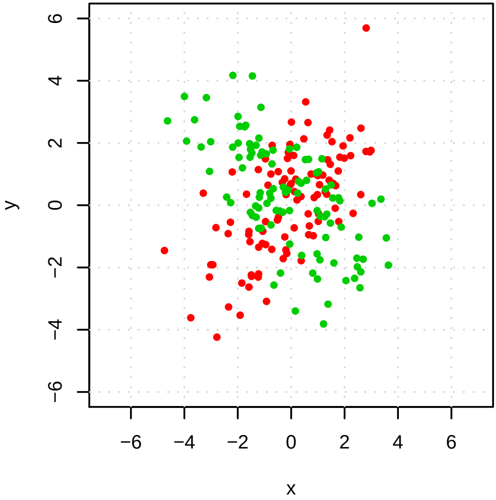
- visually
- Euclidean distance
- pattern correlation

Problems

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

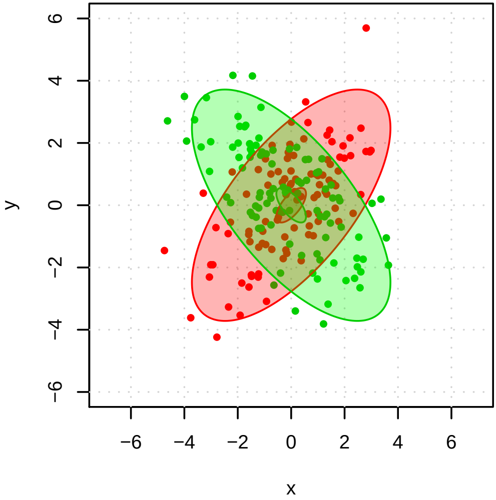
Compare WTs including shape and size

Comparing Weather Types



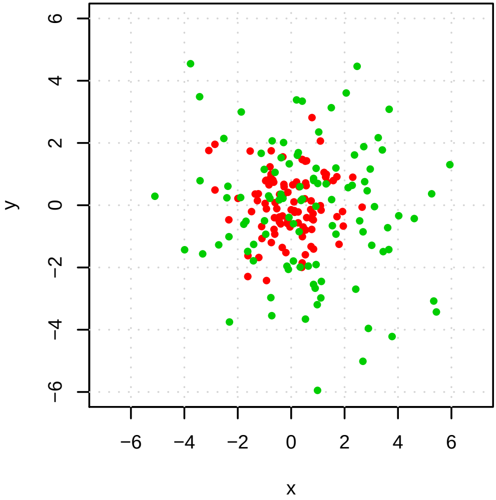
same centre, different orientation

Comparing Weather Types



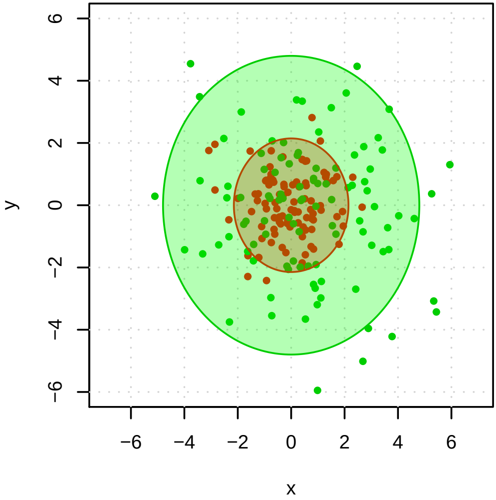
same centre, different orientation

Comparing Weather Types



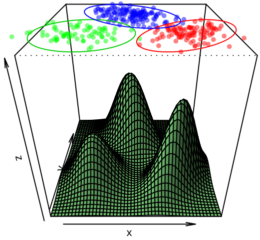
same centre, different extension

Comparing Weather Types

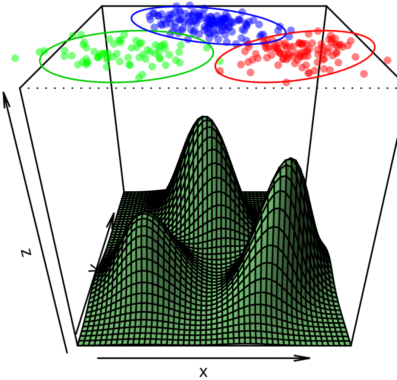


same centre, different extension

Gaussian Mixture Models



Gaussian Mixture Models

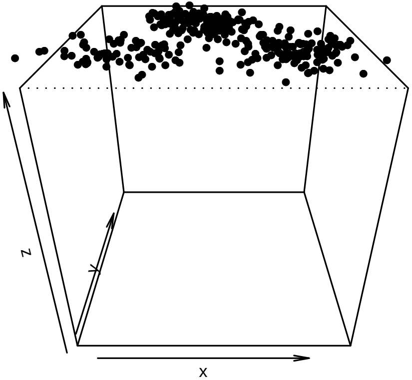


Modelling

- "truth"
- is unknown
- model selection
- estimation
- pdf, classification
- uncertainty

$$p(x \mid \boldsymbol{\theta}) = \sum_{k=1}^{3} a_k f(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

