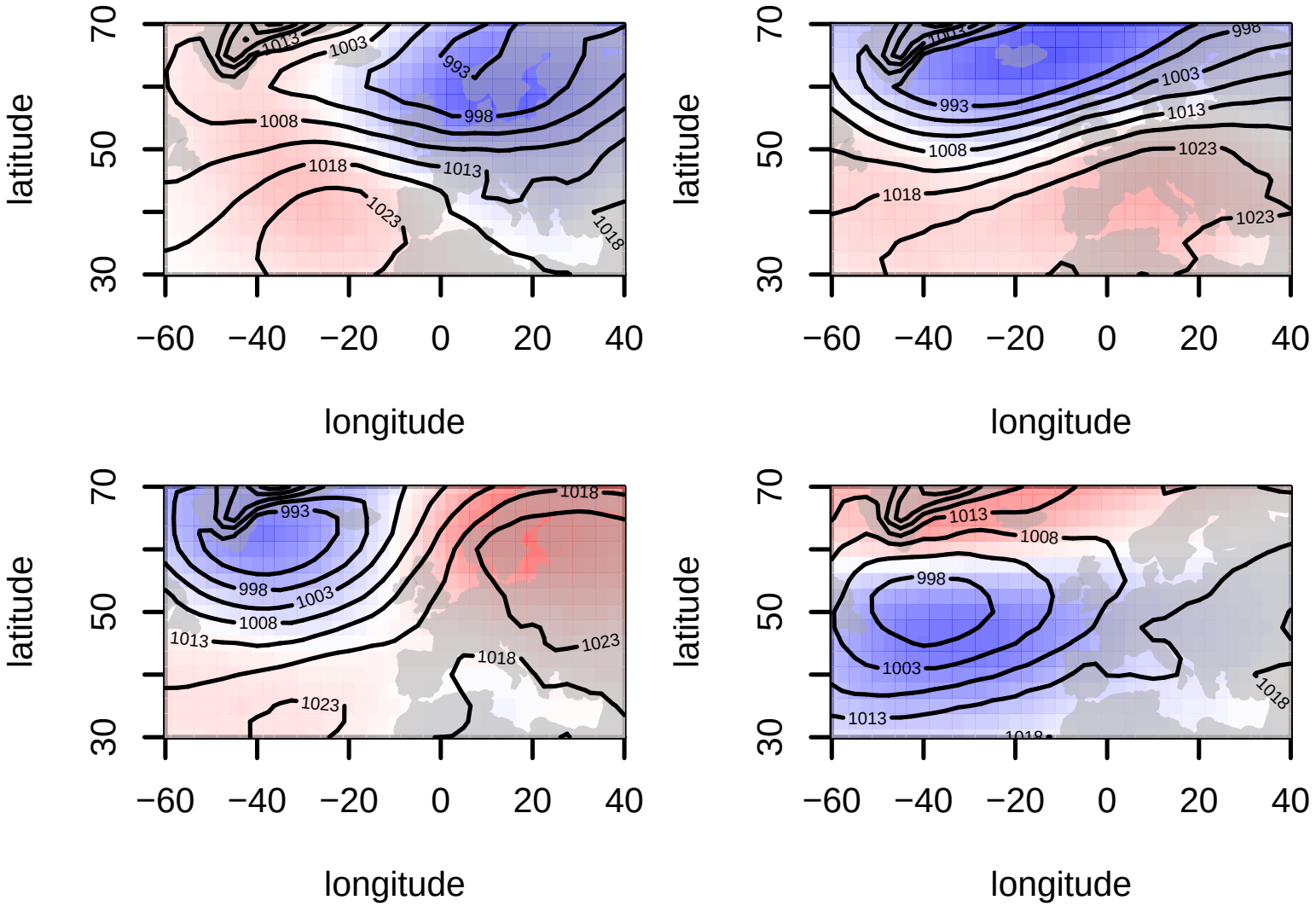
Quantifying differences in circulation patterns with probabilistic methods

Henning Rust[1]
Mathieu Vrac[2], Matthieu Lengaigne[3], Benjamin Sultan[3]

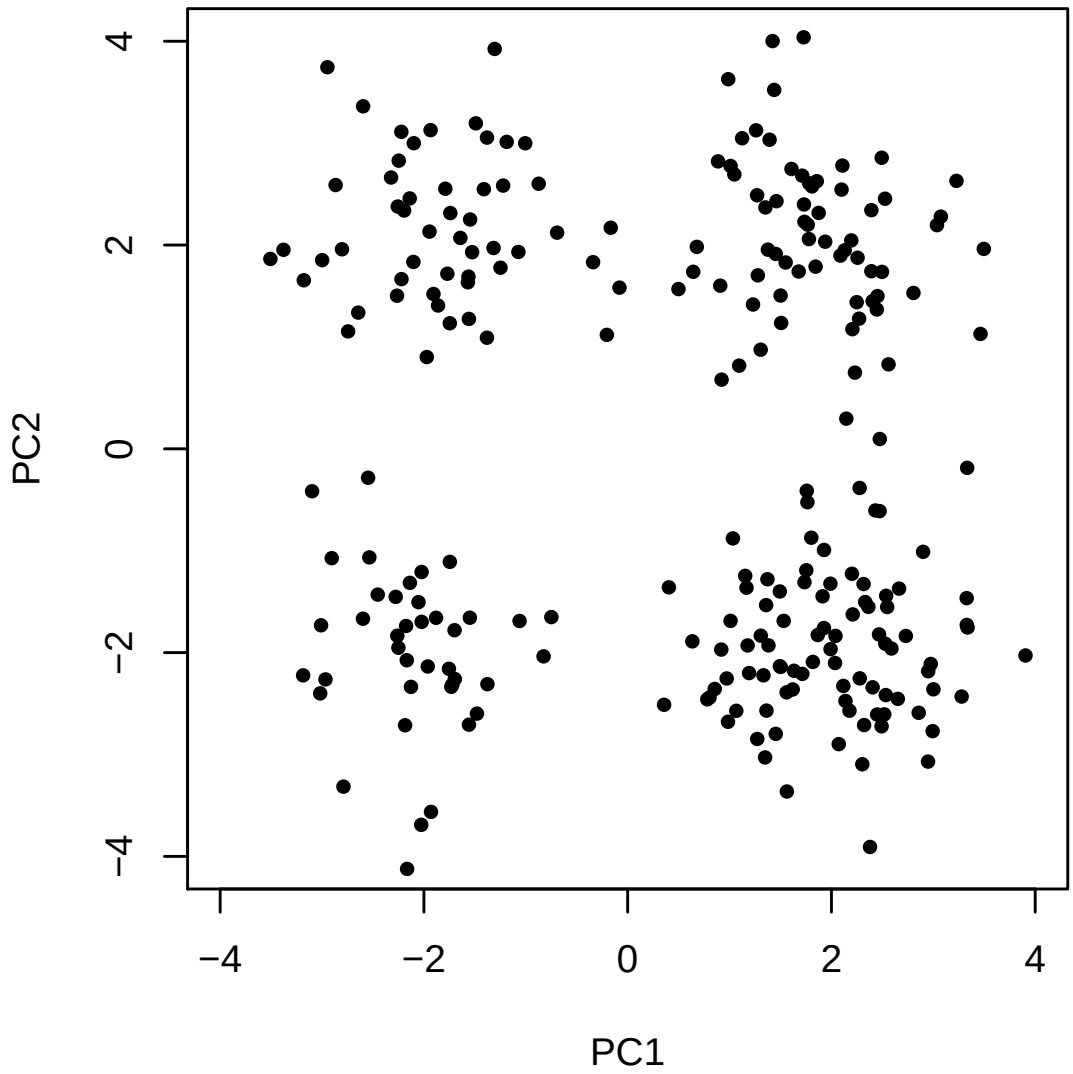
[1] IfM FU Berlin
[2] LSCE, Gif-sur-Yvette, Frankreich
[3] LOCEAN, Paris, Frankreich

