Validation and Model Weighting for PRUDENCE and ENSEMBLES Simulations

Ole B. Christensen
DMI

Key Messages

- PRUDENCE and ENSEMBLES (RT2b and RT3)
- PRUDENCE: ANOVA analysis of bias sources
- ENSEMBLES aim: Probabilistic climate change scenarios; this means weighting
  - Status at the moment: Validation of ERA40-based experiments; new observational gridded data set came out recently
  - Weighting: Added value by RCM only, or realism of fields/physical processes/weather
- Some examples of the work on weighting of models

PRUDENCE GCM-RCM Matrix

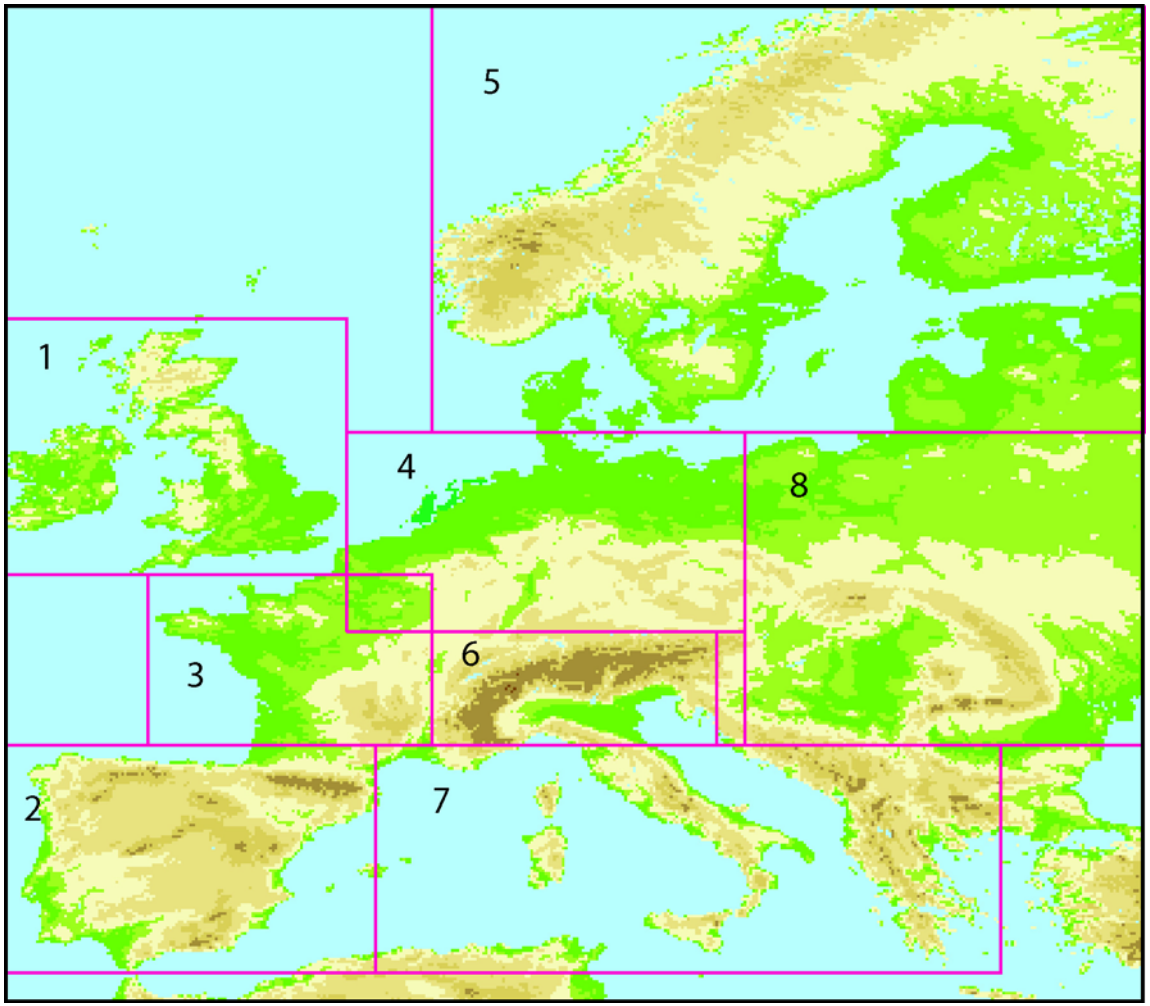
	CNRM	DMI	ETH	GKSS	HC	ICTP	KNMI	MPI	SMHI	UCM
A2+HadAM3H		3	1	1	3	1	1	1	1	1
A2+ECHAM4		1							1	
A2+ARPEGE3	1									
B2+HadAM3H					1	1			1	1
B2+ECHAM4		1							1	
B2+ARPEGE3	3									

ENSEMBLES GCM-RCM Matrix

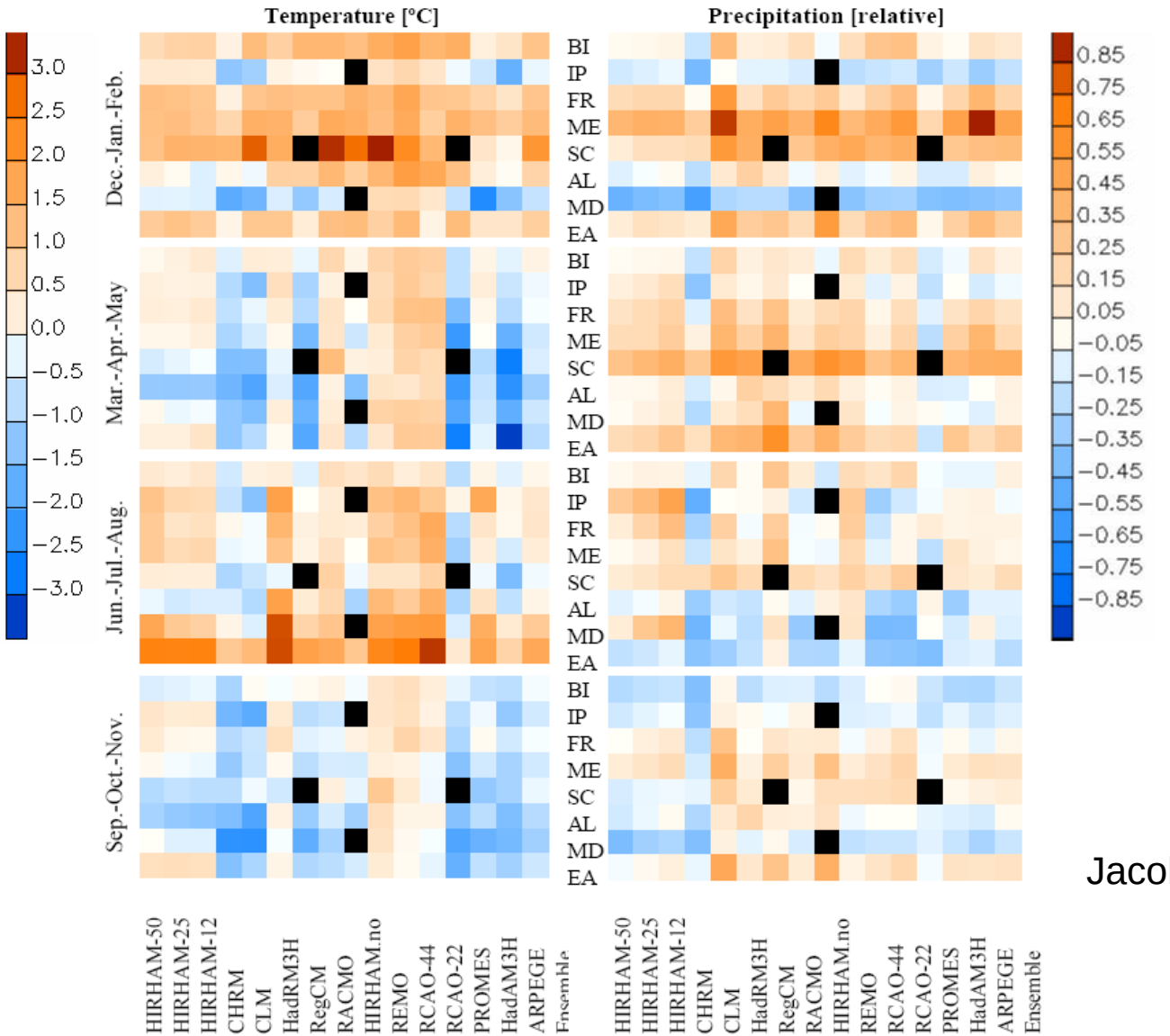
Global model / Regional model	METO-HC	MPIMET	IPSL	CNRM	NERSC	CGCM3	Total number
METO-HC	1950-2100£	1950-2100					2 (4)
MPIMET		1950-2100	1950-2050*				2
CNRM				1950-2050			2
DMI		1950-2100		1950-2050*			2
ETH	1950-2050						1
KNMI		1950-2050					1
ICTP		1950-2050					1
SMHI	1950-2050				1950-2050*		2
UCLM	1950-2050						1
C4I		1950-2050					1
GKSS**			1950-2050*				1
Met.No**					1950-2050*		1
CHMI**				1950-2050*			1
OURANOS**						1950-2050*	1
Total (1950-2050)	4 (6)	6	2	3	2	1	18

