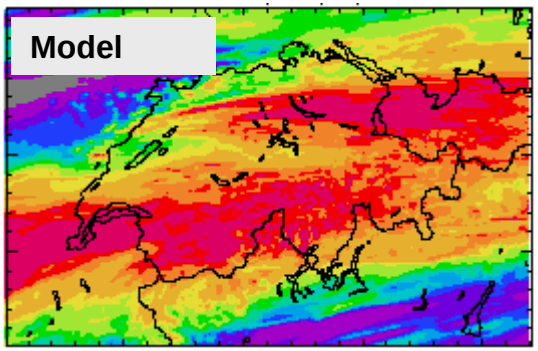
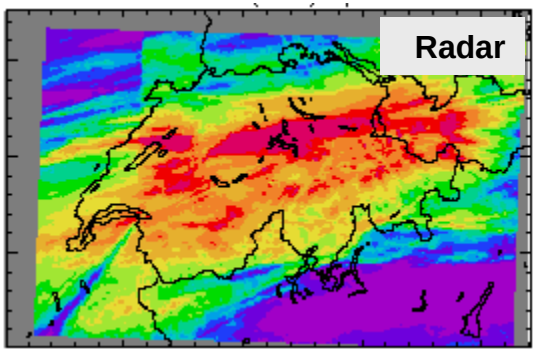
Precipitation verification during MAP D-PHASE ~~and other cases~~

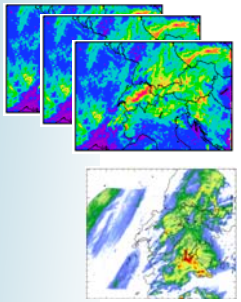
Felix Ament
Universität Hamburg





(Lyss flood, 29.8.2007)

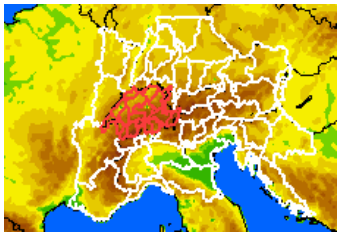
Verification – rules of the game

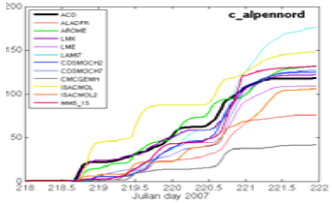


Models

OBS

domain averages

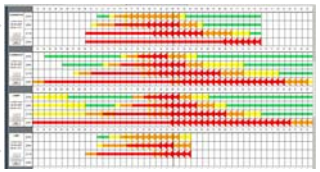




RR time series

apply warnlevels

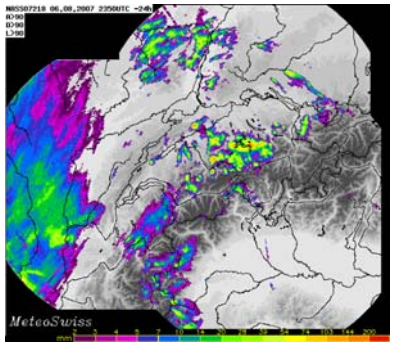
10 times a year	Twice a year	Every 10 years



Alert time series

- **Period:** Summer 2007 (June, July and August)
- **Spatial resolution:** 18 target regions in Switzerland
- **Temporal resolution:** 3 hour intervals
- **Forecast range:** Use most recent forecast, but ignore a certain cut-off time (default cut-off: 3h)
- **Models:** Almost all deterministic D-PHASE models plus IFS by ECMWF

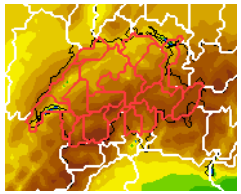
Observational data



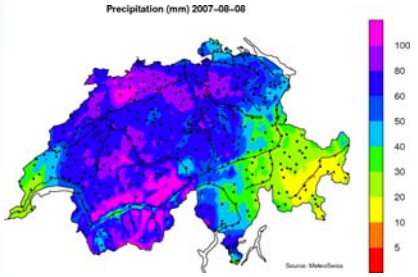
Swiss Radar composite

- 3 Radar stations
- 5 min scans accumulated to hourly estimates
- 1km resolution

Spatial average



Spatial average



Gridded rain gauge data

- Statistical interpolation + elevation correction
- Daily accumulations

Multiplicative correction to achieve match of daily accumulations

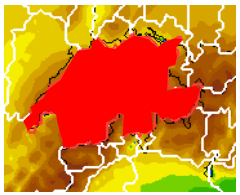
RADAR time series

- Warn regions averages
- Hourly accumulations

RADAR_CAL time series

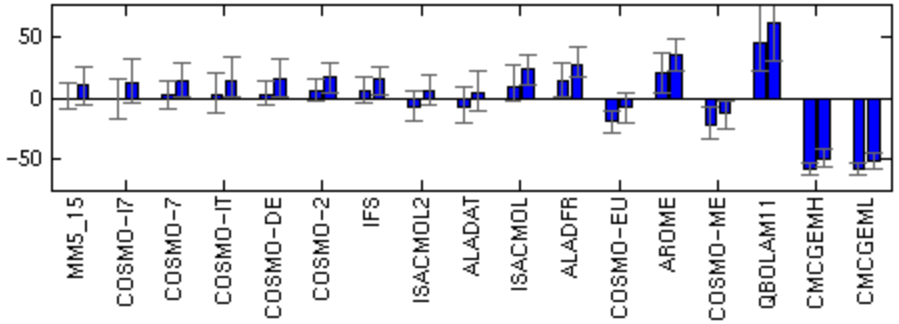
- Warn regions averages
- Hourly accumulations
- Daily sums equivalent to gridded gauge data

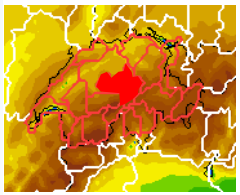
Verification of precipitation amount I



Whole Switzerland, summer 2007, relative BIAS

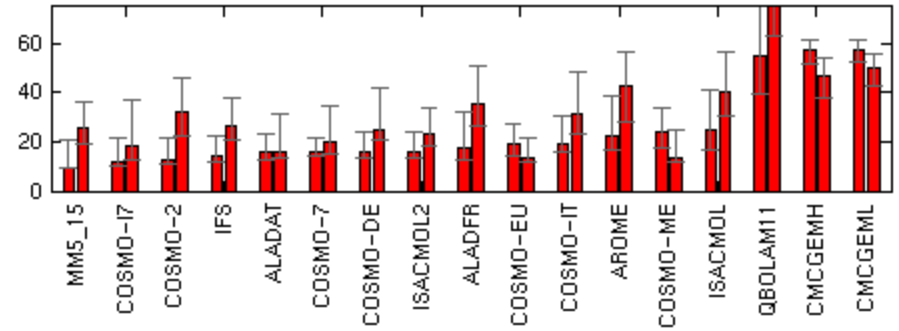
Relativ BIAS full domain (%)

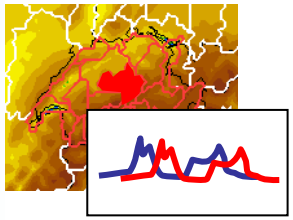




Single target region, summer 2007, relative BIAS

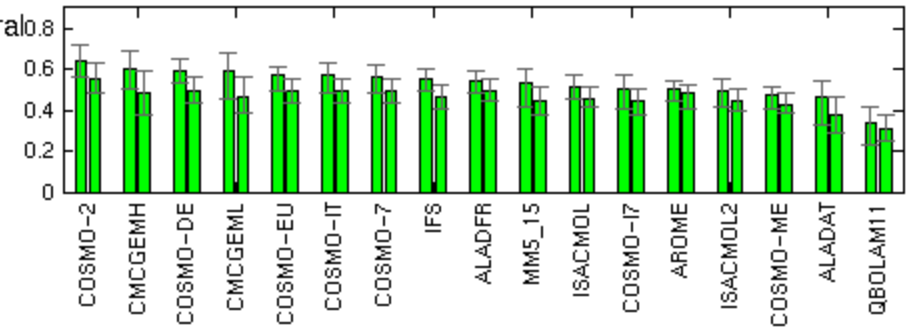
Mean |rel. BIAS| of a region (%)

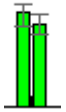




Single target region 3 hourly resolution, correlation

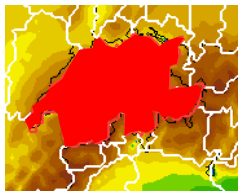
Mean temporal correlation in a region (hourly)



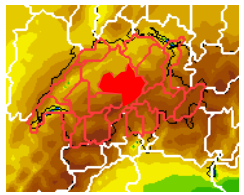
RADAR_CAL  RADAR

Verification of precipitation amount II

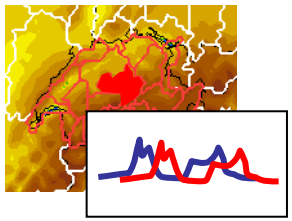
RADAR_CAL reference



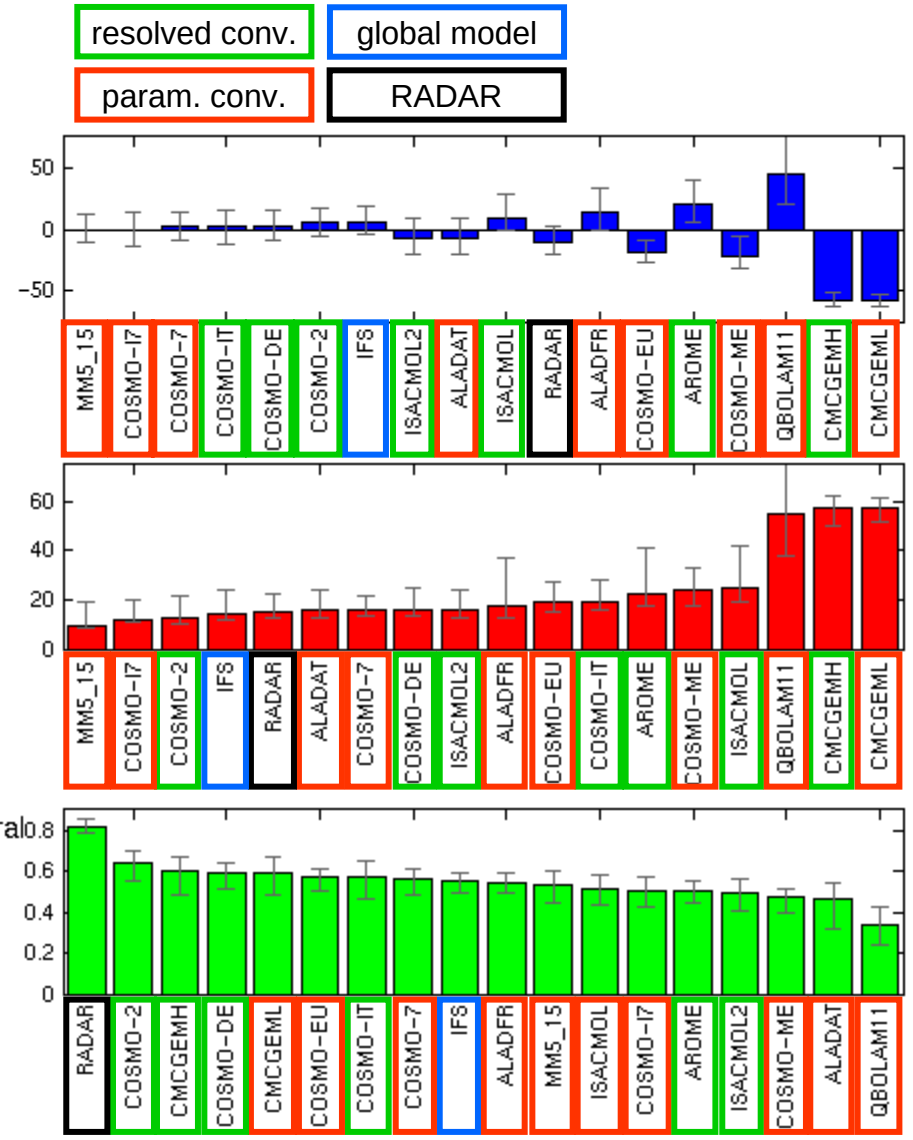
Whole Switzerland, summer 2007, relative BIAS