EMS2011, Berlin

Motivation


latitude
longitude
70
50
30
−60
−40
−20
0
20
40
1013
1003
993
998
1008
1018
1023

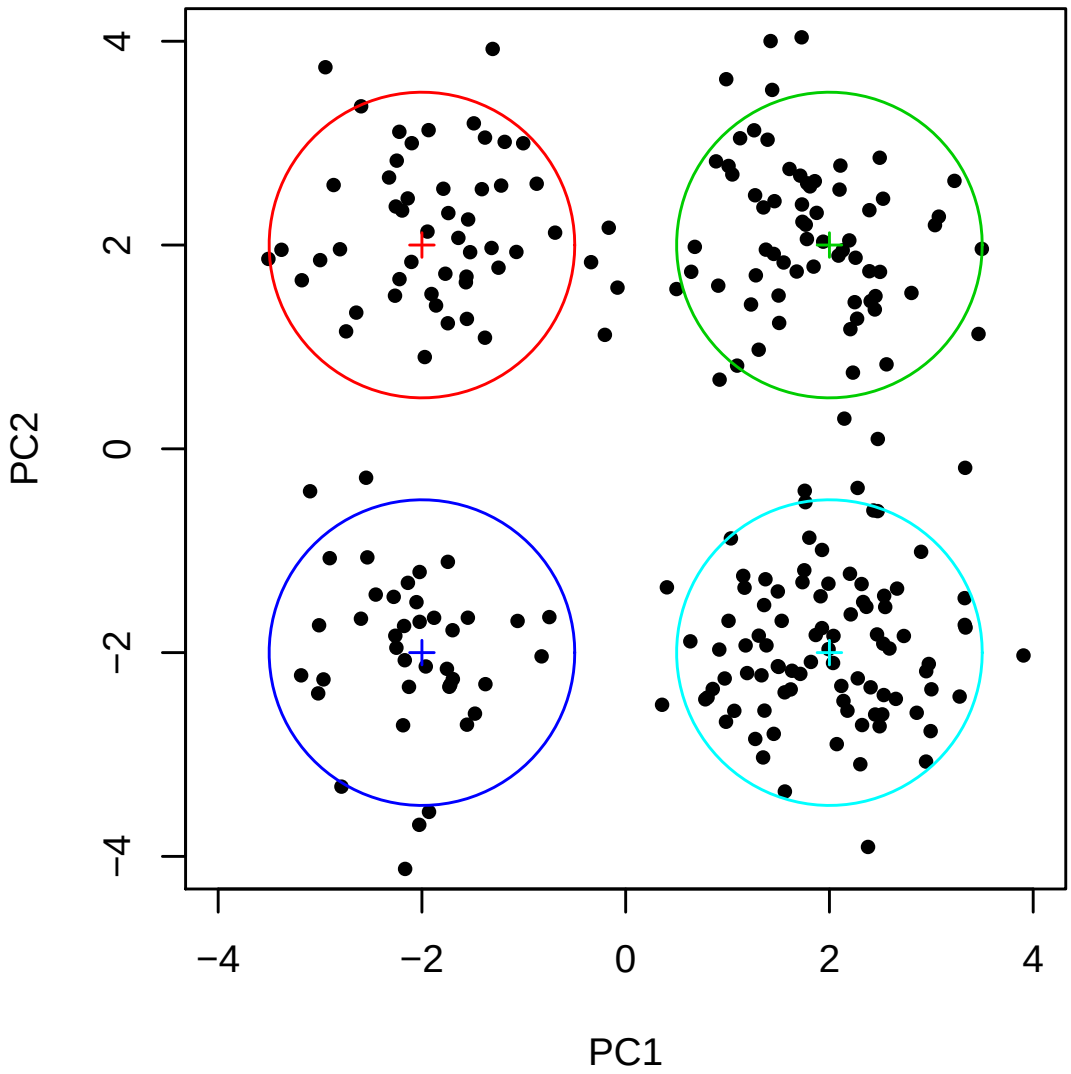
Motivation



Procedure

- reanalysis

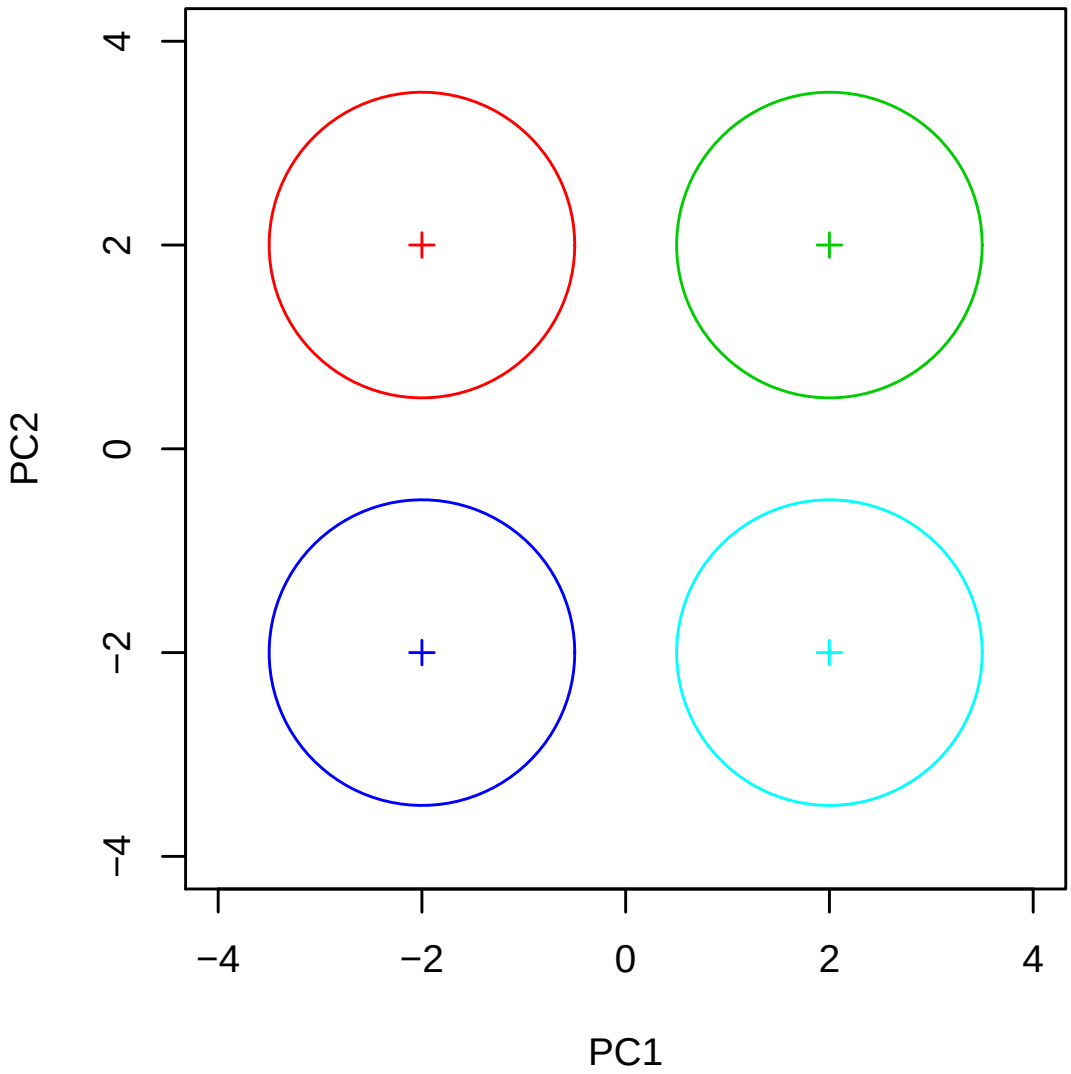
Motivation



Procedure

- reanalysis
- clusters

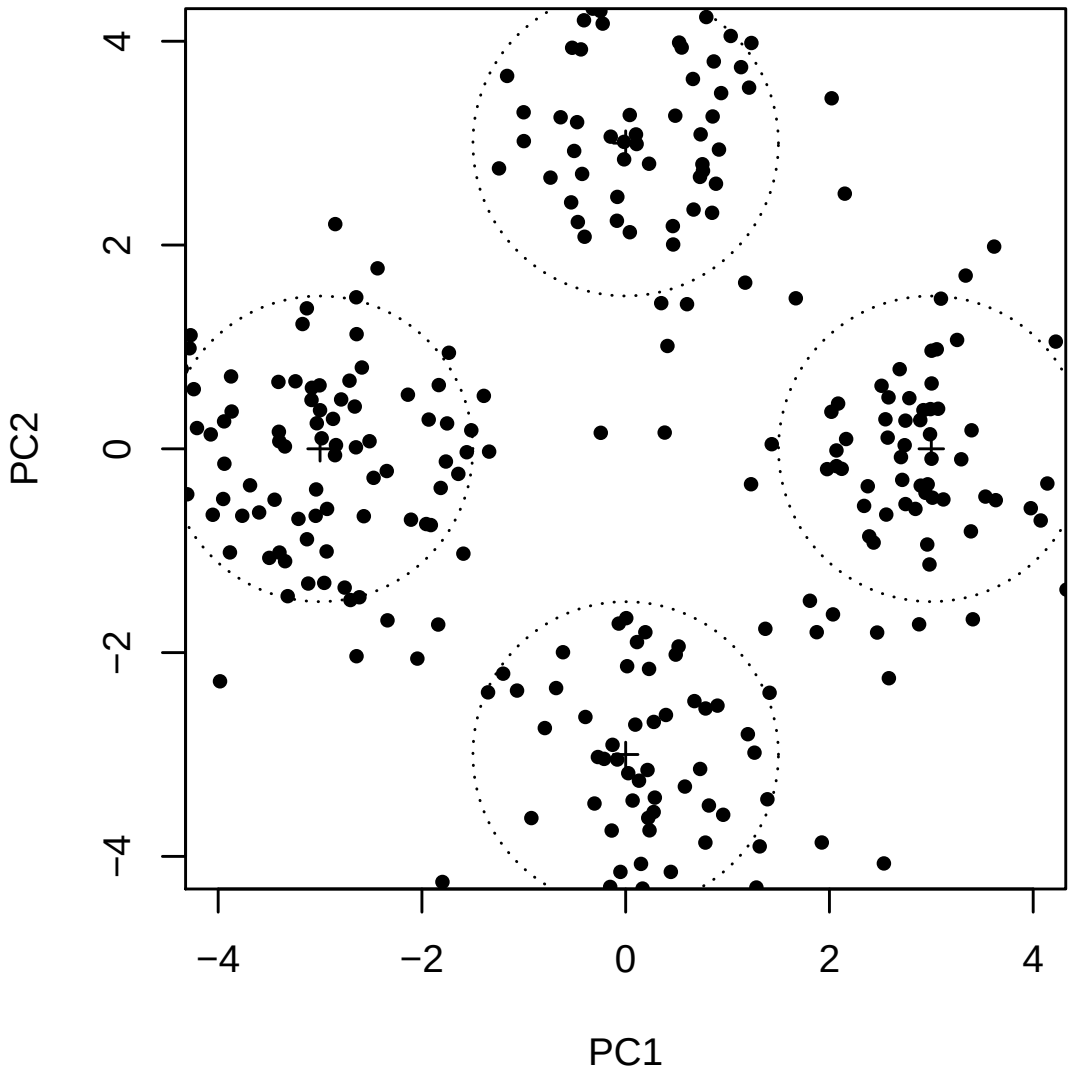
Motivation



Procedure

- reanalysis
- clusters
- CPs!

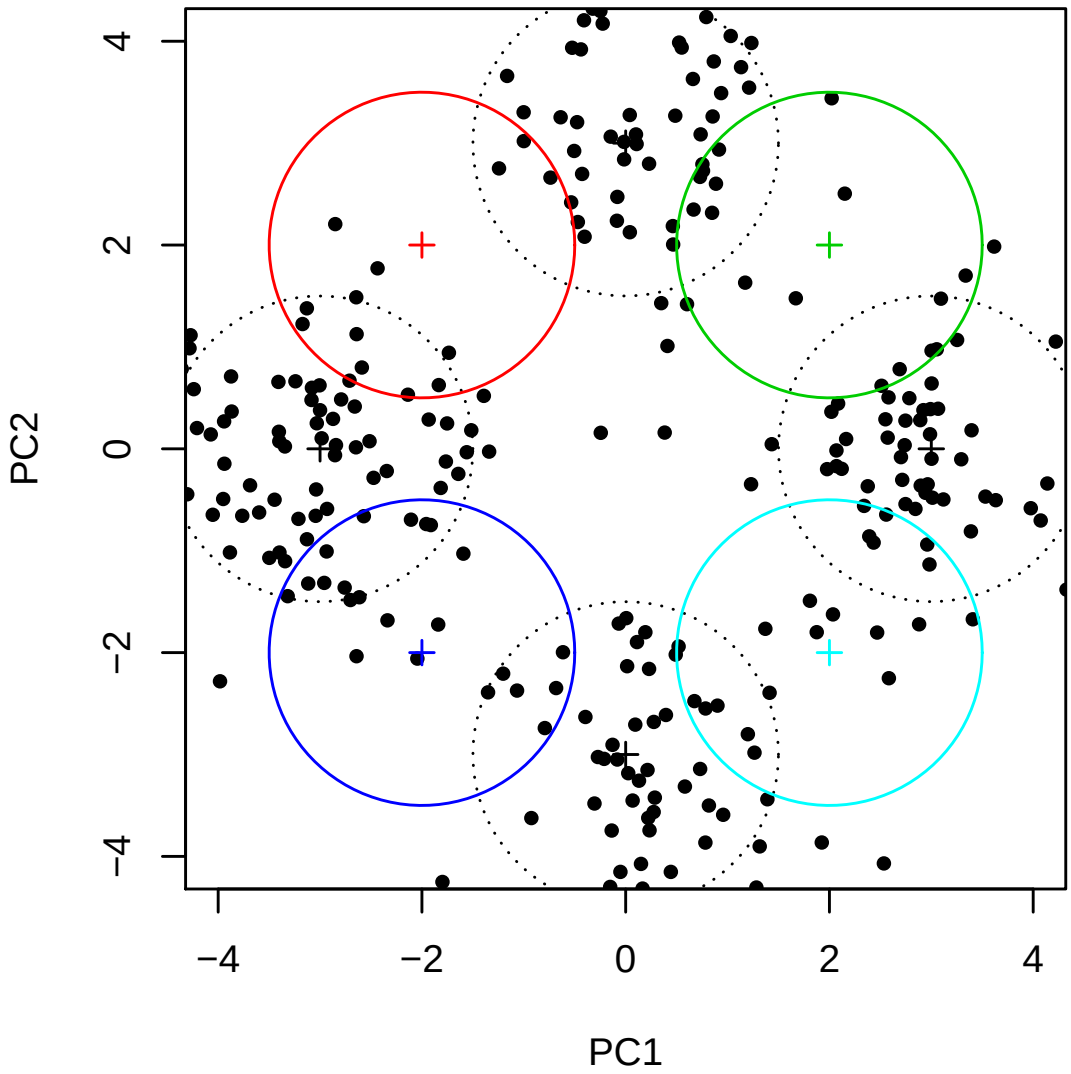
Motivation



Procedure

- reanalysis
- clusters
- CPs!
- model

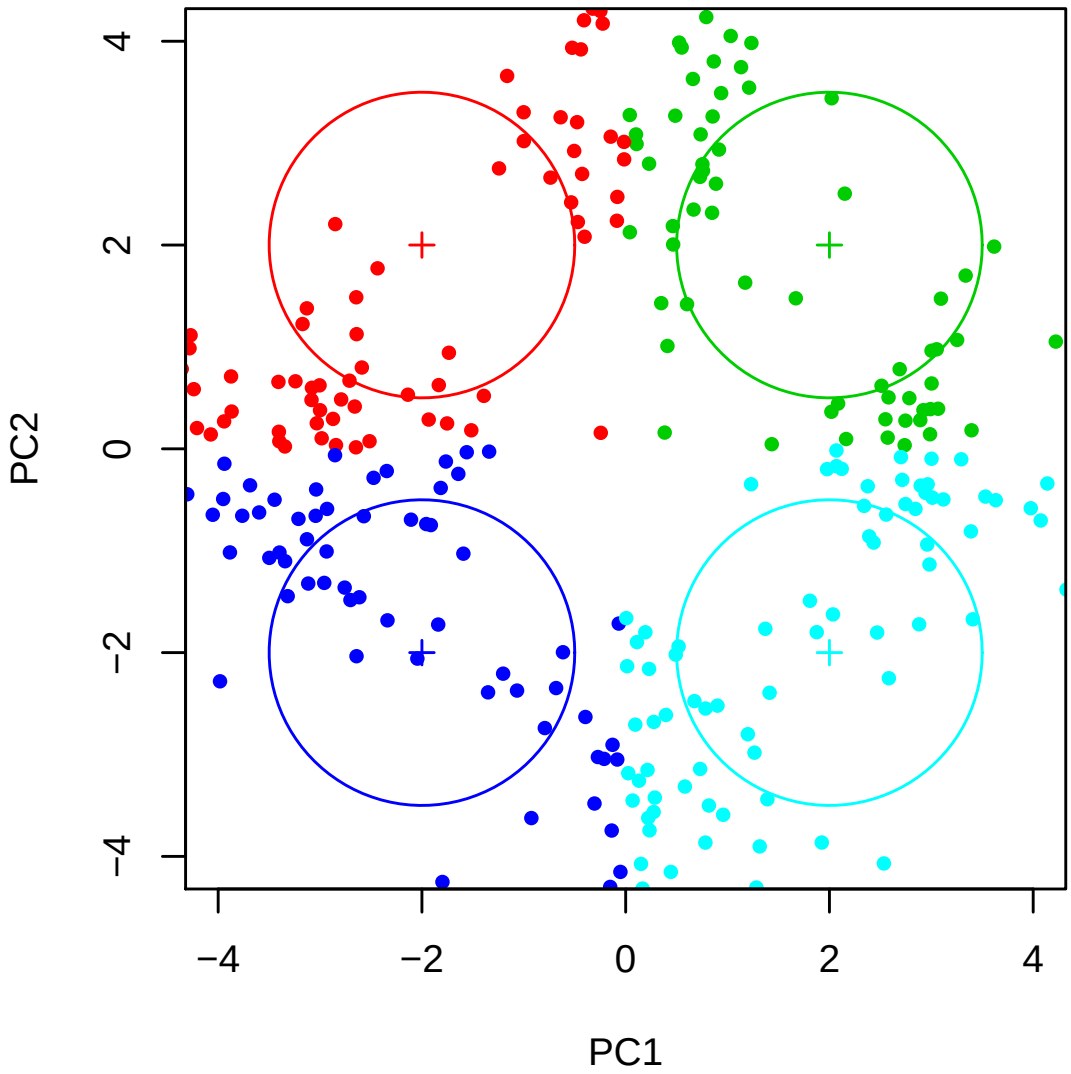
Motivation



Procedure

- reanalysis
- clusters
- CPs!
- model
- project

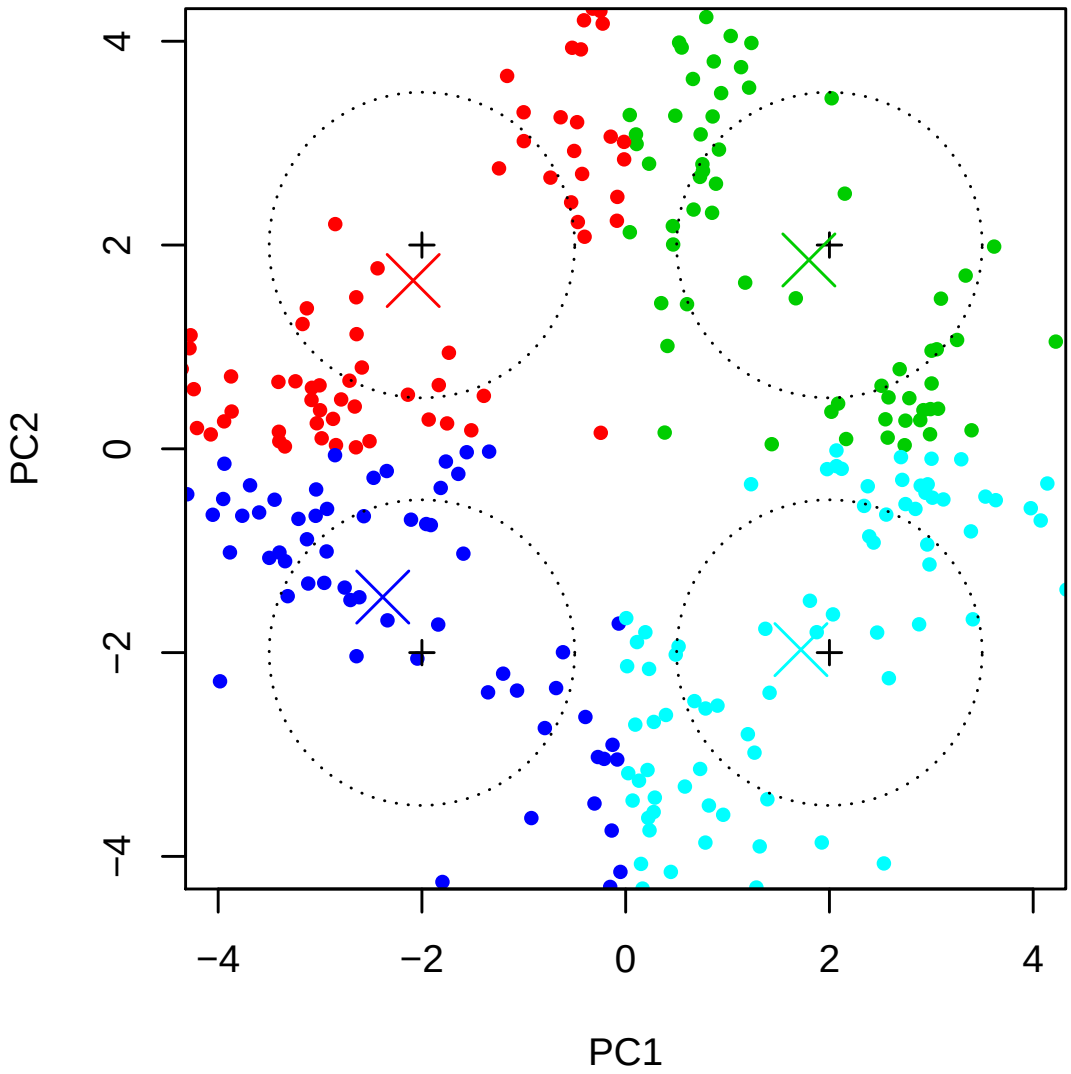
Motivation



Procedure

- reanalysis
- clusters
- CPs!
- model
- project
- associate

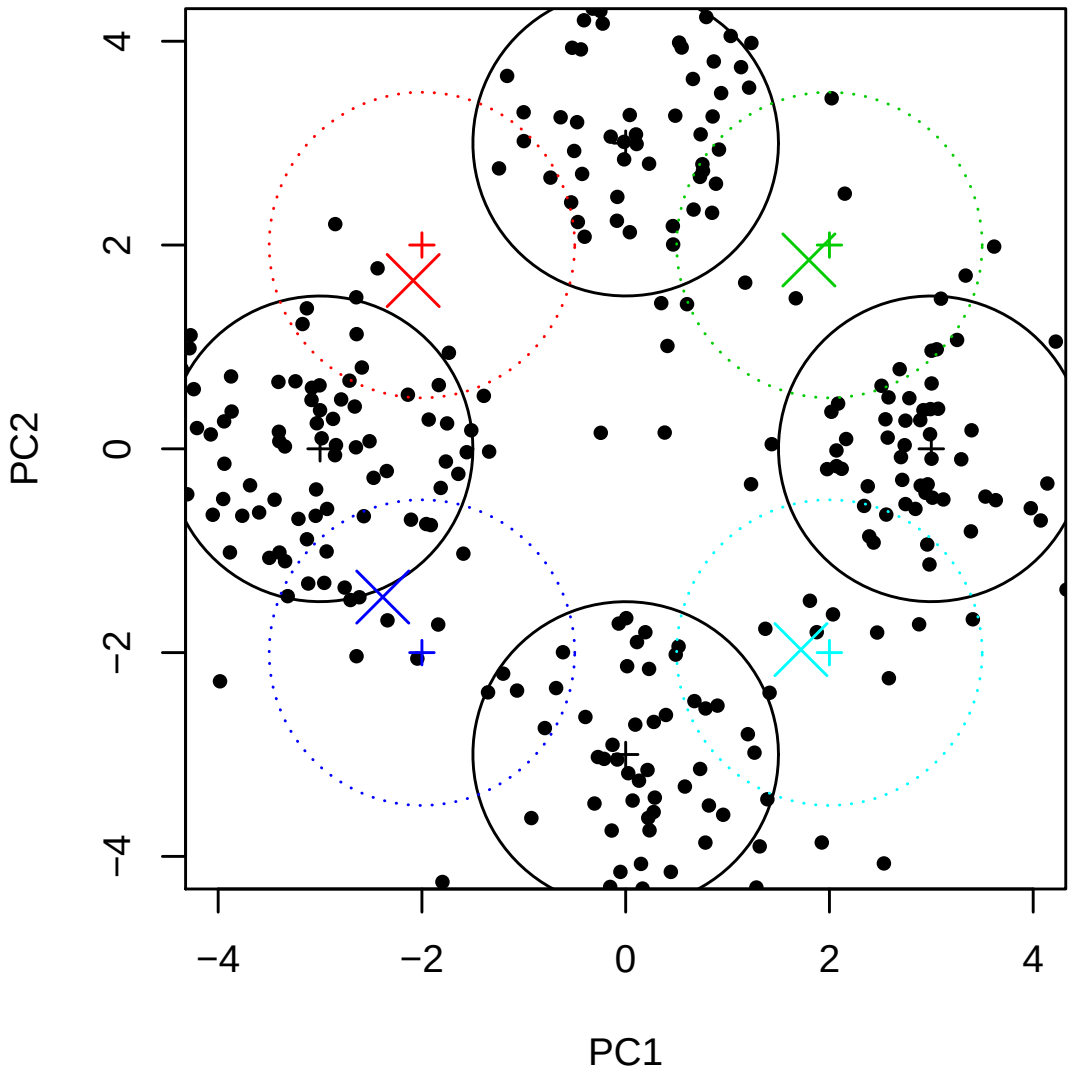
Motivation



Procedure

- reanalysis
- clusters
- CPs!
- model
- project
- associate
- centroids

Motivation



Procedure

- reanalysis
- clusters
- CPs!
- model
- project
- associate
- centroids
- meaning?

