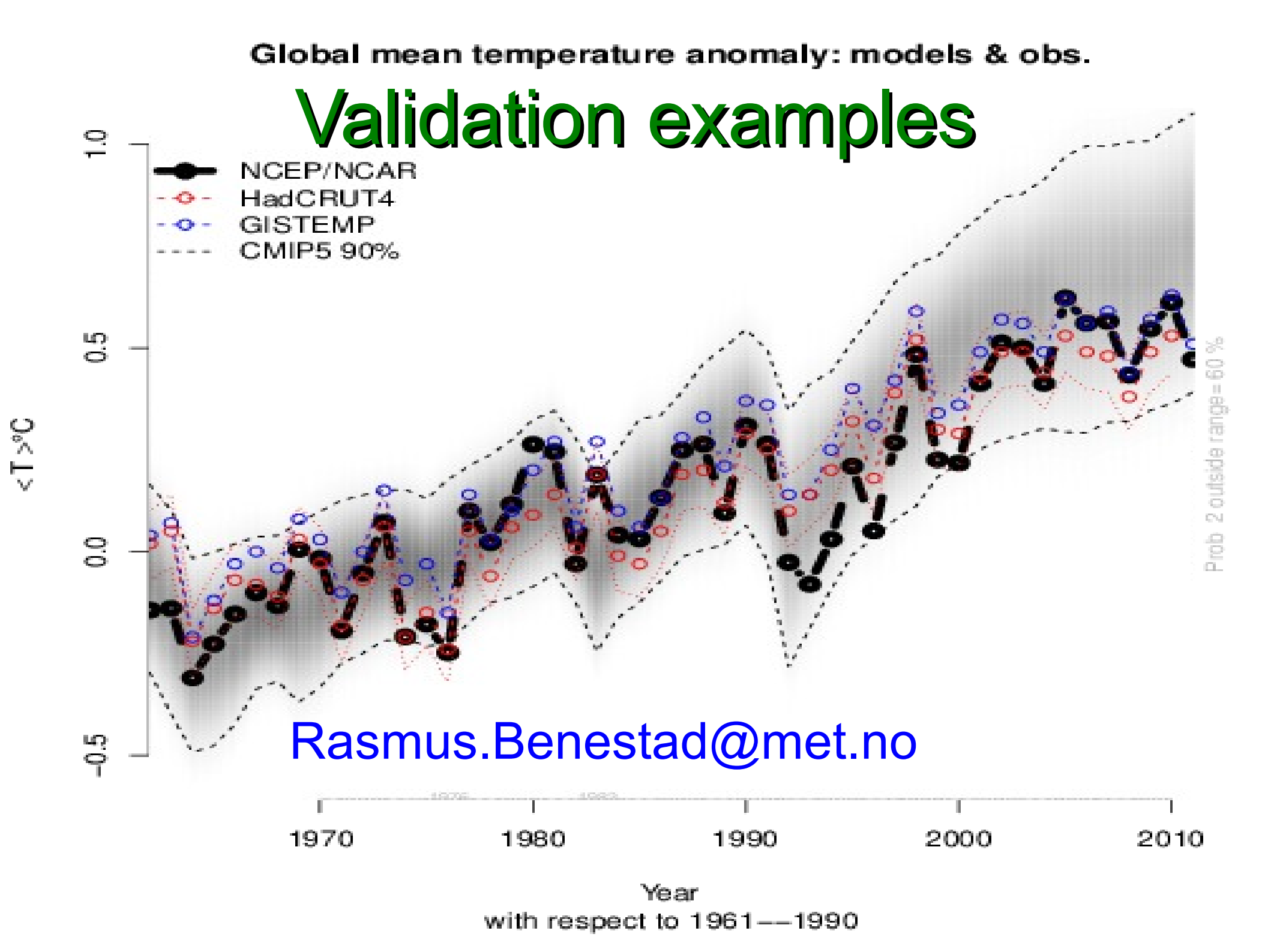
Validation examples



Rasmus.Benestad@met.no

Take-home message

- Validation should encompass **both** physics and statistics
- The validation must be purpose relevant: *evaluation of past trends important for prediction of change*.
- Validation of ensembles: are the observations distinguishable from the sample of results provided from the models?
- Validate the **entire process** (GCM + DS).

Kerr: 'Forecasting Regional Climate Change Flunks Its First Test'
Science 8 February 2013:
Vol. 339 no. 6120 p. 638
DOI: 10.1126/science.339.6120.638

Model purpose

- Past trends
- ESD models driven by observations
  - Split sample
  - Different choices

