Using principal component analysis in empirical-statistical downscaling to emphasise synoptic scales

RE Benestad, A Mezghani, KM Parding, D Chen, L Fan

EMS2016-214: Trieste, 14 Sep 2016

Twitter: @RasmusBenestad

Meteorologisk institutt 150 år

What is PCA?

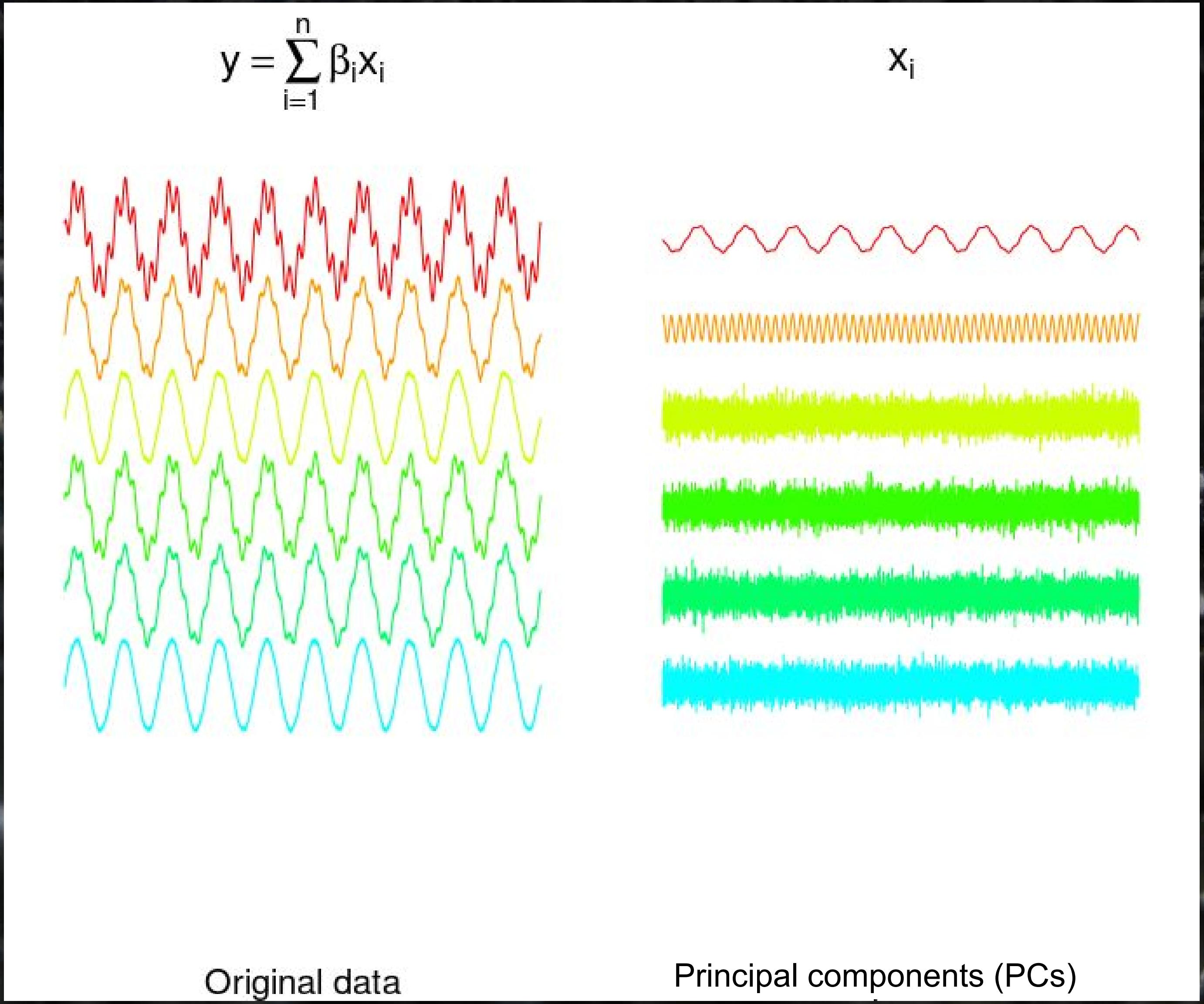
A way of organising the data according to information contents

Similar to

- Eigenvalues
- Empirical Orthogonal Functions (EOFs)
- Fourier series

Meteorologisk institutt 150 år

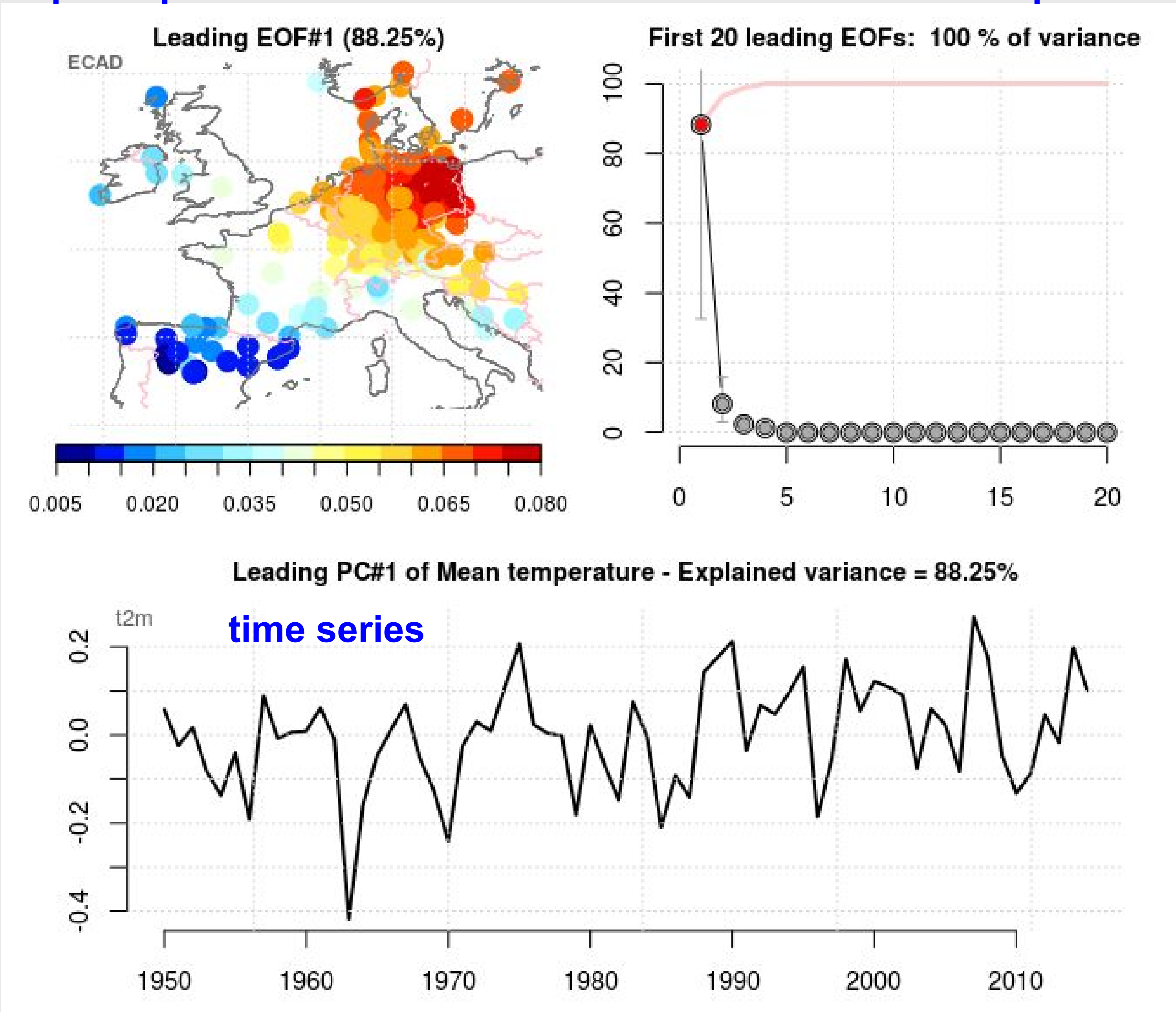
What is PCA?



$y = \sum_{i=1}^{n} \beta_i x_i$
x_i
Original data
Principal components (PCs)

spatial patterns

amplitudes



Why principal component analysis (PCA)?