Single target region, summer 2007, relative BIAS



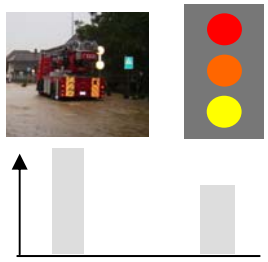
Single target region, 3 hourly resolution, correlation



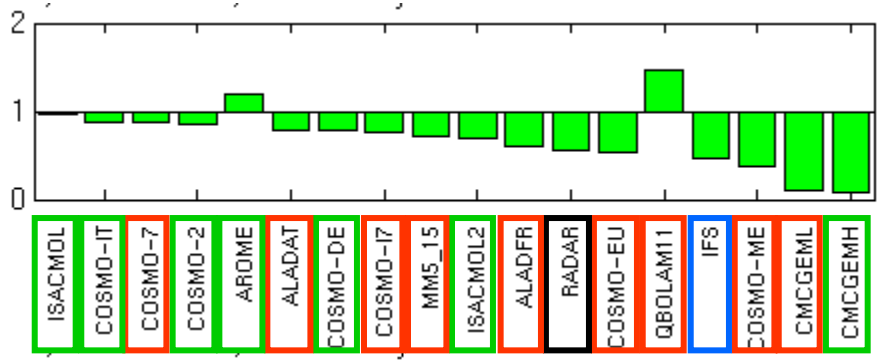
Alerts – level „yellow“, 3h intervals

(Alert level yellow = return frequency of 10 times per year)

resolved conv. | global model | param. conv. | RADAR

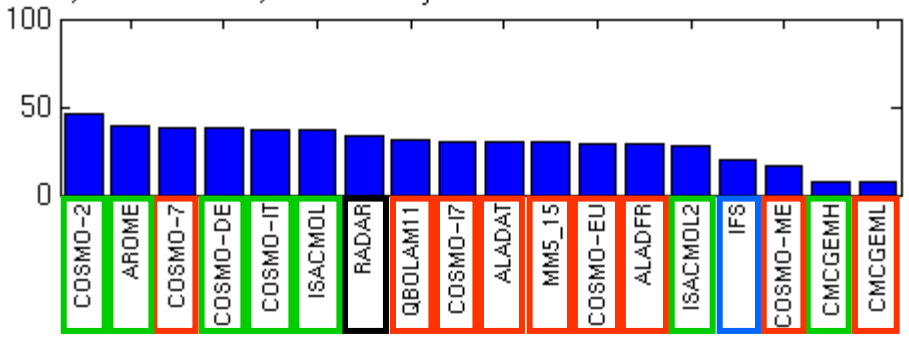


Relative frequency of an alert
(frequency bias)



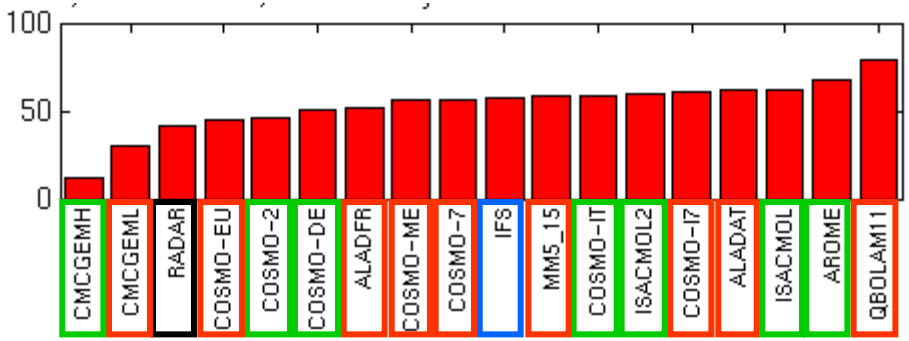
and

Probability to detect an event
(probability of detection)



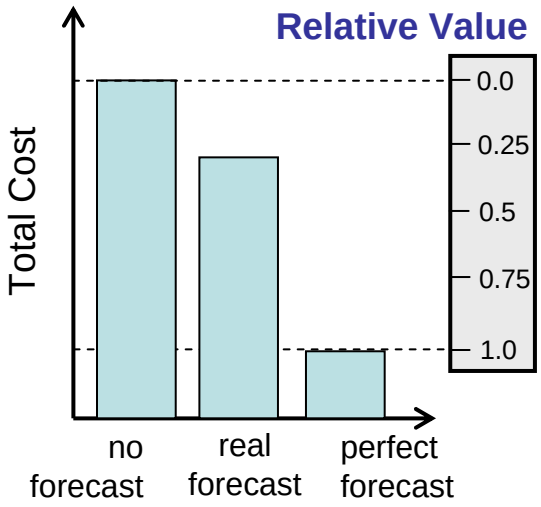
but

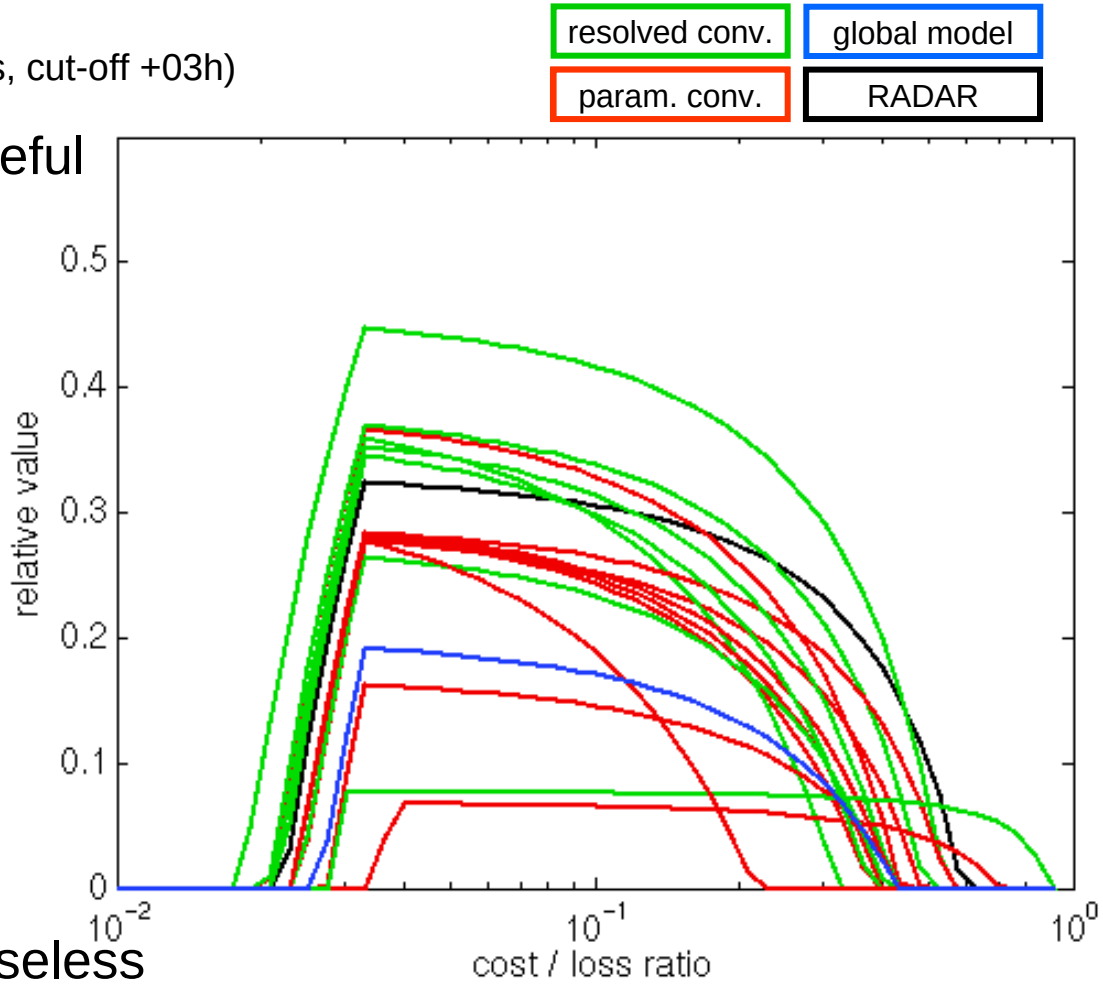
Probability to issue an false alarm
(false alarm ratio)



Relative value – Alert level „yellow“

(03h, 06h and 12h accumulations, cut-off +03h)





insensitive … sensitive …

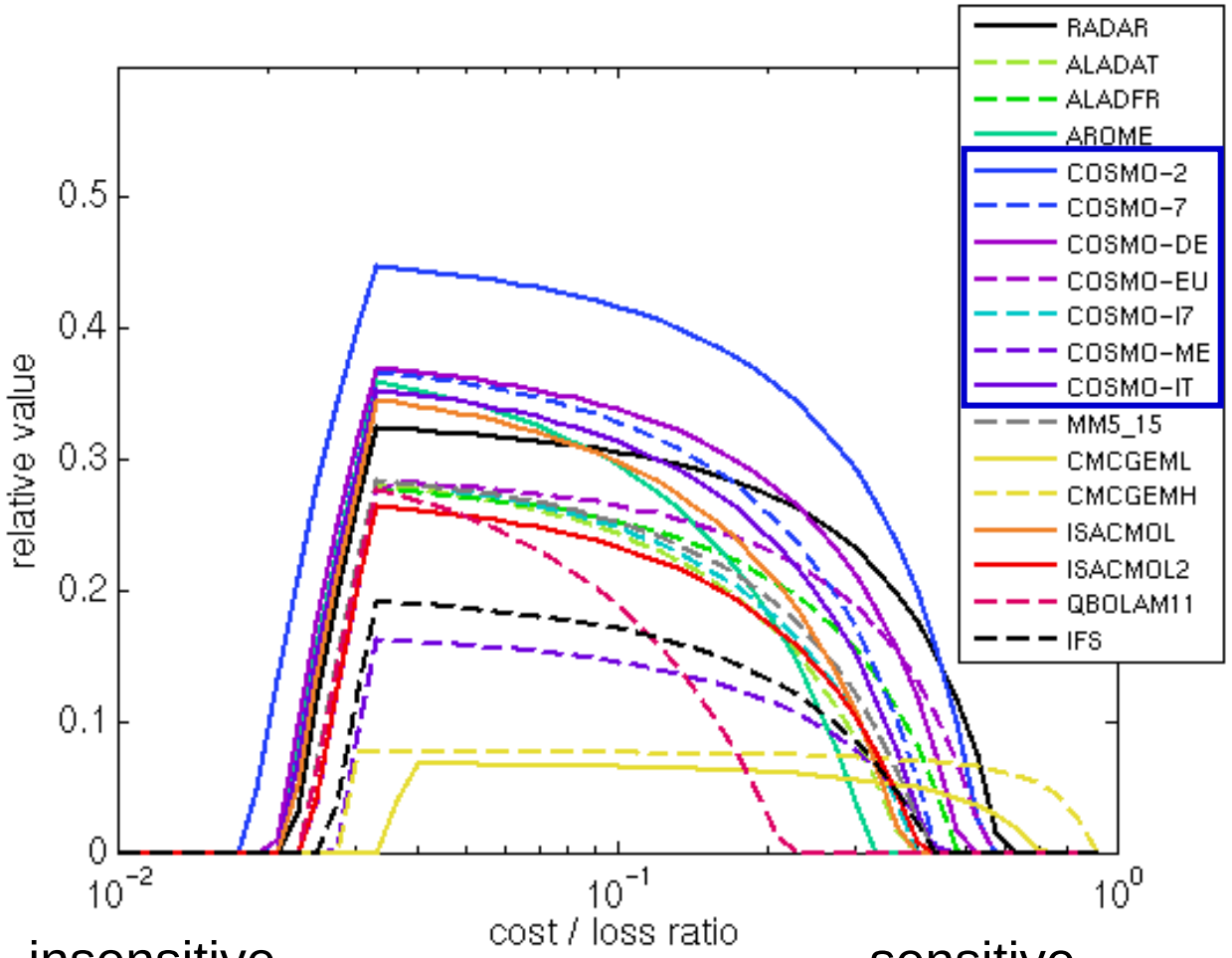
… against false alarms

Relative value – Alert level „yellow“

(03h, 06h and 12h accumulations, cut-off +03h)