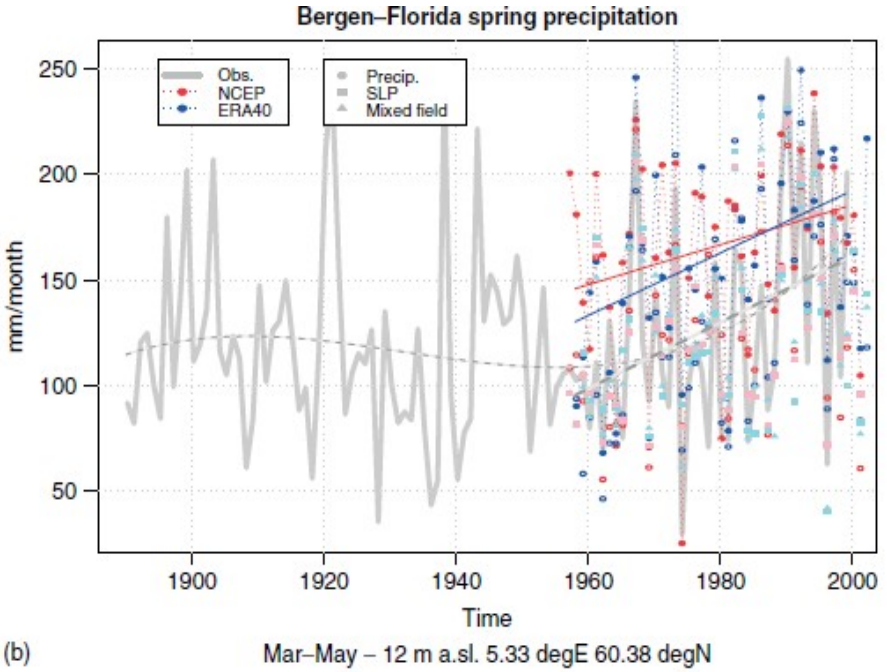
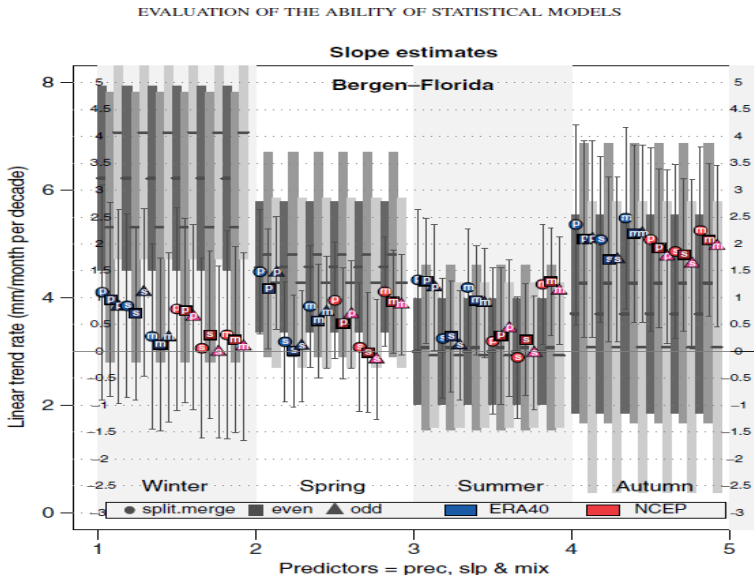


Figure 2. Time series showing the observed and 'split-merge' downscaled trends in precipitation in Bergen for the winter, spring, summer, an autumn seasons respectively. The blue and light blue symbols represent ERA40 results whereas the red and pink represent NCEP values. Circula symbols represent precipitation PMs, with filled symbols for 'split-merge' and open circles for odd and even. The square symbols mark result from SLP PMs, and triangles mark results from mixed-field PMs. Lines are only shown for 'split-merge' precipitation PMs, and only result from 'odd' and 'even' are shown for SLP and mixed-fields to avoid cluttering of the figures.