RCM 25km: In addition ERA40 simulations available

PRUDENCE domains (Rockel-boxes)

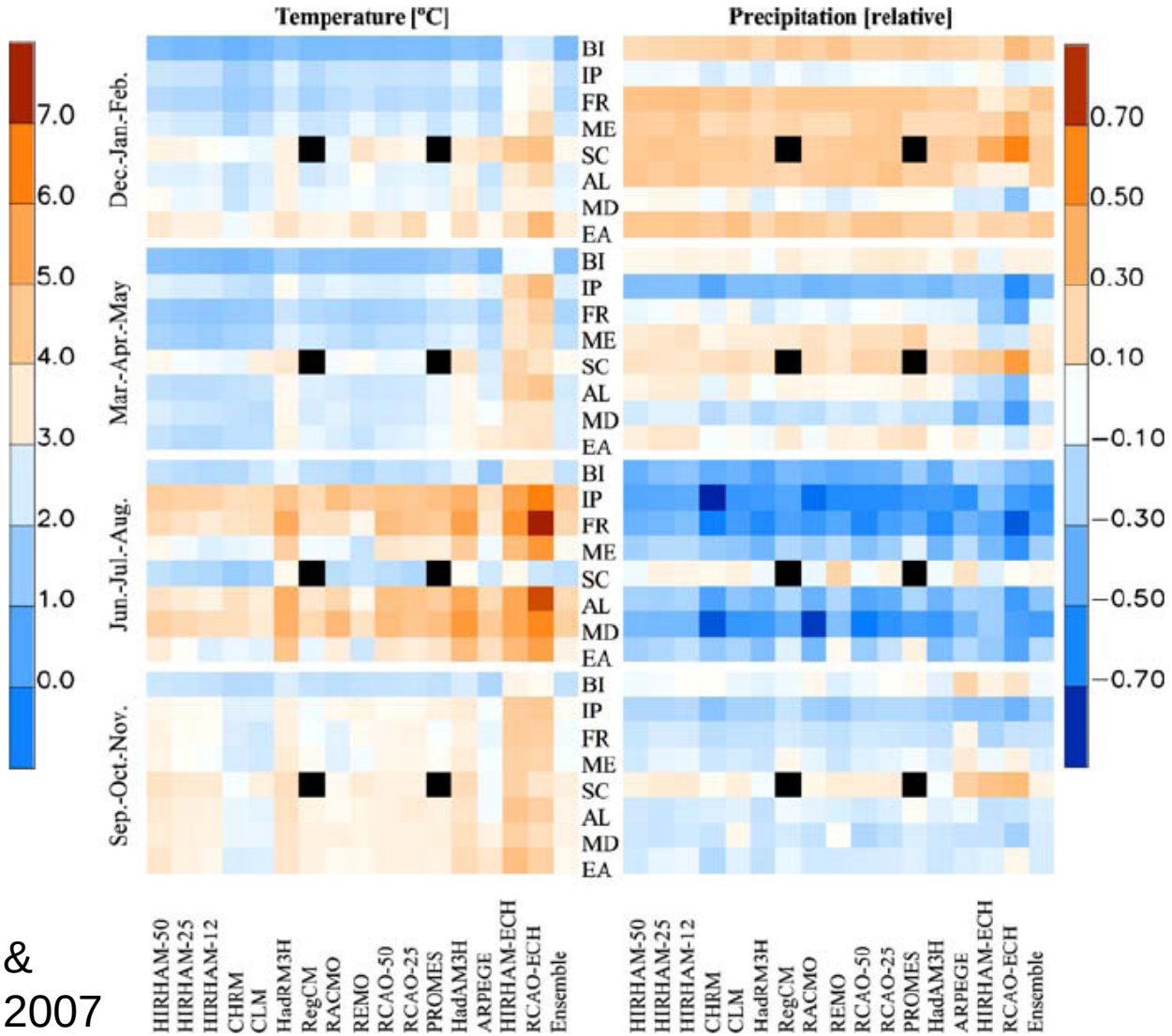


Model bias



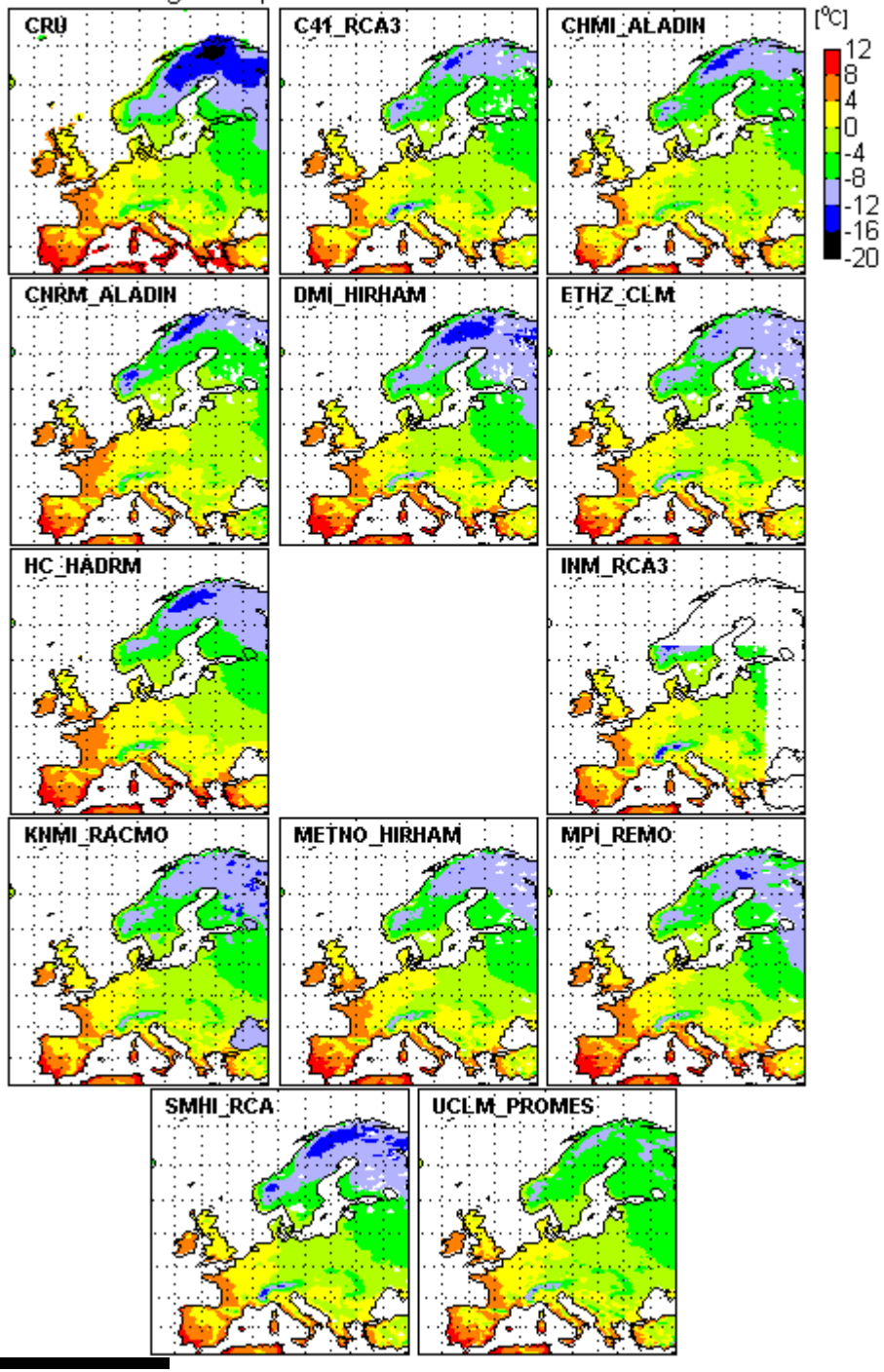