+ useful

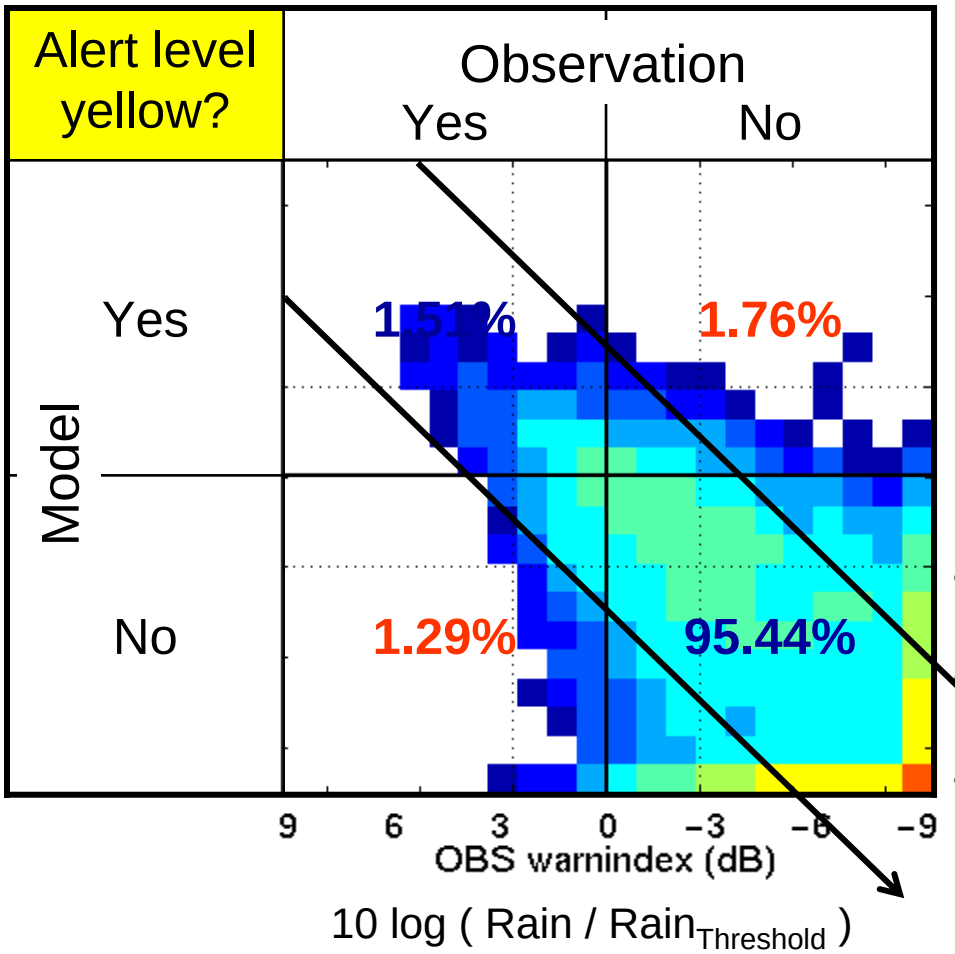
- useless



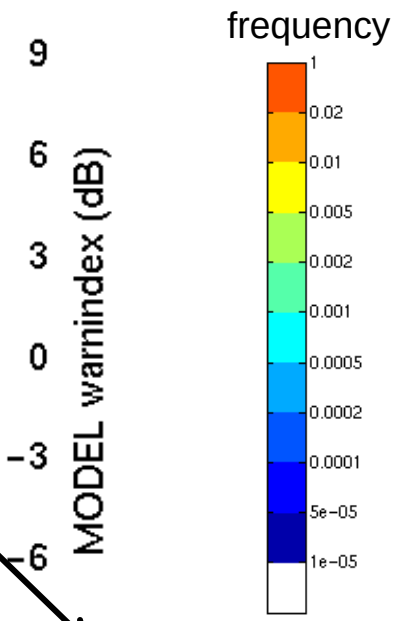
insensitive … sensitive …

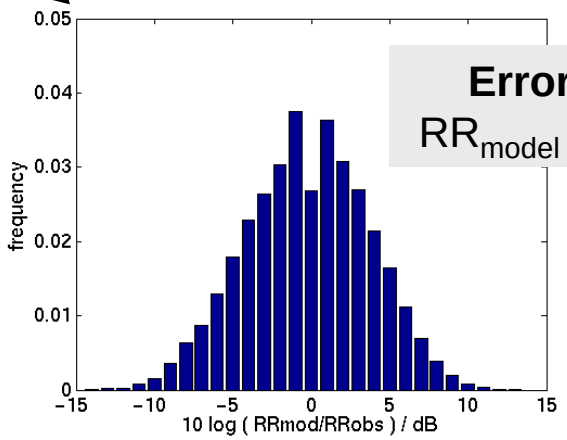
… against false alarms

Contingency table in detail



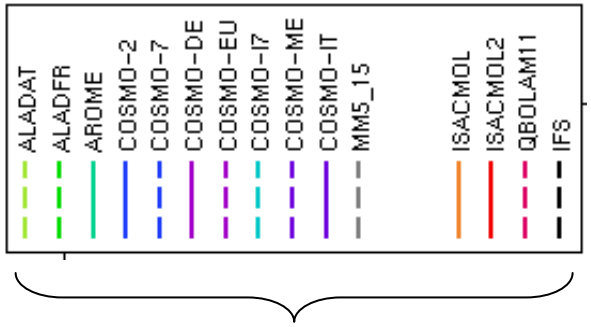
Example: COSMO-2 model





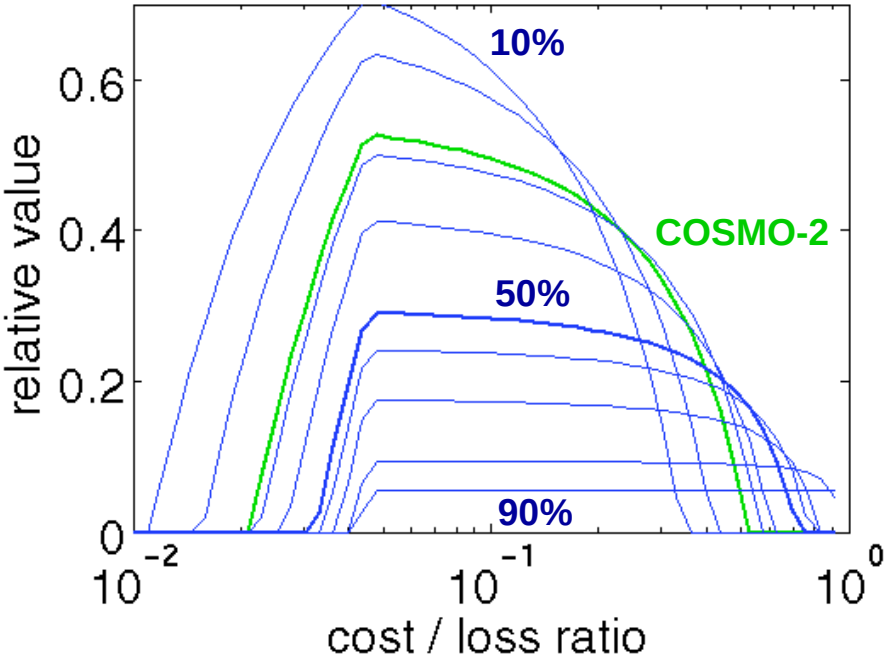
Error PDF
RR_{model} / RR_{OBS}

Think Ensemble!

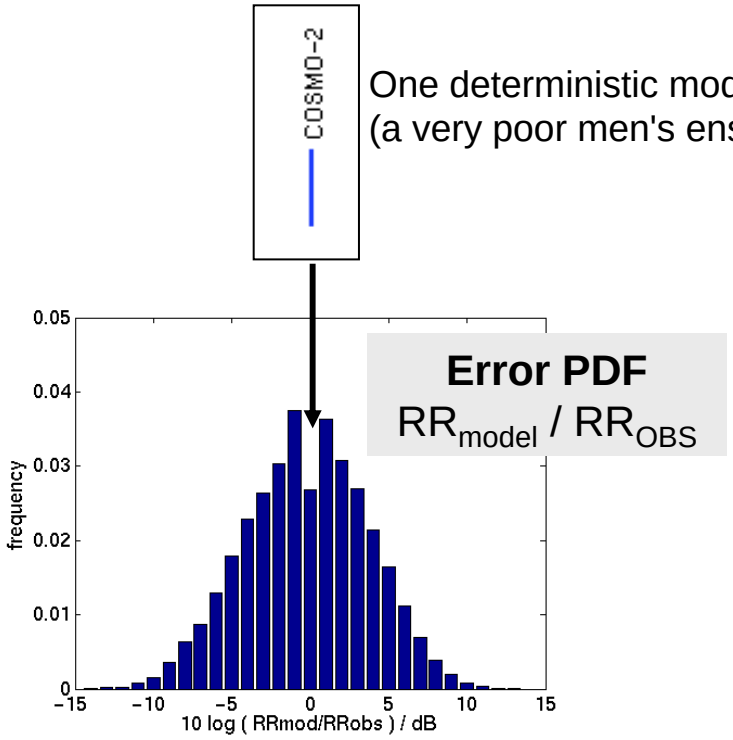


D-PHASE poor-men's ensemble

Issue an alert, if a certain fraction of models gives a warning !

