Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

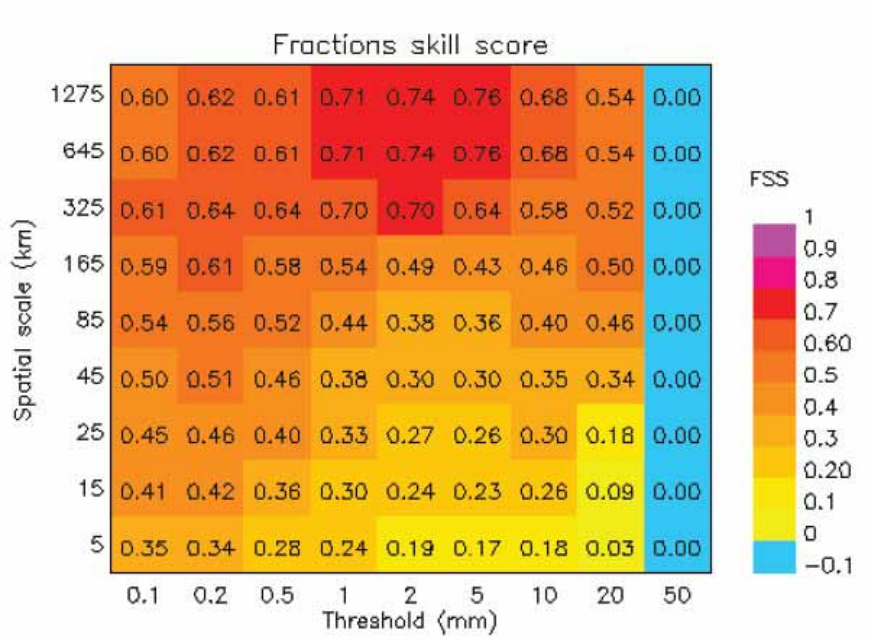
Results of fuzzy verification methods with COSMO over Switzerland and Germany

work by Felix Ament, Tatjana Bähler, Tanja Weusthoff,
Matteo Buzzi (MeteoSwiss), and Ulrich Damrath (DWD)
compiled by Francis Schubiger (MeteoSwiss)
presented by Marco Arpagaus

30th EWGLAM & 15th SRNWP meeting
7 October 2008, Madrid

Fuzzy verification

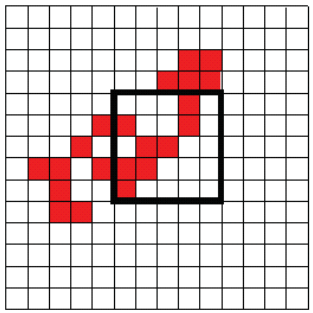
- Beth Ebert has built up a collection of existing fuzzy forecasting verification scores in a toolbox
- define scales of interest; consider “average” features within each box



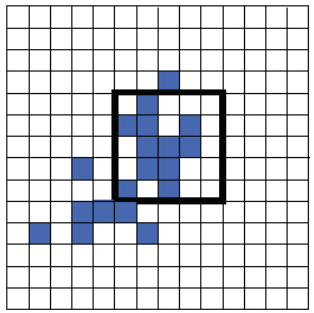
(© Beth Ebert)

Example: Fractions skill score

Compare fractional coverage in a box



observation



forecast

(© Beth Ebert)

- score depends on considered scale and threshold (defining an event)