Jacob et al. 2007

Climate change

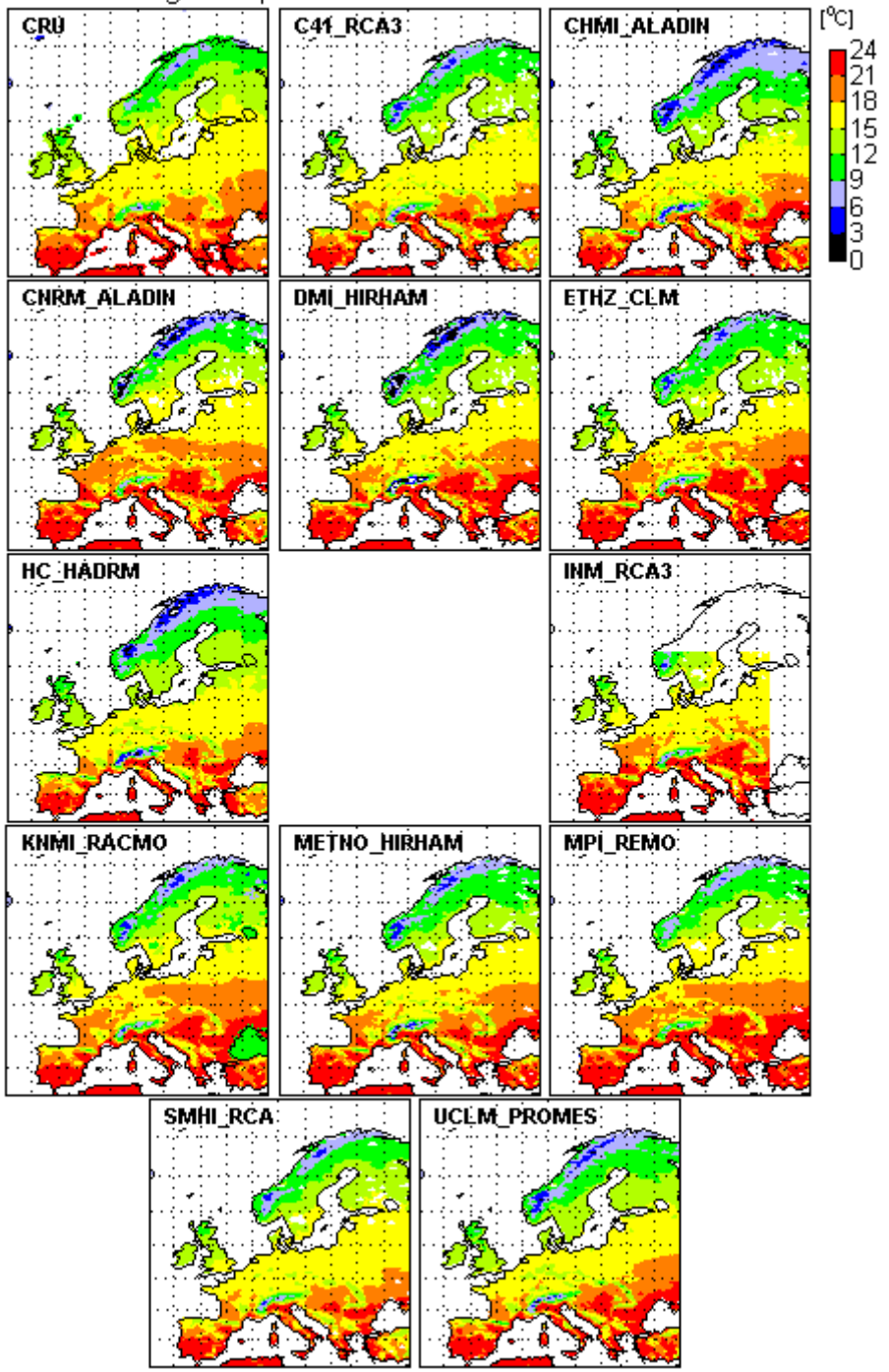


Christensen &
Christensen 2007

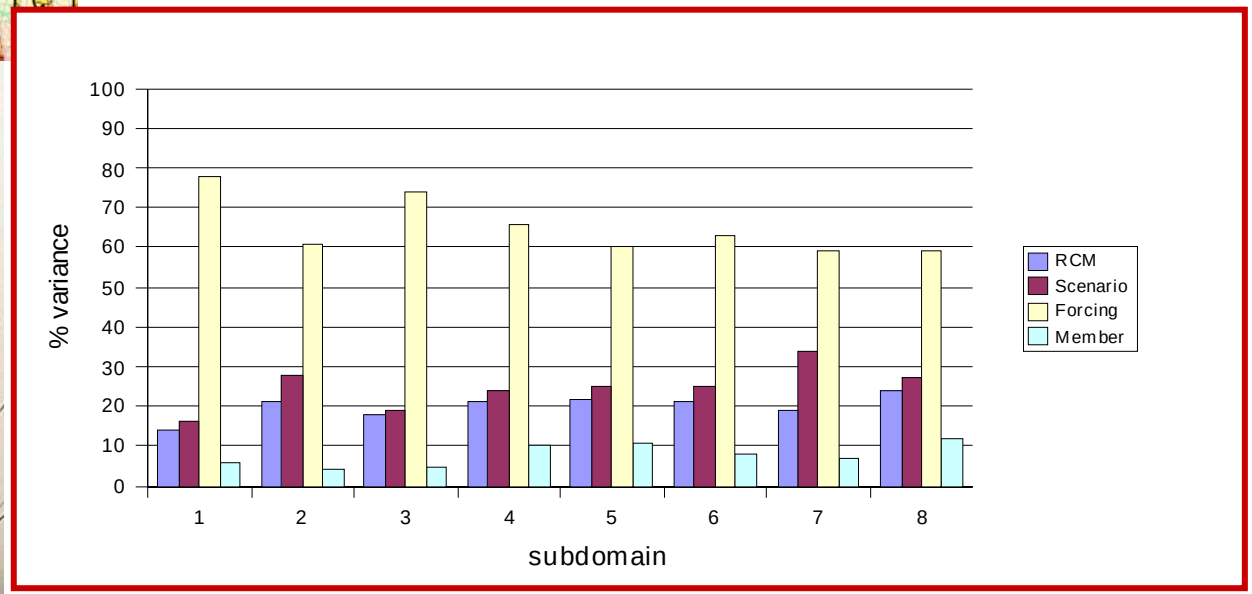
Average temperature for DJF season 1961-1990