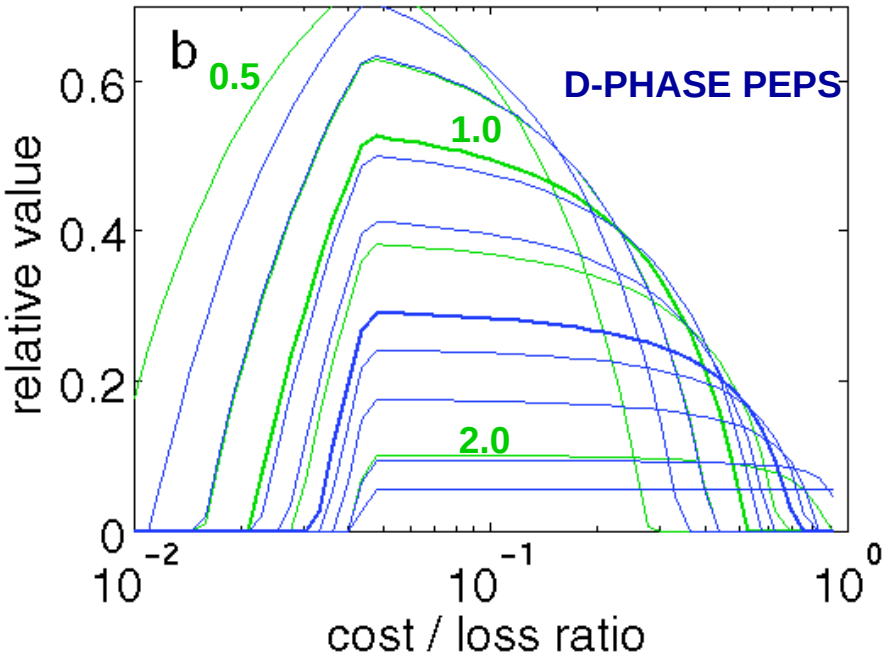


Think (ensemble) probabilistic!



One deterministic model
(a very poor men's ensemble)

Error PDF
RR_{model} / RR_{OBS}

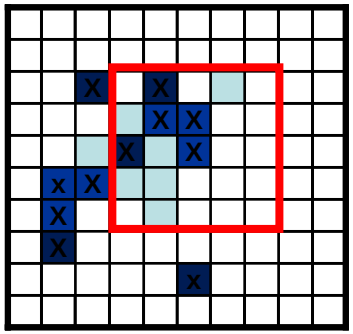
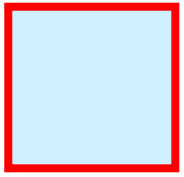
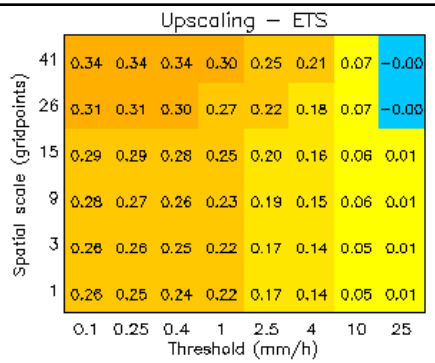
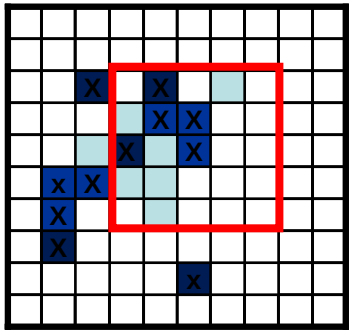
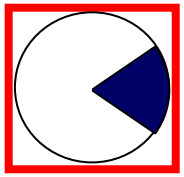
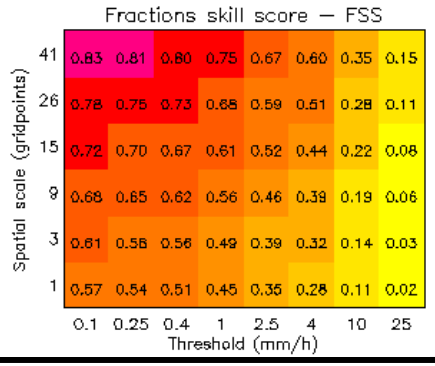


Simple calibration of COSMO-2

Multiply COSMO-2 precipitation forecasts by certain factors

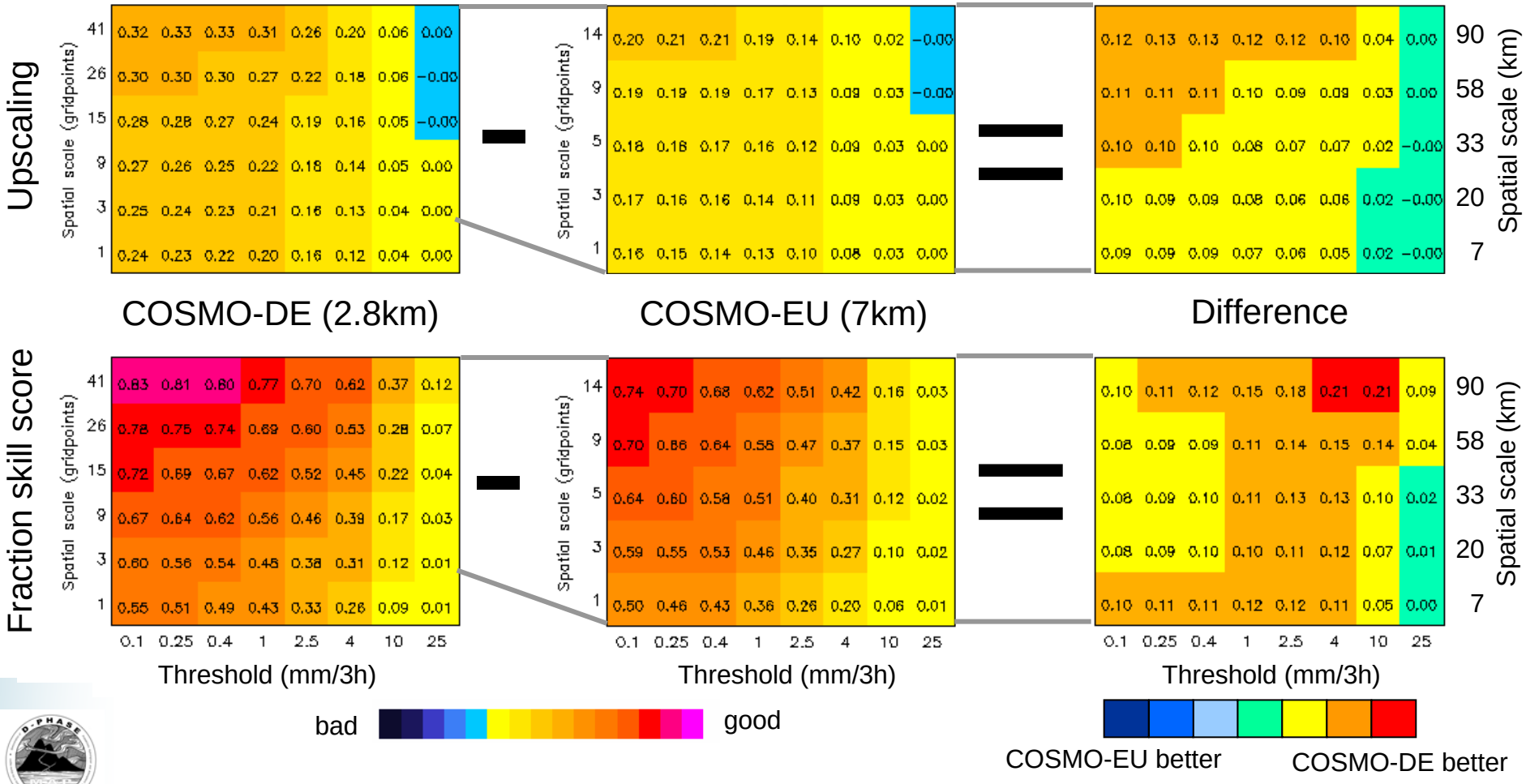
Fuzzy Verification

Verification on coarser scales than model scale: “Do not require a point wise match!“

Method	Raw Data	Fuzzyfication	Score	Example result
Upscaling		Average 	Equitable threat score	
Fraction Skill Score (Roberts and Lean, 2005)		Fractional coverage 	Skill score with reference to worst forecast	

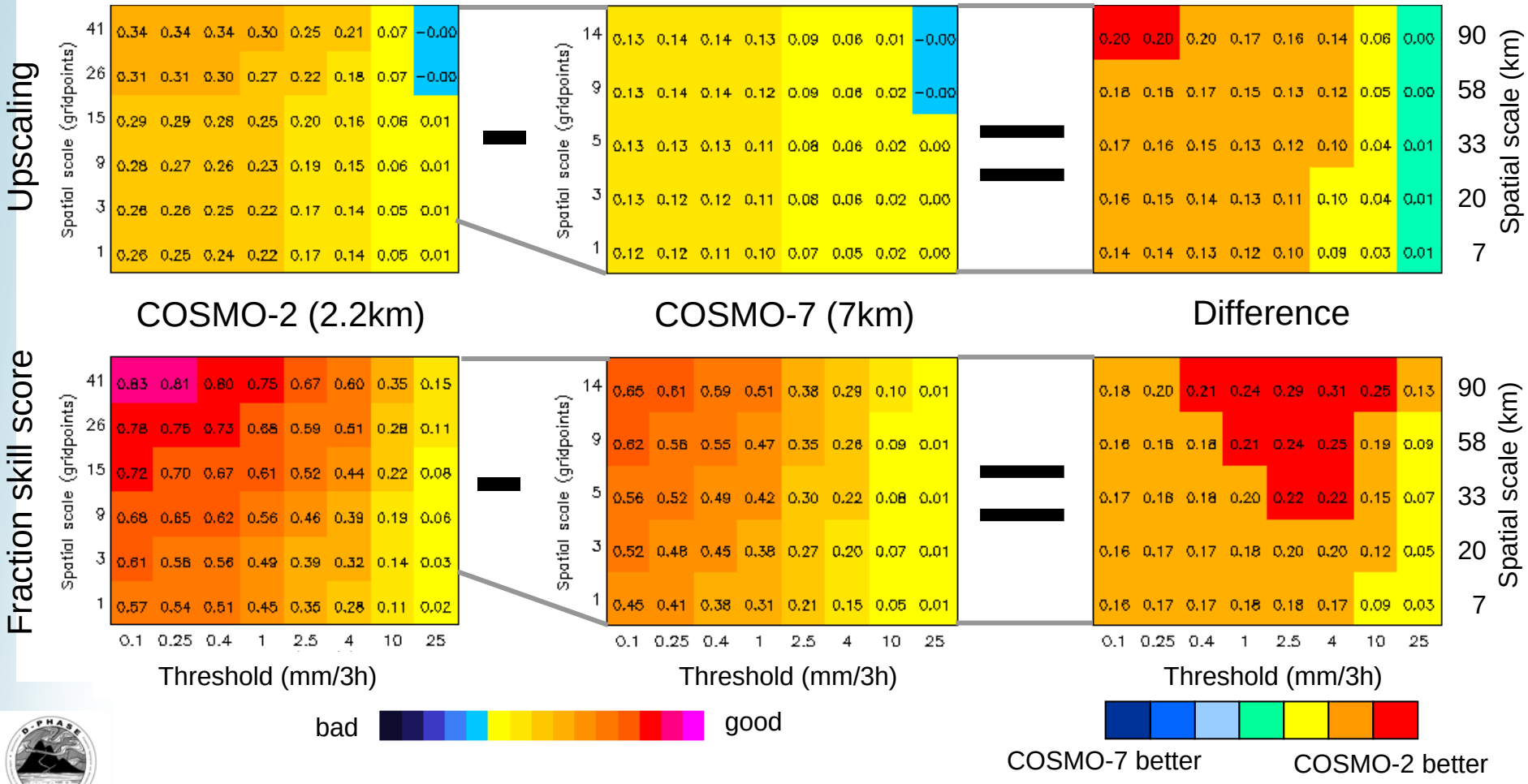
Fuzzy Verification COSMO-DE – COSMO-EU

JJA 2007, Verification against Swiss Radar Composite, 3 hourly accumulations



Fuzzy Verification COSMO-2 – COSMO-7

JJA 2007, Verification against Swiss Radar Composite, 3 hourly accumulations



Fuzzy verification of DOP - revisited

Motivation

- Testing of additional methods: Pragmatic approach, Practically perfect forecast
- How stable are the results?