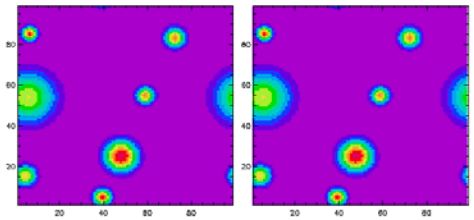
A Fuzzy Verification Toolbox

Fuzzy method	Decision model for useful forecast
Upscaling (Zepeda-Arce et al. 2000; Weygandt et al. 2004)	Resembles obs when averaged to coarser scales
Anywhere in window (Damrath 2004), **50% coverage**	Predicts event over minimum fraction of region
Fuzzy logic (Damrath 2004), **Joint probability** (Ebert 2002)	More correct than incorrect
Multi-event contingency table (Atger 2001)	Predicts at least one event close to observed event
Intensity-scale (Casati et al. 2004)	Lower error than random arrangement of obs
Fractions skill score (Roberts and Lean 2005)	Similar frequency of forecast and observed events
Practically perfect hindcast (Brooks et al. 1998)	Resembles forecast based on perfect knowledge of observations
Pragmatic (Theis et al. 2005)	Can distinguish events and non-events
CSRR (Germann and Zawadzki 2004)	High probability of matching observed value
Area-related RMSE (Rezacova et al. 2005)	Similar intensity distribution as observed

Ebert, E.E., 2007: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. Meteorol. Appls., submitted.
Toolbox available at http://www.bom.gov.au/bmrc/wefor/staff/eee/fuzzy_verification.zip