Redundant information

Signal enhancement: emphasis large scales

Covariance preservation

Computation time

Orthogonality

How true are these reasons?

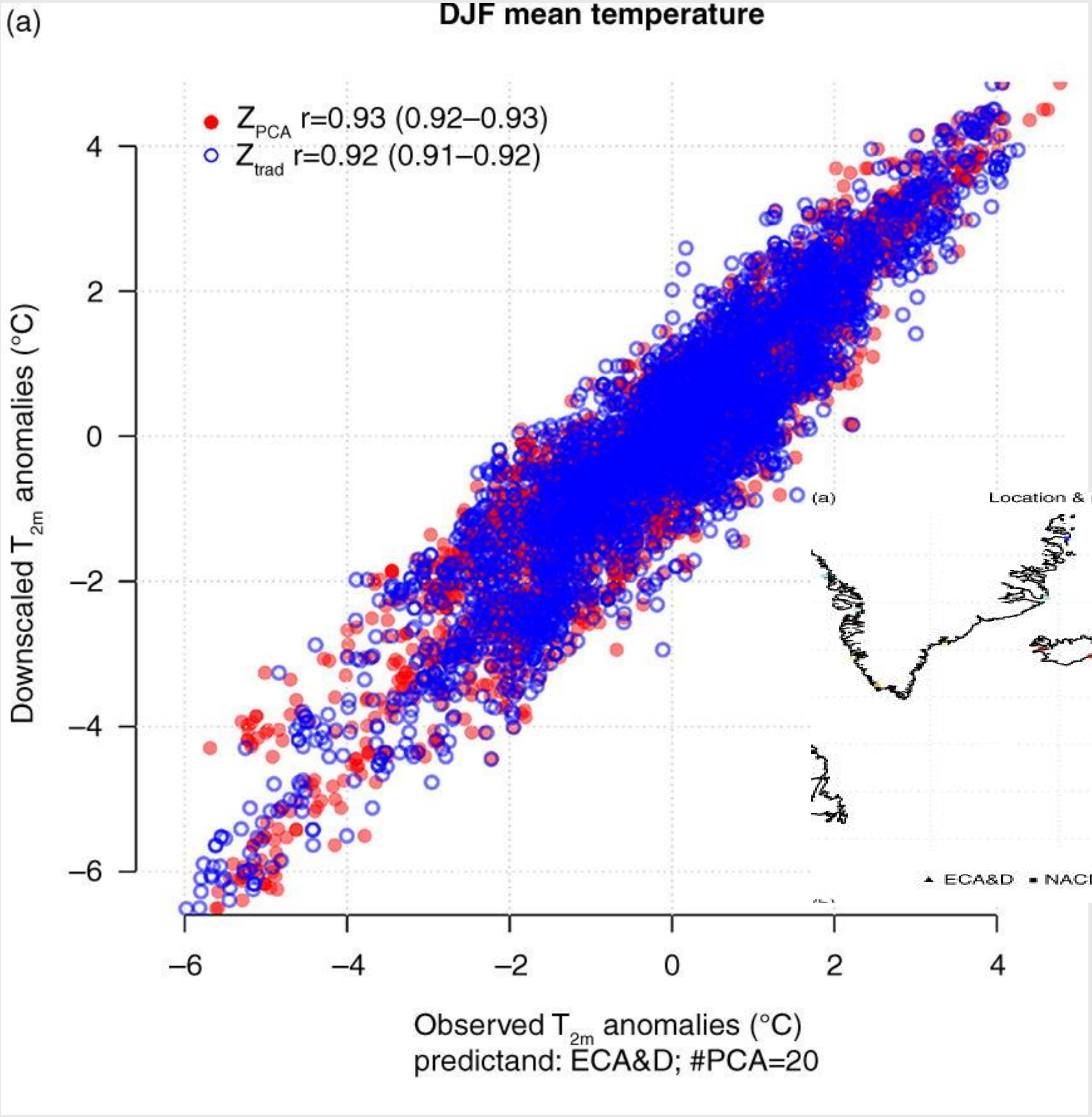
Computation speed

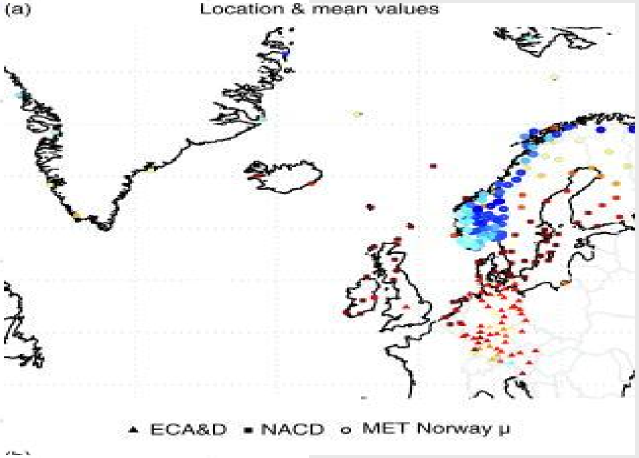
- Apply multiple regression to single principal components.
- Fast: 4 leading modes rather than ~400 stations
  - especially for large ensembles

Computation speed

- Apply multiple regression to single principal components.
- Fast: 4 leading modes rather than ~400 stations
  - especially for large ensembles
- Works for station groups and gridded data

Signal: Restructure data with information first





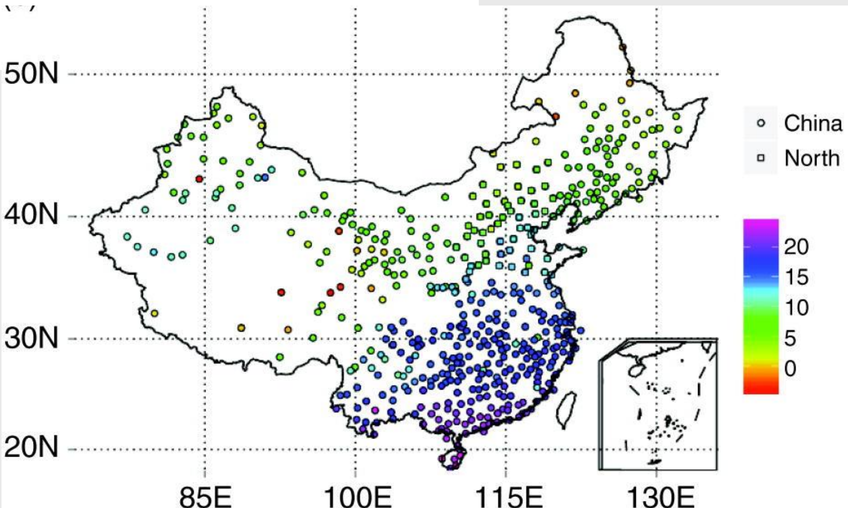
Tellus A 2015, **67**, 28326, http://dx.doi.org/10.3402/tellusa.v67.28326