Configuration

- Same as before, but
  with slightly different thresholds and inclusion of all situations
  (not only rainy ones)

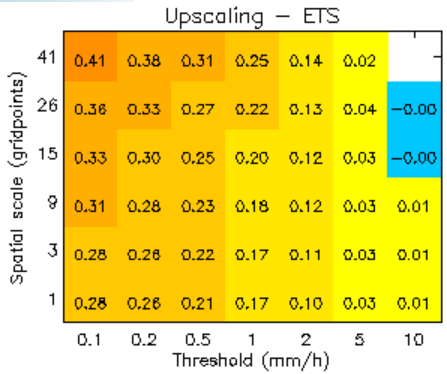
Sensitivities to be tested

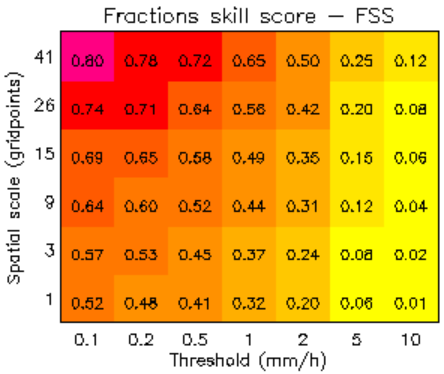
- Vary the results from month to month?
- Impact of lead time?
- Impact of accumulation period?
- Sampling error (bootstrap resampling)?