Motivation

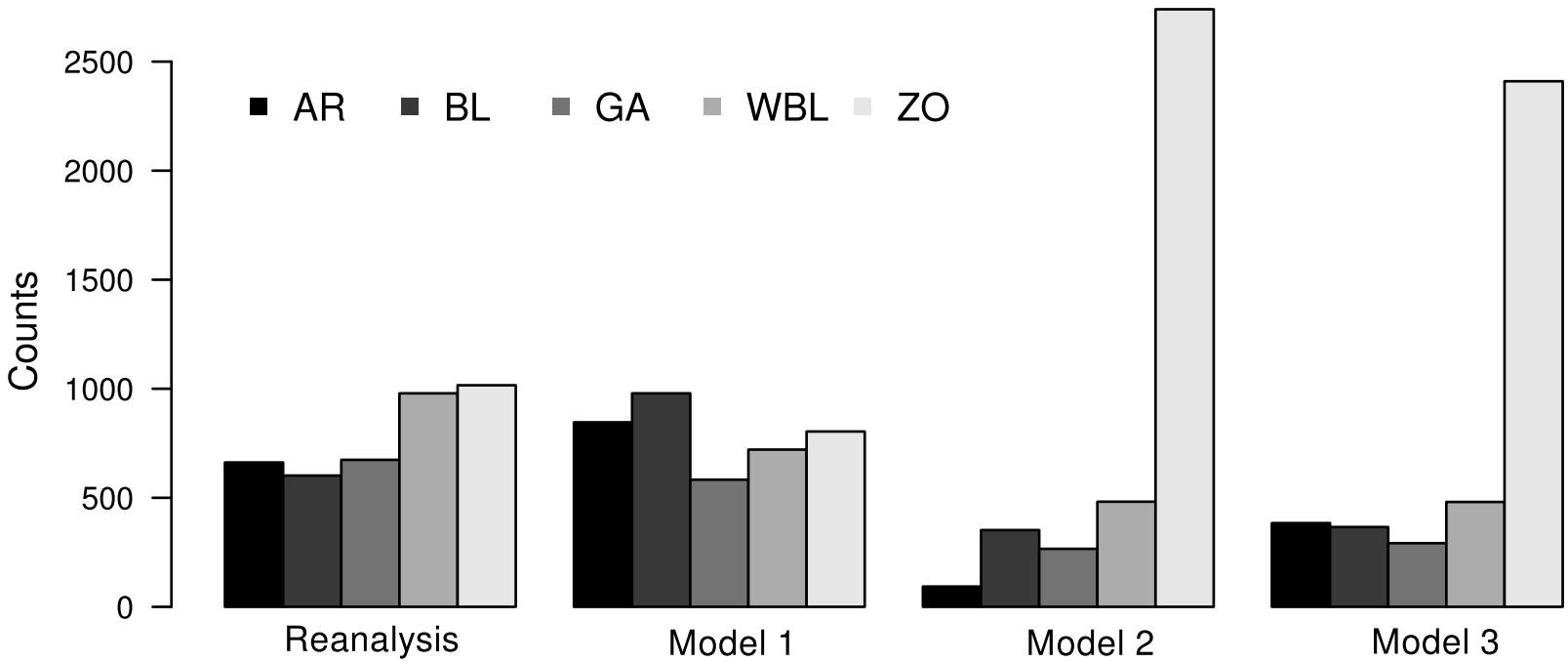
Comparing circulation patterns

- projecting onto reanalysis patterns
- comparing centroids

Motivation

Comparing circulation patterns

- projecting onto reanalysis patterns
- comparing centroids
- counting frequencies



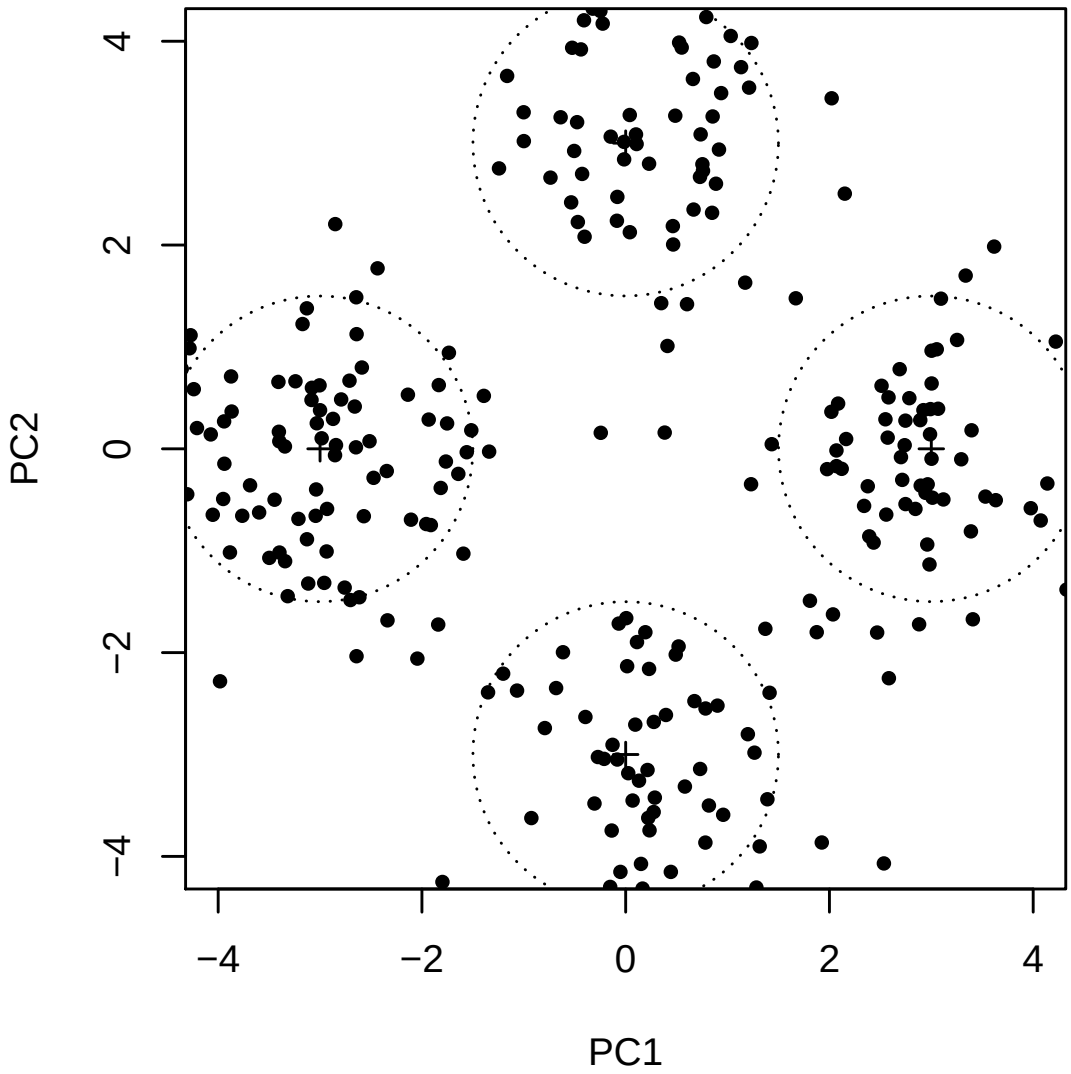
Motivation

Comparing circulation patterns

- projecting onto reanalysis patterns
- comparing centroids
- counting frequencies

other possibilities?