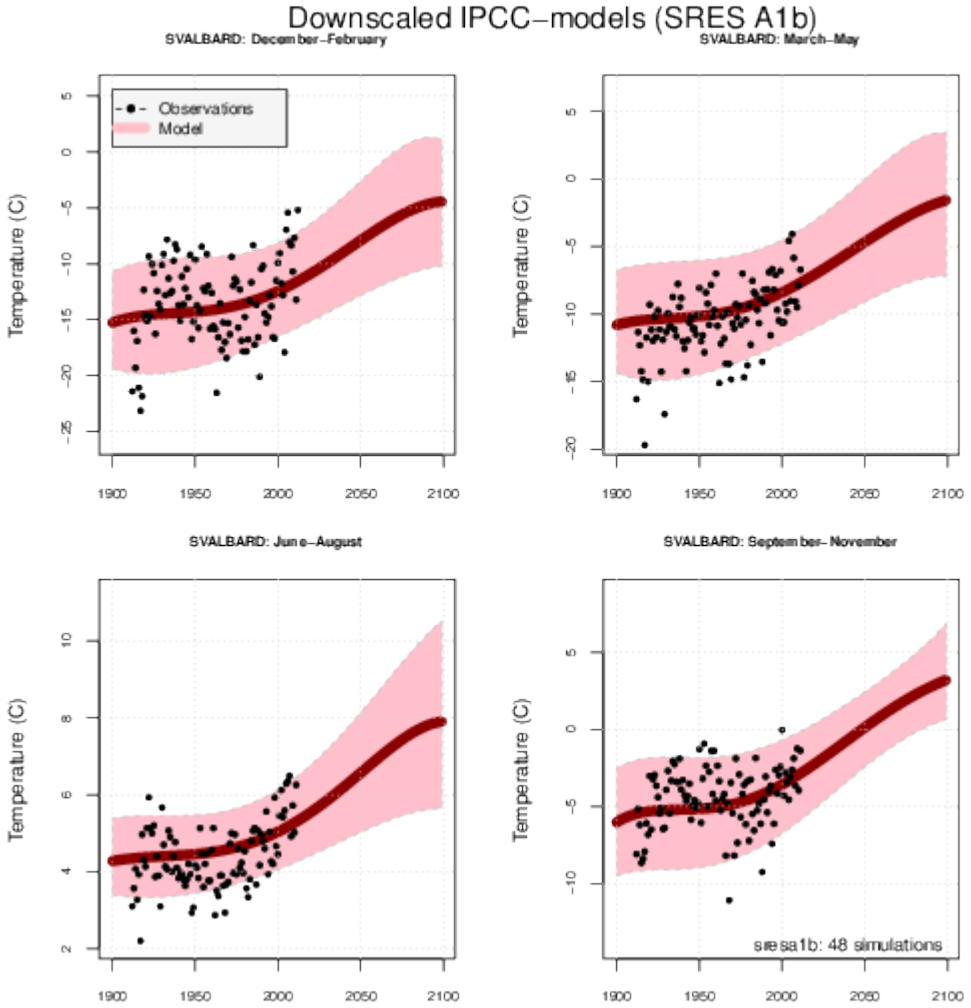


Benestad, R.E., Hanssen-Bauer, I., Førland, E.J., 2007. An Evaluation of Statistical Models for Downscaling Precipitation and Their Ability to Capture Long-Term Trends. *International Journal of Climatology* 27, 649–665.

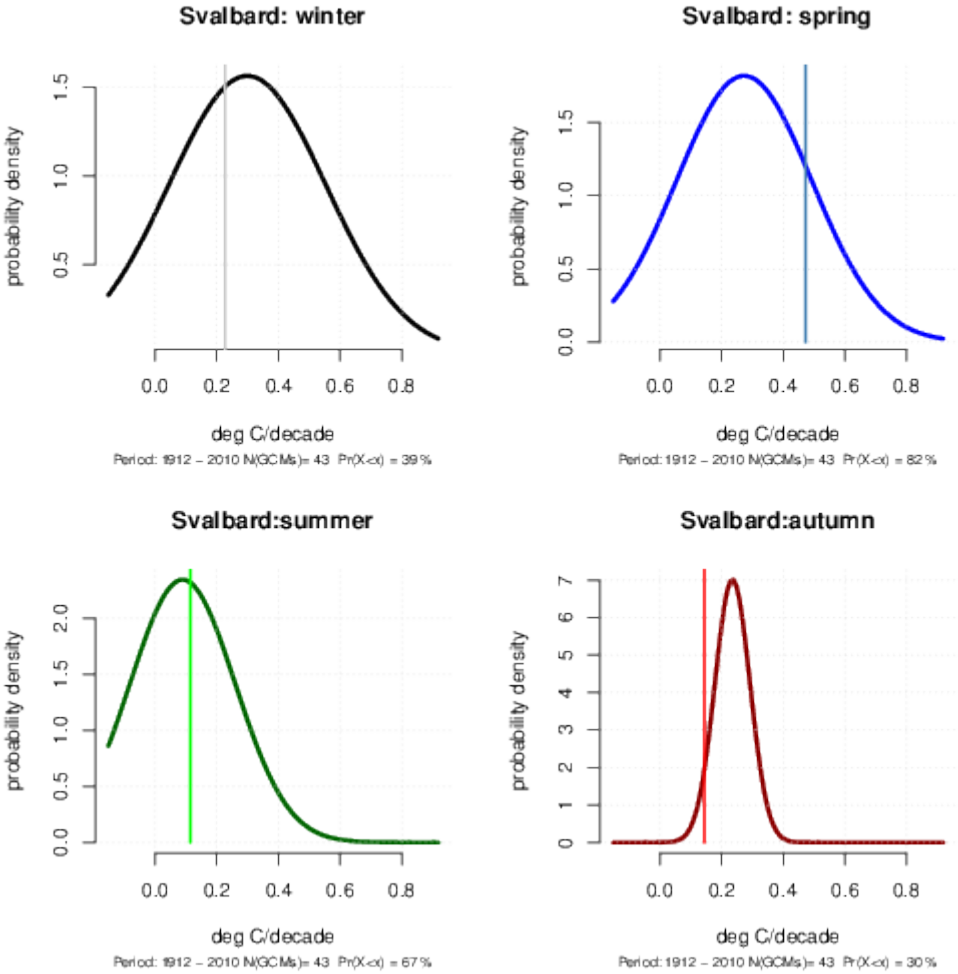
Do ensembles give realistic description?