Average temperature for JJA season 1961-1990

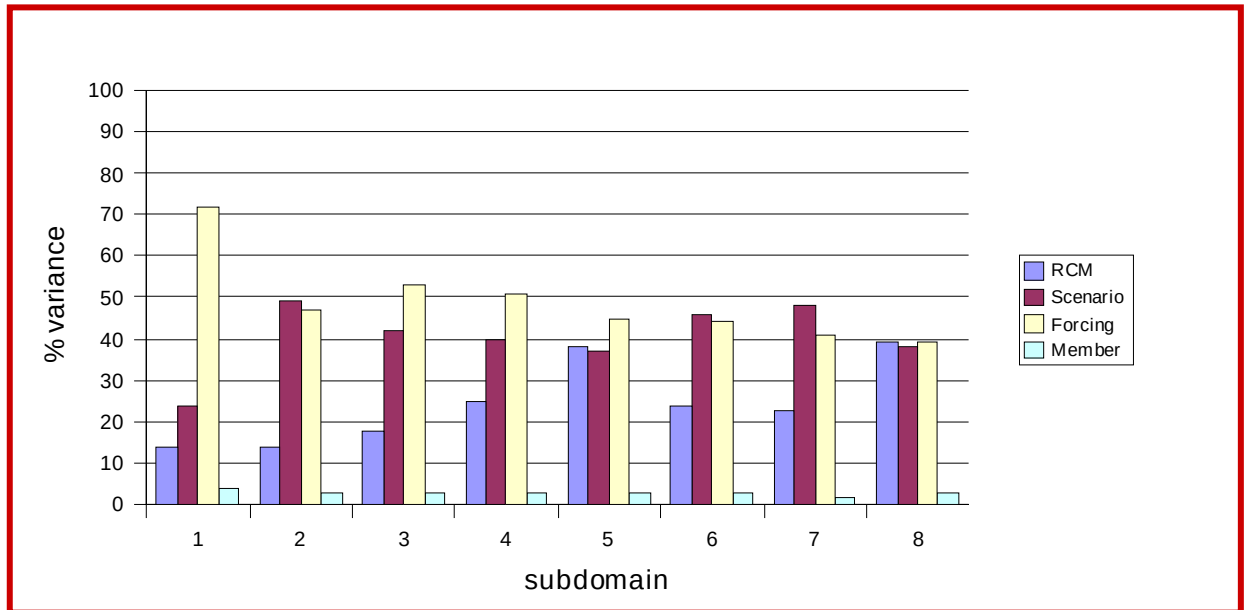


Temperature change – sources of uncertainty



DJF

Depends on driving model

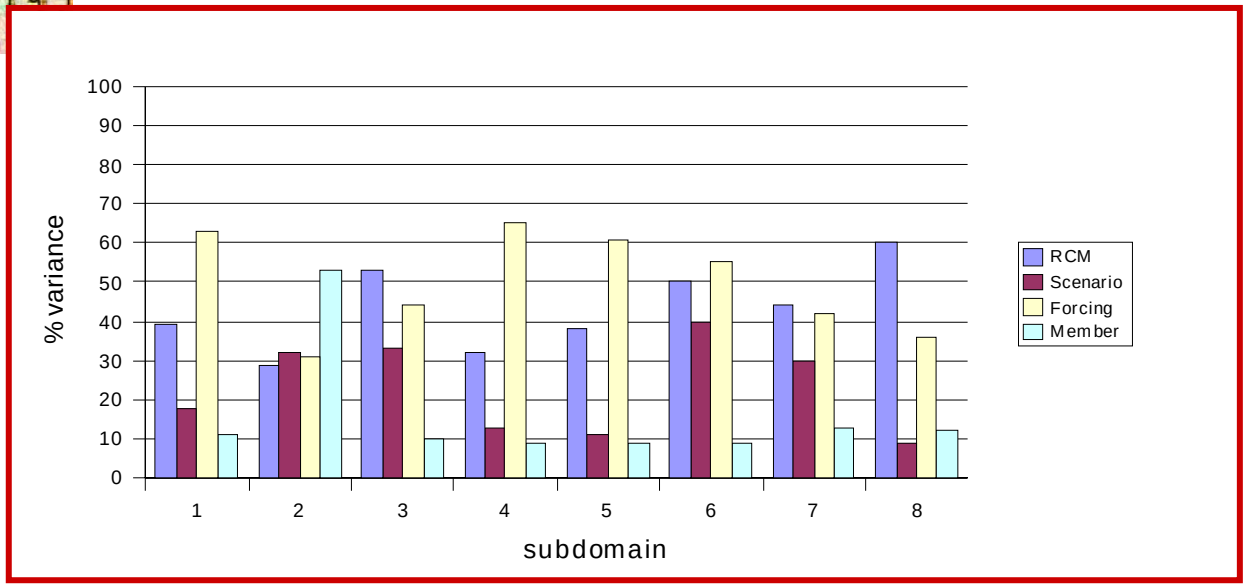


JJA

Also on RCM and scenario

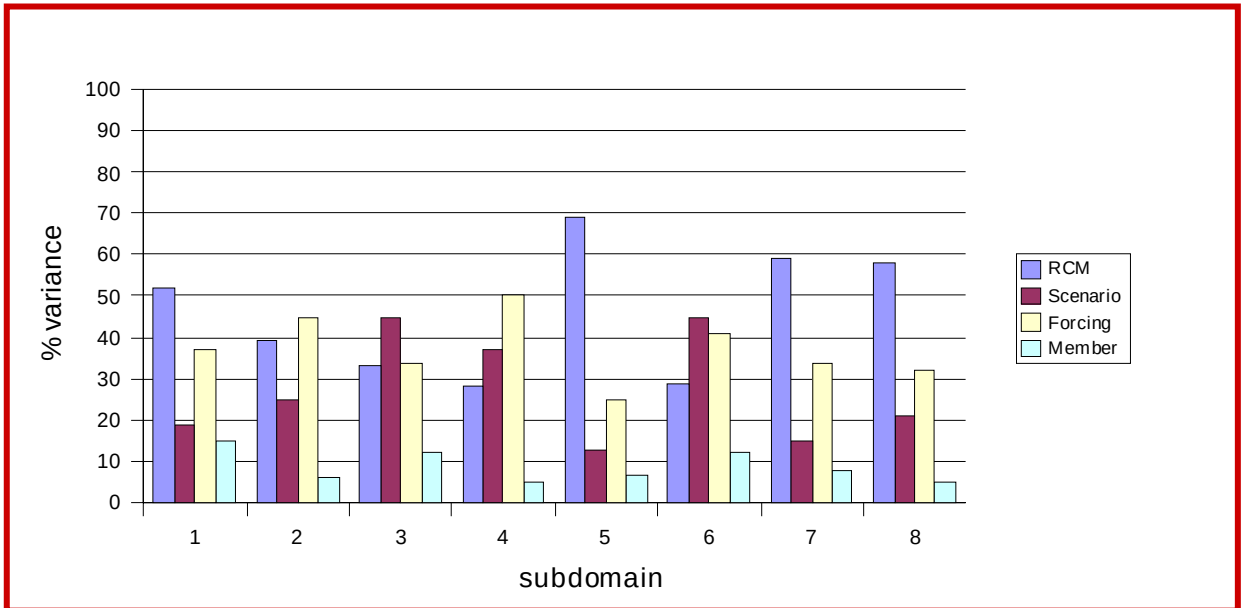
Déqué et al. 2007

Precipitation change – sources of uncertainty



DJF

Driving GCM and RCM

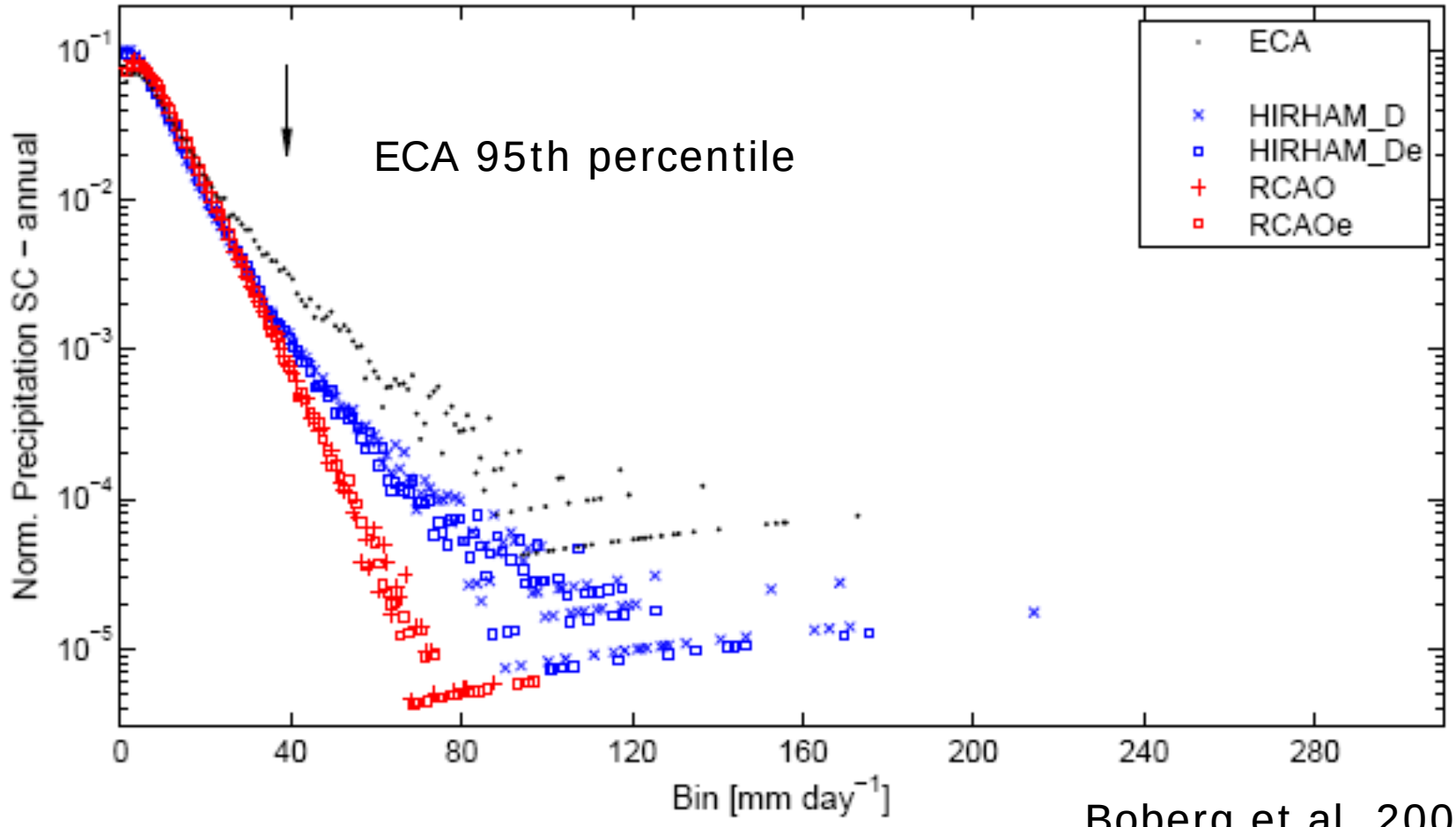


JJA

RCM quite important

Déqué et al. 2007

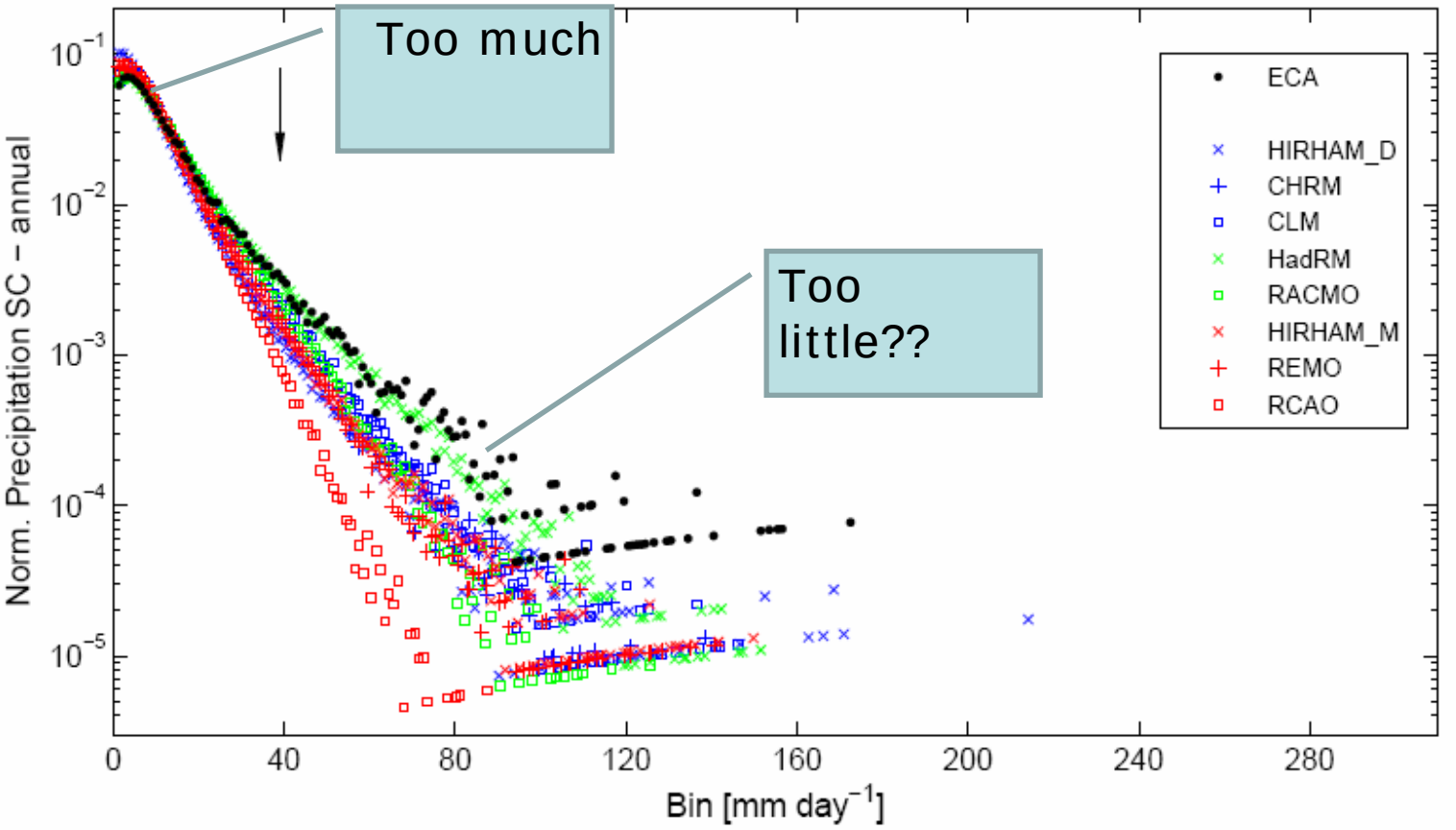