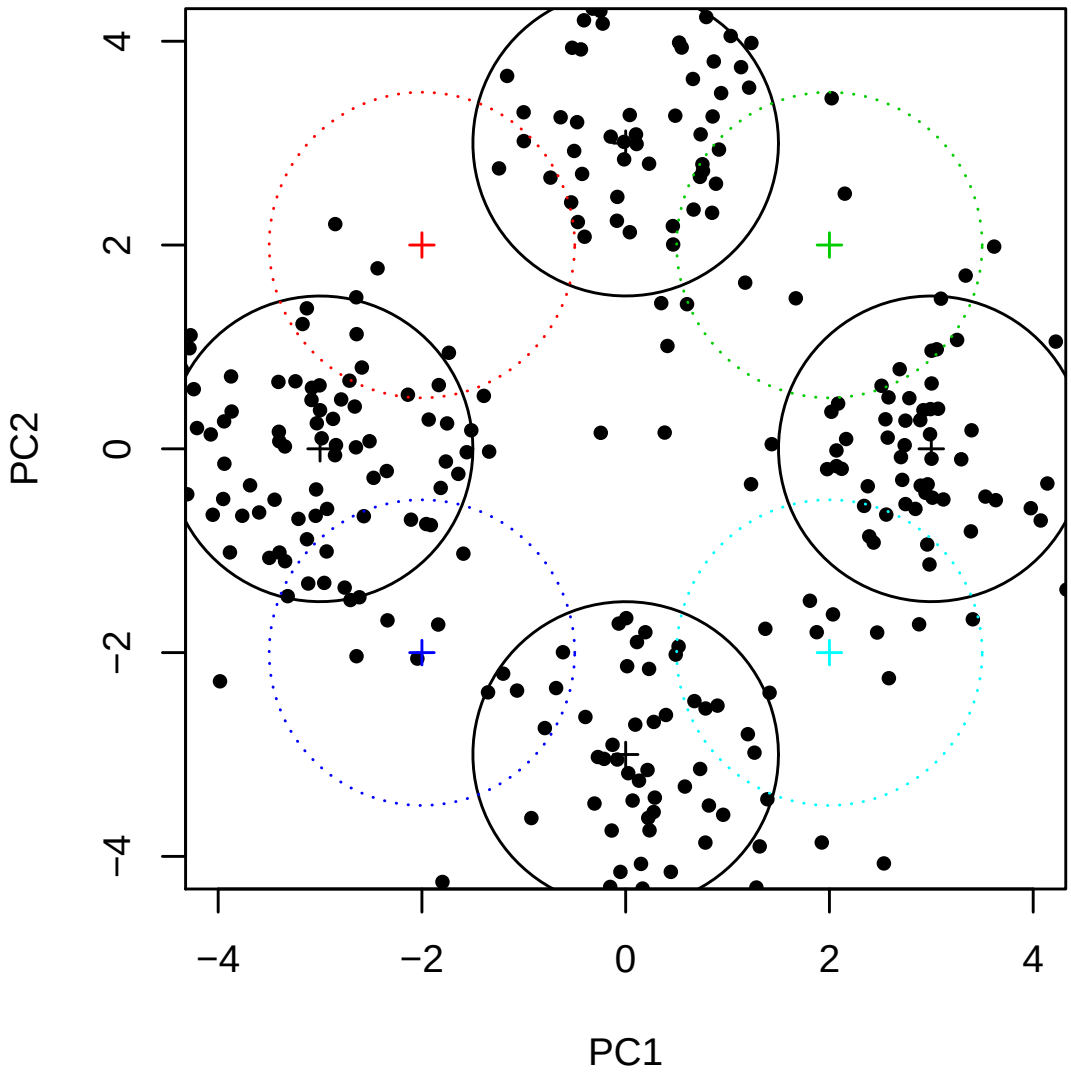
Motivation



Procedure

- cluster separately

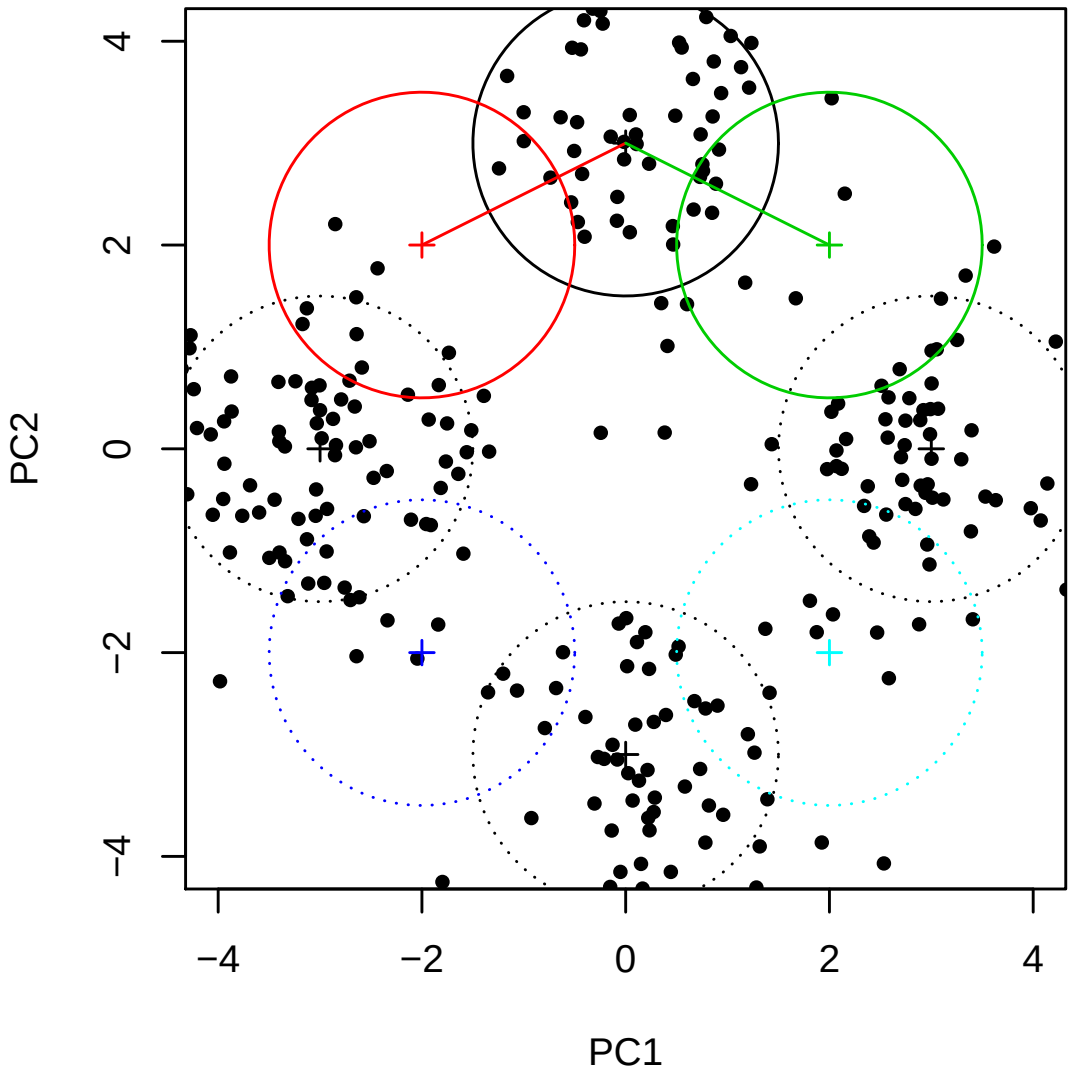
Motivation



Procedure

- cluster separately
- compare clusters

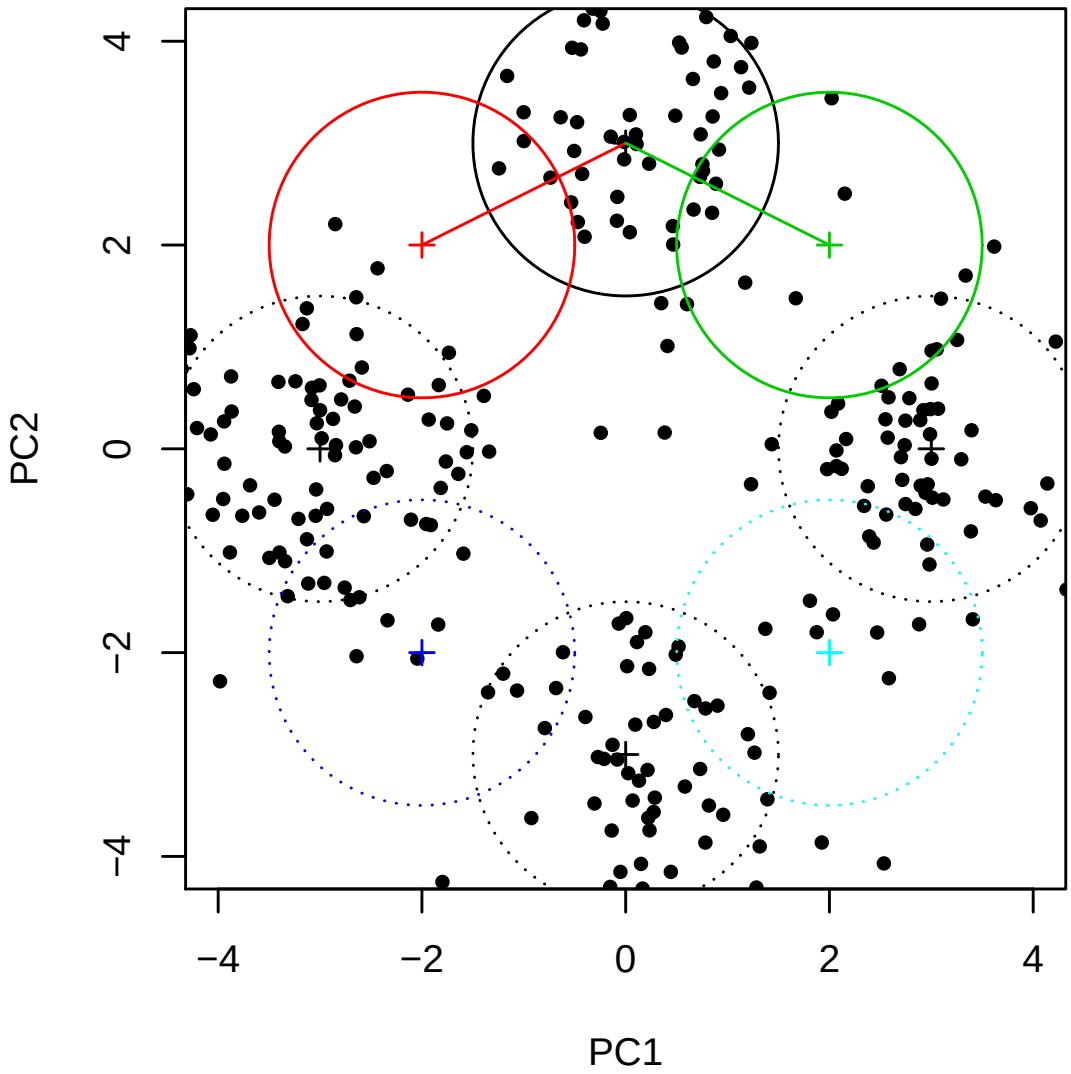
Motivation



Procedure

- cluster separately
- compare clusters
- e.g. means

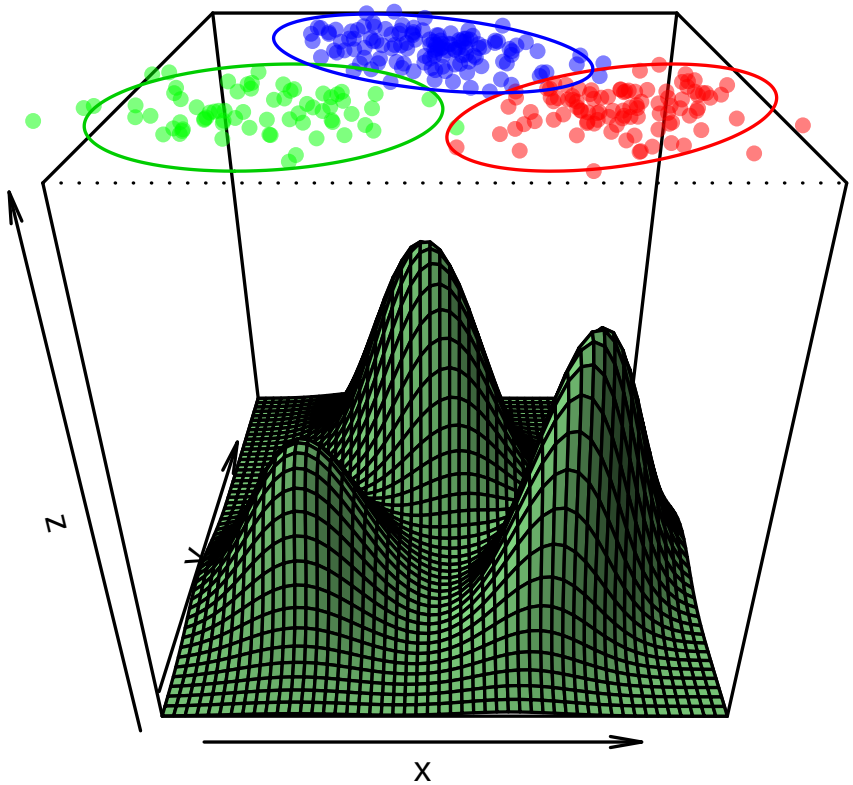
Motivation



Procedure

- cluster separately
- compare clusters
- e.g. means
- allows for other ways

Cluster analysis: Gaussian mixture models

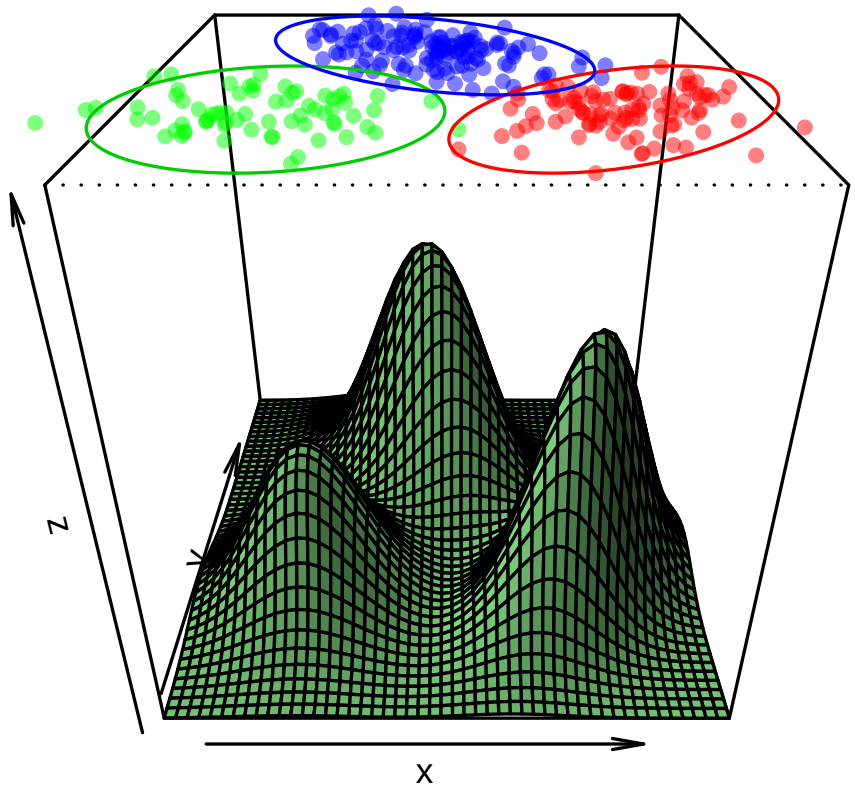


Modelling

- mixture of Gaussians

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

Cluster analysis: Gaussian mixture models



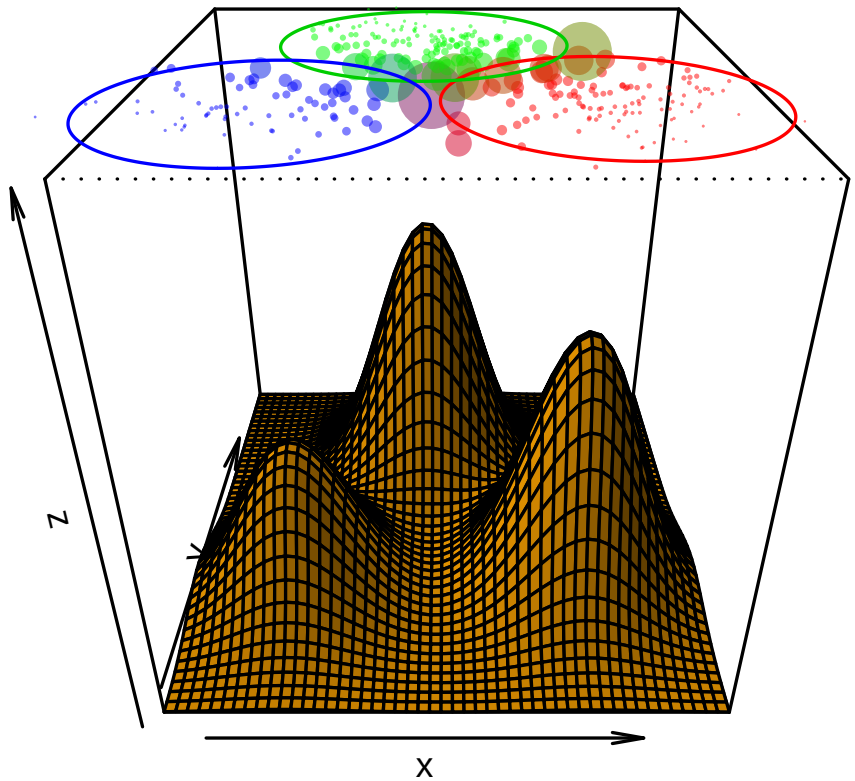
Modelling

- mixture of Gaussians
- parameter estimation

Parameter estimates (EM):

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

Cluster analysis: Gaussian mixture models



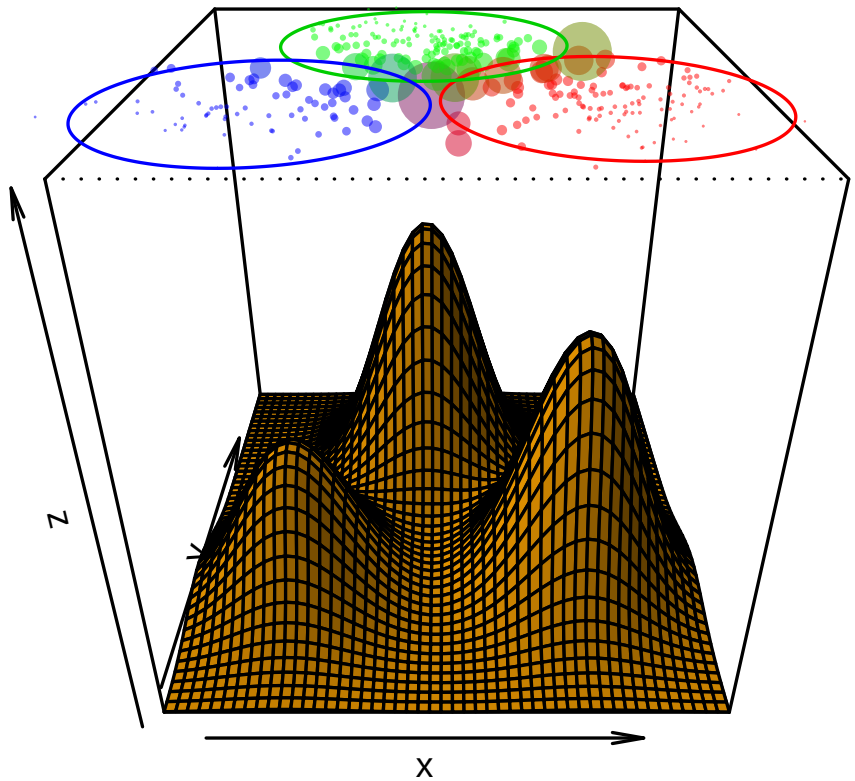
Modelling

- mixture of Gaussians
- parameter estimation
- classification with uncertainty

Classification uncertainty for $\mathbf{x_i}$:

$$\bar{P}(\mathbf{x}_i) = 1 - P(\mathbf{x_i} \in \mathrm{Cl}_{\mathbf{x_i}} \mid \mathbf{x_i})$$

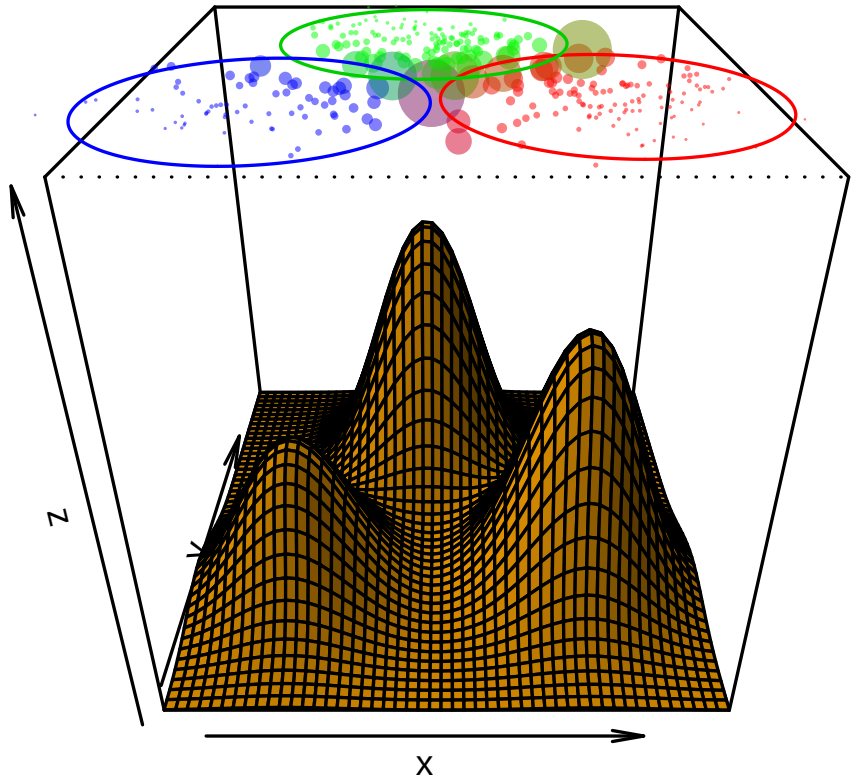
Cluster analysis: Gaussian mixture models



Modelling

- mixture of Gaussians
- parameter estimation
- classification with uncertainty
- ellipsoidal shapes