Signal: Sensitivity to the choice of large-scale domain

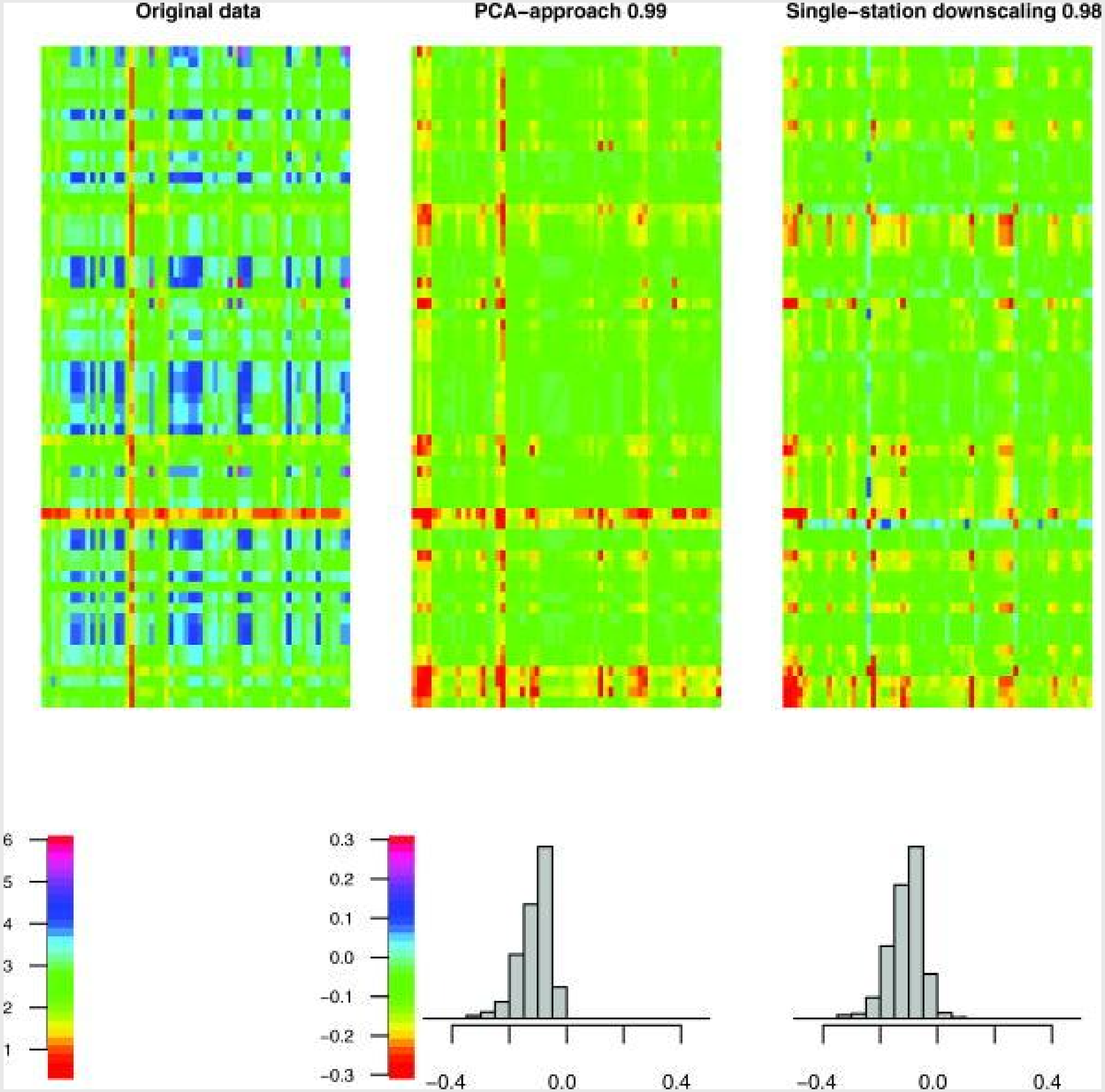
Table 2. Skill scores for the PCA-based and traditional ESD for China

Area (km^2)	27 800 000	31 800 000	35 800 000	34 400 000	39 300 000	44 200 000
DJF						
PCA	**0.87**	**0.86**	**0.87**	**0.87**	**0.86**	**0.86**
trad.	0.83	0.82	0.82	0.84	0.83	0.82
JJA						
PCA	**0.71**	0.69	**0.72**	**0.73**	**0.70**	**0.70**
trad.	0.70	**0.71**	0.70	0.68	0.65	0.64
N_{PCA}	20	20	20	20	20	20
Domain						
W (°E)	70.00	65.00	60.00	70.00	65.00	60.00
E (°E)	140.00	145.00	150.00	140.00	145.00	150.00
S (°N)	15.00	15.00	15.00	10.00	10.00	10.00
N (°N)	55.00	55.00	55.00	60.00	60.00	60.00
N_{PCA}	20	20	20	20	20	20

The entries are correlation scores from the cross-validation. The predictor was the NCEP/NCAR reanalysis and predictands included 499 stations with complete time series over the period 1961–2012. The highest score in each of the comparisons is shown in bold.