- Trends from GCMs + ESD.

Example of DS results

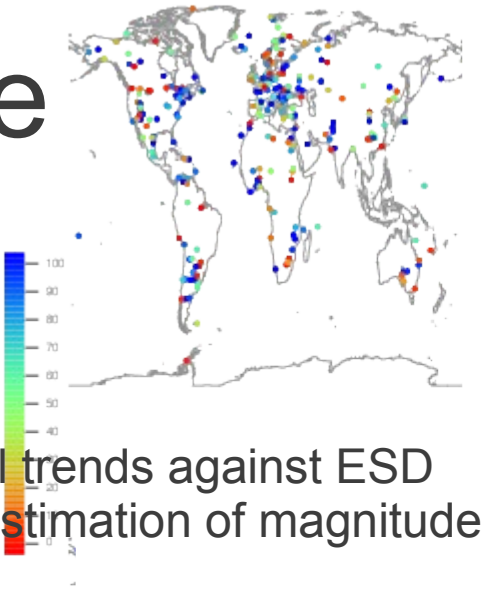


Comparison of observed past trends with trends derived from ESD

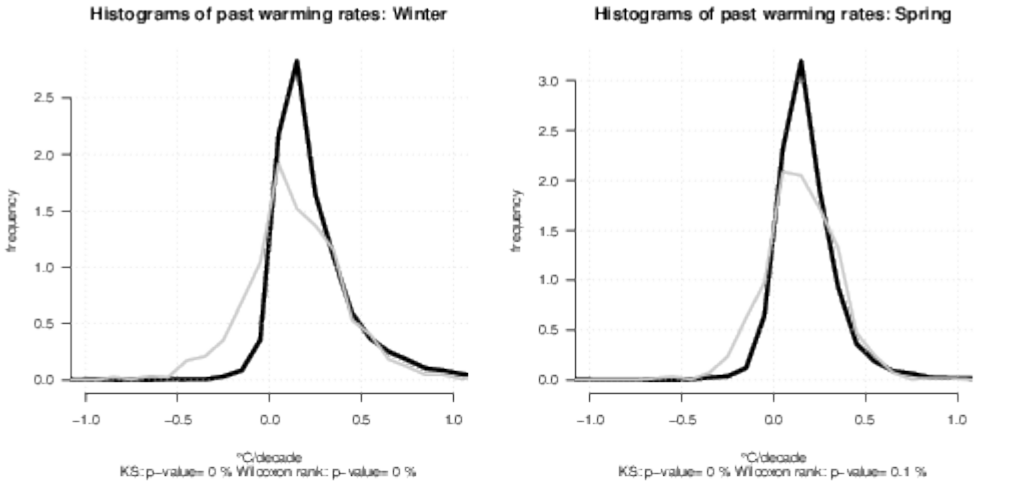


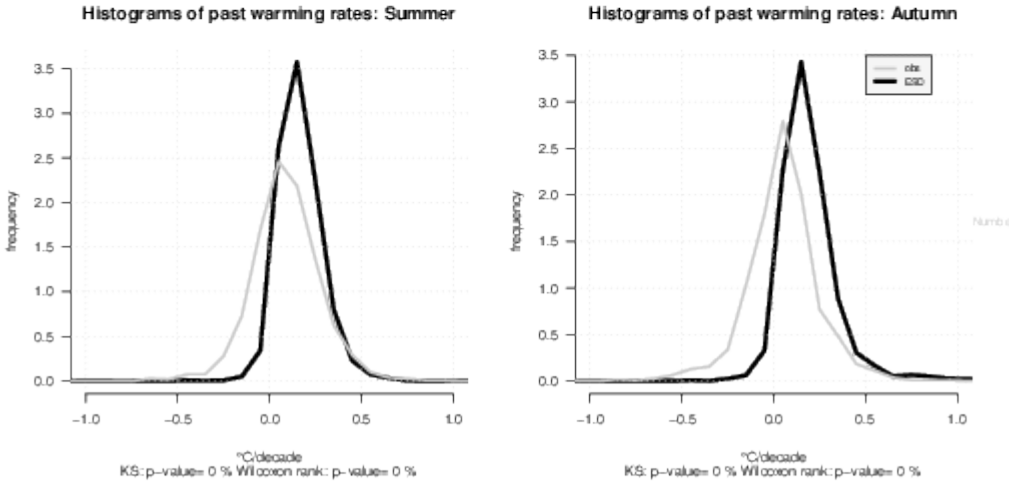
Downscaled trends world-wide

- Evaluate: GCM-ensembles + ESD

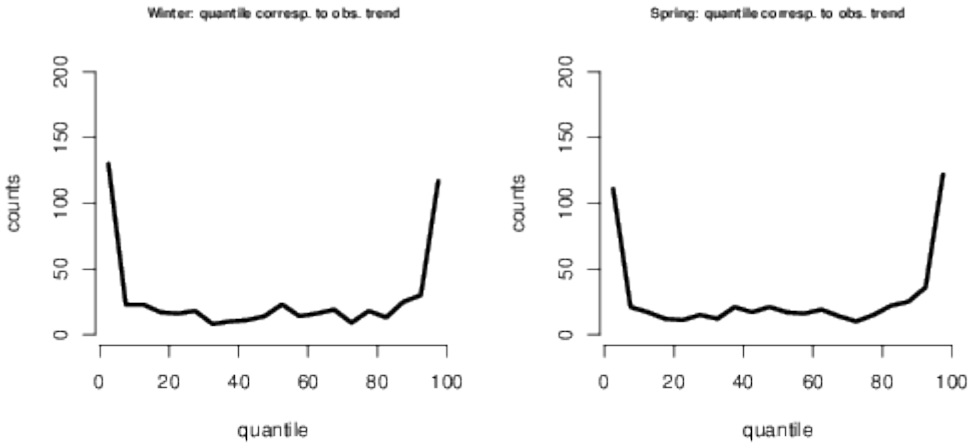