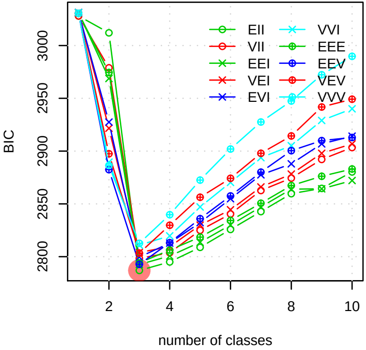
Gaussian Mixture Models



Modelling

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

Gaussian Mixture Models

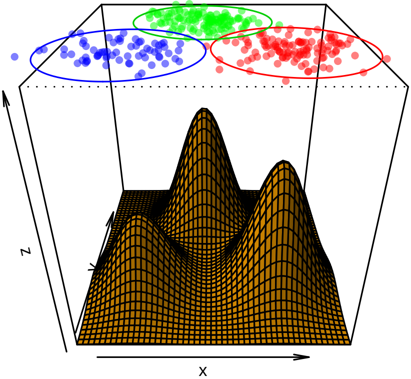


Modelling

- "truth"
- is unknown
- model selection
- estimation
- pdf, classification
- uncertainty

$$\mathrm{BIC} = -2 \log L(\boldsymbol{\theta}; \mathbf{x}) + m \log N,$$

$\boldsymbol{\theta}$: parameter vector, m: number of parameters



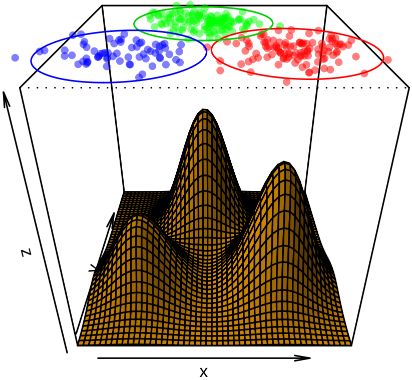
Modelling

- "truth"
- is unknown
- model selection
- estimation
- pdf, classification
- uncertainty

Estimated parameters (EM):

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\mu}}_3, \hat{\boldsymbol{\Sigma}}_1, \hat{\boldsymbol{\Sigma}}_2, \hat{\boldsymbol{\Sigma}}_3, \hat{a}_1, \hat{a}_2, \hat{a}_3)$$

Gaussian Mixture Models

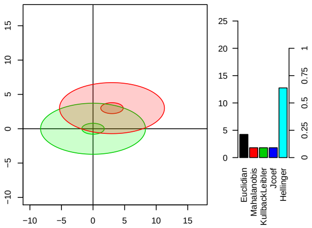


Modelling

- "truth"
- is unknown
- model selection
- estimation
- pdf, classification
- uncertainty

$$p(\mathbf{x} \mid \hat{\theta}) = \sum_{k=1}^{3} \hat{a}_k f(\mathbf{x}; \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k), \quad \mathrm{Cl}_{\mathbf{x_i}} \in \{1, 2, 3\}$$

Similarity Measures



Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
   $d^2_{\mathrm{Eucl}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2 = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbb{1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$
2. Mahalanobis distance:
   $d_{\mathrm{Mah}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}} = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$
3. Kullback-Leibler discrimination (KL):
   $d_{\mathrm{KL}}(P,Q) = I(P \mid Q) = \int_{\mathbb{R}} \log\left(\frac{q(x)}{p(x)}\right) q(z)\mathrm{d}x$
4. J-coefficient:
   $d_{\mathrm{J}}(P,Q) := (I(P \mid Q) + I(Q \mid P))/2$
5. Hellinger coefficient (s=1/2):
   $d_{\mathrm{H}}(P,Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0,1]$

Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
$$d^2_{\mathrm{Eucl}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2 = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbb{1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$
2. Mahalanobis distance:
$$d_{\mathrm{Mah}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}} = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$
3. Kullback-Leibler discrimination (KL):
$$d_{\mathrm{KL}}(P,Q) = I(P \mid Q) = \int_{\mathbb{R}} \log\left(\frac{q(x)}{p(x)}\right) q(z)\mathrm{d}x$$
4. J-coefficient:
$$d_{\mathrm{J}}(P,Q) := (I(P \mid Q) + I(Q \mid P))/2$$
5. Hellinger coefficient (s=1/2):
$$d_{\mathrm{H}}(P,Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0,1]$$

Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
$$d^2_{\mathrm{Eucl}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2 = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbb{1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$

2. Mahalanobis distance:
$$d_{\mathrm{Mah}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}} = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$

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

4. J-coefficient:
$$d_{\mathrm{J}}(P,Q) := (I(P \mid Q) + I(Q \mid P))/2$$

5. Hellinger coefficient (s=1/2):
$$d_{\mathrm{H}}(P,Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0,1]$$

Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
$$d^2_{\mathrm{Eucl}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2 = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbb{1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$
2. Mahalanobis distance:
$$d_{\mathrm{Mah}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}} = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$
3. Kullback-Leibler discrimination (KL):
$$d_{\mathrm{KL}}(P,Q) = I(P \mid Q) = \int_{\mathbb{R}} \log\left(\frac{q(x)}{p(x)}\right) q(z)\mathrm{d}x$$
4. J-coefficient:
$$d_{\mathrm{J}}(P,Q) := (I(P \mid Q) + I(Q \mid P))/2$$
5. Hellinger coefficient (s=1/2):
$$d_{\mathrm{H}}(P,Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0,1]$$

Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
$$d^2_{\mathrm{Eucl}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2 = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbb{1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$

2. Mahalanobis distance:
$$d_{\mathrm{Mah}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}} = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$$

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

4. J-coefficient:
$$d_{\mathrm{J}}(P,Q) := (I(P \mid Q) + I(Q \mid P))/2$$

5. Hellinger coefficient (s=1/2):
$$d_{\mathrm{H}}(P,Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0,1]$$

Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
$d^2_{\mathrm{Eucl}}(P, Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2 = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbb{1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$

2. Mahalanobis distance:
$d_{\mathrm{Mah}}(P, Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}} = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$