RCM-characteristic results



Boberg et al. 2007

-Note that we compare grid averages with point observations

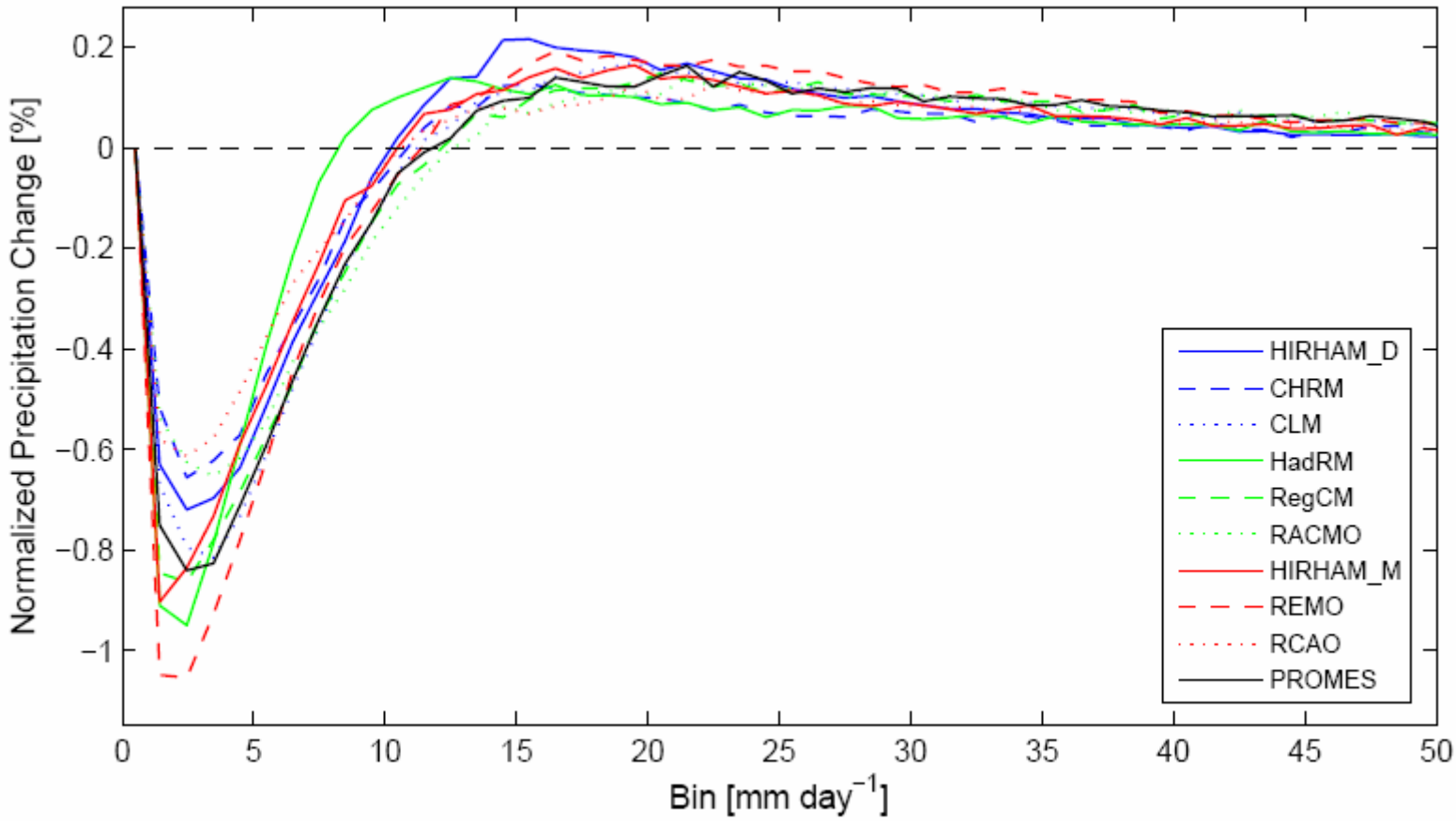
Normalized Probability Density Functions (PDFs) can be used to show the range and intensity of precipitation



Overlap of normalized area ~90%

-Note that we compare grid averages with point observations

Difference in the PDF for control and scenario (PRUDENCE)



A skill score metric –Prerequisite for probabilistic statements

Cumulative minimum of two distributions.