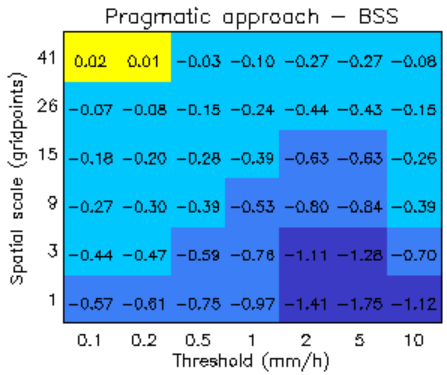
Results of COMSO-2 and COSMO-7

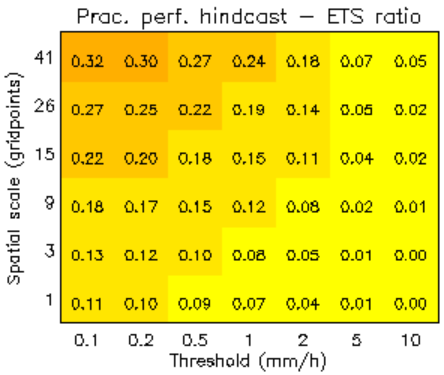
JJA 07, cut-off 03h, accumulation 03h

COSMO-2

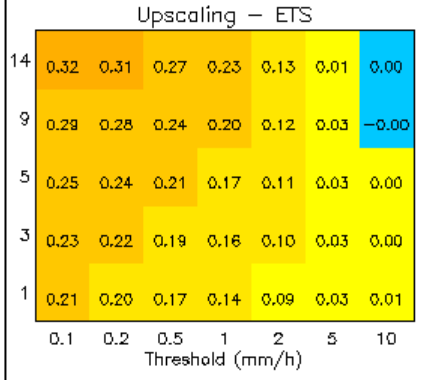


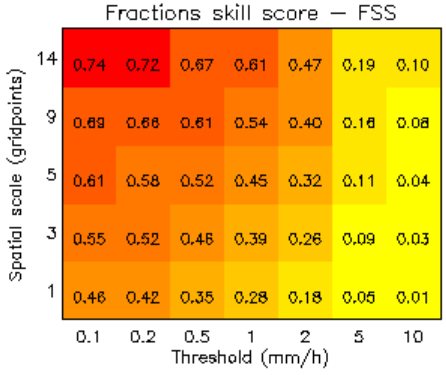


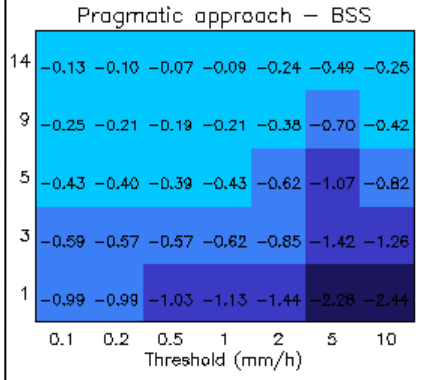


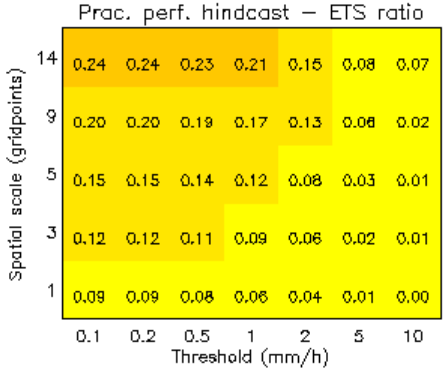


COSMO-7



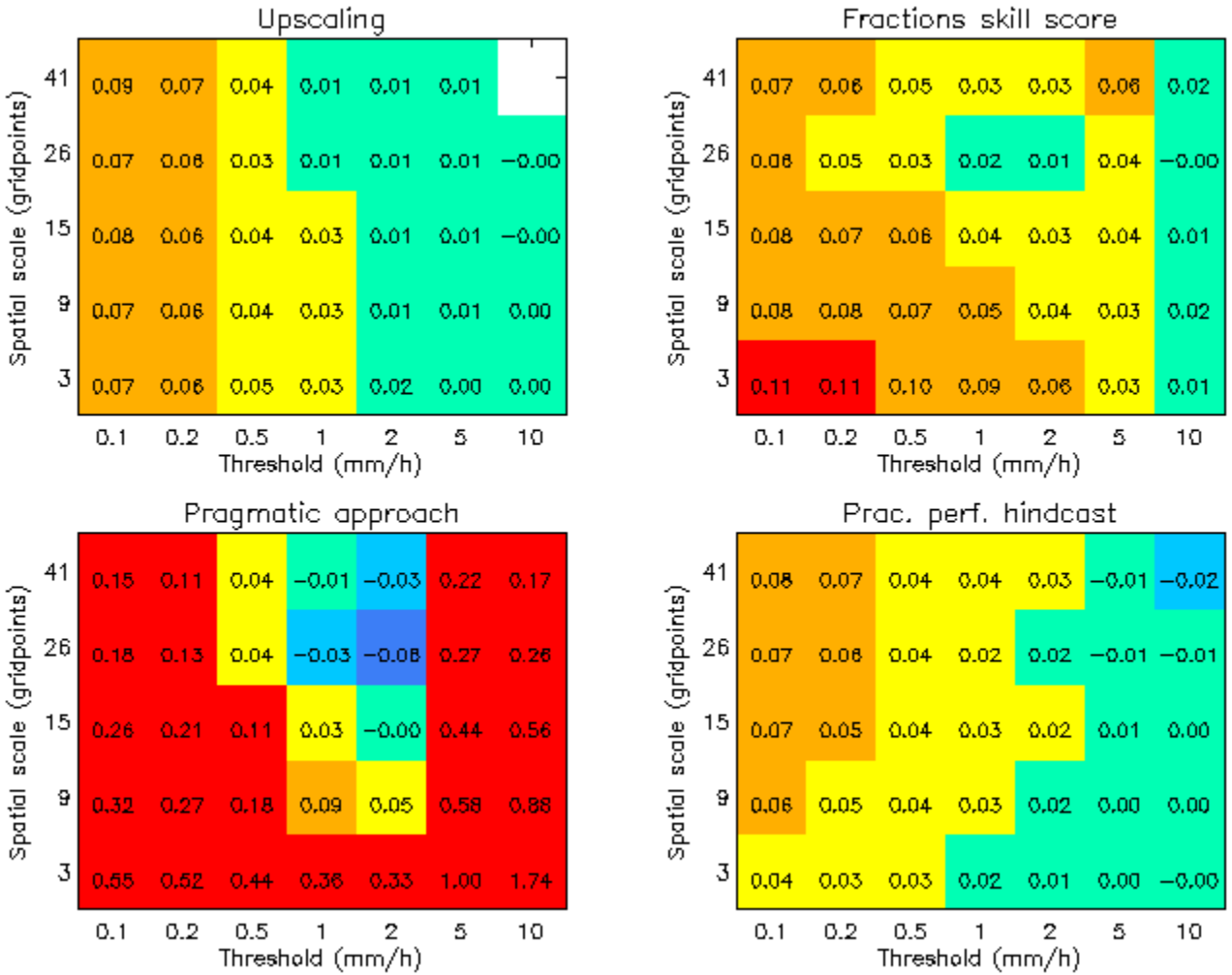






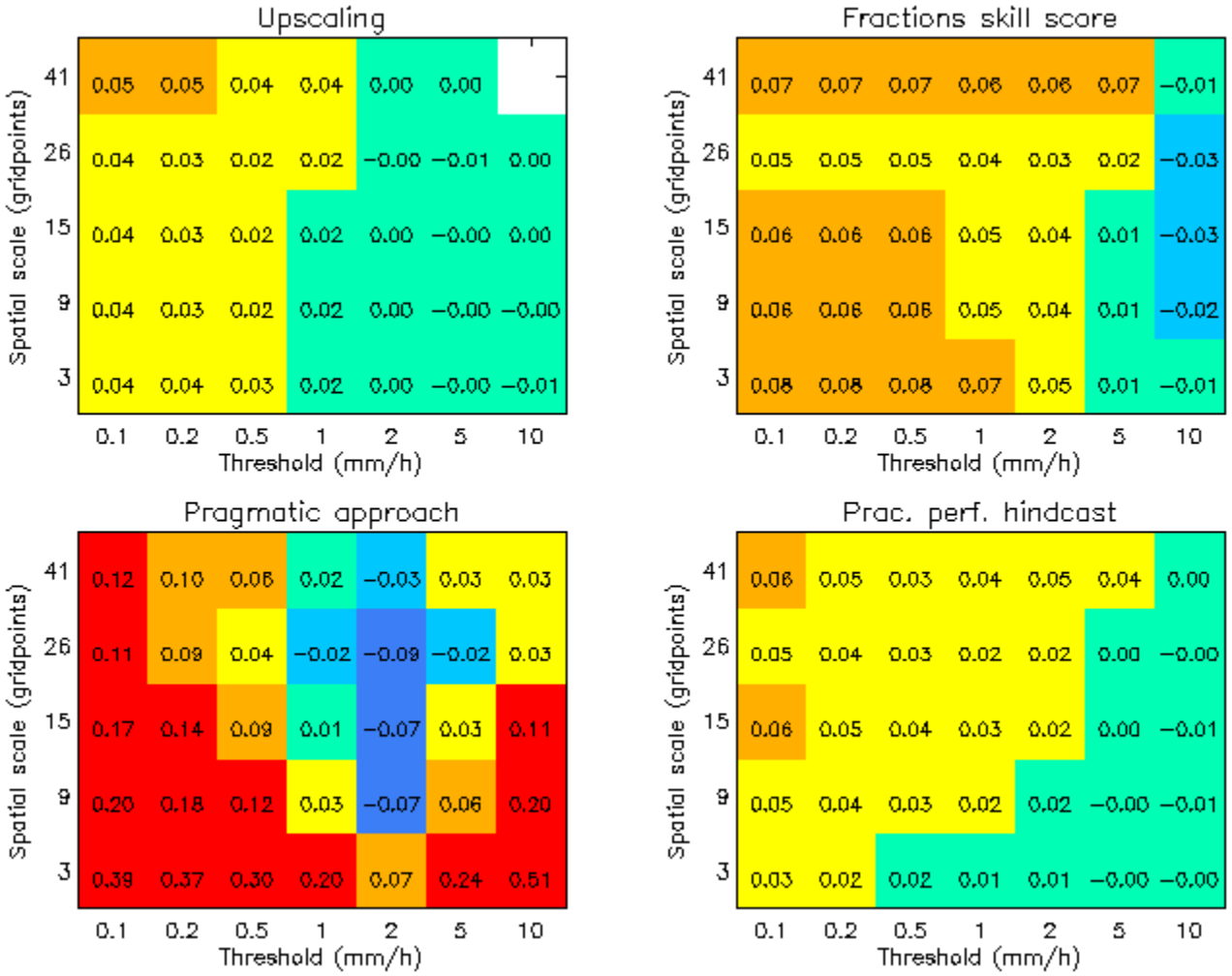
Difference COMSO-2 and COSMO-7

JJA 07, cut-off 03h, accumulation 03h



Difference of COMSO-DE and COSMO-EU

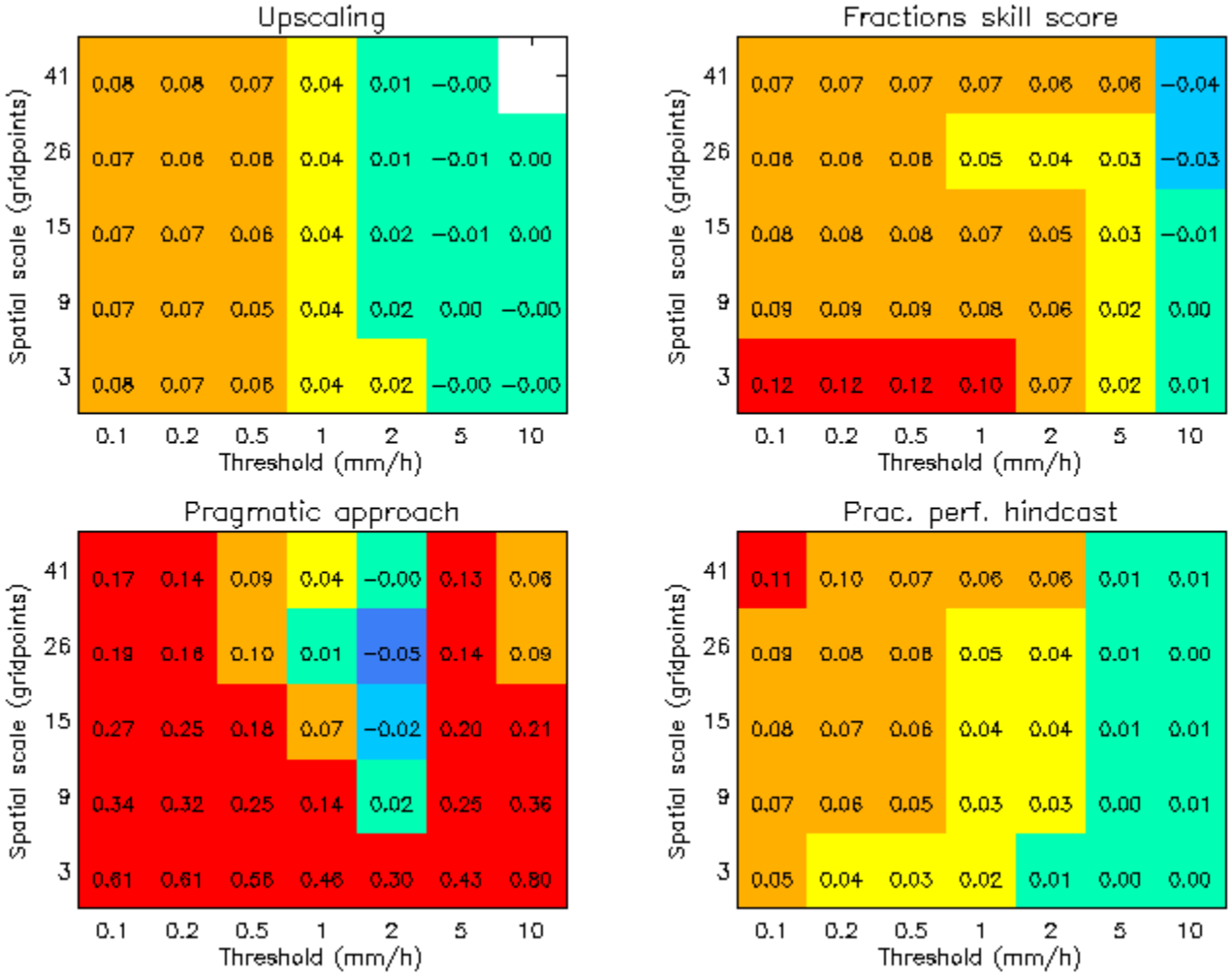
JJA 07, cut-off 03h, accumulation 03h



Felix.Ament@zmaw.de

Difference COMSO-2 and COMSO-7

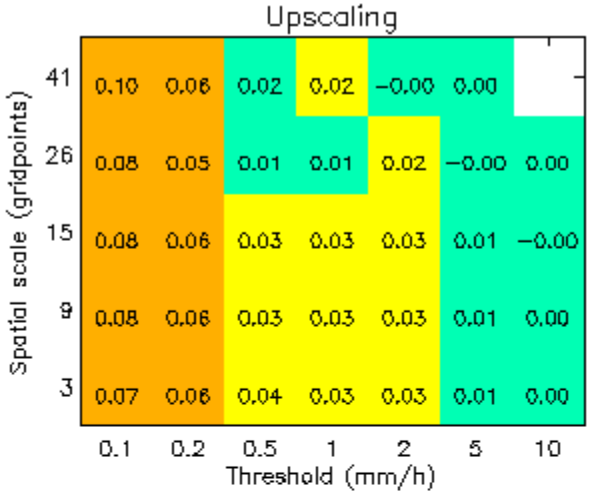
June 07, cut-off 03h, accumulation 03h

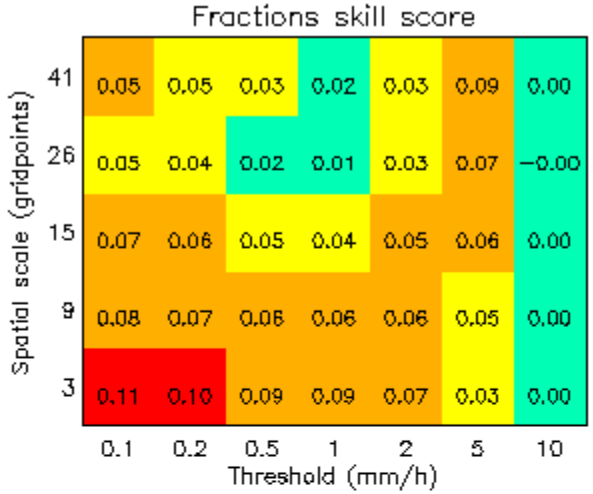


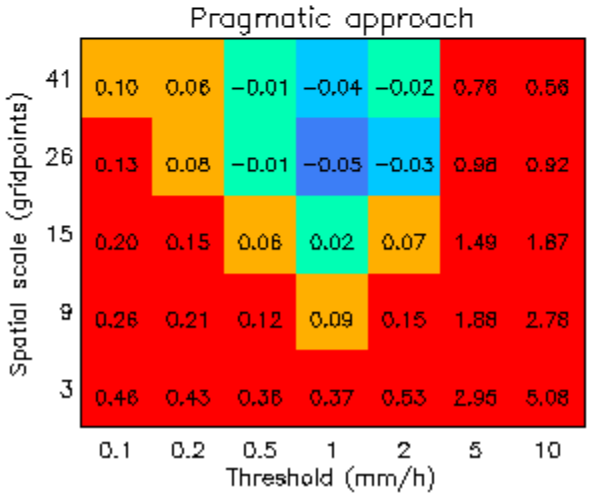
Felix.Ament@zmaw.de

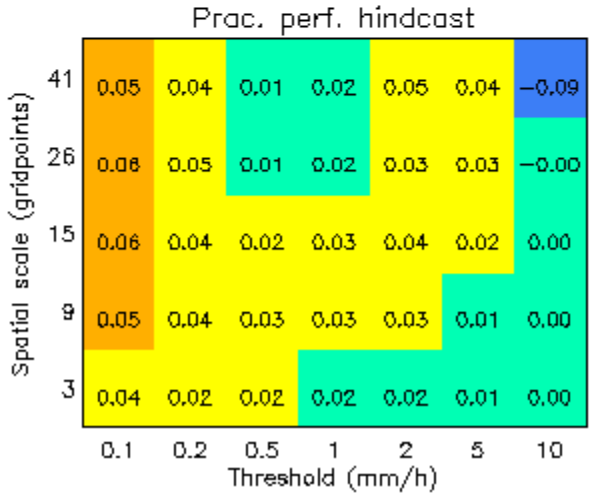