Cluster analysis: Gaussian mixture models

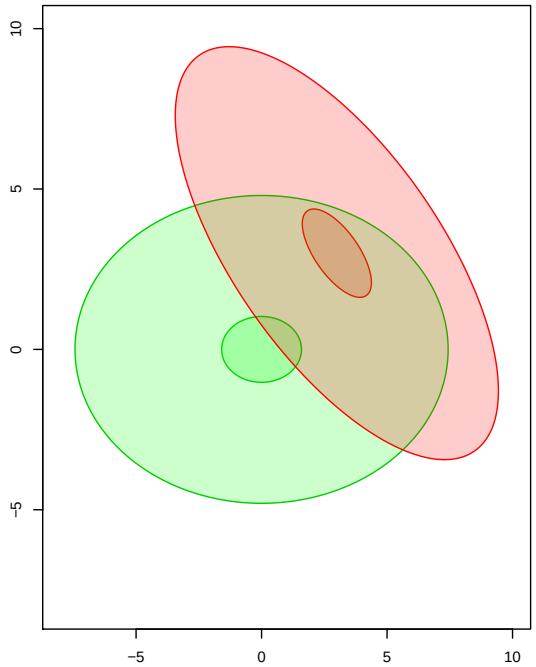


Modelling

- mixture of Gaussians
- parameter estimation

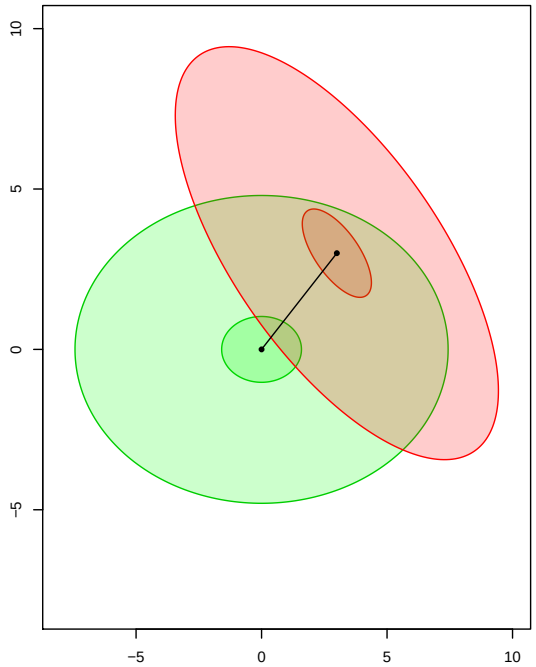
- classification with uncertainty
- ellipsoidal shapes
- other means of comparison

Difference/Similarity Measures for (Gaussian) pdfs



Difference/Similarity Measures for (Gaussian) pdfs

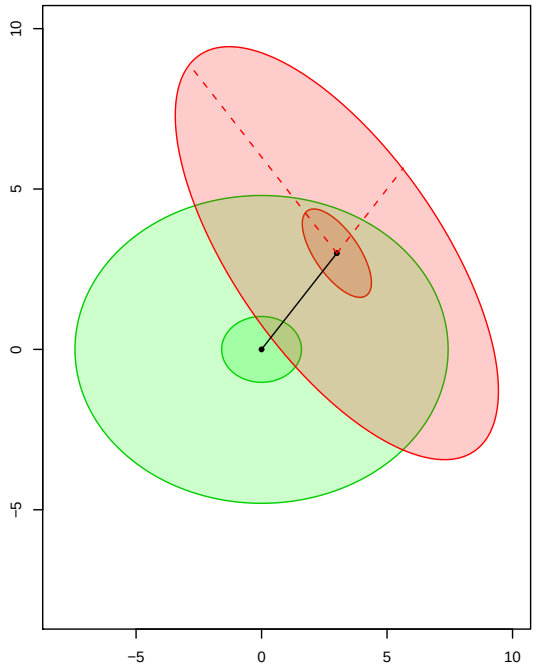
1. Euclidian distance:
   $d^2_{\mathrm{Eucl}}(P, Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2$



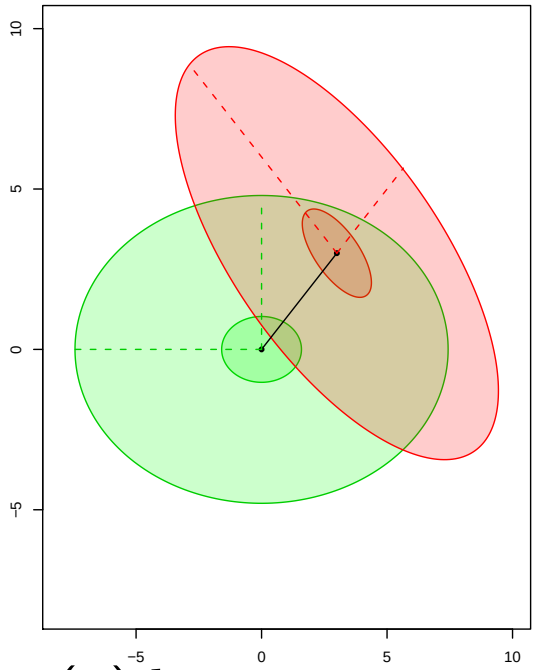
Difference/Similarity Measures for (Gaussian) pdfs

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

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



Difference/Similarity Measures for (Gaussian) pdfs

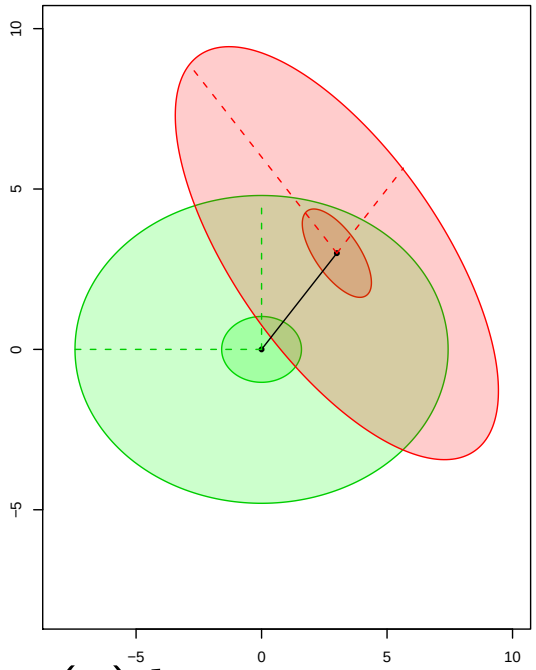


1. Euclidian distance:
   $d^2_{\mathrm{Eucl}}(P, Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2$

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

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

Difference/Similarity Measures for (Gaussian) pdfs



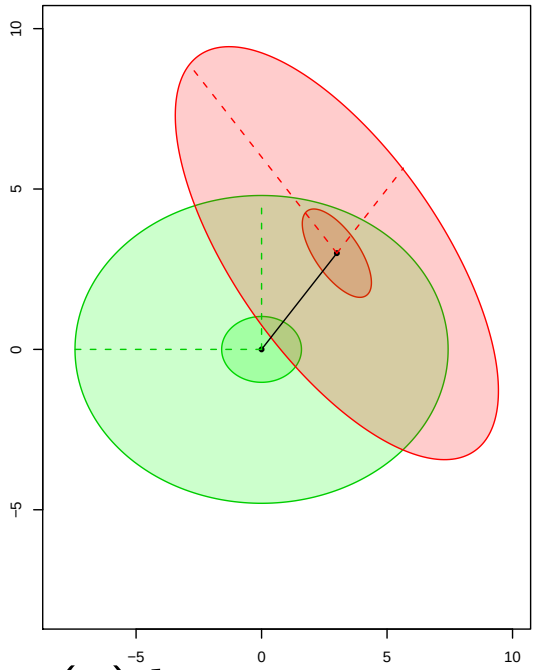
1. Euclidian distance:
   $d^2_{\mathrm{Eucl}}(P, Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2$

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

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

Difference/Similarity Measures for (Gaussian) pdfs



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

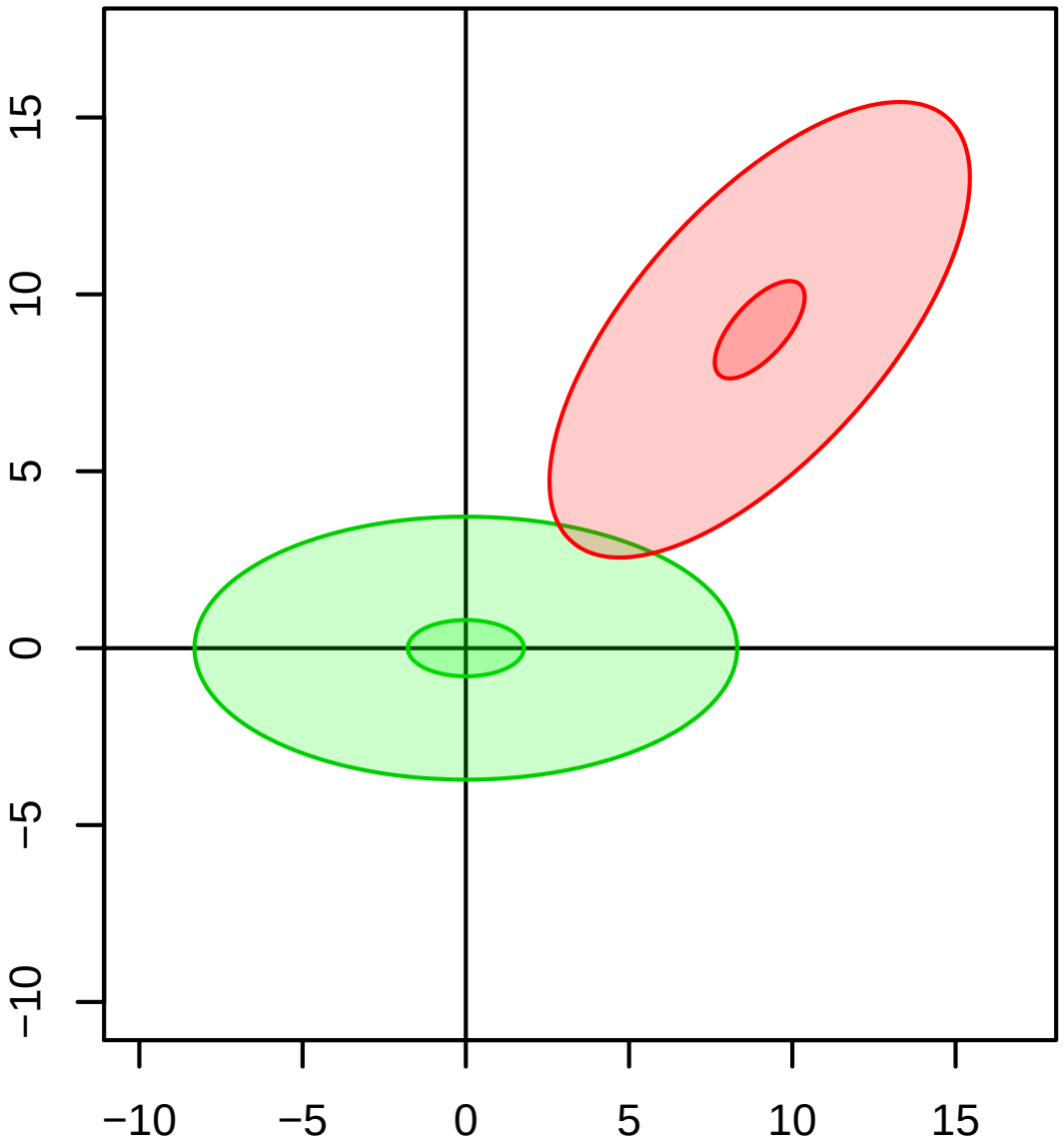
2. Mahalanobis distance:
   $d_{\mathrm{Mah}}(P,Q) = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\|^2_{\boldsymbol{\Sigma}_p^{-1}}$

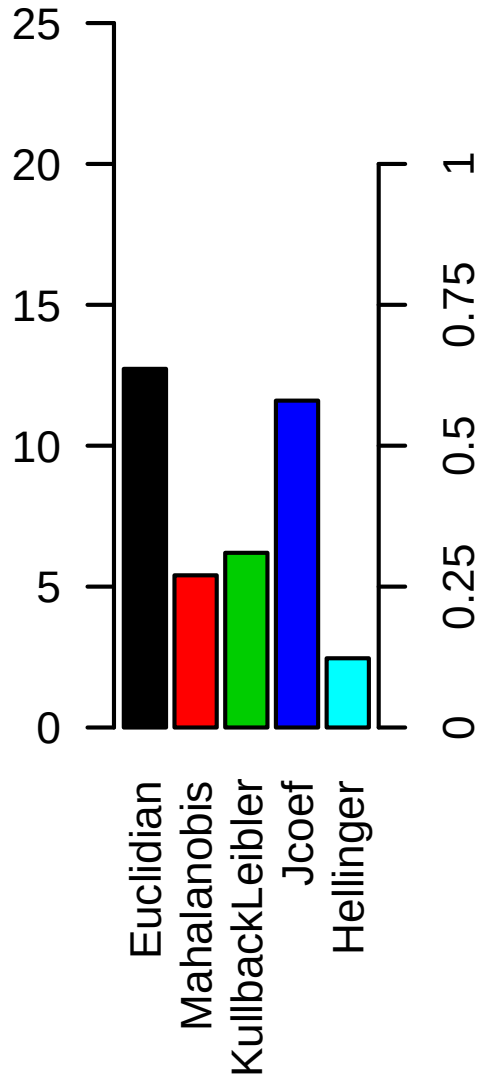
3. Kullback-Leibler discrimination (KL):
   $d_{\mathrm{KL}}(P,Q) = I(P \mid Q) = \int_{\mathbb{R}} \log\left(\frac{q(x)}{p(x)}\right) q(z)\mathrm{d}x$

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

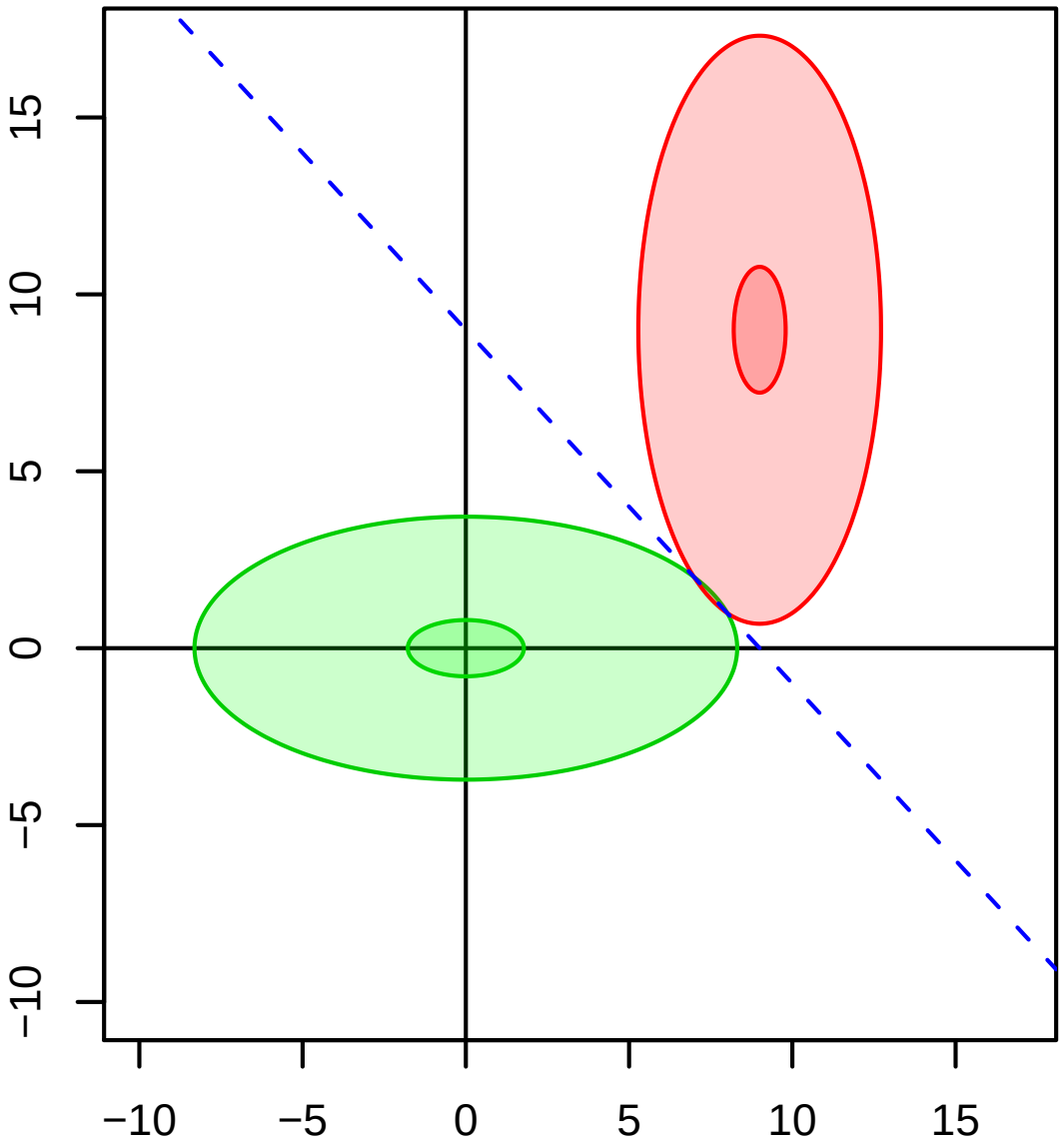
5. Hellinger coefficient (s=1/2):
   $d_{\mathrm{H}}(P,Q) = \int_{\mathbb{R}} q(x)^s p(x)^{(1-s)}\mathrm{d}x, \quad d_{\mathrm{H}} \in [0,1]$