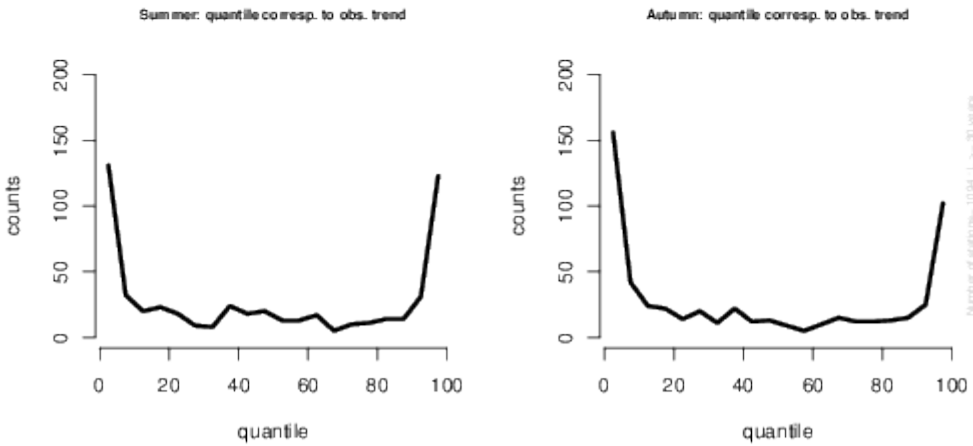
Trend statistics for the world





Ranking of local trends against ESD results: under-estimation of magnitude





Realistic extremes?

- Confidence intervals and outliers – binomial distribution
- +Standard PDF-tests.

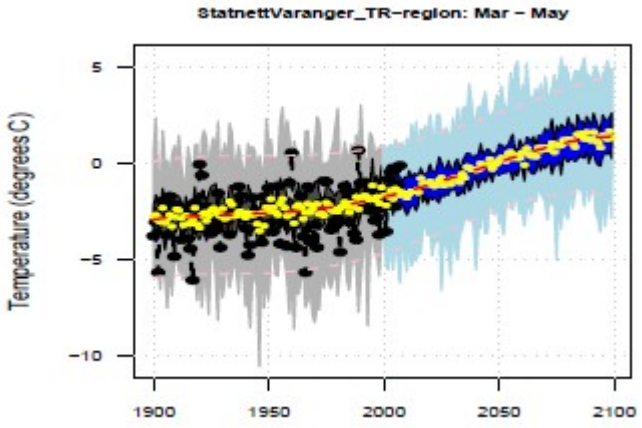
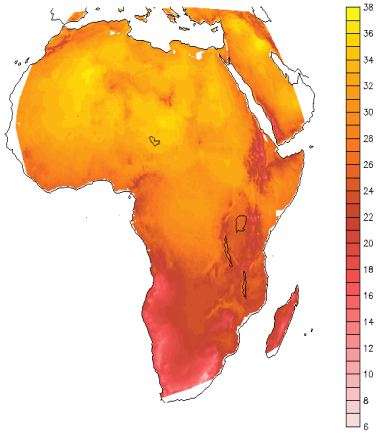