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

4. J-coefficient:
$d_{\mathrm{J}}(P, Q) := (I(P \mid Q) + I(Q \mid P))/2$

5. Hellinger coefficient (s=1/2):
$d_{\mathrm{H}}(P, Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0, 1]$

Similarity Measures for (Gaussian) pdfs

1. Euclidian distance:
$d^2_{\mathrm{Eucl}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2 = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbb{1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$

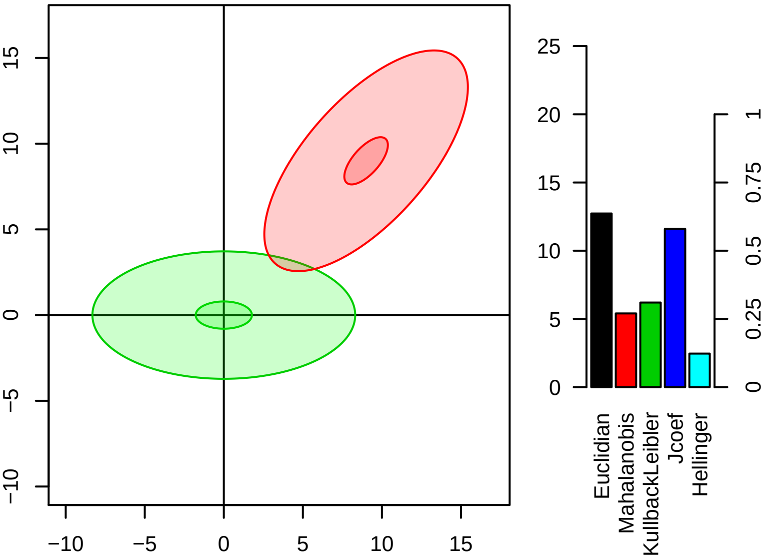
2. Mahalanobis distance:
$d_{\mathrm{Mah}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}} = (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)$

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

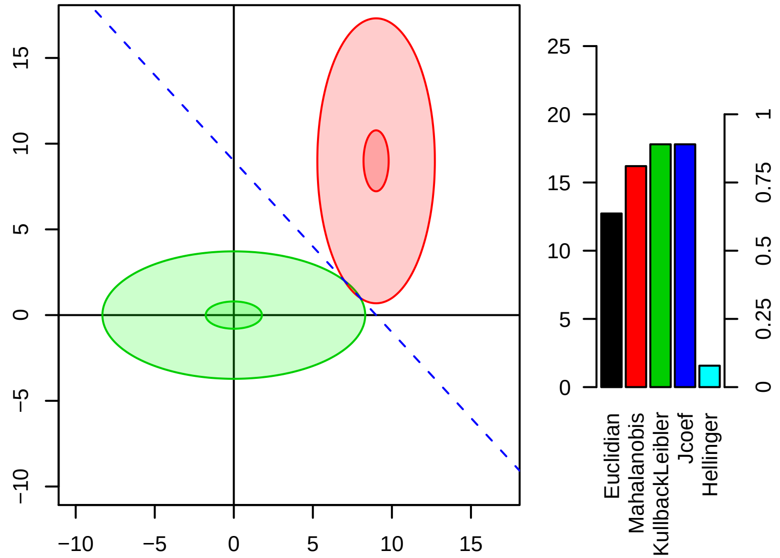
4. J-coefficient:
$d_{\mathrm{J}}(P,Q) := (I(P \mid Q) + I(Q \mid P))/2$