Difference COMSO-2 and COMSO-7

July 07, cut-off 03h, accumulation 03h



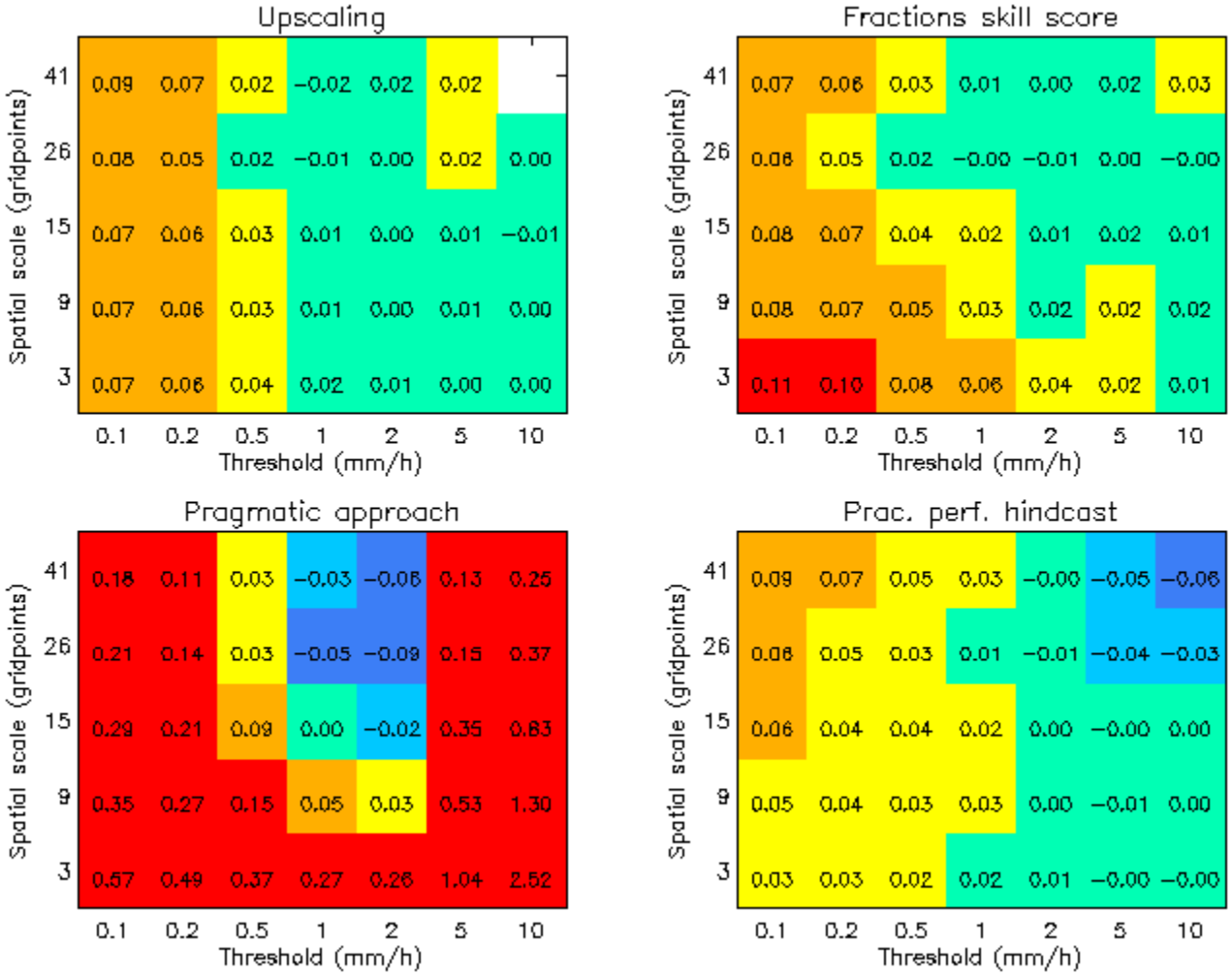






Difference COMSO-2 and COMSO-7

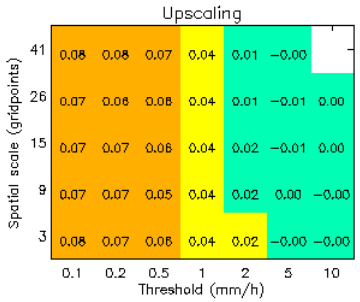
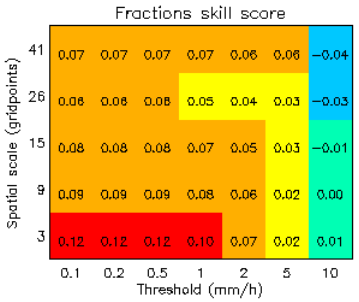
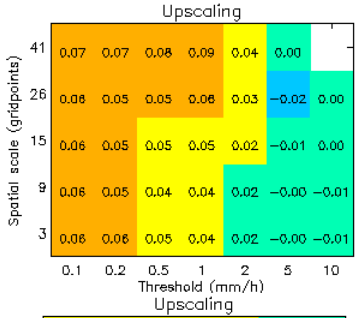
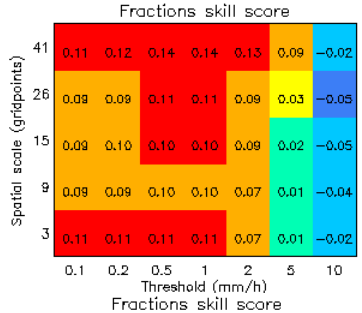
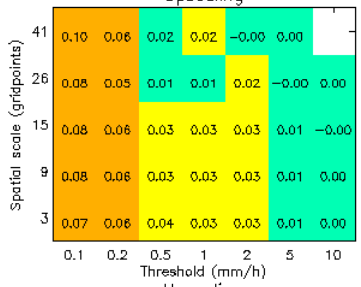
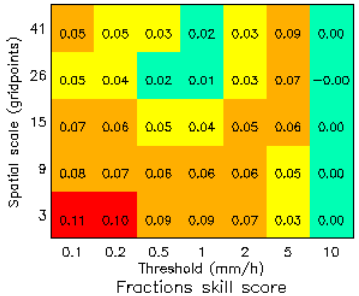
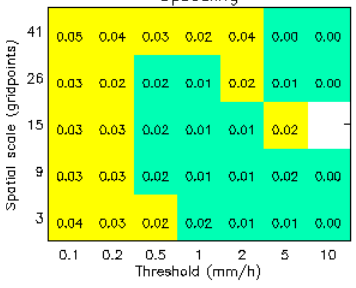
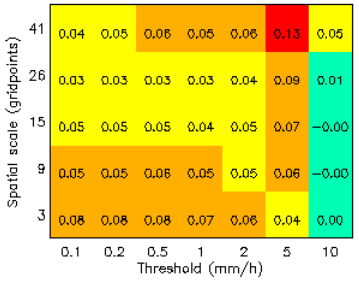
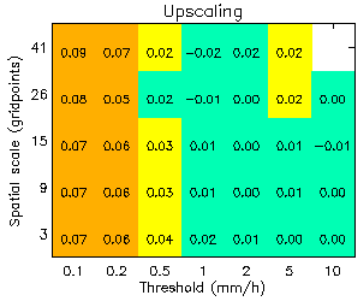
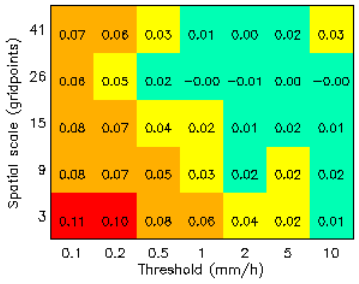
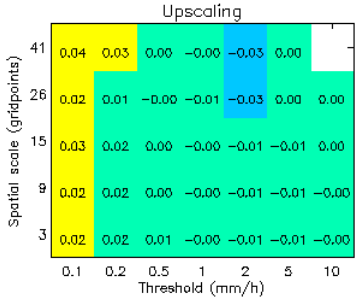
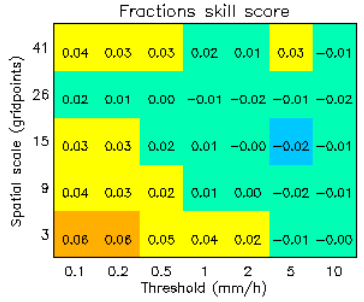
August 07, cut-off 03h, accumulation 03h



Felix.Ament@zmaw.de

Monthly dependency

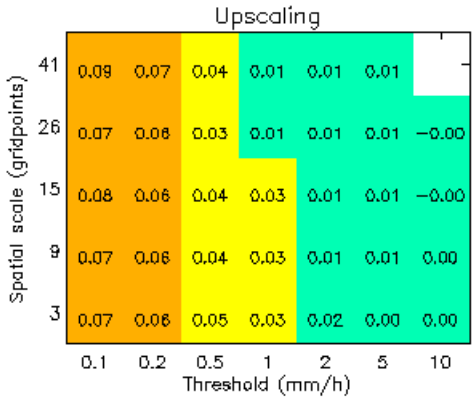
cut-off 03h, accumulation 03h

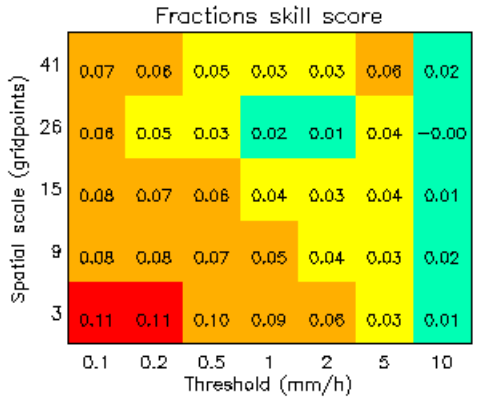
	COSMO-2 - COSMO-7		COSMO-DE - COSMO-EU	
June				
July				
August				