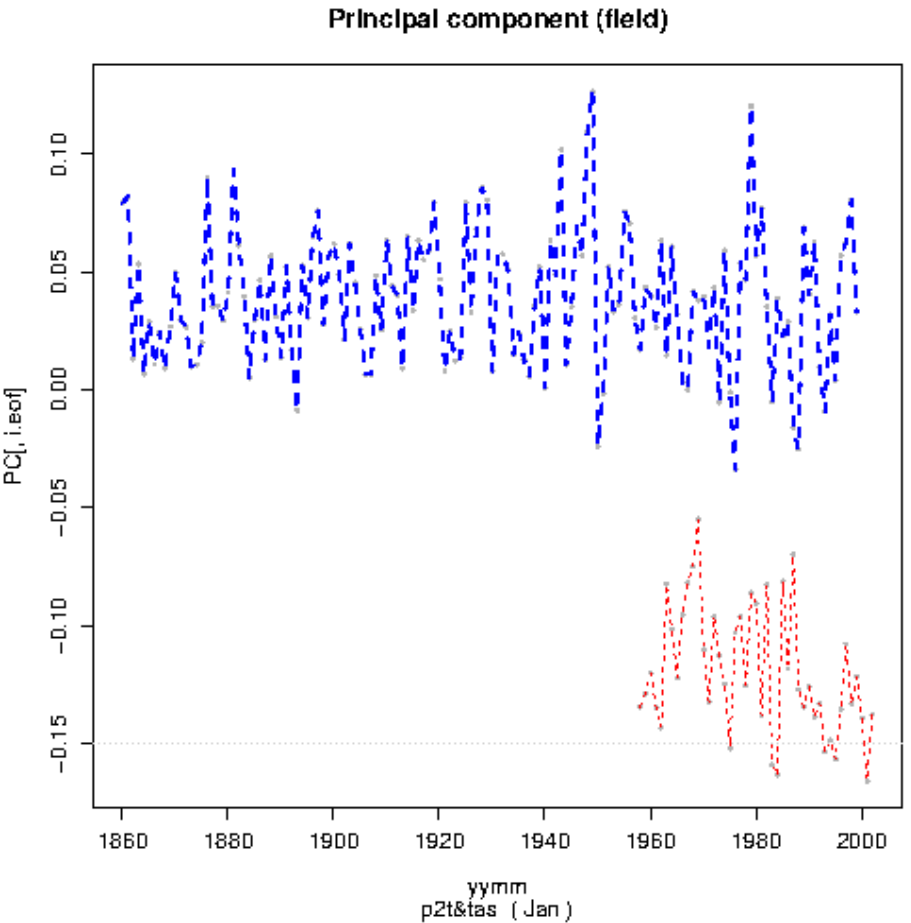
Table 2: *Counts of times when observed temperatures exceed the polynomial fit to* $q_{0.95}$ *for the past or drop below* $q_{0.05}$. *The 90% confidence intterval was estimated using a binomial distribution of the same sample size as the observations,* $p = 0.05$ *is 1–8 (Sample size=108). The cases where the count is outside the 90% confidence intterval are shown in* bold font.

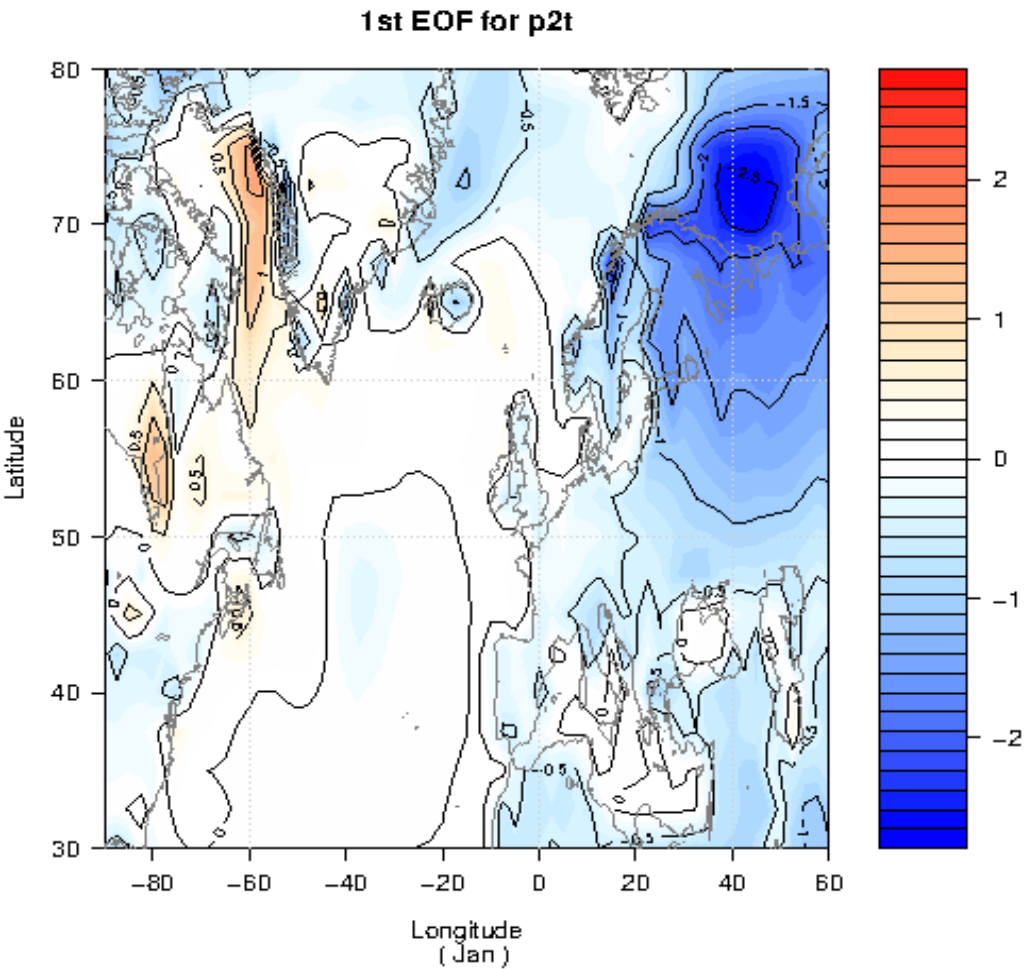
		Winter	Spring	Summer	Autumn
Region 1	over $q_{0.95}$	4	2	10	7
	under $q_{0.05}$	1	8	10	2
Region 2	over $q_{0.95}$	0	1	3	1
	under $q_{0.05}$	1	3	9	1
Region 3	over $q_{0.95}$	1	2	5	5
	under $q_{0.05}$	1	6	8	5
Region 4	over $q_{0.95}$	3	3	12	5
	under $q_{0.05}$	1	6	8	3
Region 5	over $q_{0.95}$	0	2	3	1
	under $q_{0.05}$	1	2	8	0
Region 6	over $q_{0.95}$	9	9	15	8
	under $q_{0.05}$	9	13	12	6