5. Hellinger coefficient (s=1/2):
$d_{\mathrm{H}}(P,Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0,1]$

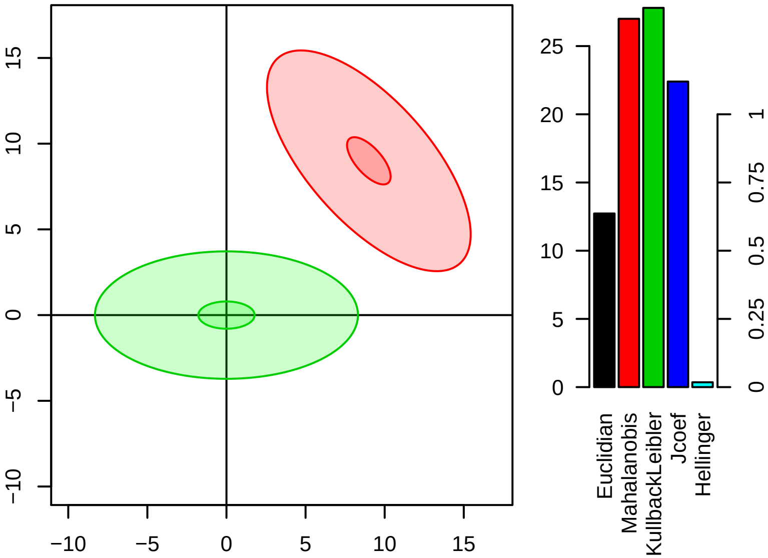
Similarity Measures, Example



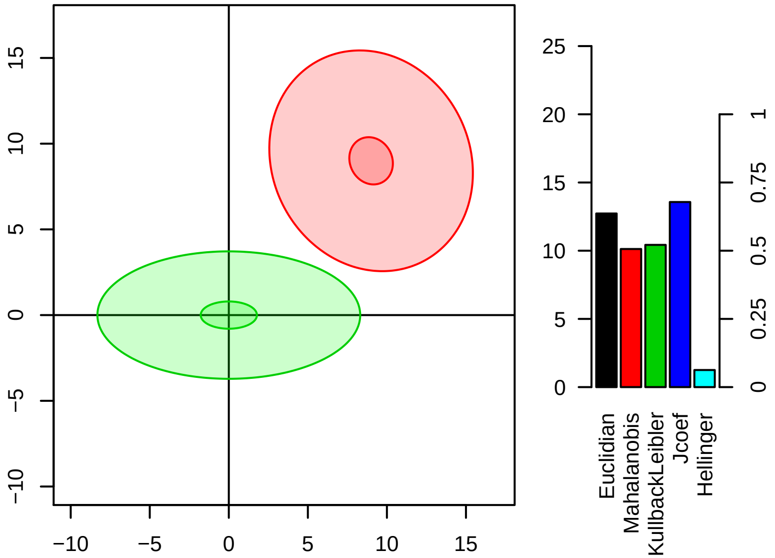
Similarity Measures, Example



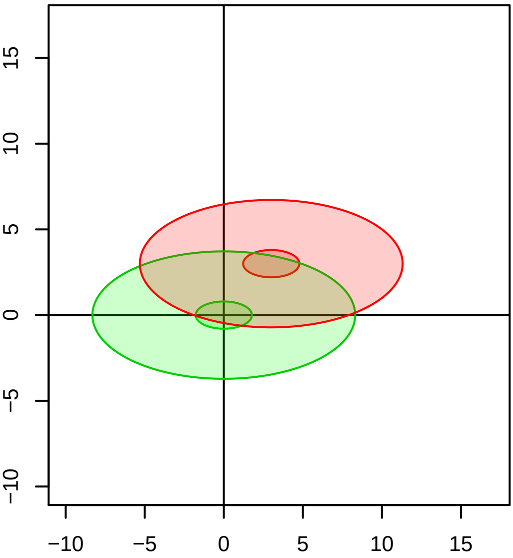
Similarity Measures, Example

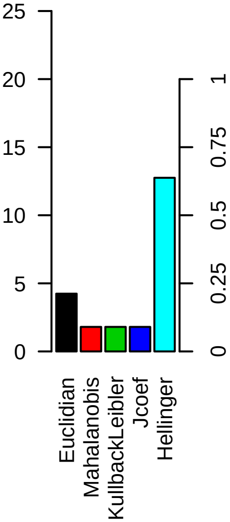


Similarity Measures, Example

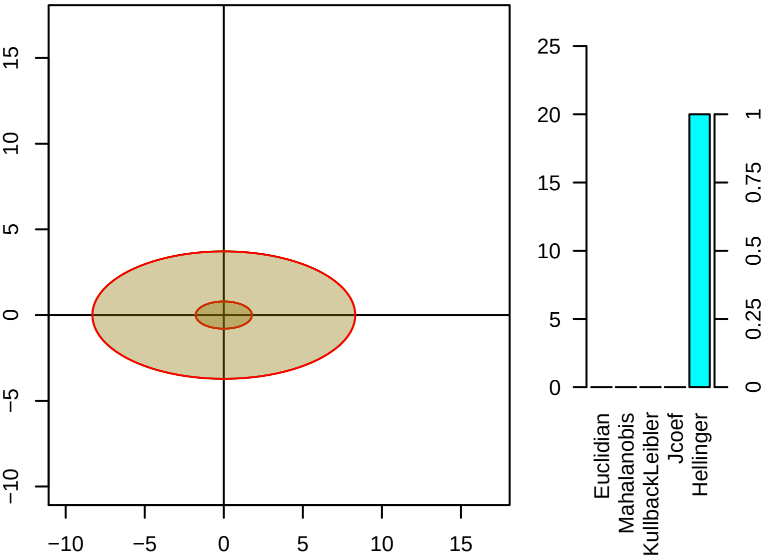


Similarity Measures, Example

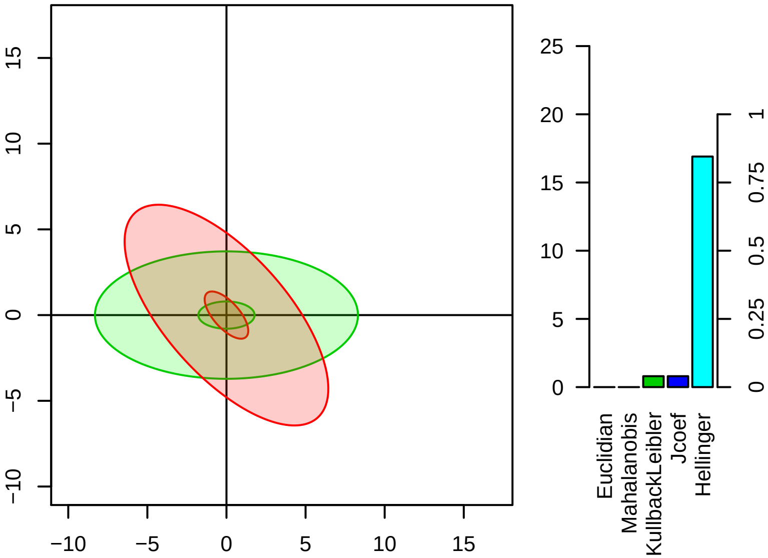




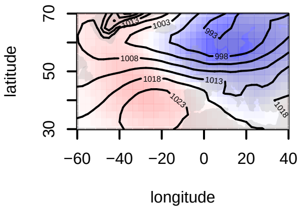
Similarity Measures, Example



Similarity Measures, Example



Case Study

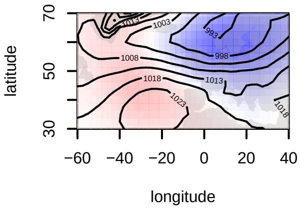


Comparing North-Atlantic WTs

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

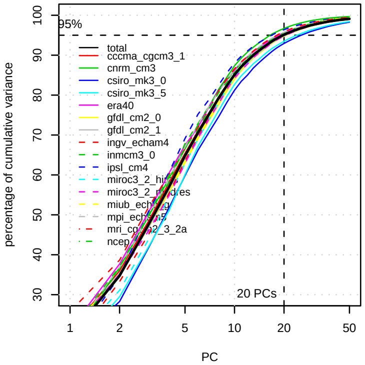
14 IPCC Models

- 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%



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 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