Difference COSMO-2 – COSMO-7

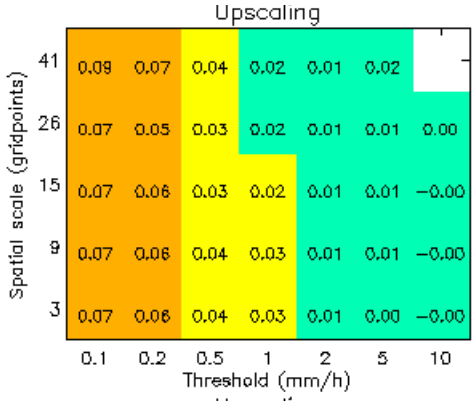
JJA 07, accumulation 03h

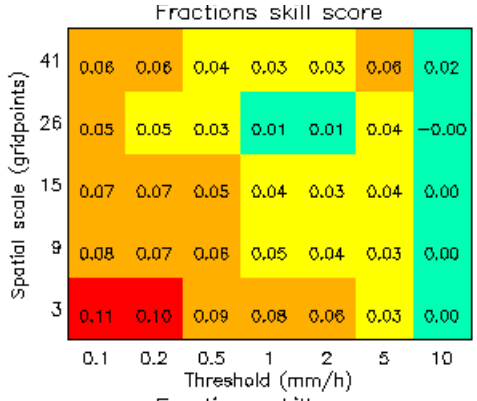
cut-off 03h



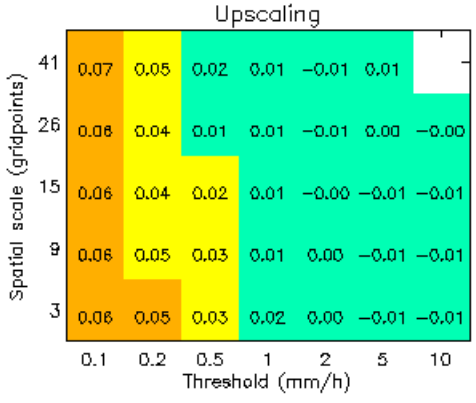


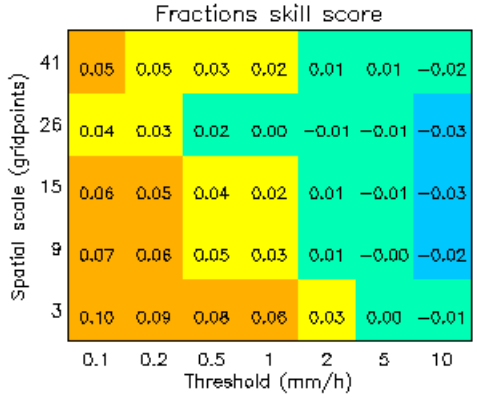
cut-off 06h





cut-off 12h





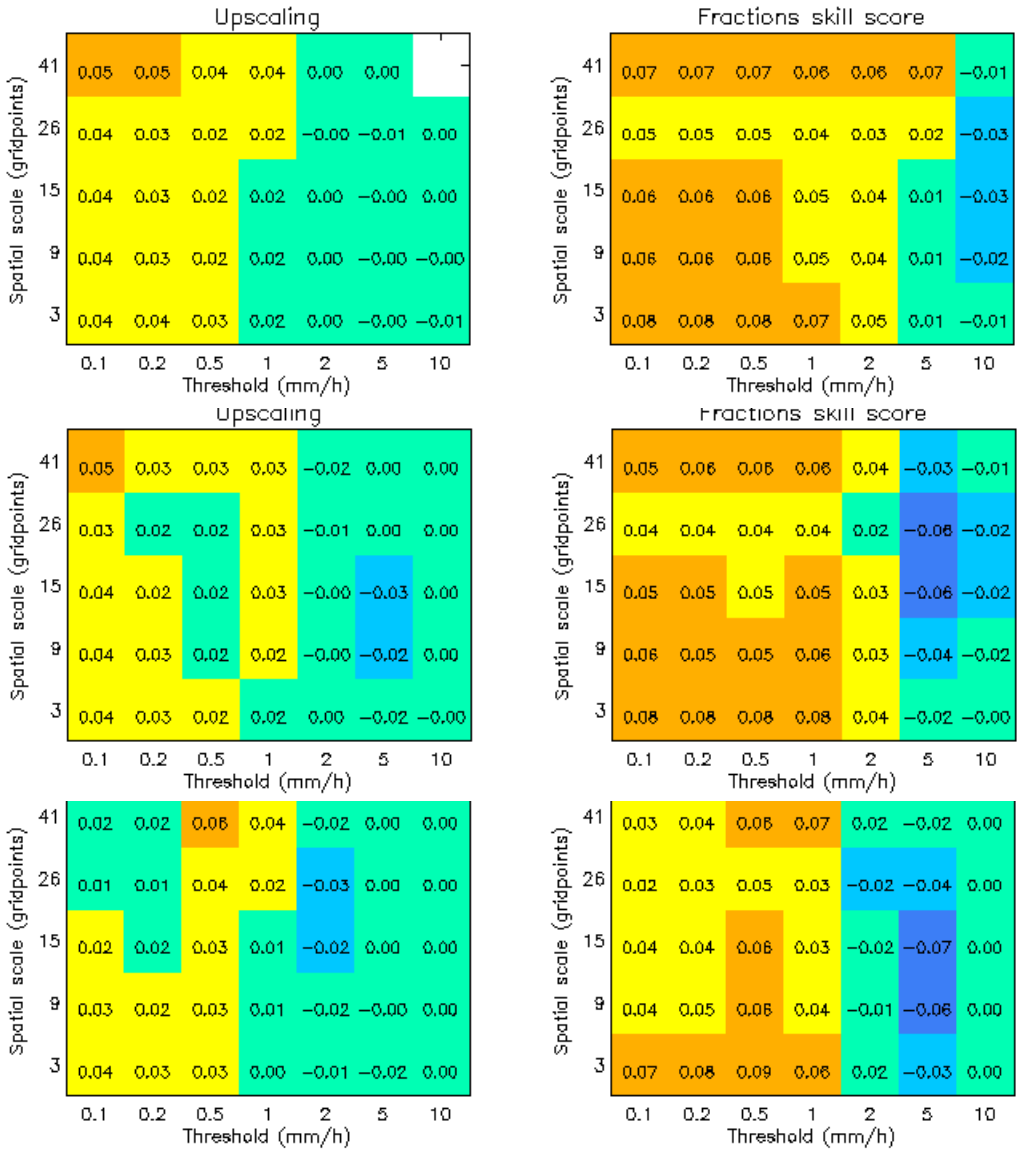
Difference COSMO-DE – COSMO-EU

JJA 07, cut-off 03h

accumulation 03h

accumulation 06h

accumulation 12h

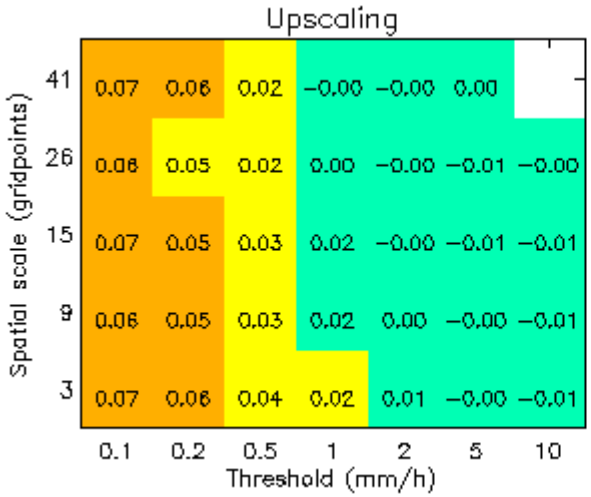


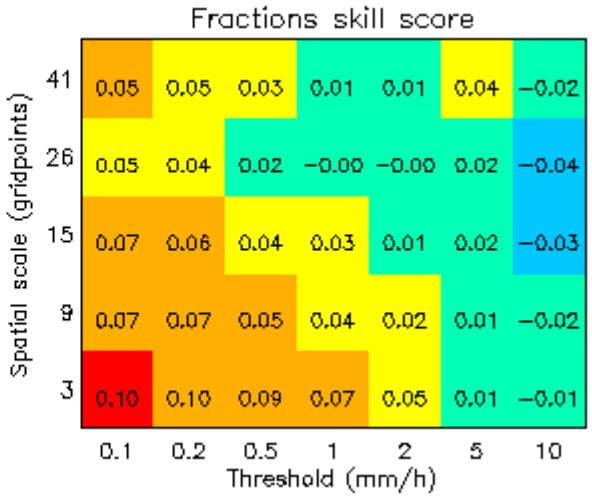
Felix.Ament@zmaw.de

Difference COSMO-2 – COSMO-7 - bootstrap

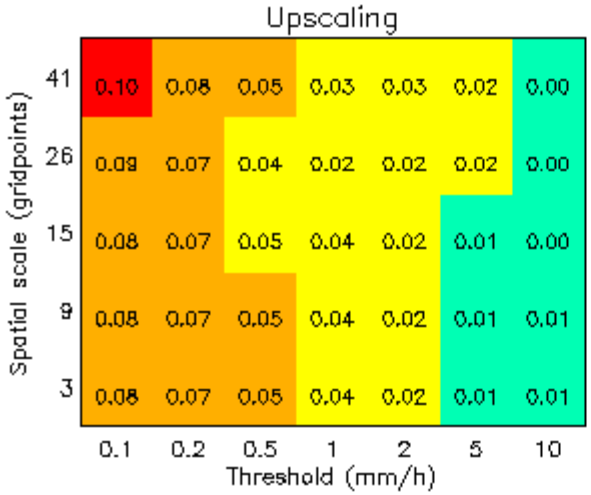
JJA 07, accumulation 03h, cut-off 03h, 100 samples

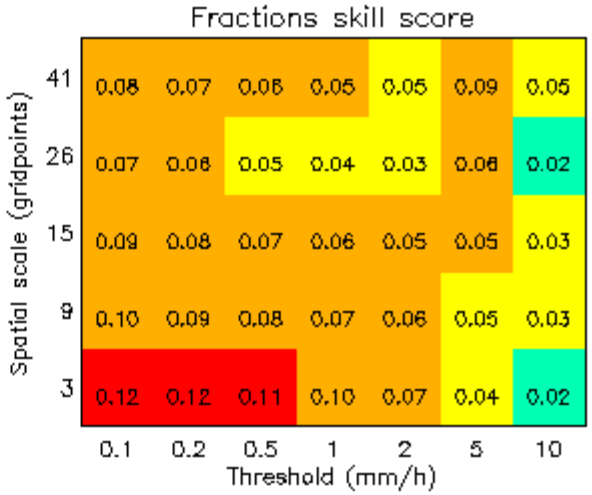
"worst" 5%





"best" 5%

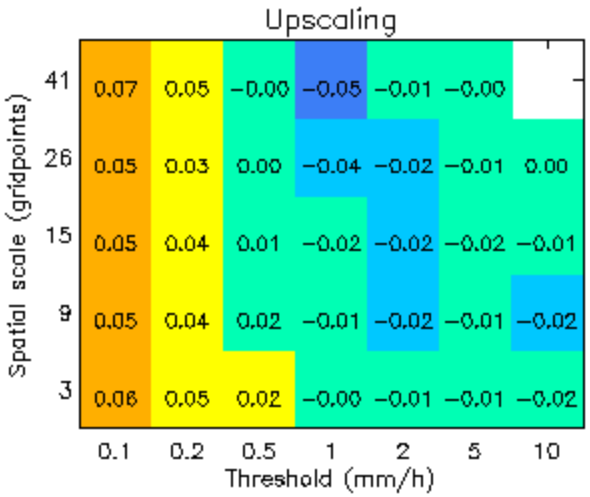


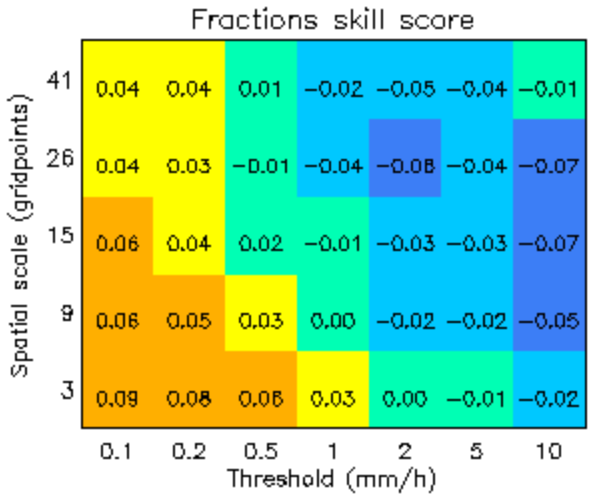


Difference COSMO-2 – COSMO-7 - bootstrap

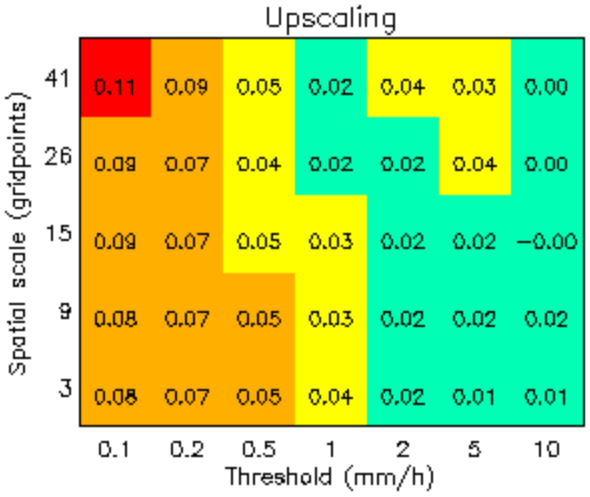
August 07, accumulation 03h, cut-off 03h, 100 samples

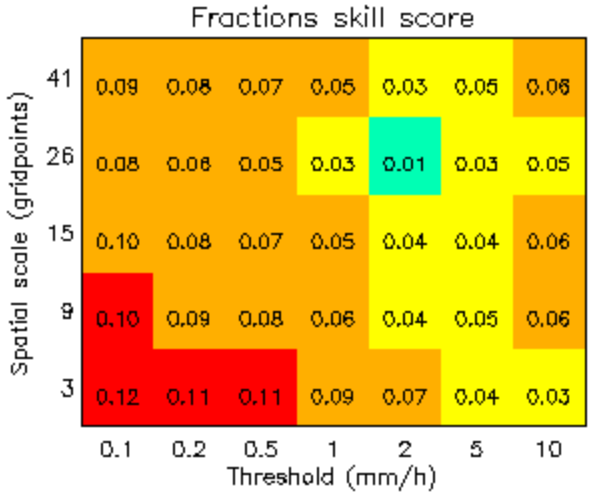
“worst” 5%





“best” 5%





Conclusions

DOP verification

- Observational error is considerably large
- Quality of precipitation warnings is poor (D-PHASE set-up: ~50% POD, ~50% FAR).
- However COSMO models give very good results compared to other D-PHASE models.
- Most high-resolution models outperform their host models and the IFS.
- Probabilistic forecasts are useful! Ensembles are not the only way – verification results might help!

Fuzzy verification

- Results of Upscaling and Fraction skill score are reasonable
- Difference between version I and II has to be explained
- However, results are surprisingly stable, even on a monthly bases
- Overall better results for high-res models
- This benefits shrinks slowly, if large lead times or accumulation periods are considered
- Great variation from month to month – results depend strongly the weather during the verification episode