Spatial structures

- Utilise geographical dependencies
  - Regression
- Common EOFs.

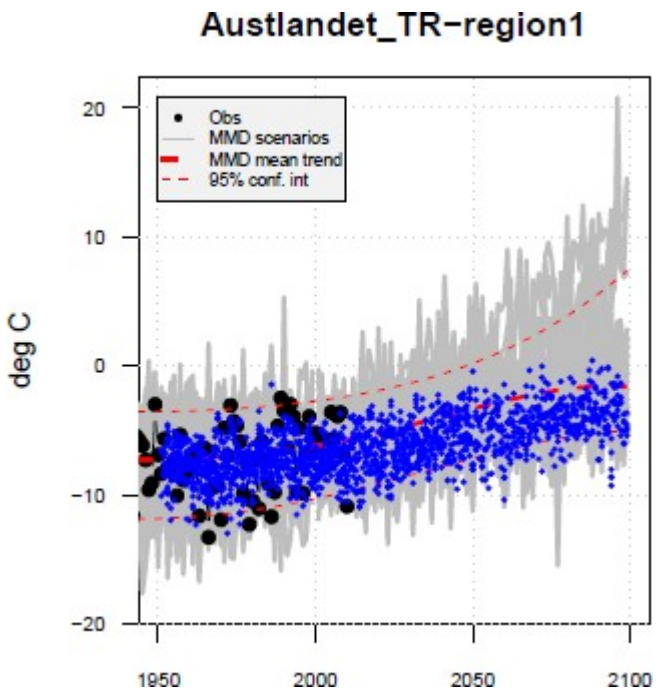






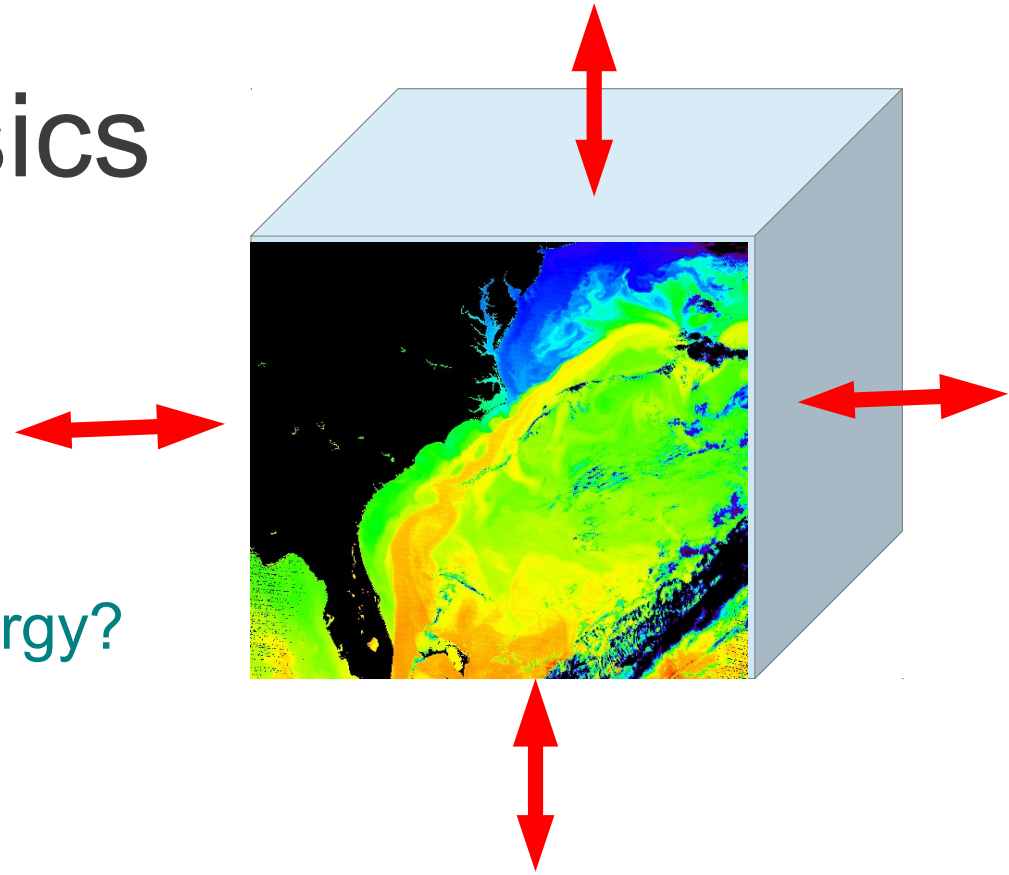
Consistent results?