Covariance structure

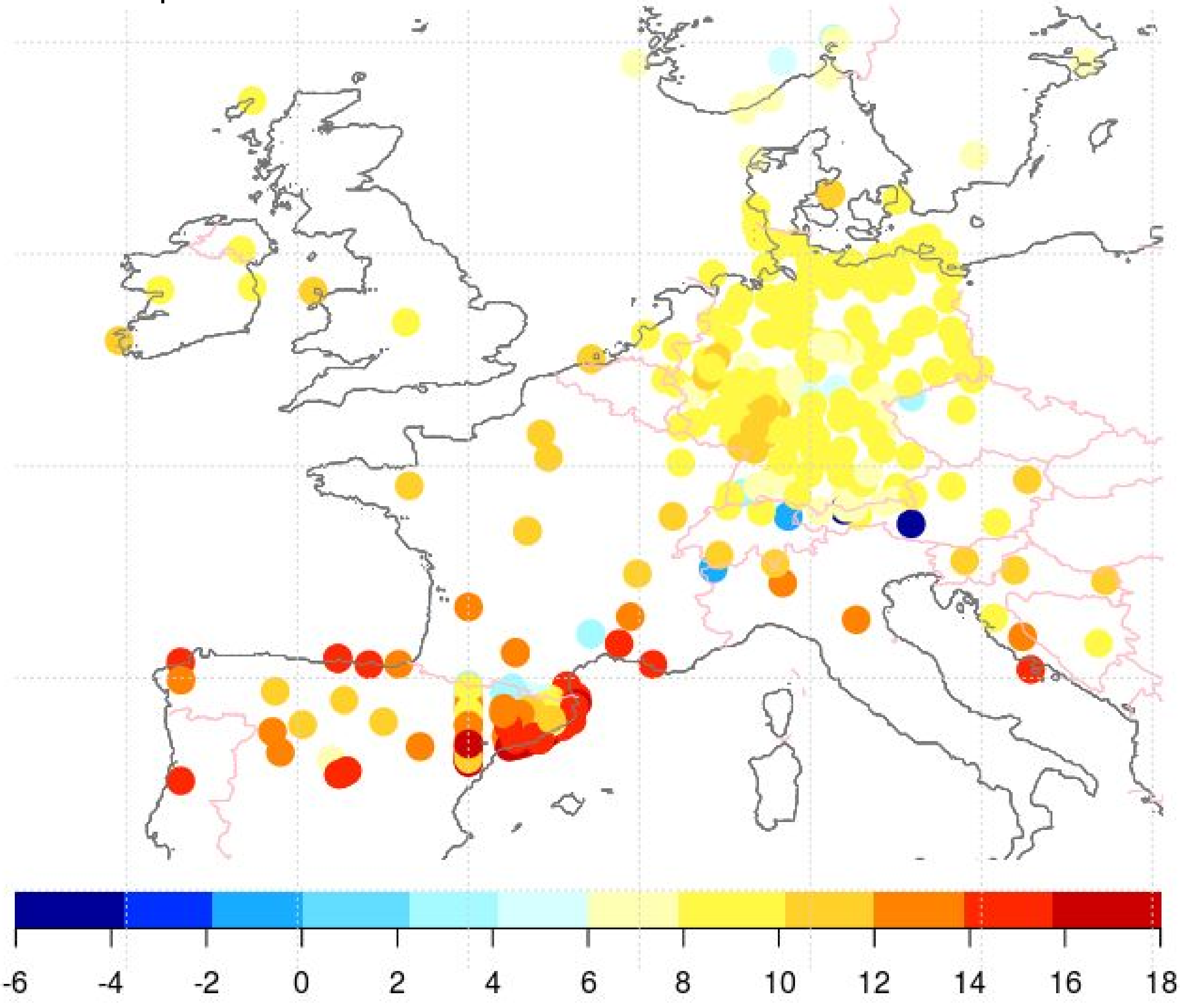


Requirements for PCA-based downscaling

- No missing data
- Suitable distribution
- Not too spatially spread

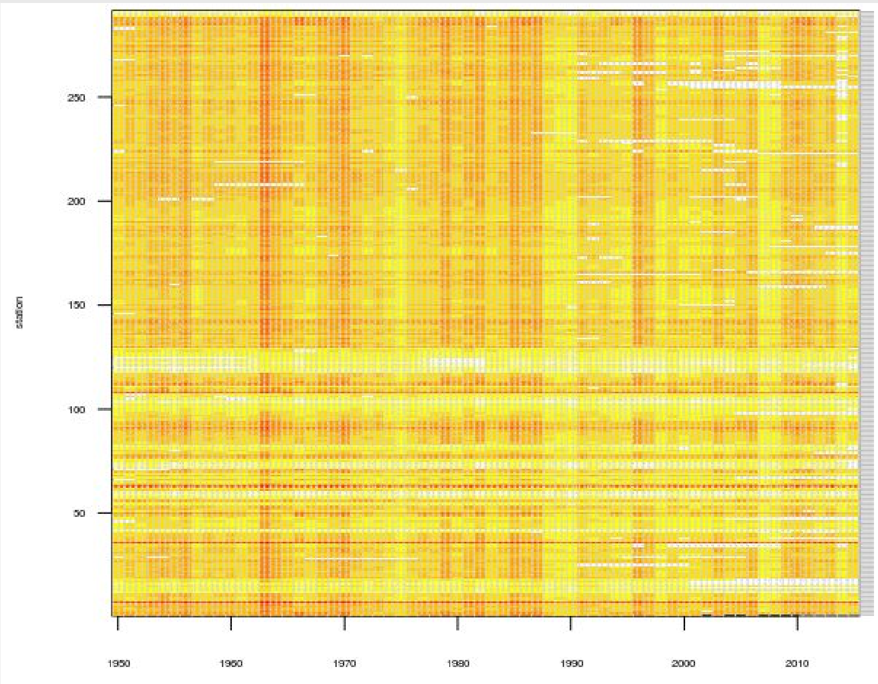
‘pcafill’ (esd R-package)

ECA&D temperature station network

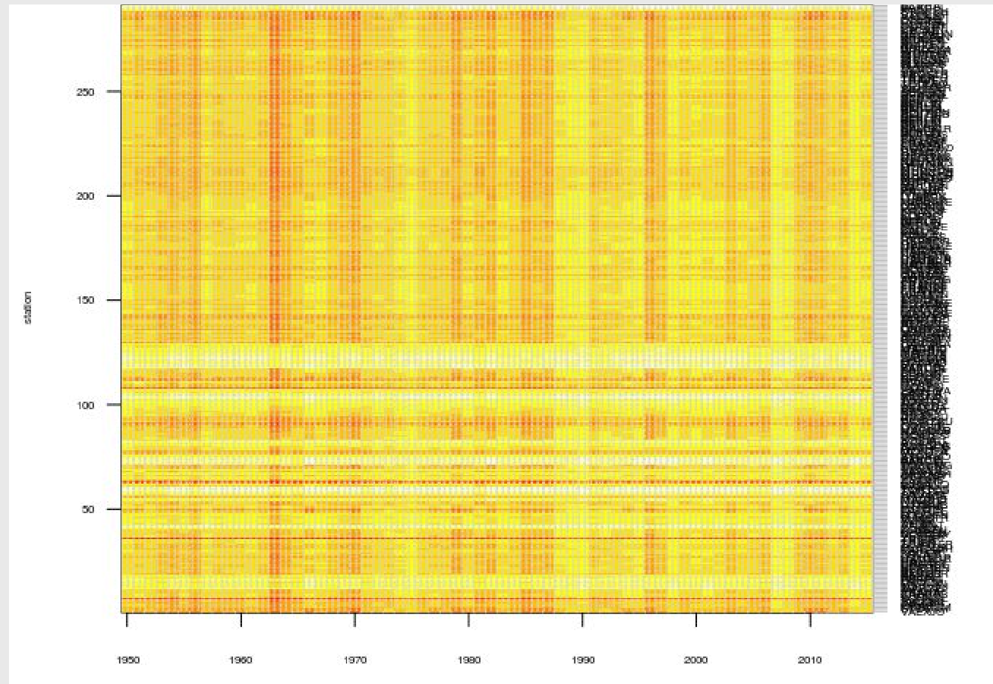


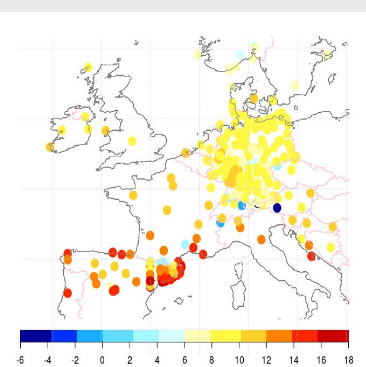
‘pcafill’ (esd R-package)

Data matrix with missing data (white spaces)