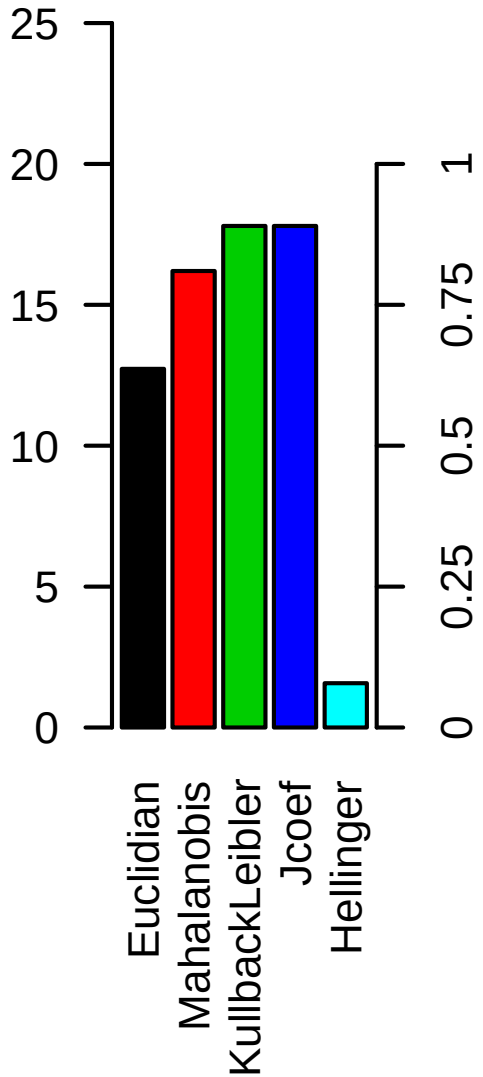
Difference and similarity measures for pdfs



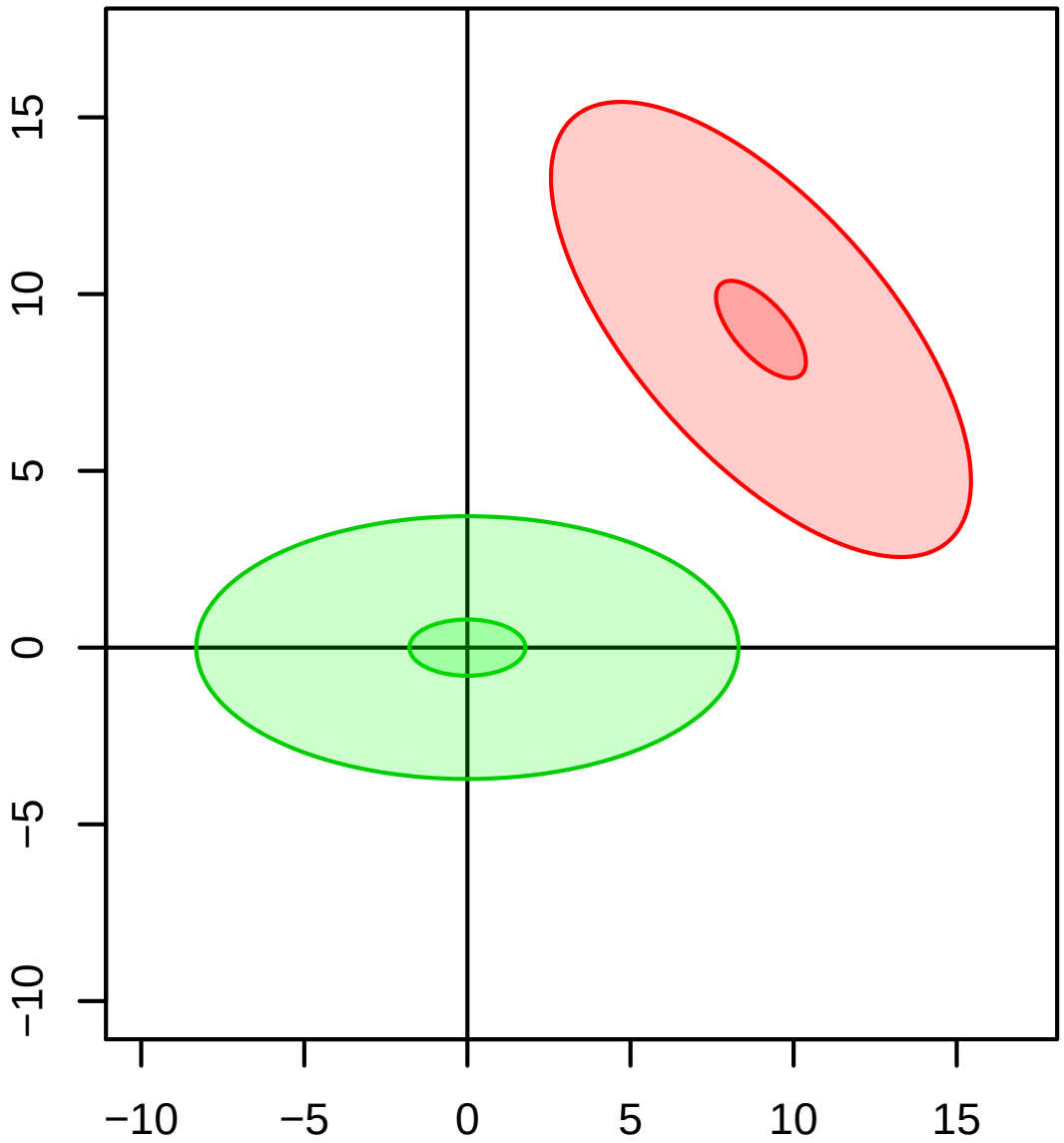


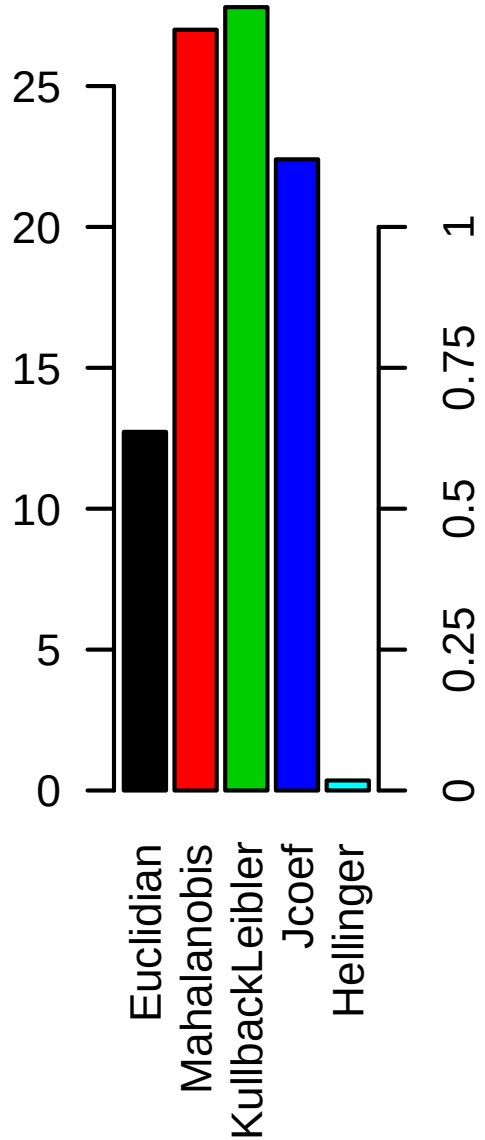
Difference and similarity measures for pdfs



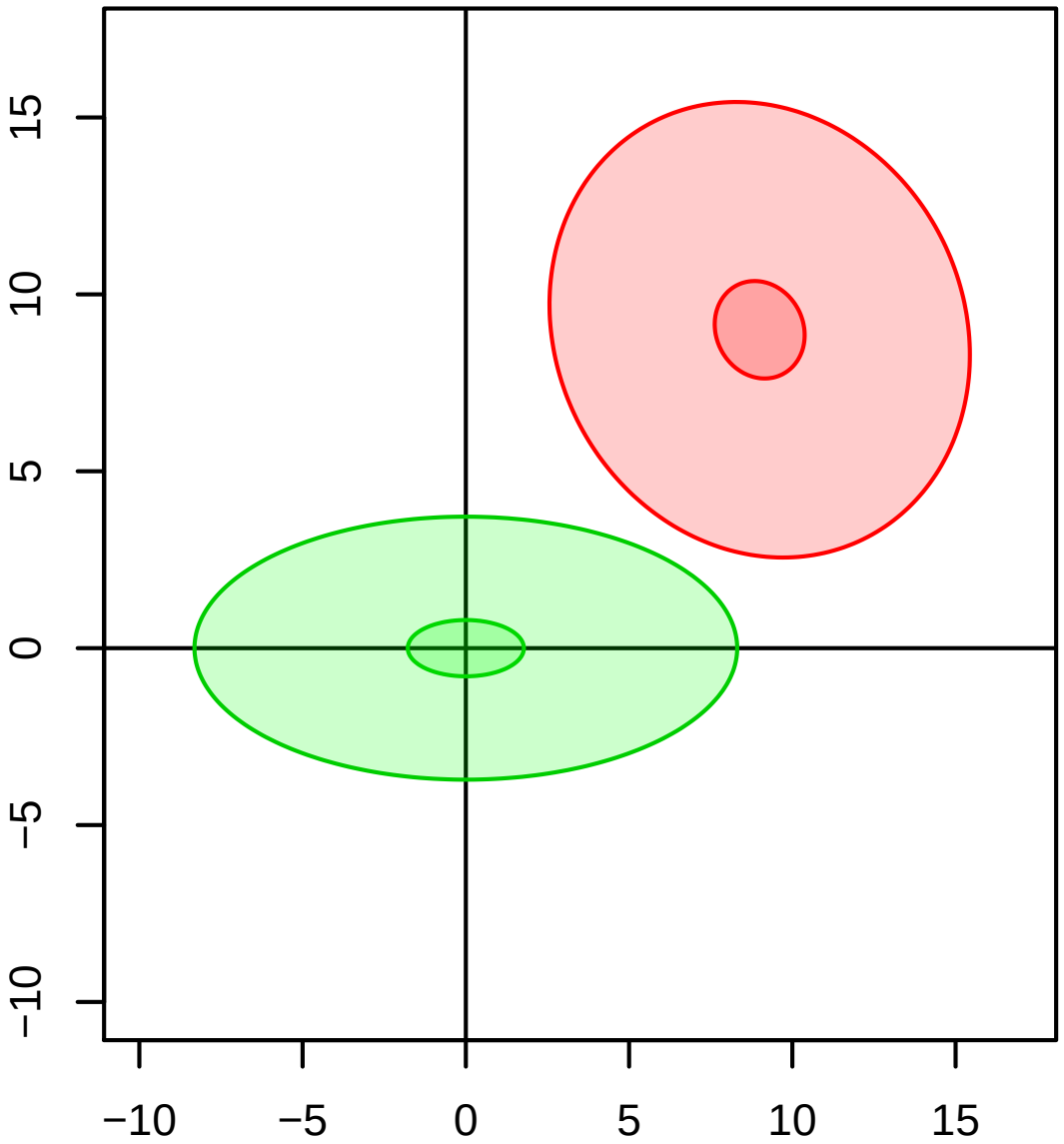


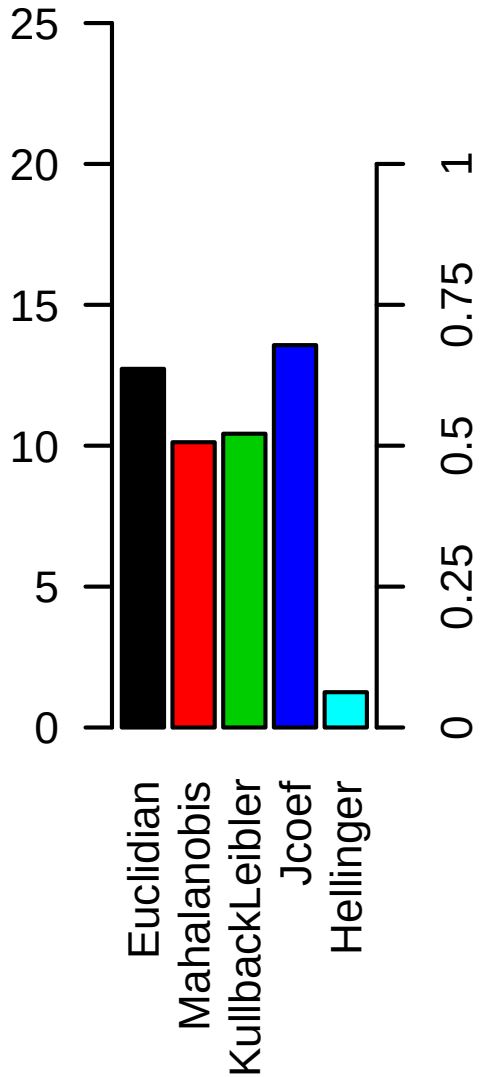
Difference and similarity measures for pdfs



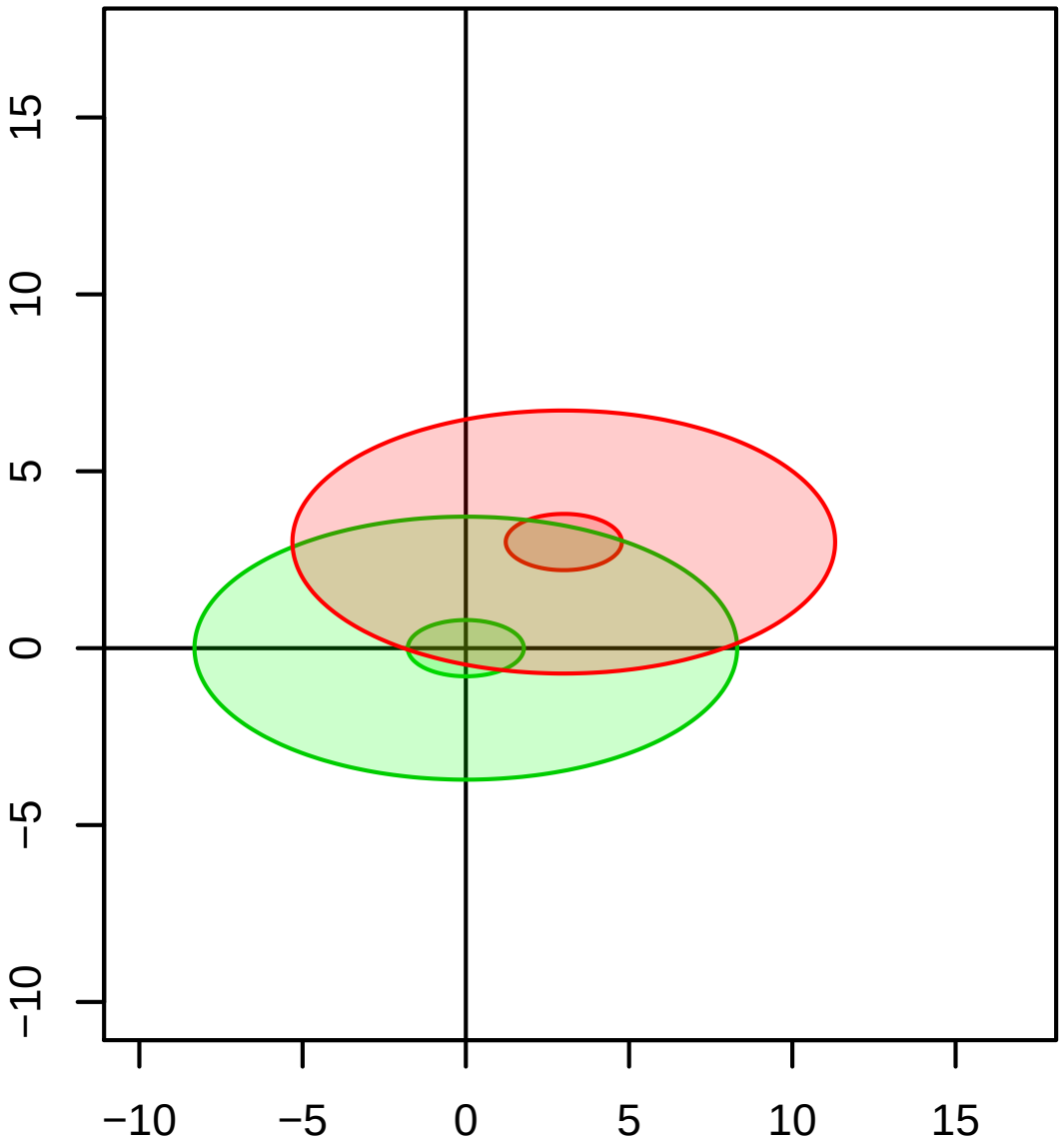


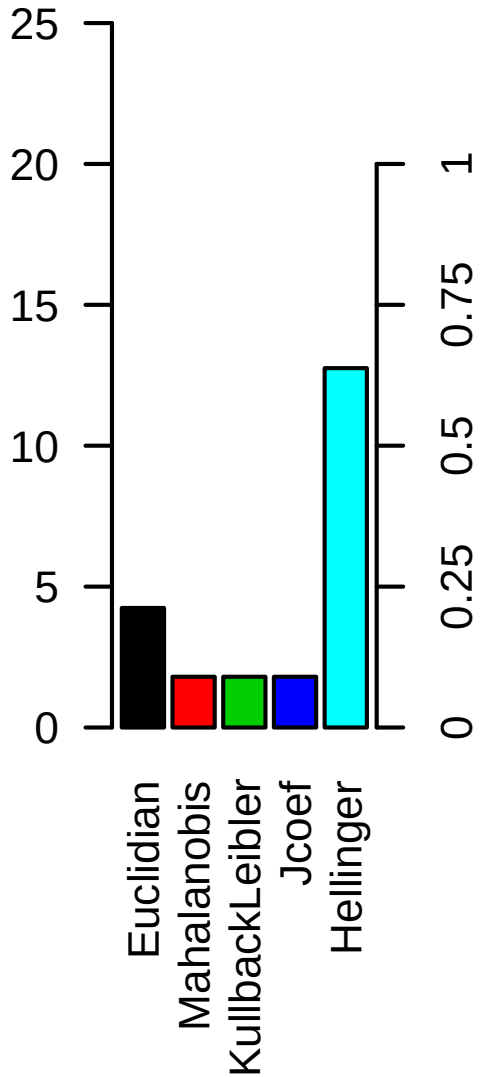
Difference and similarity measures for pdfs