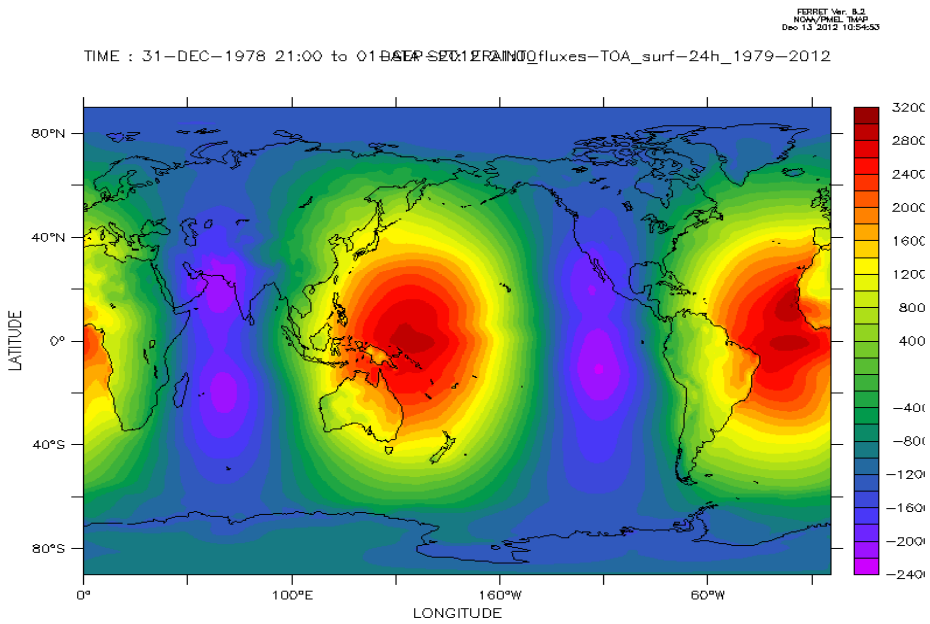
- Compare RCM – ESD
- Downscaled and GCM – aggregated
  - When up-scaled (aggregated), the DS results should agree with the GCM results



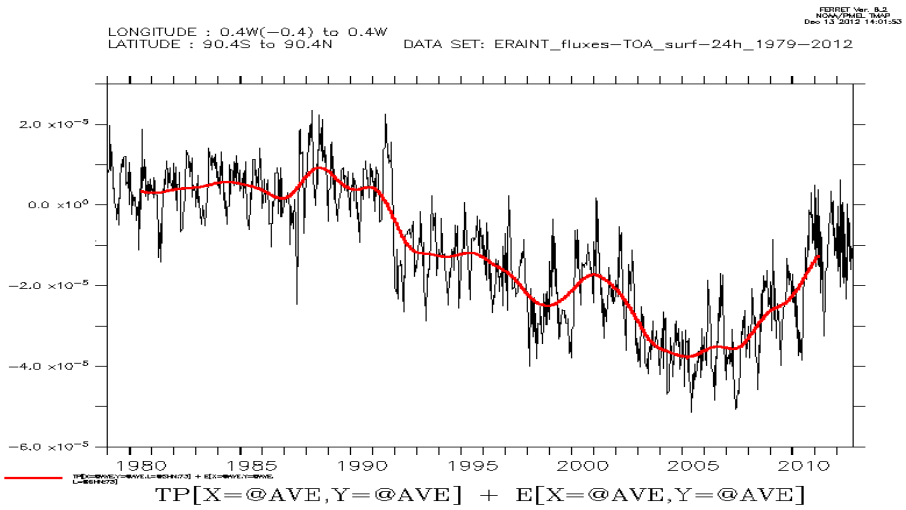
Physics

- Closure problems and energy/mass budgets
  - Consistent between nested models?
  - Conservation of mass and energy?
- Sensitivity tests.





TSR[L=@AVE] + TTR[L=@AVE]



TP[X=@AVE,Y=@AVE] + E[X=@AVE,Y=@AVE]

Recommendations

- Not sufficient to look at only downscaled results.
  - Hewitson's 4 moments: history, phenomena, GCM, and DS results.
- Need to examine both statistics as well as physics.
- Take into account the purpose of the model.
- Independent samples: good observations.
- What information is sought by the end-users?
- What is the minimum skillful scale?