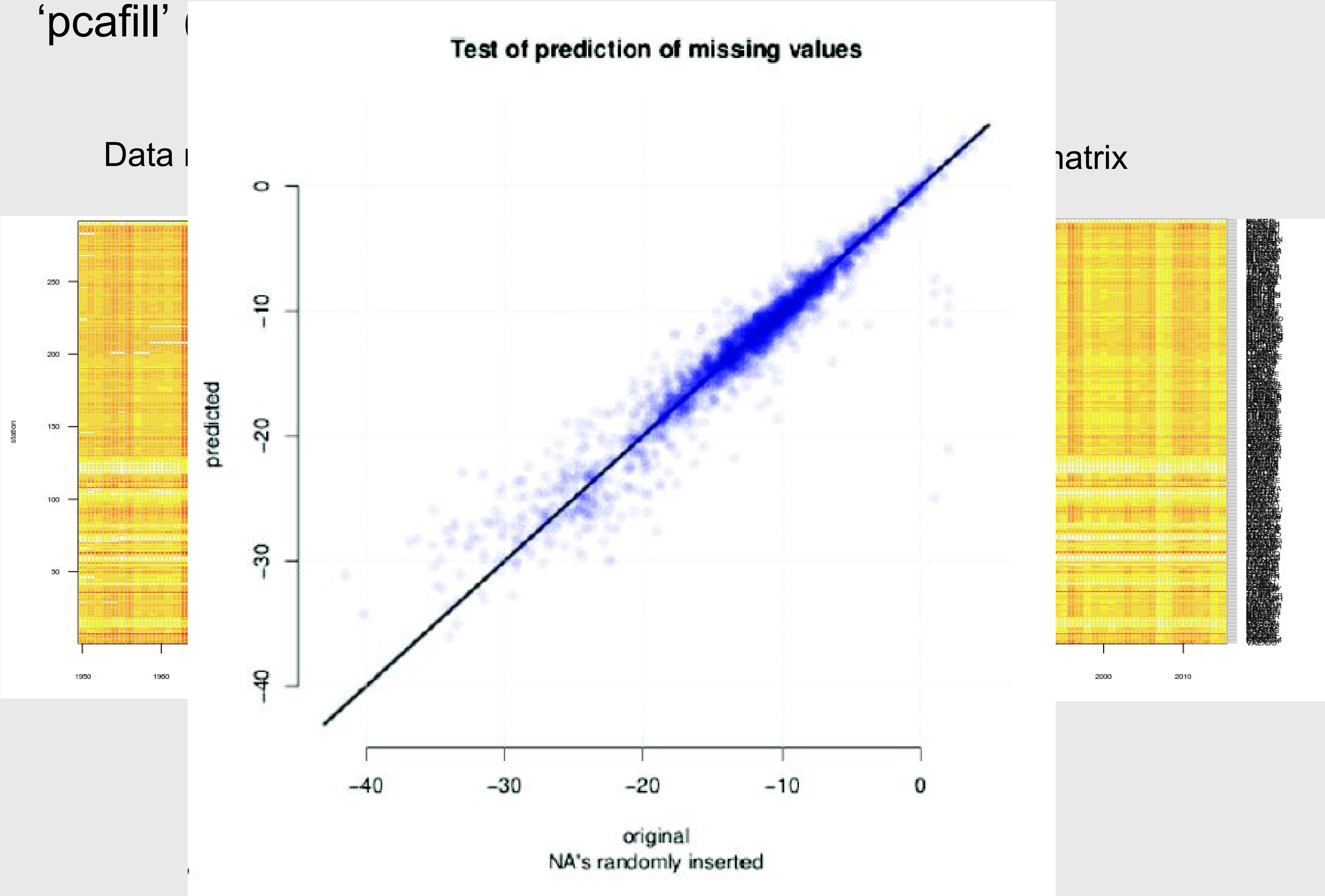
Complete data matrix



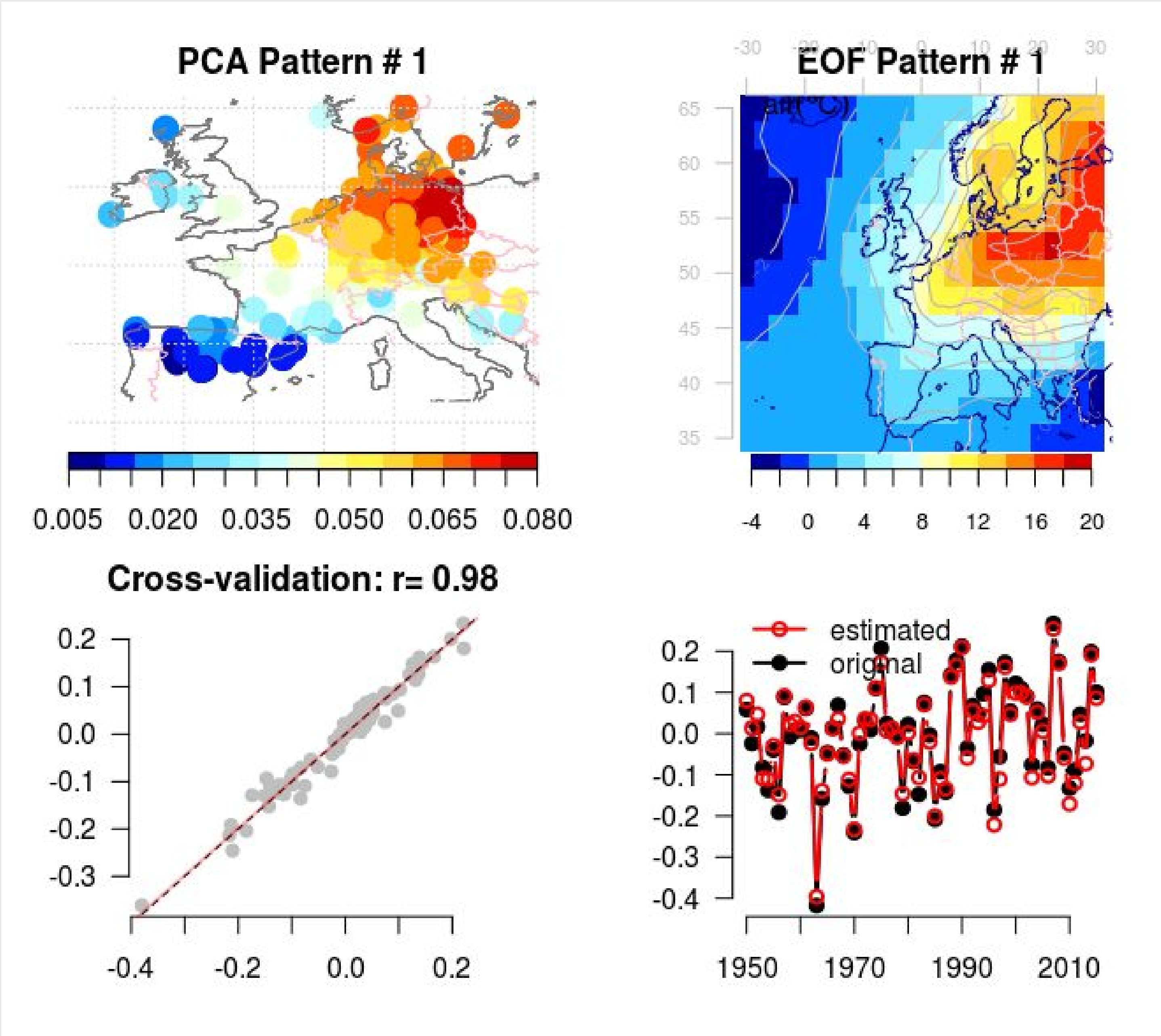


‘pcafill’ (

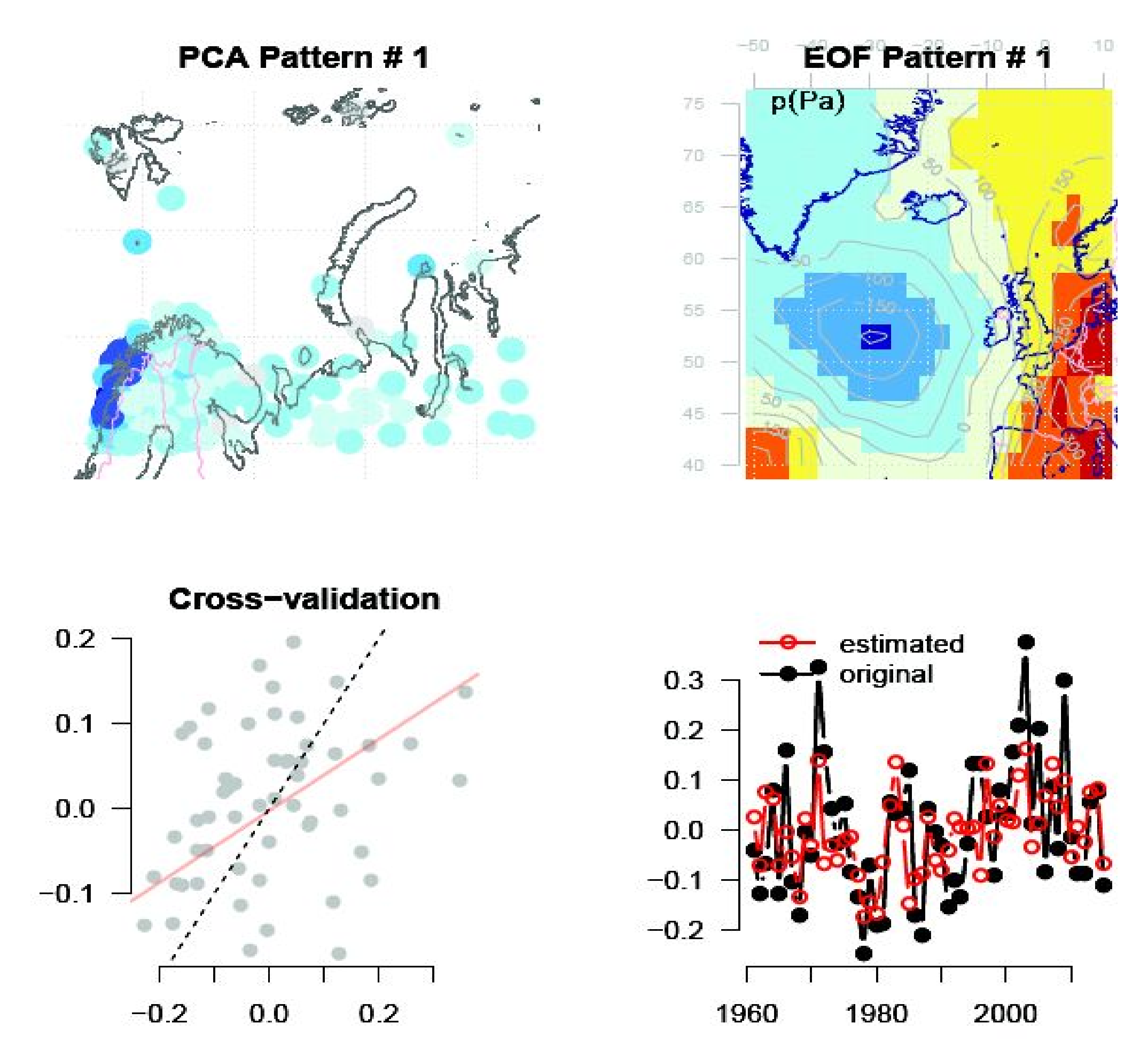
Data r atrix



Downscaling: diagnostics for leading PCA

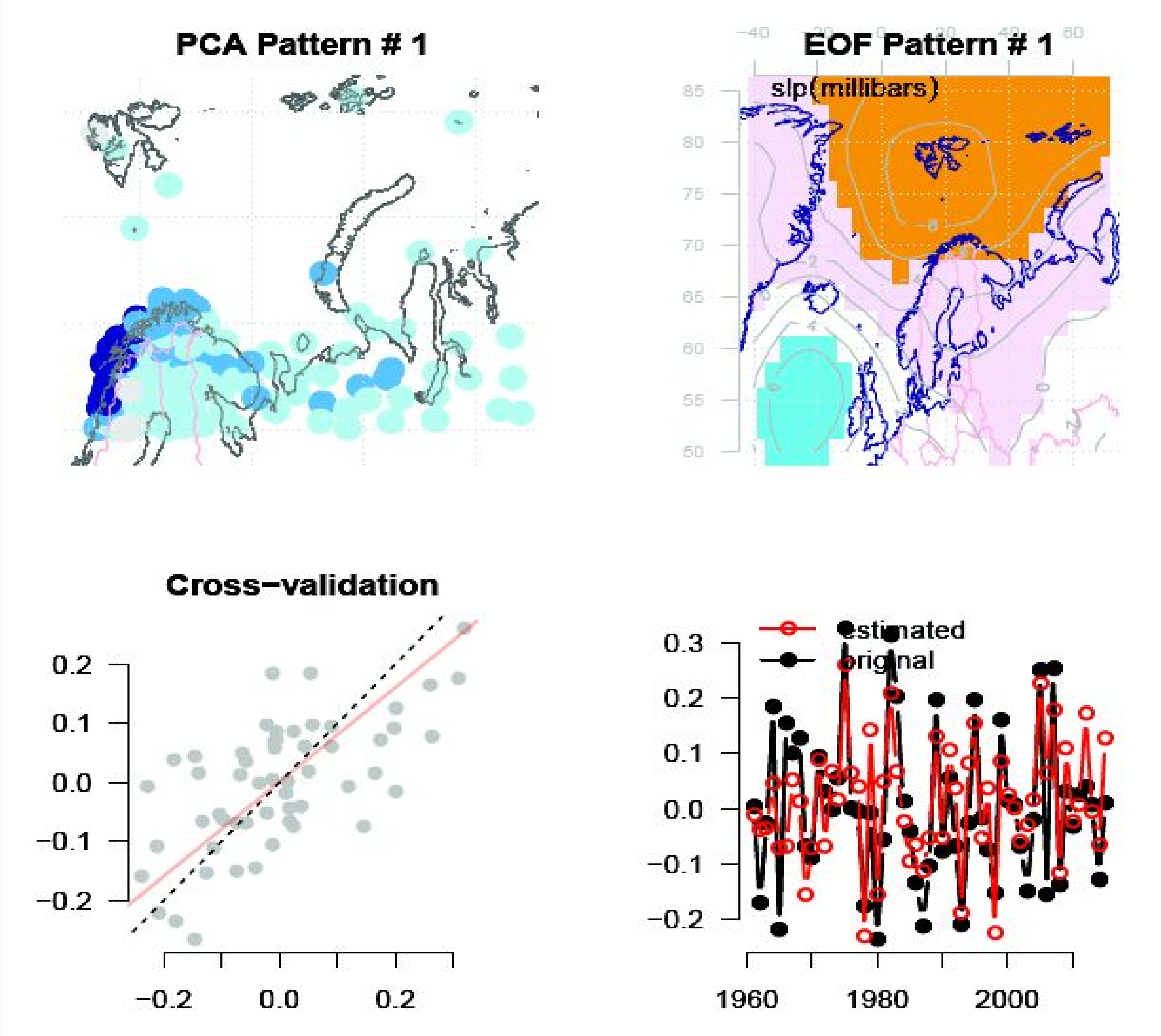


“Warm season” wet-day mean precip μ



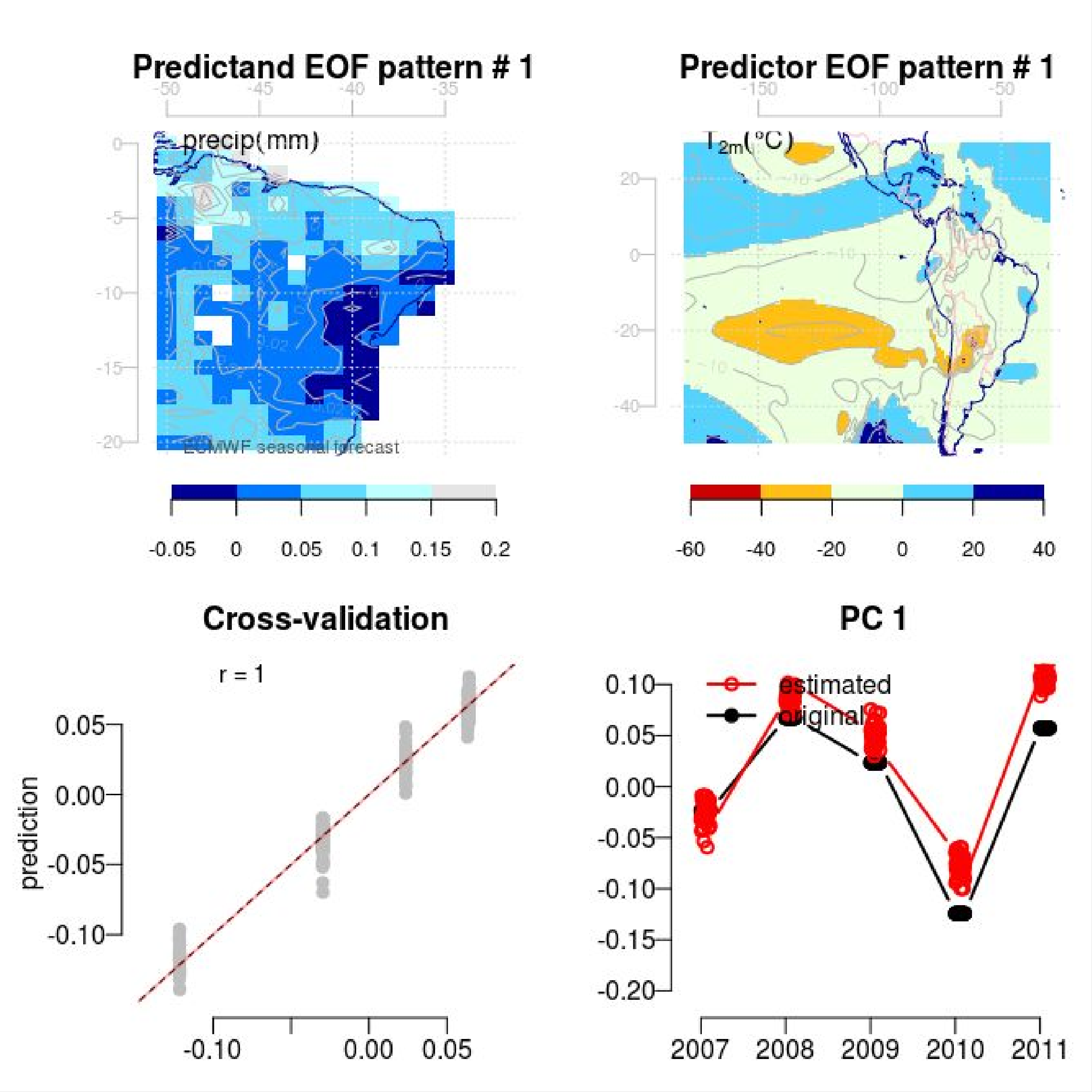
Institute