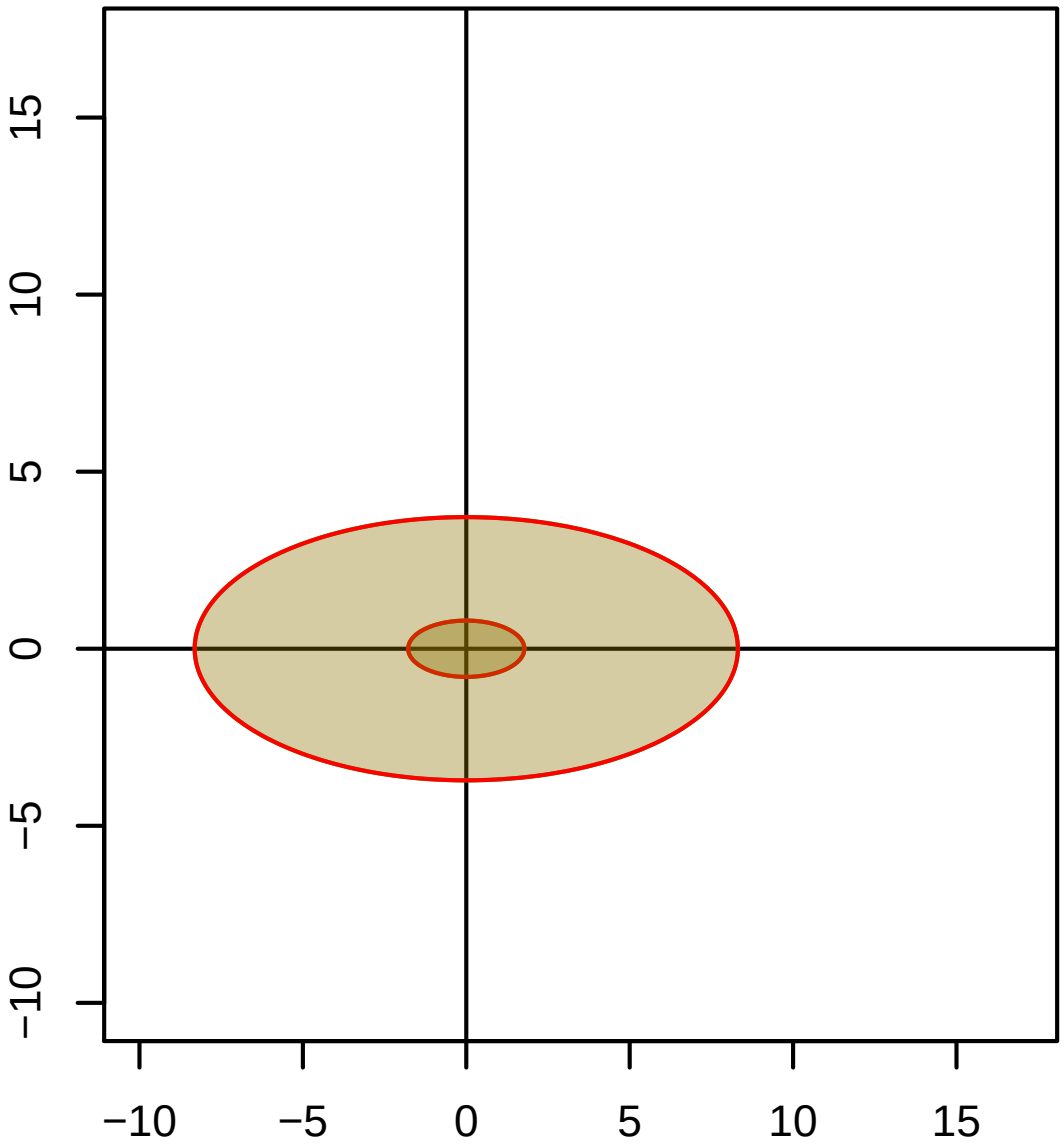


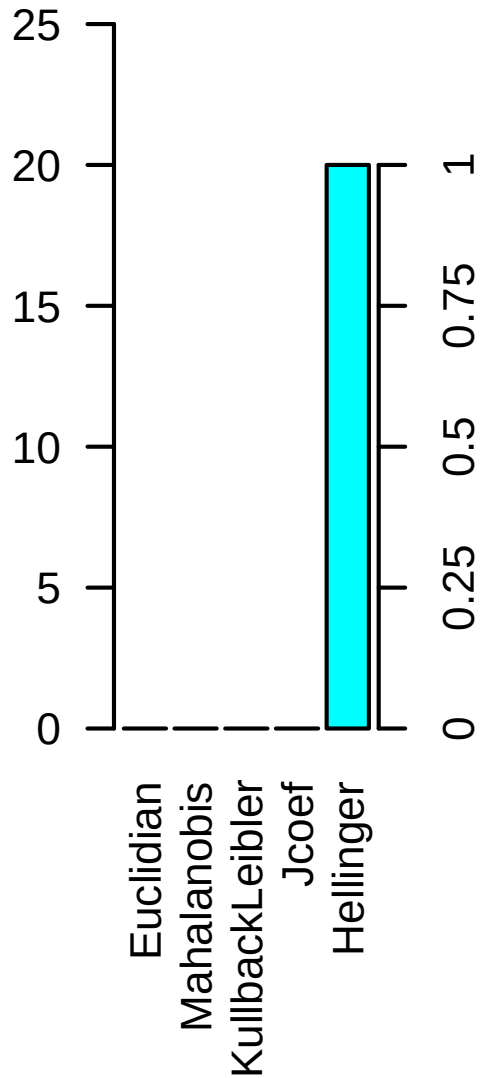
Difference and similarity measures for pdfs



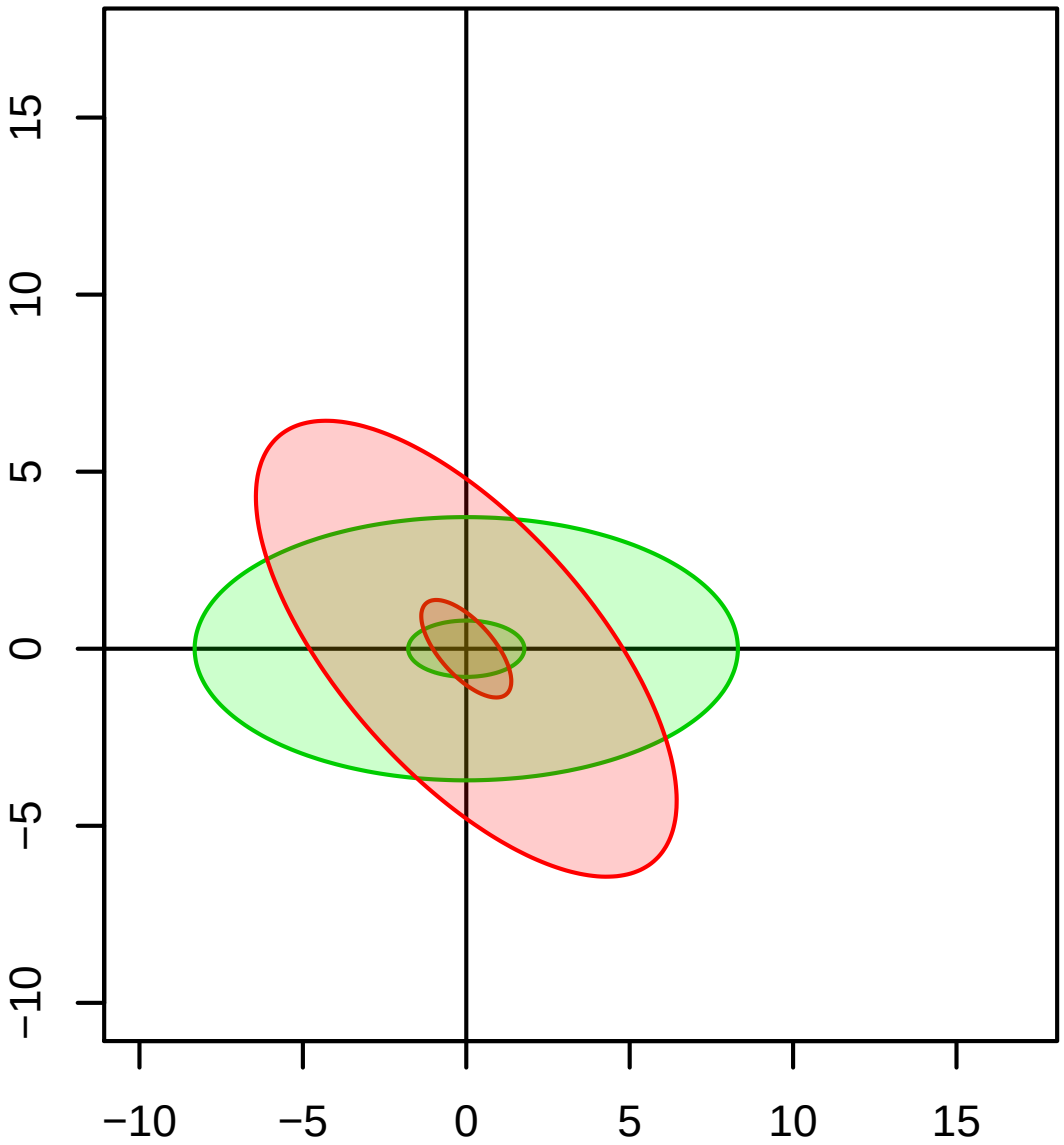


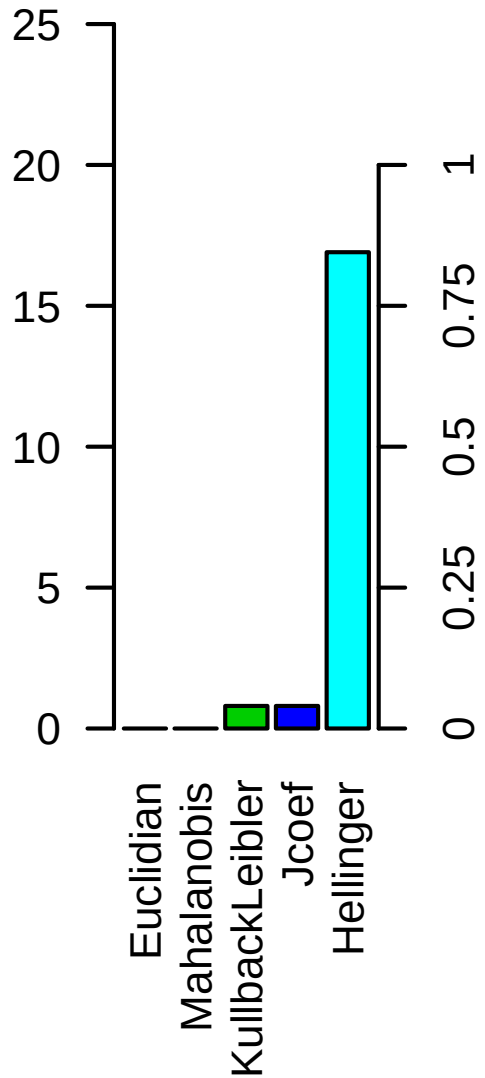
Difference and similarity measures for pdfs





Difference and similarity measures for pdfs

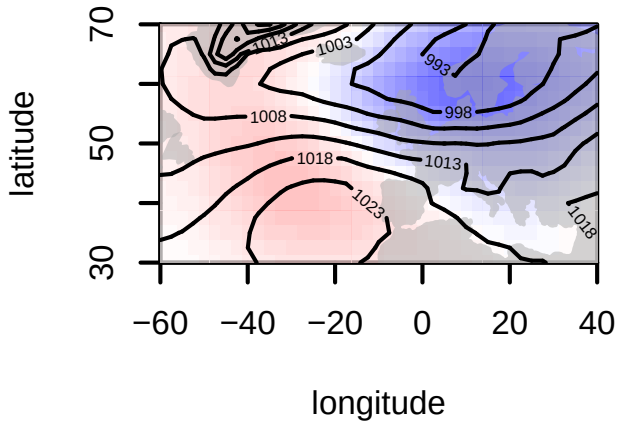




North Atlantic region

Setting

- North Atlantic region
- daily SLP anomalies
- 1975 – 2000, NDJFM



North Atlantic region

Setting

- North Atlantic region
- daily SLP anomalies
- 1975 – 2000, NDJFM
- data sets, interpolated to NCEP/NCAR grid

Reanalyses

- NCEP/NCAR
- ERA-40

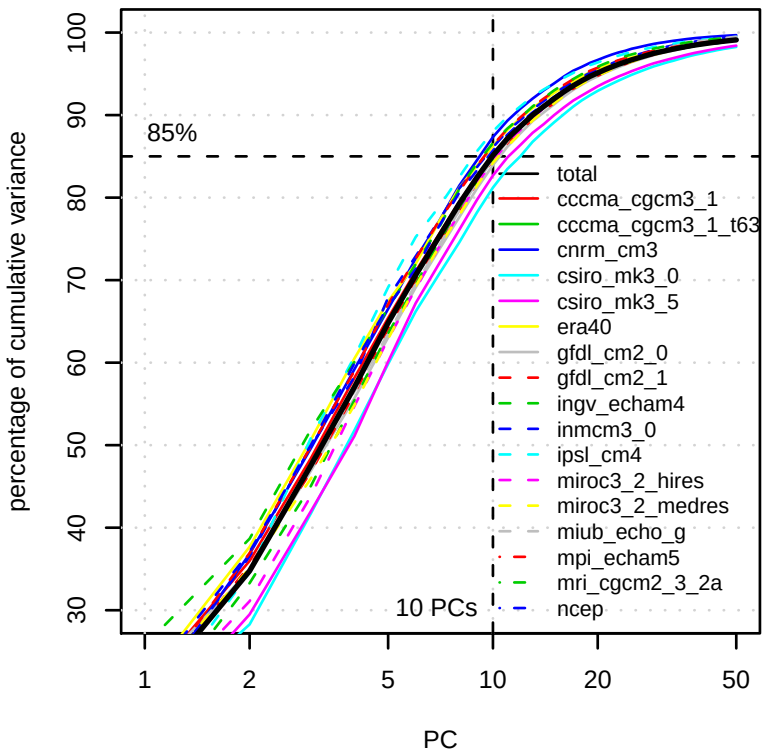
14 IPCC Models (AR4)