$$S_{\text{score}} = \sum_{1}^{n} \text{minimum}(Z_m, Z_o)$$

Takes a value between 0 and 1

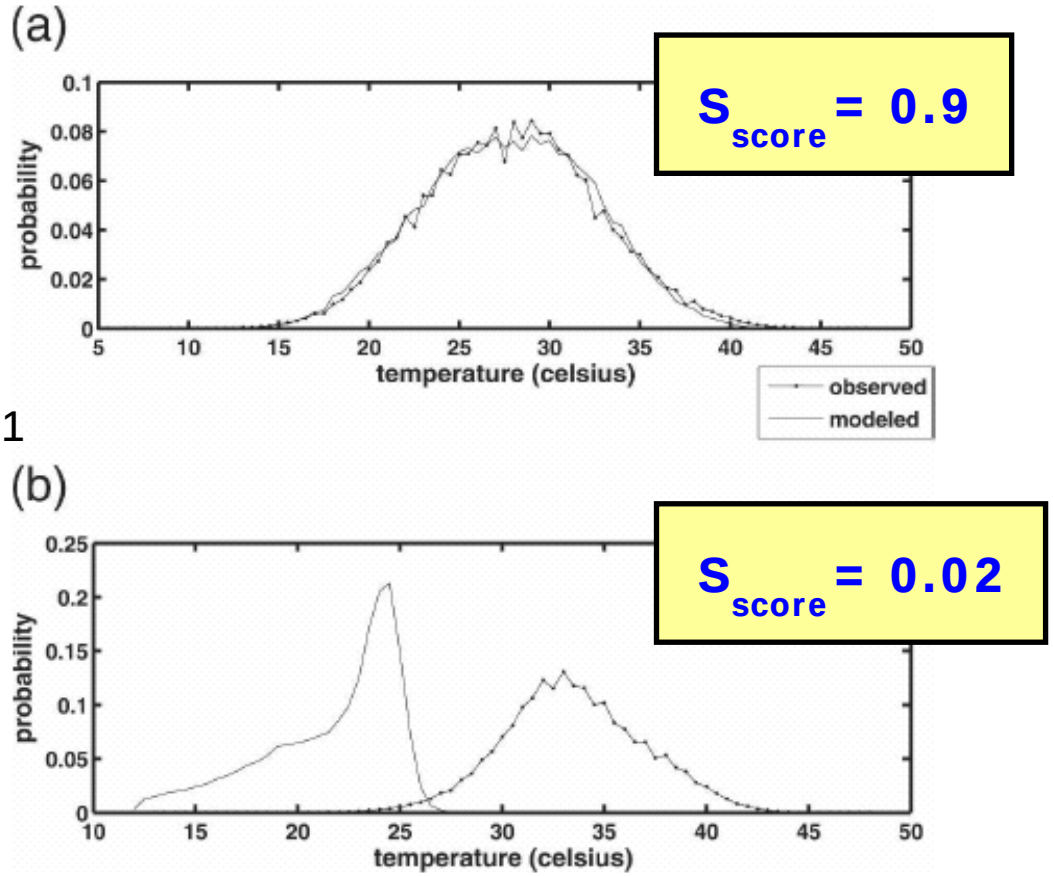


FIG. 3. Diagrams of modeled vs observed PDF illustrating the total skill score in (a) a near-perfect skill score test (0.9) and (b) a very poor skill score (0.02).

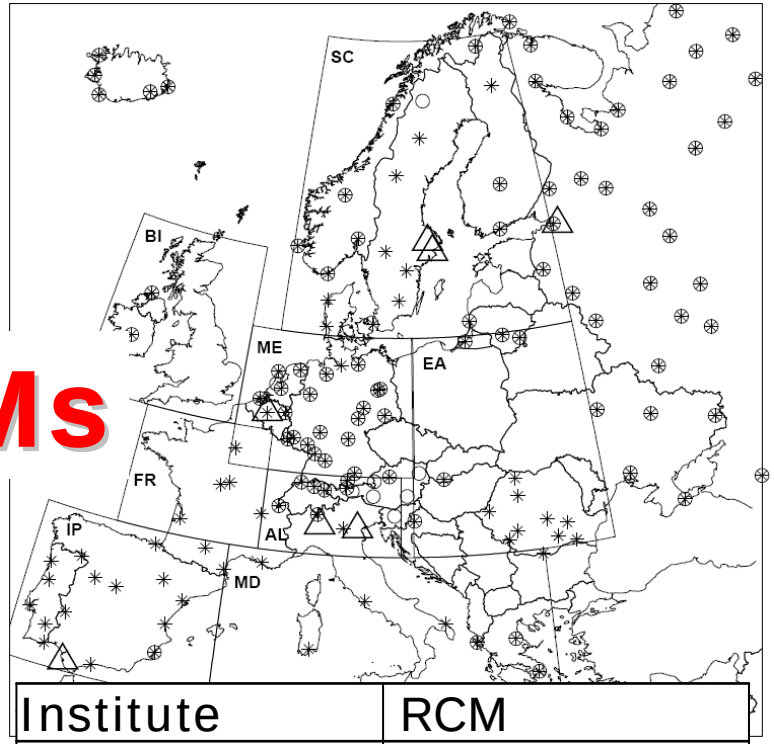
Perkins et al., *J. Clim., 20,* 2007

Observations, models and regions

- **Data**:
  - Observations from ECA set are used
  - 8 ENSEMBLES RCMs that have been run at 25 km are included

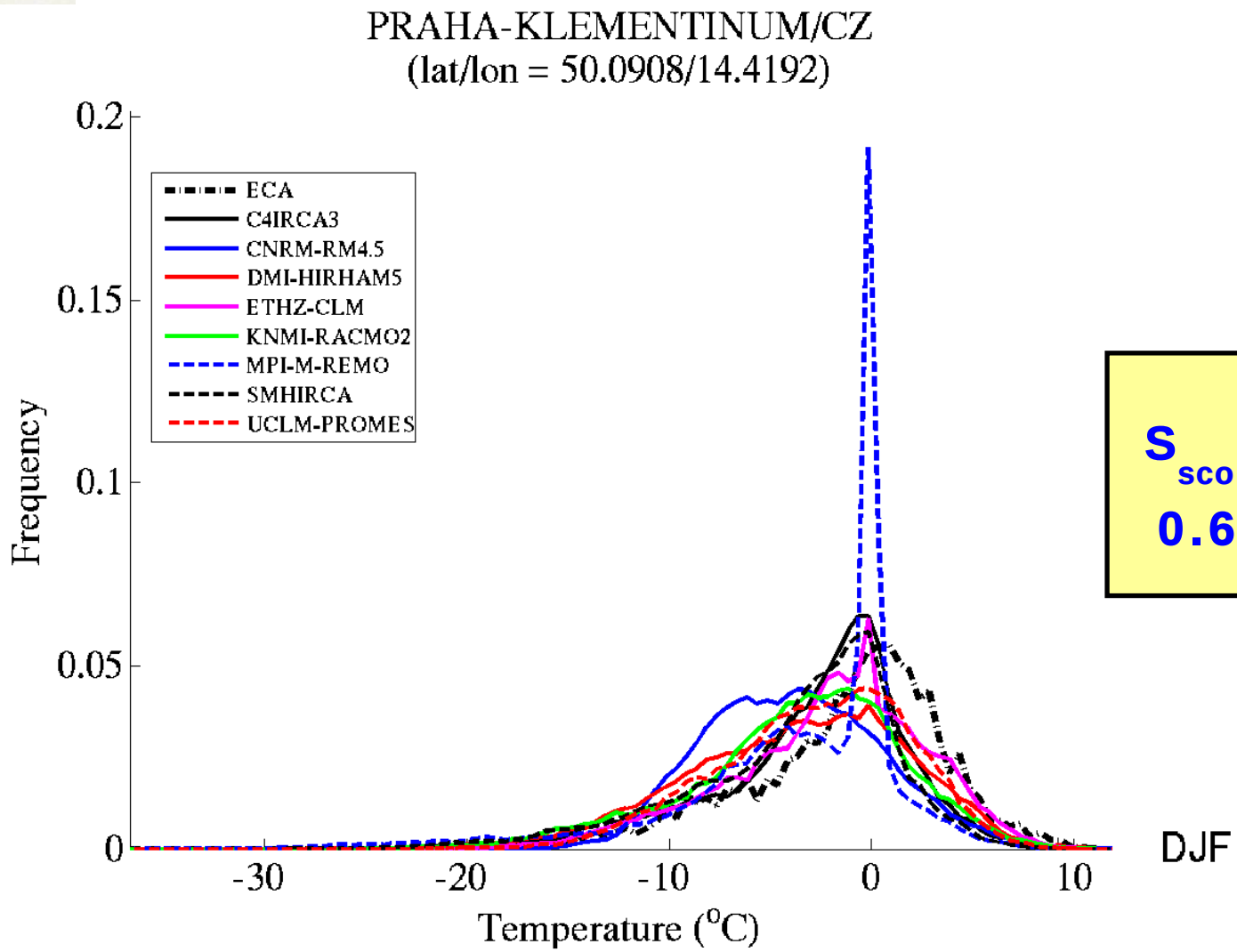
8 ENSEMBLES RCMs

  - Binning of data in 0.5K (mm/day) intervals
  - Calculation of skill scores for each ECA station, no adjustment due to differing altitudes between model and observations
  - Skill scores for all available stations within 8 areas are averaged (In each area only model grid points with a land fraction larger than 0.9 are included)