“Warm season” wet-day frequency f_w



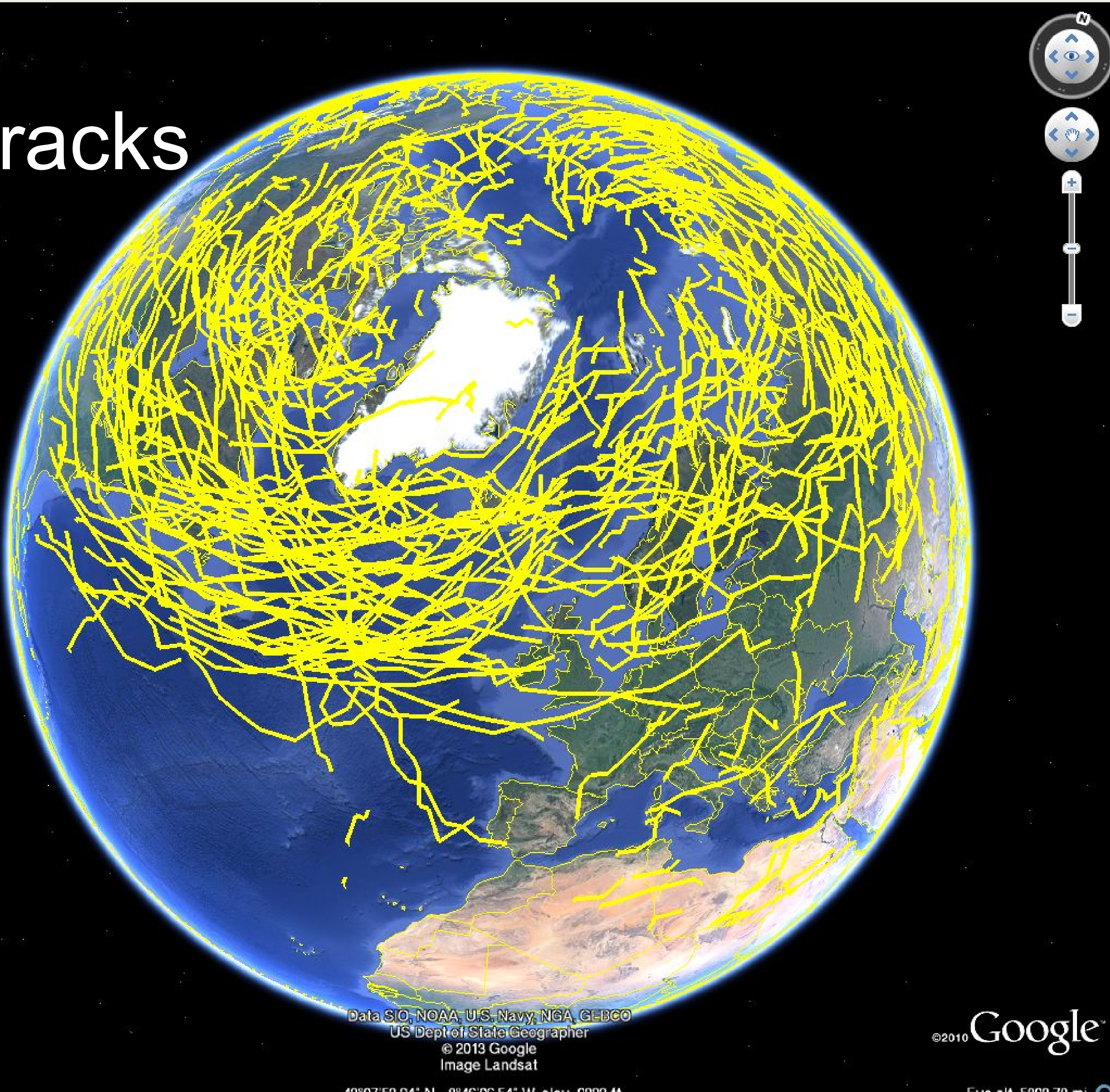
stitute

Model output statistics (MOS) in seasonal prediction ensemble

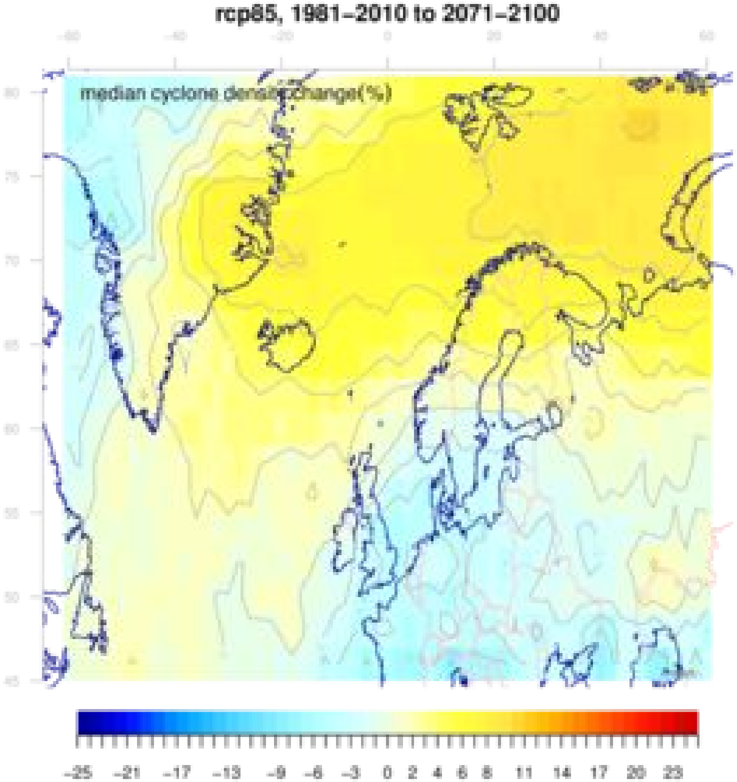


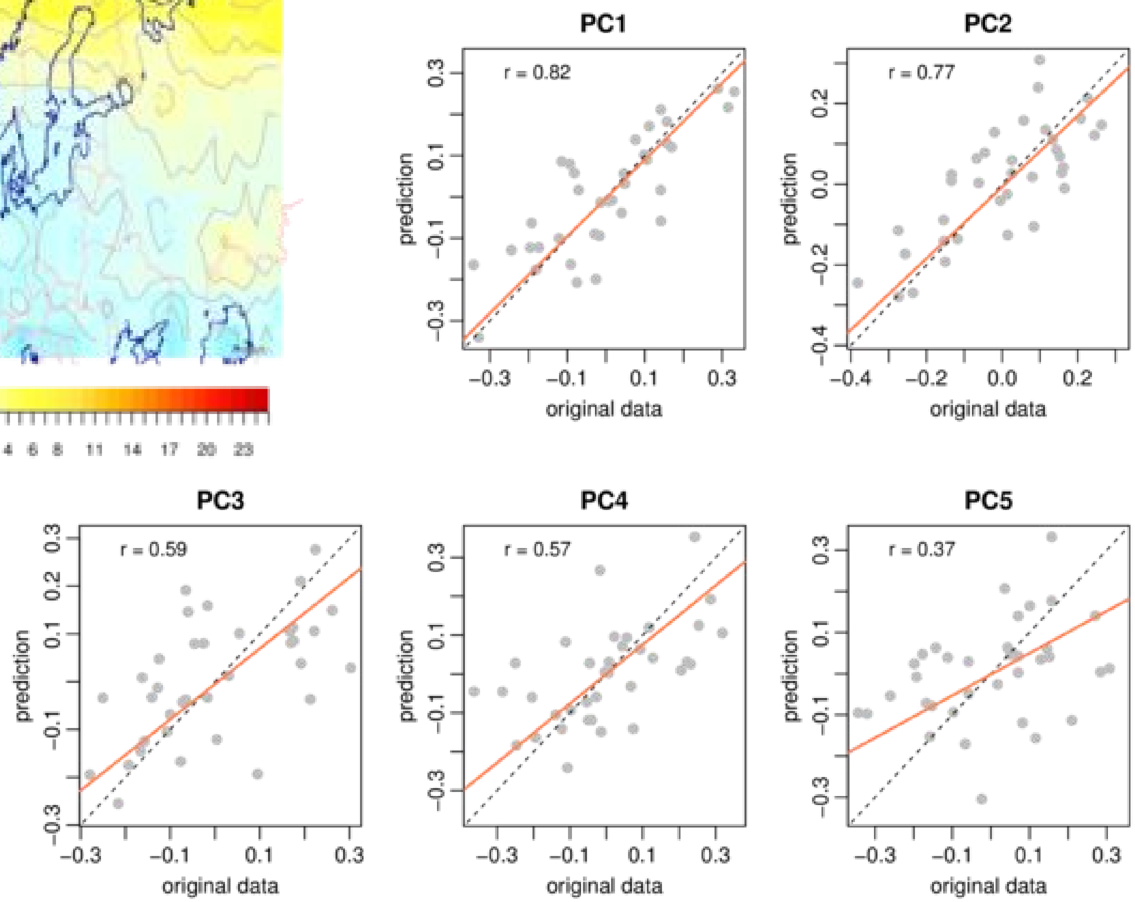
Norwegian Meteorological Institute

Storm tracks



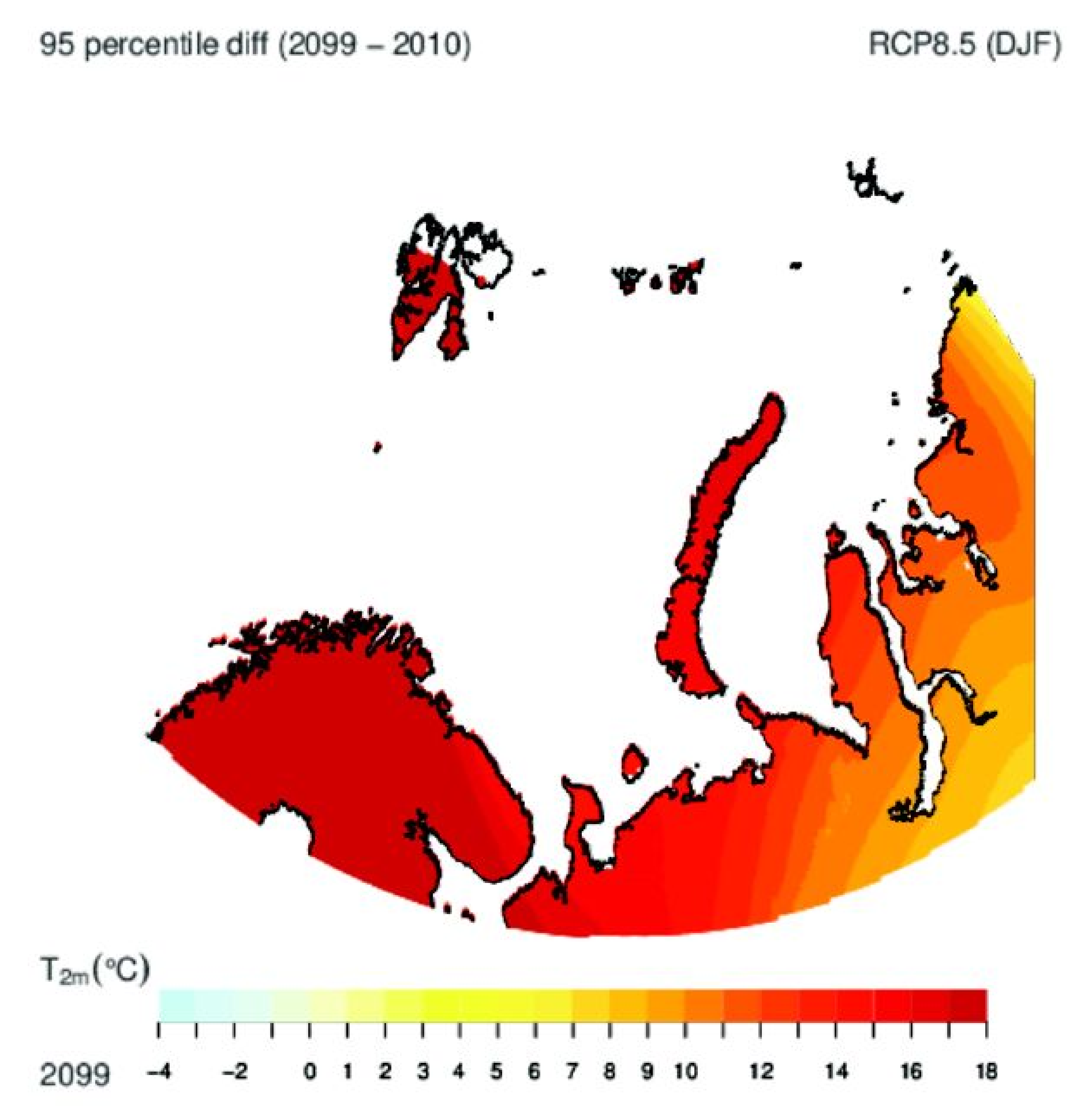
“Warm season” wet-day mean precip μ





Easier post-processing