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

1. force 5 spherical clusters ($\Sigma_k = \sigma_k \mathbb{1}$)
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

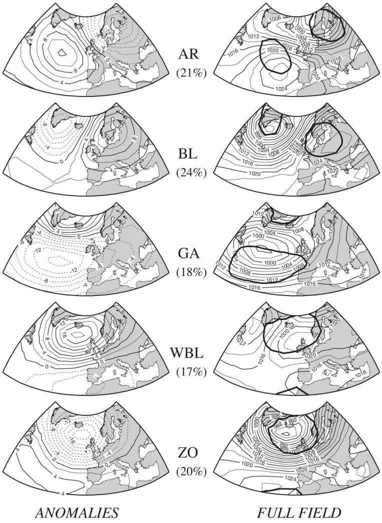
Reproduce k-Means Result (Plaut&Simmonet, 2001)

1. force 5 spherical clusters ($\Sigma_k = \sigma_k \mathbb{1}$)
2. define 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

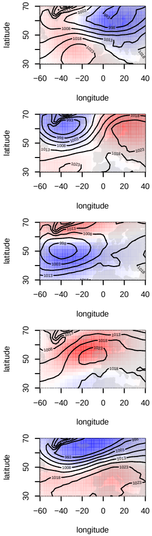
Reproduce k-Means Result (Plaut&Simmonet, 2001)

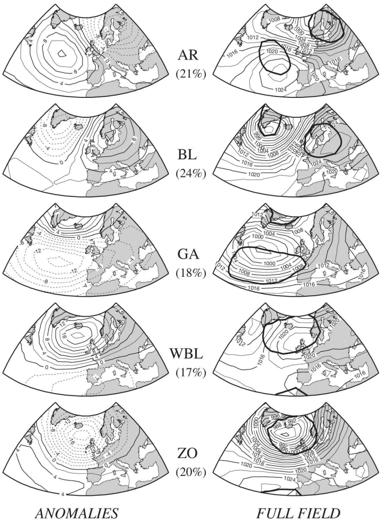
1. force 5 spherical clusters ($\Sigma_k = \sigma_k \mathbb{1}$)
2. define WTs on NCEP/NCAR
3. associate NCEP/NCAR means to reference
4. define WTs on ERA-40/GCMs
5. associate GCM WTs to NCEP/NCAR WTs



from Plaut&Simonnet, (2001)

Defining Weather Types with Gaussian Mixtures



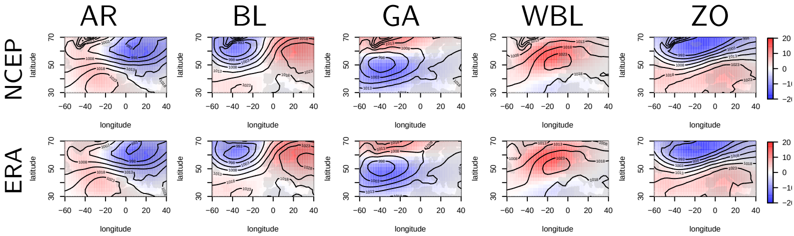


from Plaut&Simonnet, (2001)

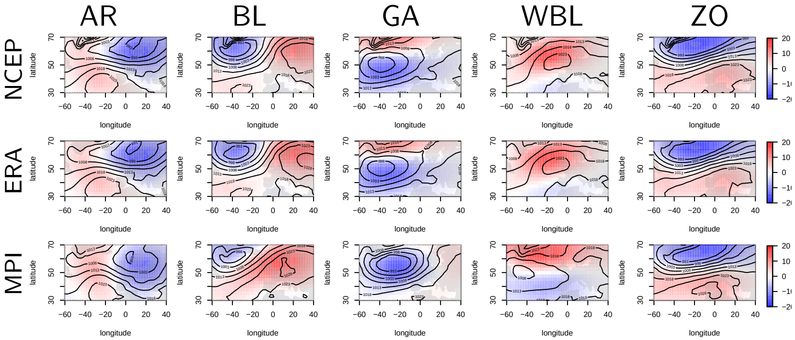
Reproduce *k*-Means Result (Plaut&Simmonet, 2001)

1. force 5 spherical clusters ($\Sigma_k = \sigma_k \mathbb{1}$)
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