Institute	RCM
C4I	RCA3
CNRM	RM4.5
DMI	HIRHAM5
ETH	CLM
KNMI	RACMO3
MPI-met	REMO
SMHI	RCA3
UCLM	PROMES

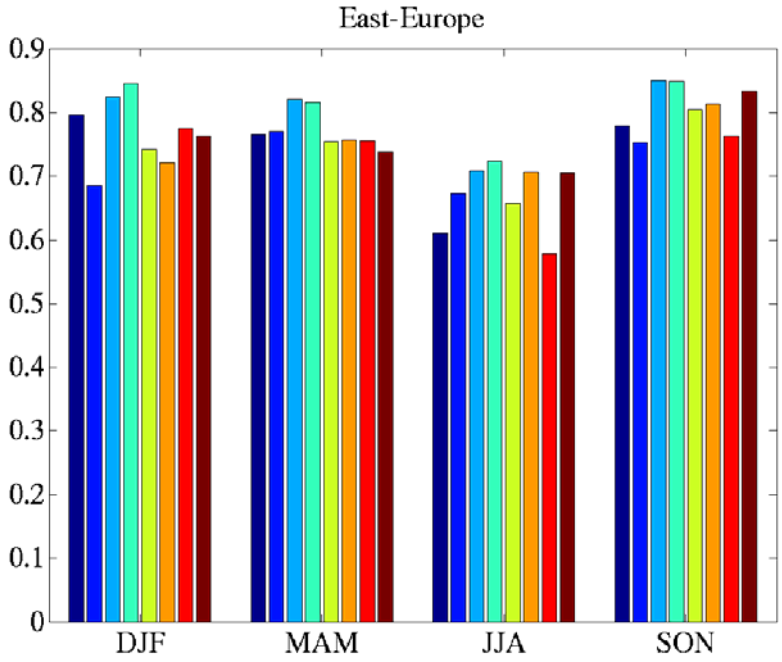
A winter example: Frequency distributions of daily T_{min}



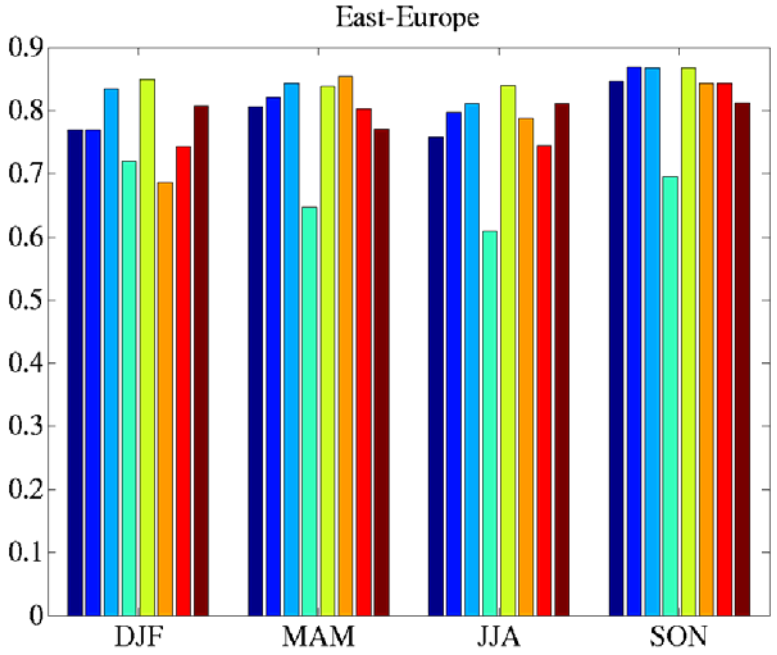
S_{score} between 0.68 and 0.87

Area aggregated skill scores: Eastern Europe

T_{min}



T_{max}

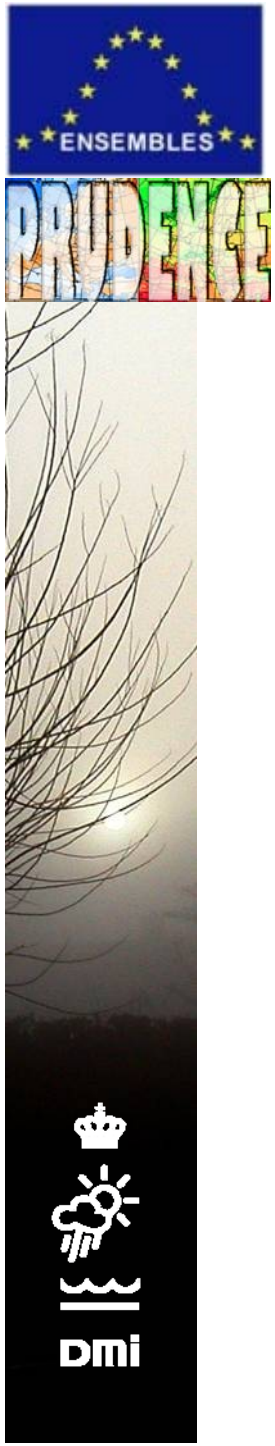


Robustness

	Full year - 1st rank [%]								
Model	BI	IP	FR	ME	SC	AL	MD	EA	EUx
HIRHAM_D	0.14								
CHRM	19.64								
CLM	4.31	11.01	76.80	4.34	19.78	2.90	27.37		89.94
HadRM	23.18	88.84	16.95	24.07	72.34	69.10	0.75	8.01	10.06
RegCM	17.06		6.06	71.59				91.99	
RACMO	15.79	0.15			7.88	21.45			
HIRHAM_M	1.70		0.05			0.07			
REMO	10.63								
RCAO	6.51								
PROMES	1.04		0.14			6.48	71.88		

10,000 Monte Carlo bootstrap trials were performed for all ten models in each region and the percentage of occurrences of first place ranking is recorded in columns 2 through 10.

So, looking at a single variable, the skill score seems rather robust



Summary

- This is still preliminary work. Perspectives for assigning weights to individual RCMs are under exploration.
  - Preliminary results indicate that the present generation of European RCMs are of comparable quality, so weights that seriously up- or down-grade individual models are currently not identified
  - Different models have different kinds of errors
- Ideas to expand on the GCM/RCM matrix beyond its present stage using statistical approaches are being pursued
- Some open questions: Weighting schemes based on model performance, and hence PDFs, can be constructed, BUT
  - What are the relevant variables to use? Anything about climate sensitivity in the weights? Realism of processes?
  - Should we use the same weights for different target variables? The role of the RCM is different…
  - What to do with GCM/RCM combinations? $w(RCM) \cdot w(GCM)$ or $w(GCM,RCM)$?