PCA quick to grid: only a few patterns



LatticeKrig
Covariate: z

Norwegian Meteorological Institute

Twitter: @RasmusBenestad

Meteorologisk institutt 150 år

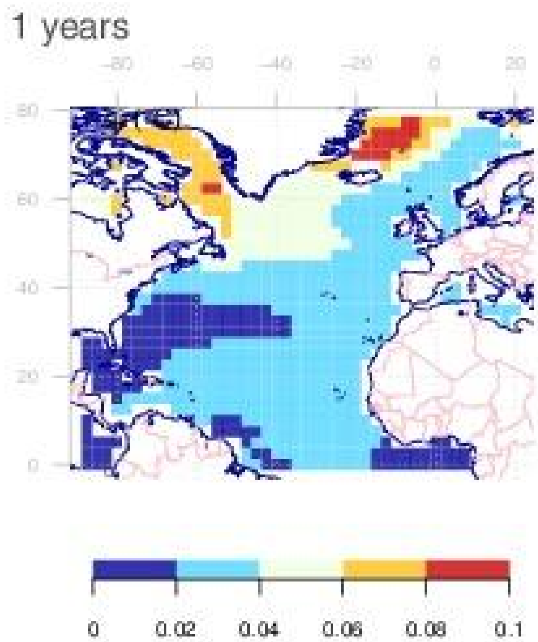
Thank you for your attention!

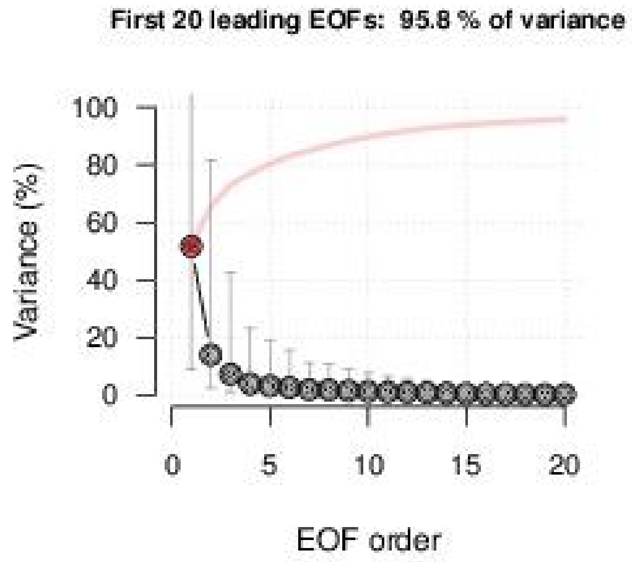
Tellus A 2015, 67, 28326, http://dx.doi.org/10.3402/tellusa.v67.28326