- 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

North Atlantic region

Setting

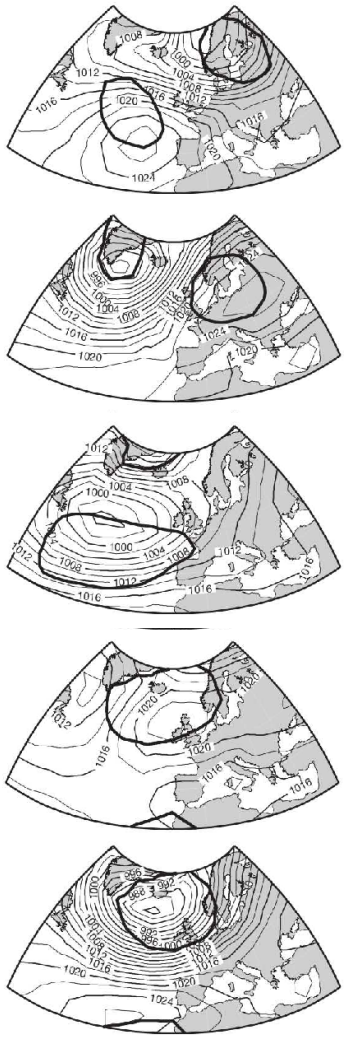
- North Atlantic region
- daily SLP anomalies
- 1975 – 2000, NDJFM
- data sets, interpolated to NCEP/NCAR grid
- common PCA 85% (10PCs)



North Atlantic region

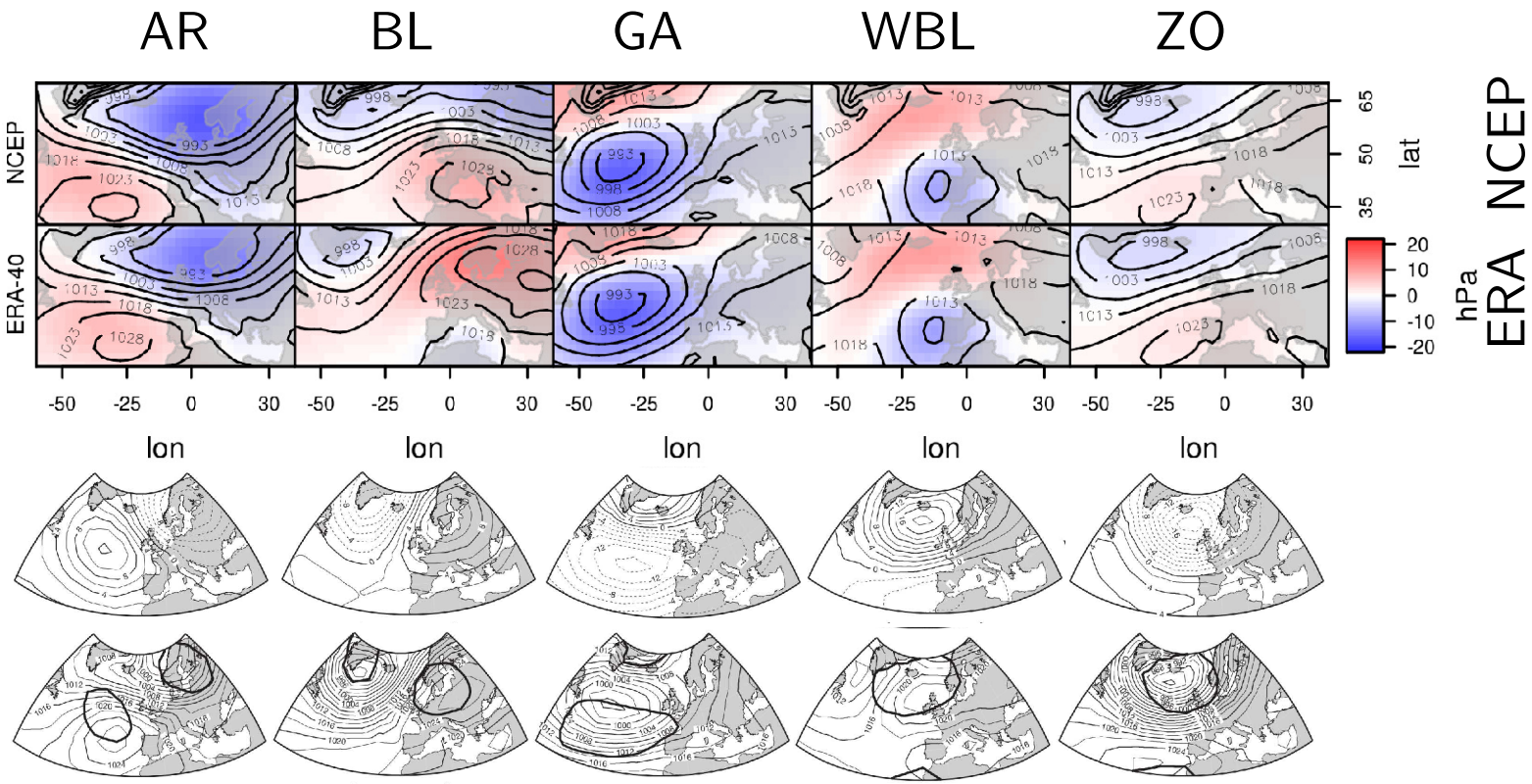
Setting

- North Atlantic region
- daily SLP anomalies
- 1975 – 2000, NDJFM
- data sets, interpolated to NCEP/NCAR grid
- common PCA 85% (10PCs)
- Clusteranalysis use 5 Clusters for comparison



Plaut, Simonnet (2001)

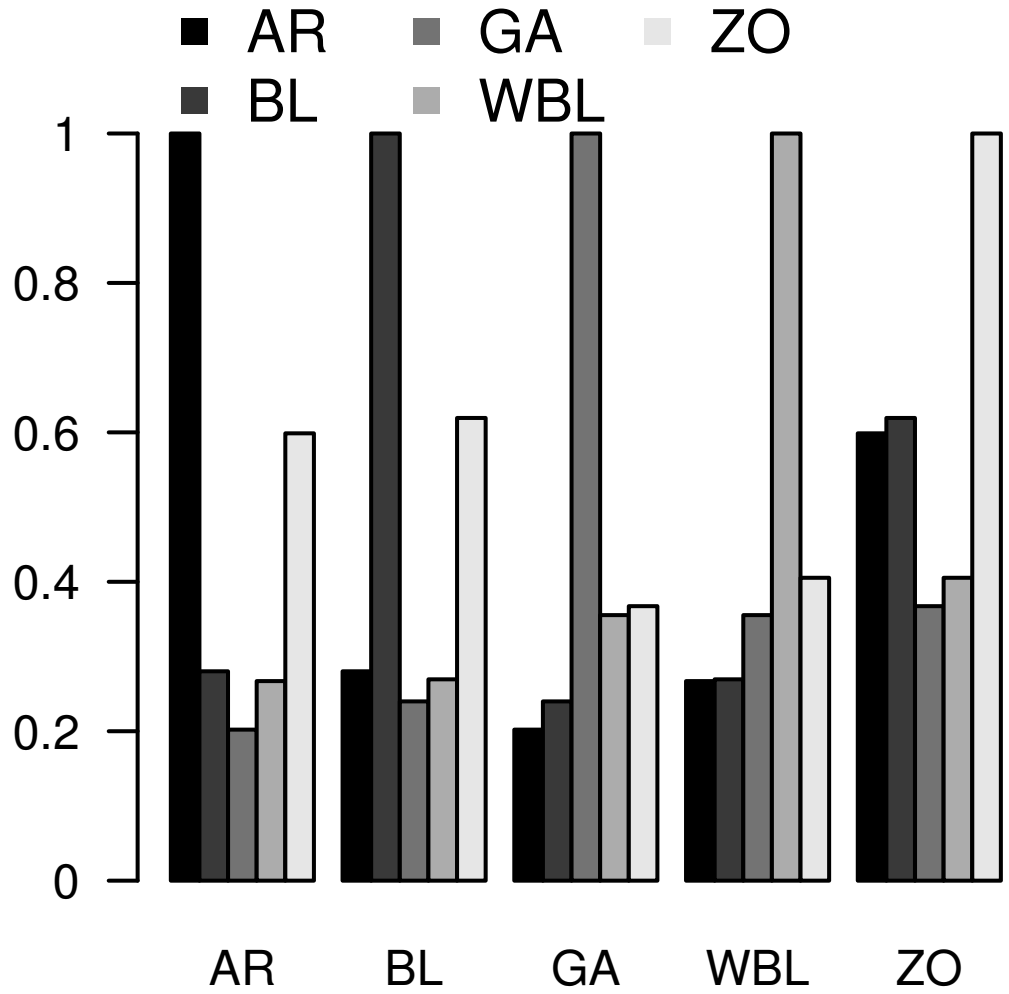
Comparison with PS01, centroids



from Plaut and Simonnet, *Clim.Res.* (2001)

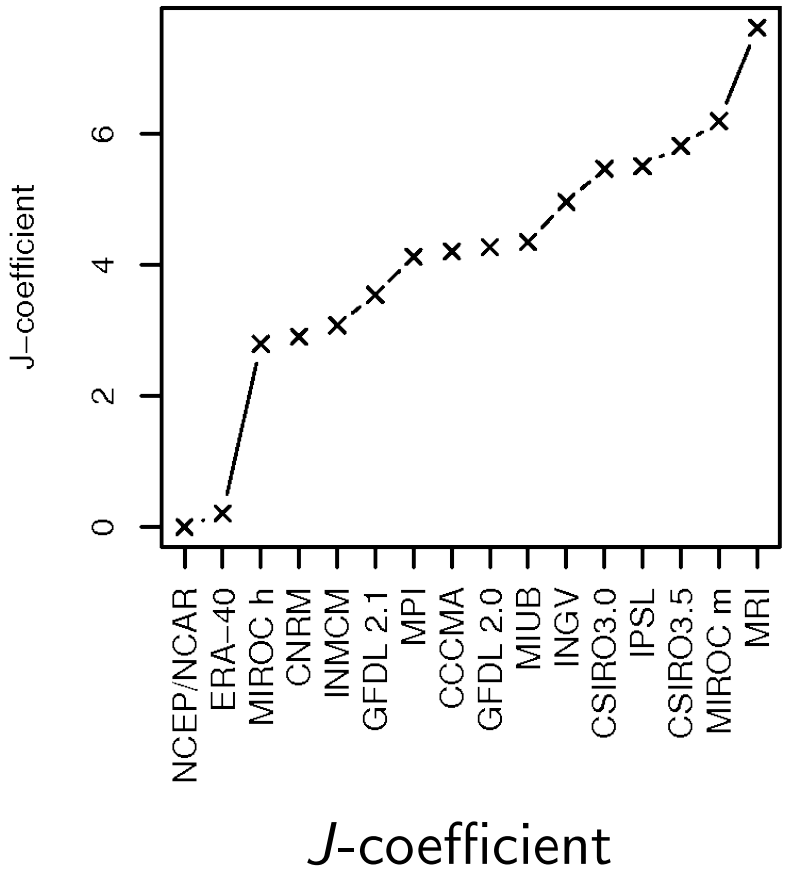
Results

Cluster configuration NCEP/NCAR (Hellinger)



Results

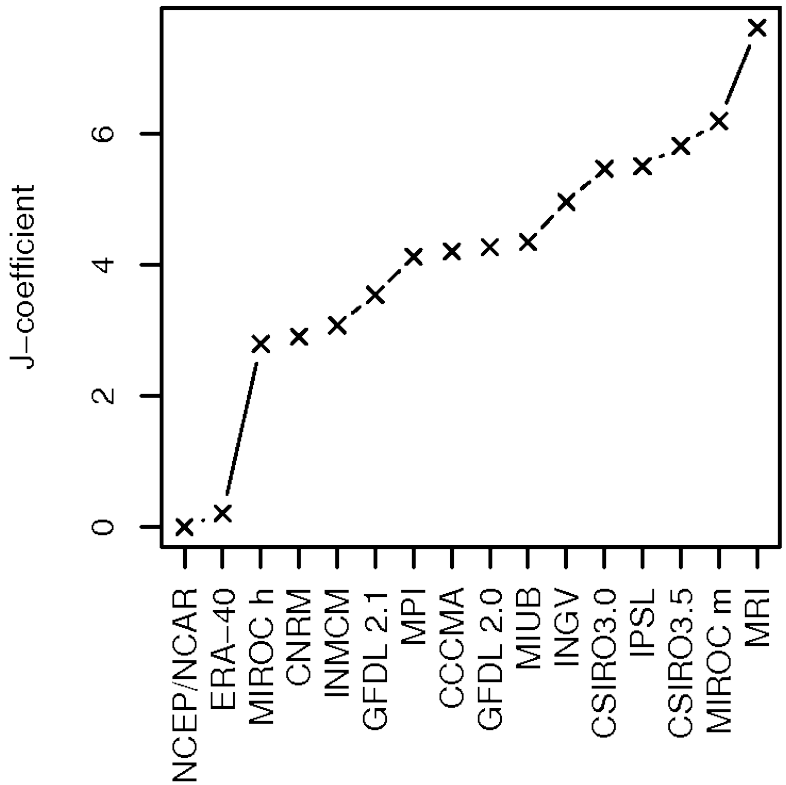
GA: many GCMs with NCEP/NCAR



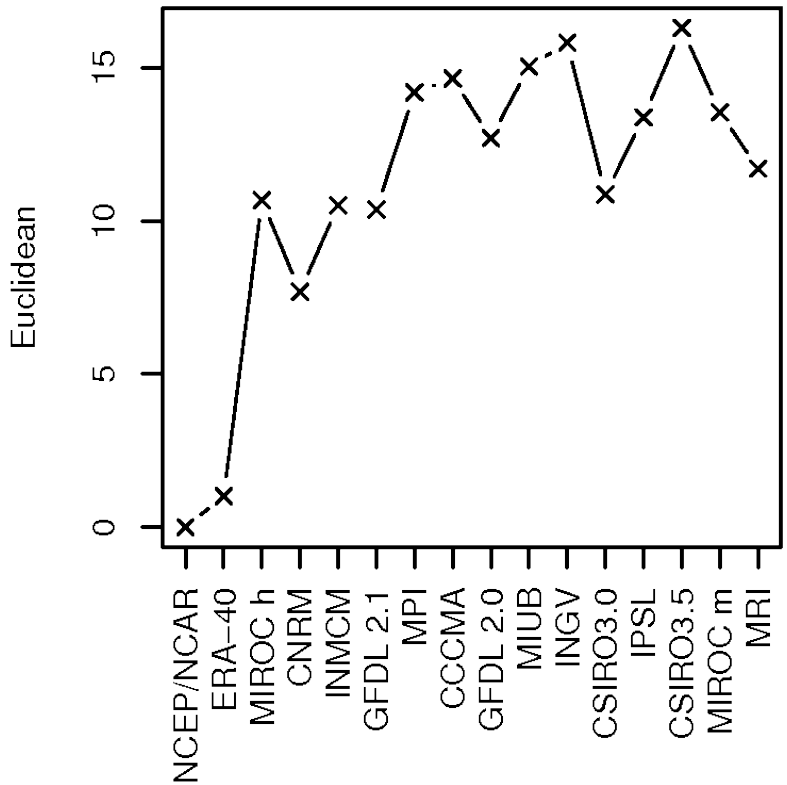
Ellipsoidal GCM

Results

GA: many GCMs with NCEP/NCAR



J-coefficient



Euclidian distance

Ellipsoidal GCM

Summary

- Gaussian mixture model clustering
  - non-spherical clusters
  - pdf-based

Summary

- Gaussian mixture model clustering
  - non-spherical clusters
  - pdf-based
- quantitative cluster differences
  - explore state space
  - means, variances (size, shape)
  - compare many models

- Mahalanobis
- Kullback-Leibler
- *J*-coefficient
- Hellinger coeff.

Summary

- Gaussian mixture model clustering
  - non-spherical clusters
  - pdf-based
- quantitative cluster differences
  - explore state space
  - means, variances (size, shape)
  - compare many models
- North Atlantic region
  - ZO overlap with AR/BL, GA isolated
  - MIROC h, smallest distance (difference is CP dependent)
  - GA: J-coeff $\neq$ Euclidian

- Mahalanobis
- Kullback-Leibler
- J-coefficient
- Hellinger coeff.

References

- H. W. Rust et al.
  Quantifying differences in circulation patterns based on probabilistic models: AR4 multimodel comparison for the North Atlantic
  J.Climate 23:6573-6589 (2010)

- R-packages `gaussDiff` and `mclust` on CRAN
  `http://cran.r-project.org`