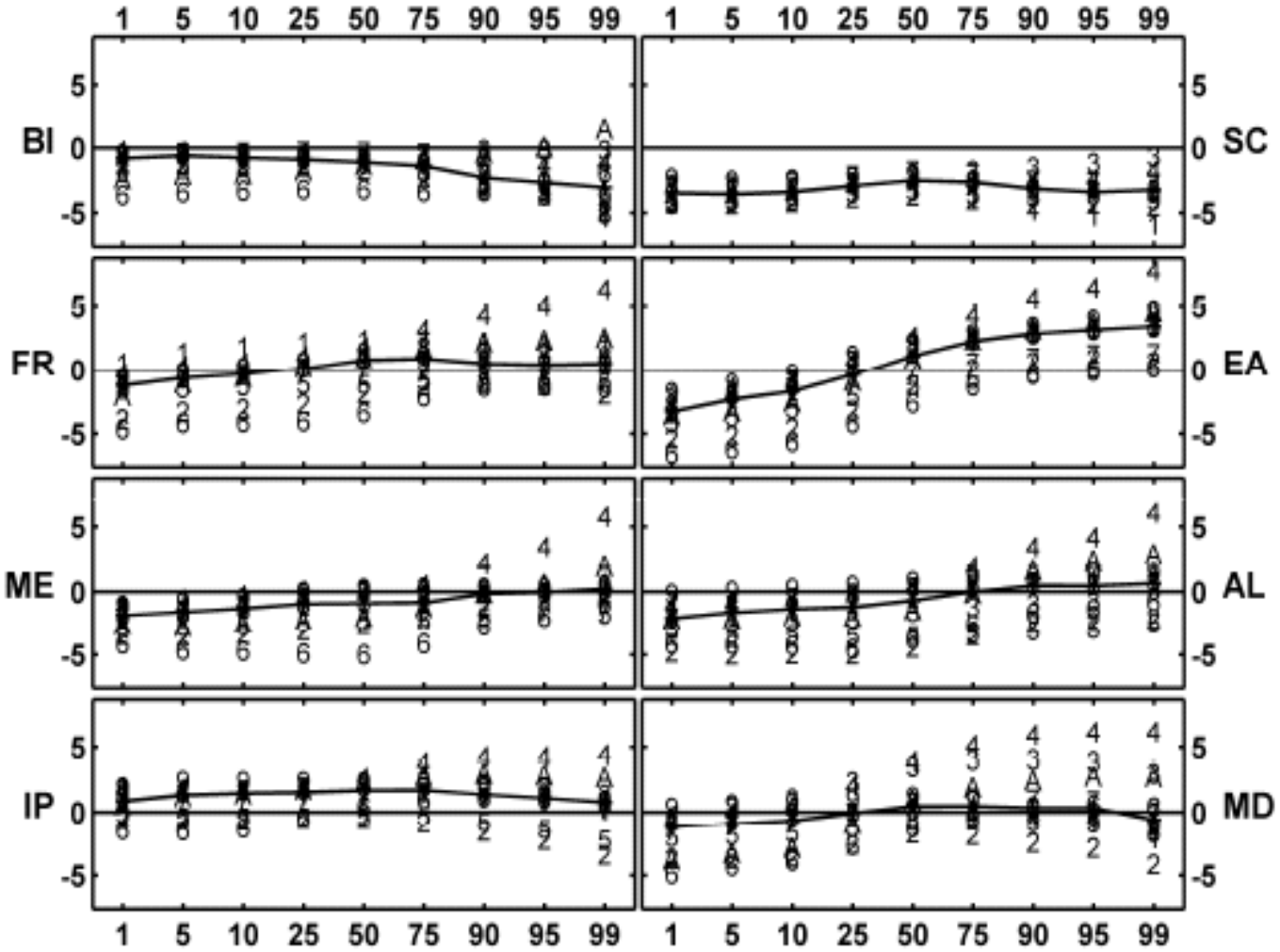
Available data

- PRUDENCE public data archive http://prudence.dmi.dk
  - A2 and B2 time slice experiments (1961-1990 vs. 2071-2100).
  - 21 fields, many experiments, around 500GB data. netCDF and OpenDAP interface
- ENSEMBLES RT3/RT2b data archive will soon be public at http://ensemblesrt3.dmi.dk
  - ERA-40 reanalysis in 50km and 25km resolution
  - Also 100-year or 150-year transient downscaling runs in 25km resolution, 1950-2050 or 1950-2100.
  - Same interface, around 120 daily fields, 7TB data and still growing

PRUDENCE biases as a function of percentile

- Biases in extremes larger than biases in means/medians
- PRUDENCE experience shows biases in the 5th percentile of daily minimum temperatures in winter and at the 95th percentile of daily maximum temperature during summer is smaller than ±3°C (±5°C) when averaged over most (all) European sub-regions.

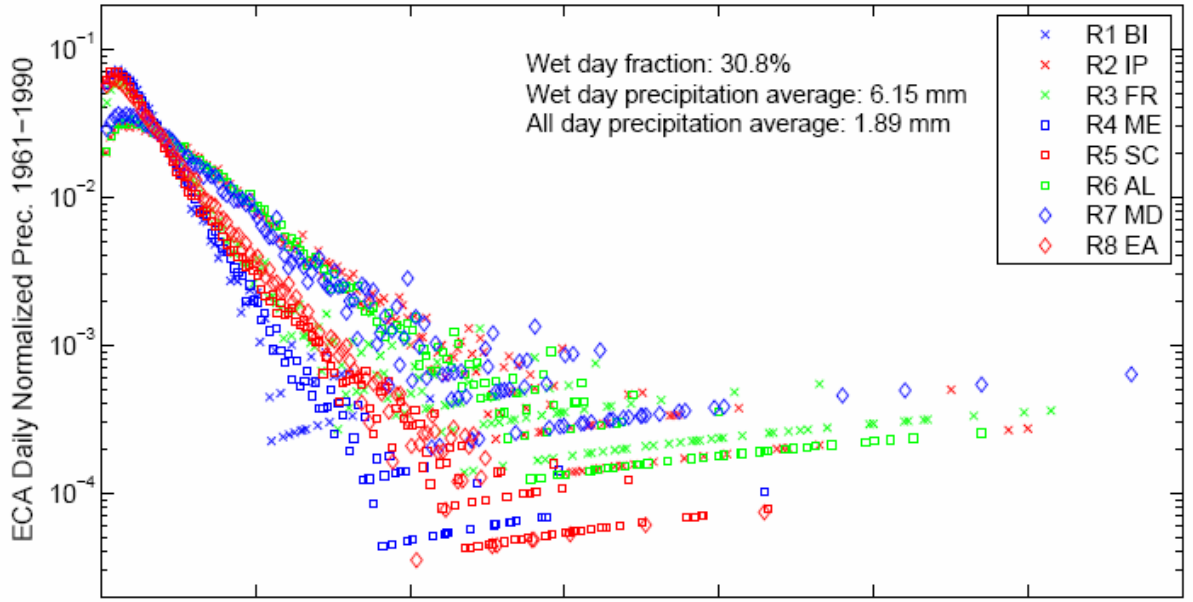
Results are highly dependent on choice of RCM

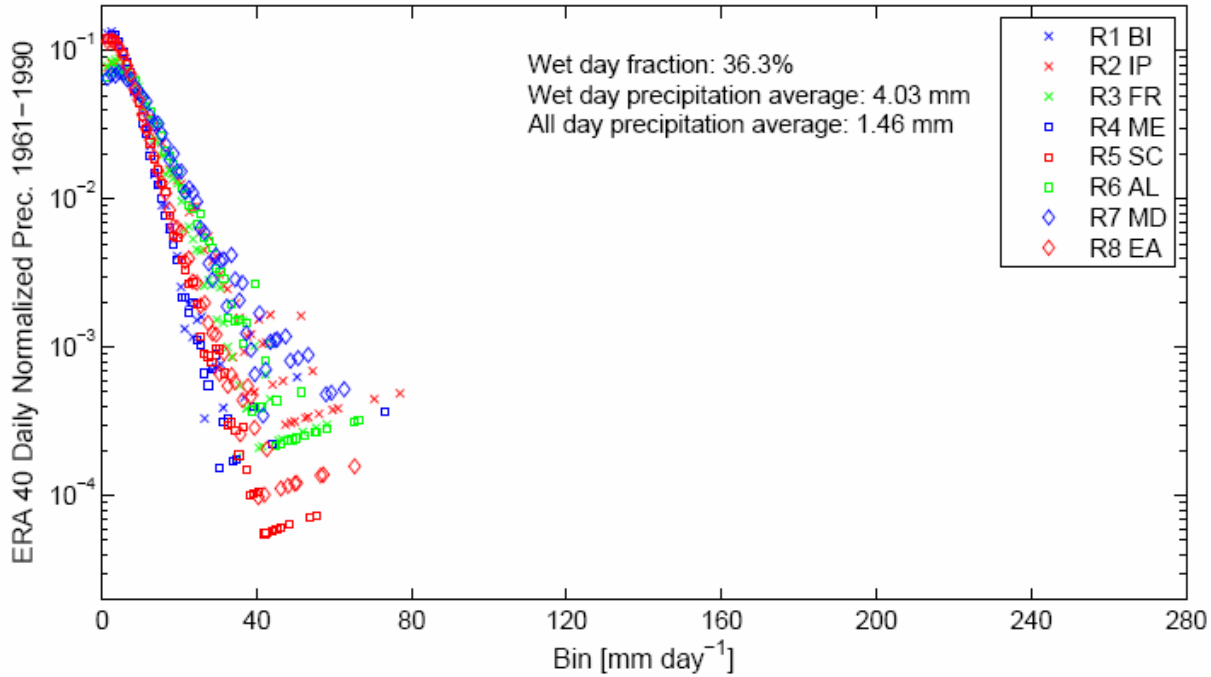


JJA, T_{2m}max, biases at different percentiles

- 10 RCMs in the common PRUDENCE experiment:
- One GCM (HadAM3H)
- Same time period (1961-1990)

Kjellström et al., 2007
Climatic Change





OBSERVATIONS VS ERA40

The regional added value (ICTP)

- On monthly time scale compare with CRU
- A large scale signal was first identified by carrying out a 5x5 grid point running spatial average of the PRUDENCE fields. This yields a signal roughly at a scale of 250 km, which is typical of the ENSEMBLES GCMs.
- The large scale signal is then subtracted from the full field. The anomaly signal derived in this way constitutes the mesoscale signal.