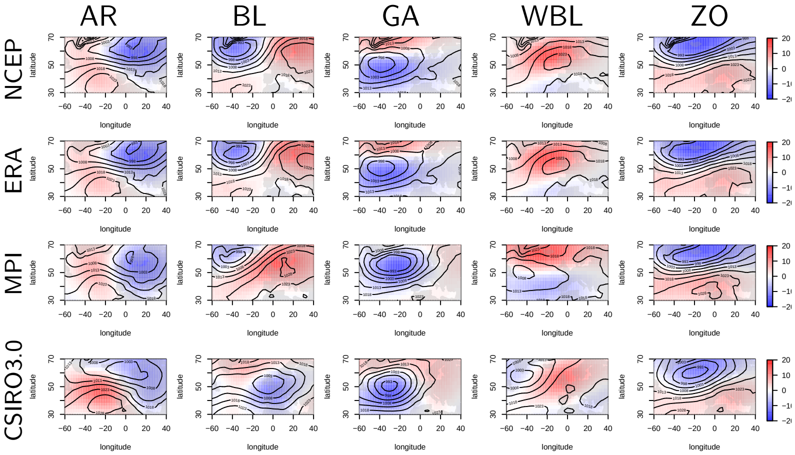
Weather Type Mean Values



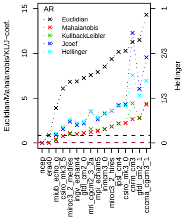
Weather Type Mean Values



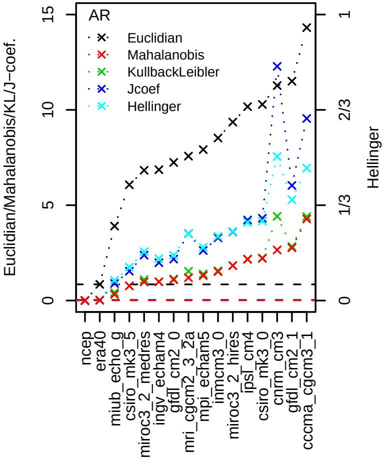
Weather Type Mean Values



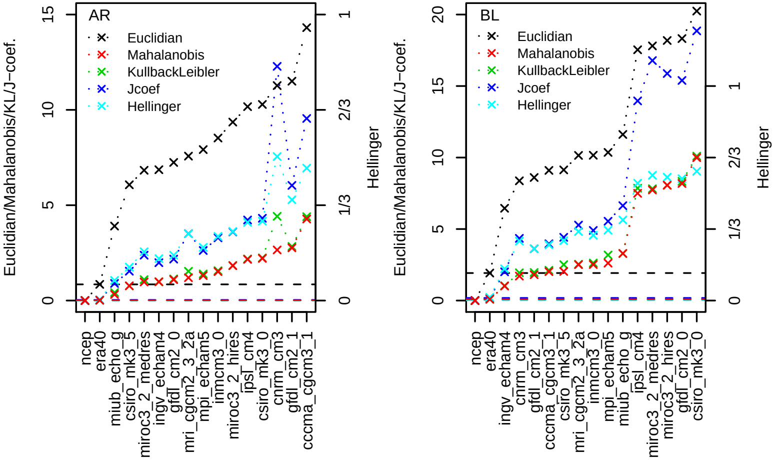
Difference to NCEP/NCAR WTs



Quantifying Differences to NCEP/NCAR



Quantifying Differences to NCEP/NCAR



Ranking

Hellinger Coefficient ($1 - d_{\mathrm{H}}$)

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

4 GCMs remain the top 4 across all measures

Euclidean Mahalanobis Kullback-Leibler *J*-Coefficient

WT differences

Henning Rust

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

Table

- 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

WT differences

Henning Rust

- WTs
- Mixtures
- Similarity
- Case Study
- 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

Table

- 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

Table

- 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

Table

- 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

Table

- 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

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
- ...

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
- ...

Disclaimer

I do not want to advocate 5 spherical clusters for the NA region!

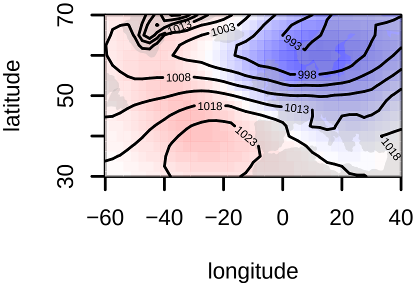
- Gaussian mixtures + BIC

Appendix

Common PCA

PCA

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



Comparison

Common PCA

PCA

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

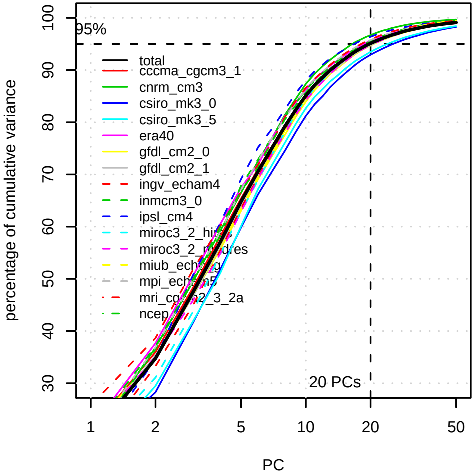
model	date	grid1	grid2	...	grid250
ERA-40	01/01/1975	0.376	0.435	...	1.344
ERA-40	01/02/1975	0.276	0.335	...	1.244
ERA-40	...	...	...	...	...
ERA-40	12/31/2000	0.276	0.335	...	1.244
NCEP	01/01/1975	0.376	0.435	...	1.344
NCEP	01/02/1975	0.276	0.335	...	1.244
NCEP	...	...	...	...	...
NCEP	12/31/2000	0.276	0.335	...	1.244
CCCMA	01/01/1975	0.376	0.435	...	1.344
CCCMA	01/02/1975	0.276	0.335	...	1.244
CCCMA	...	...	...	...	...
CCCMA	12/31/2000	0.276	0.335	...	1.244
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
MRI	01/01/1975	0.376	0.435	...	1.344
MRI	01/02/1975	0.276	0.335	...	1.244
MRI	...	...	...	...	...
MRI	12/31/2000	0.276	0.335	...	1.244

Comparison

Common PCA

PCA

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



Comparison

Association of WTs

What GCM WT is Associated to NCEP WTs?

Mapping:

model	$\rightarrow$	NCEP
Π_1	$\rightarrow$	1
Π_2	$\rightarrow$	2
	$\vdots$	
Π_5	$\rightarrow$	5

Use Mapping Minimising Sum of Distances

$$\Pi_0 = \arg \min_{\Pi \in \Pi\{1,2,3\}} \sum_{i=1}^{5} d(\mathrm{WT}^i_{\mathrm{NCEP}}, \mathrm{Cl}^{\Pi_i}_{\mathrm{GCM}}),$$

Π: permutation out of all possible permutations of $\{1, 2, 3\}$.