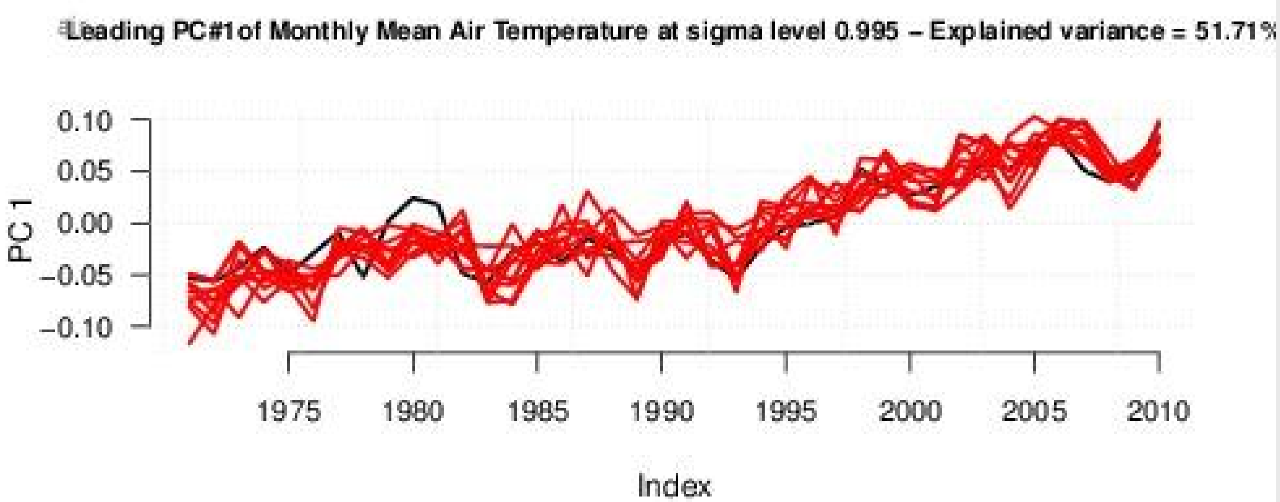
Benestad, Rasmus; Parding, Kajsa; Isaksen, Ketil, Mezghani, Abdelkader "Climate change and projections for the Barents region: what is expected to change and what will stay the same?", ERL-102170.R2 (accepted)

https://github.com/metno/esd_Rmarkdown/tree/master/CORDEX

MOS in decadal prediction ensemble







Norwegian Meteorological Institute