Comparison

Association of WTs

What GCM WT is Associated to NCEP WTs?

Mapping:

model	$\rightarrow$	NCEP
Π_1	$\rightarrow$	1
Π_2	$\rightarrow$	2
	$\vdots$	
Π_5	$\rightarrow$	5

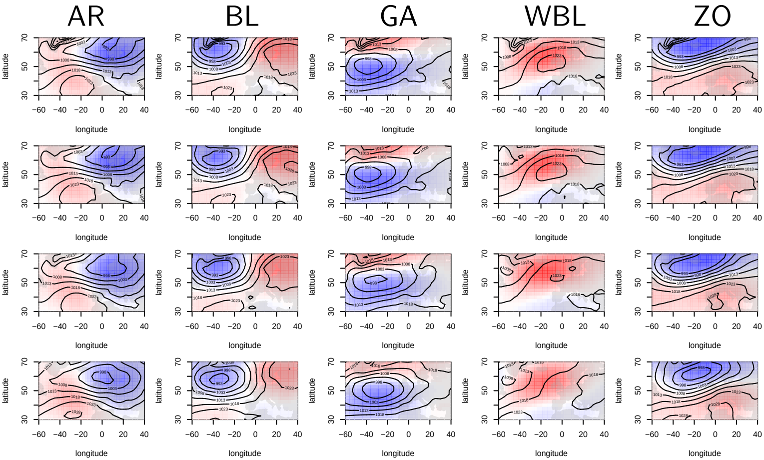
Use Mapping Minimising Sum of Distances

$$\Pi_0 = \arg \min_{\Pi \in \Pi\{1,2,3\}} \sum_{i=1}^{5} d(\mathrm{WT}^i_{\mathrm{NCEP}}, \mathrm{Cl}^{\Pi_i}_{\mathrm{GCM}}),$$

Π: permutation out of all possible permutations of $\{1, 2, 3\}$.

◂ Comparison

Centroids of Projection onto NCEP/NCAR WTs



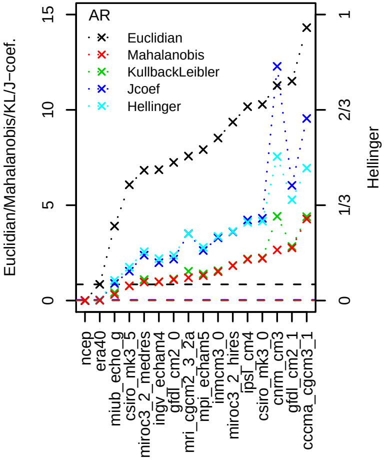
NCEP/NCAR – ERA-40 – MPI ECHAM5 – CSIRO MK3.0

Characteristics of Measures

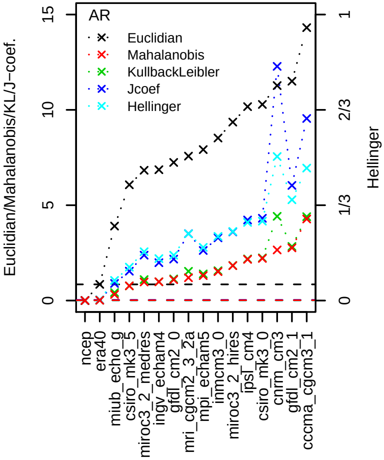
Summary Table

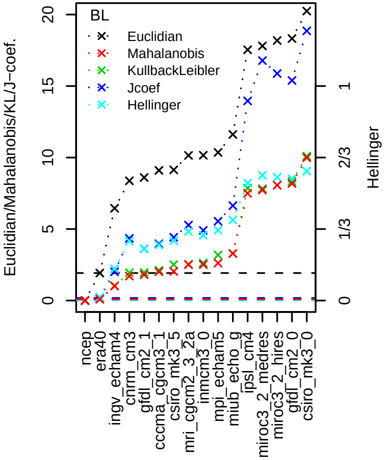
Measure	Symm.	Characteristics
Euclidean	yes	distance in means
Mahalanobis	no	distance in means, metric depending on one covariance matrix
Kullback-Leibler	no	metric depends on both covariance matrices
J-coefficient	yes	symmetrised KL
Hellinger(s=1/2)	yes	measures “overlap”

Quantifying Differences to NCEP/NCAR

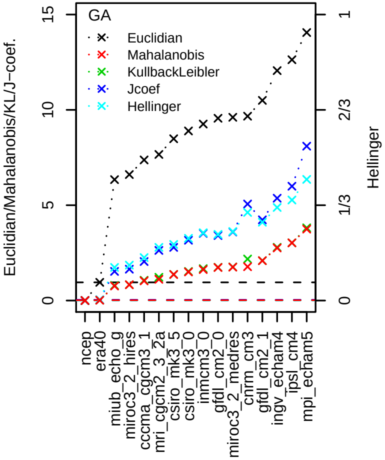


Quantifying Differences to NCEP/NCAR





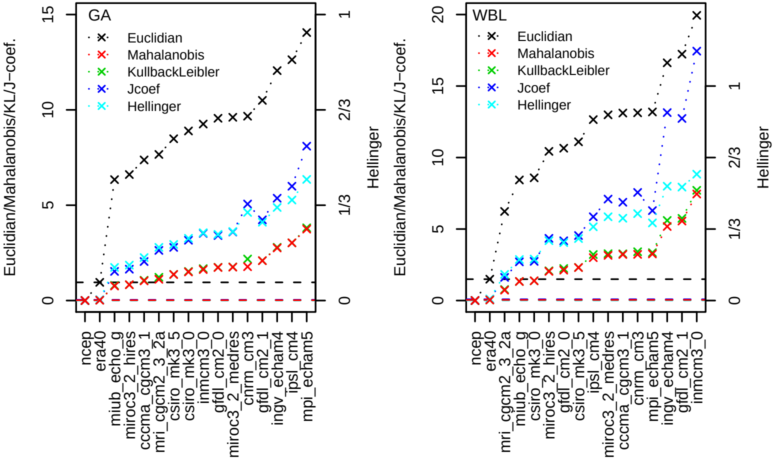
Quantifying Differences to NCEP/NCAR



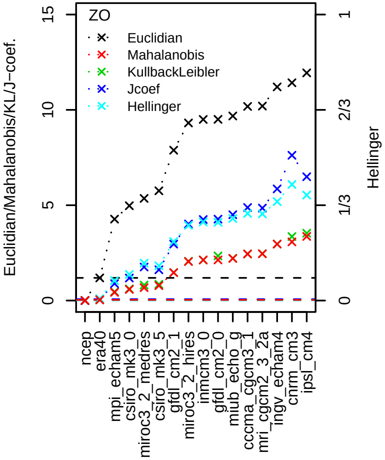
Quantifying Differences to NCEP/NCAR

WT differences

Henning Rust



Quantifying Differences to NCEP/NCAR



WT differences

Henning Rust

Ranking

Euclidean Distance

	Weather Type					
	AR	BL	GA	WBL	ZO	Mean(Std)
Reanalysis						
NCEP	0.000	0.000	0.000	0.000	0.000	0.000(0.000)
ERA-40	0.848	1.188	1.493	1.922	0.954	1.281(0.436)
GCM						
MIUB ECHO G	3.906	9.680	8.432	11.614	6.342	7.995(2.983)
CSIRO MK3.5	6.071	5.756	11.099	9.136	8.486	8.110(2.226)
MRI CGCM2.3	7.575	10.199	6.231	10.148	7.662	8.363(1.748)
MPI ECHAM5	7.917	4.270	13.188	10.360	14.056	9.958(3.997)

Ranking

Mahalanobis Distance

	Weather Type					Mean(Std)
	AR	BL	GA	WBL	ZO	
Reanalysis						
NCEP	0.000	0.000	0.000	0.000	0.000	0.000(0.000)
ERA-40	0.015	0.033	0.042	0.090	0.017	0.039(0.030)
GCM						
CSIRO MK3.5	0.770	0.781	2.308	2.035	1.366	1.452(0.706)
MIUB ECHO G	0.318	2.209	1.332	3.289	0.763	1.582(1.187)
MRI CGCM2.3	1.198	2.453	0.727	2.511	1.113	1.600(0.824)
MPI ECHAM5	1.308	0.430	3.258	2.617	3.747	2.272(1.377)

Ranking

Kullback-Leibler Divergence

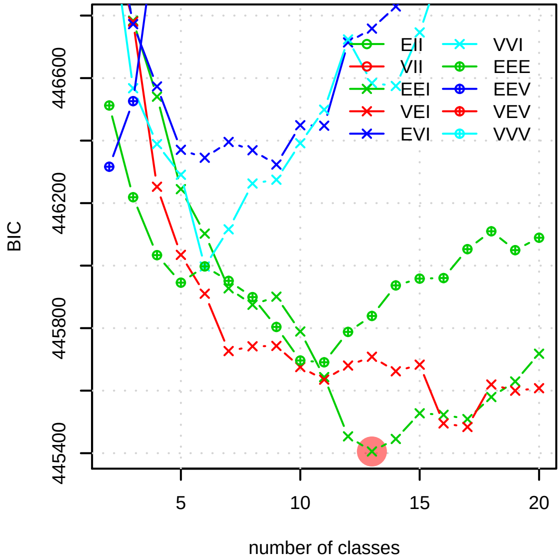
	Weather Type					
	AR	BL	GA	WBL	ZO	Mean(Std)
Reanalysis						
NCEP	0.000	0.000	0.000	0.000	0.000	0.000(0.000)
ERA-40	0.015	0.037	0.043	0.091	0.018	0.041(0.030)
GCM						
CSIRO MK3.5	0.770	0.853	2.314	2.511	1.371	1.564(0.811)
MIUB ECHO G	0.418	2.214	1.335	3.290	0.792	1.610(1.157)
MRI CGCM2.3	1.534	2.456	0.771	2.539	1.219	1.704(0.774)
MPI ECHAM5	1.396	0.459	3.347	3.184	3.813	2.440(1.438)

Ranking

J-Coefficient

	Weather Type					
	AR	BL	GA	WBL	ZO	Mean(Std)
Reanalysis						
NCEP	0.000	0.000	0.000	0.000	0.000	0.000(0.000)
ERA-40	0.030	0.073	0.086	0.181	0.036	0.081(0.060)
GCM						
CSIRO MK3.5	1.540	1.614	4.546	4.419	2.787	2.981(1.458)
MIUB ECHO G	0.896	4.501	2.706	6.634	1.527	3.253(2.336)
MRI CGCM2.3	3.506	4.850	1.617	5.277	2.628	3.576(1.521)
MPI ECHAM5	2.626	0.886	6.288	5.545	8.096	4.688(2.899)

BIC


BIC
446600
446200
445800
445400
5
10
15
20
number of classes
EII
VII
EEI
VEI
EVI
VVI
EEE
EEV
VEV
VVV