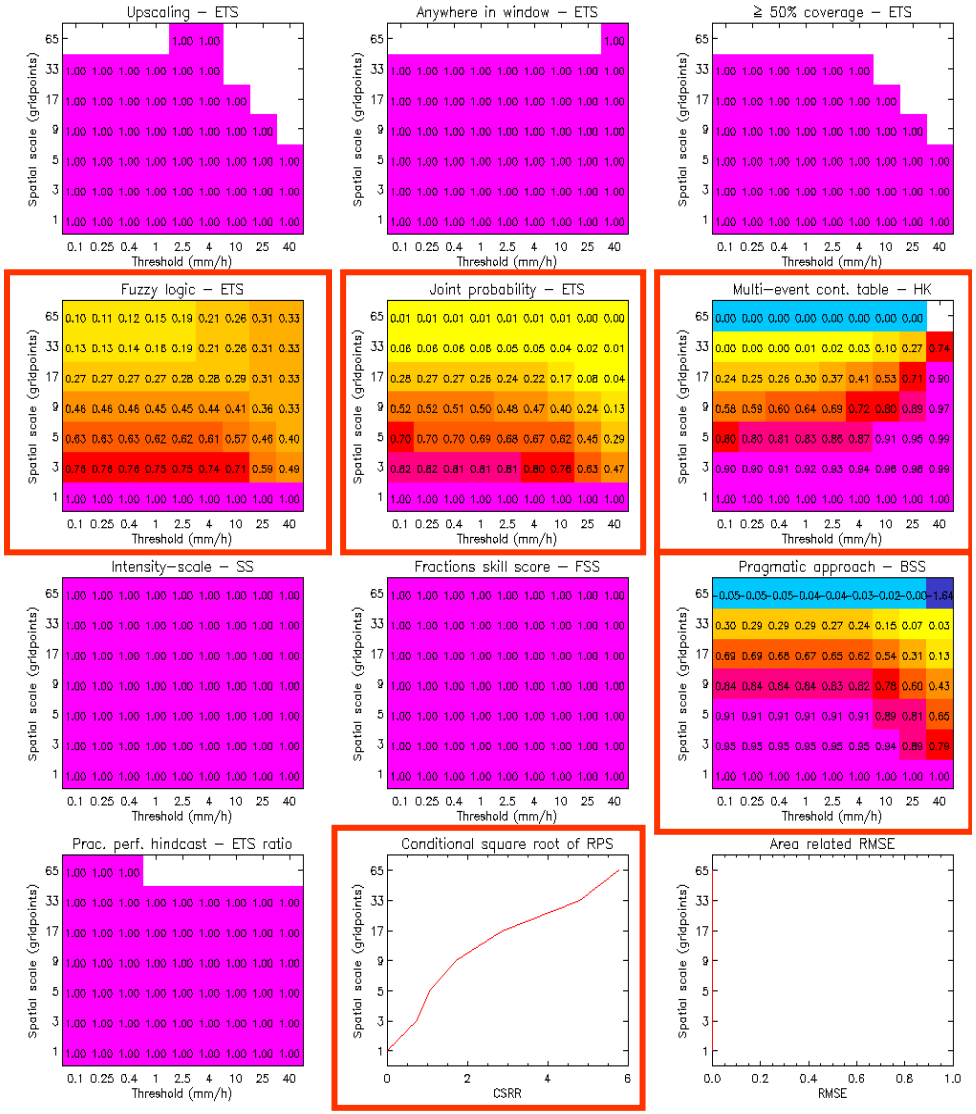
Testbed: Perfect forecast

observation = forecast



All scores should equal 1.00 **!**

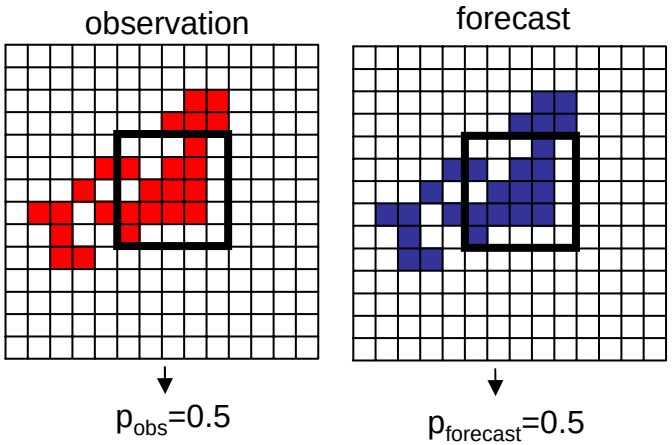
But, in fact, 5 out of 12 do not!



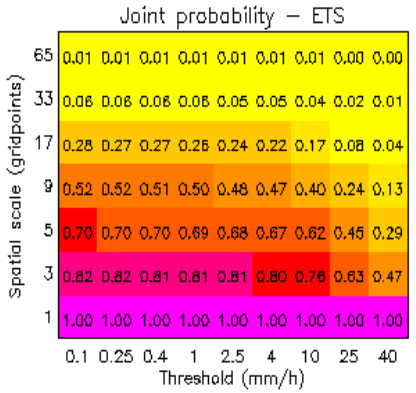
F. Ament & T. Bähler, MeteoSwiss

Effect of „Leaking“ Scores

Some methods assume no skill at scales below window size!



An example:
Joint probability method



→

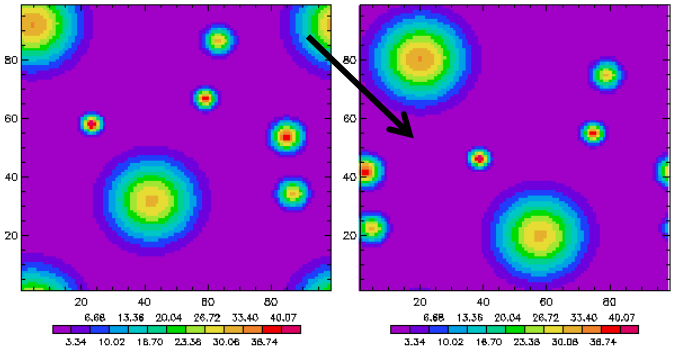
Forecast

OBS	yes	no
yes	0.25	0.25
no	0.25	0.25

→ Not perfect!

F. Ament & T. Bähler, MeteoSwiss

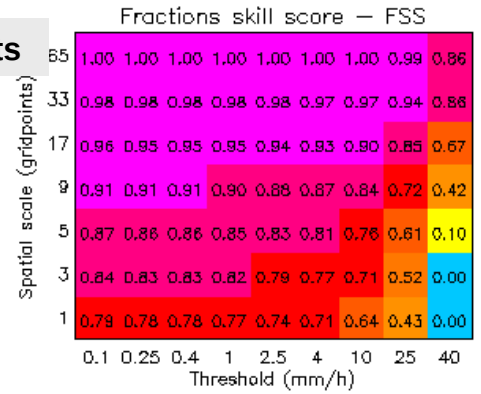
Testbed: Spatial Translation



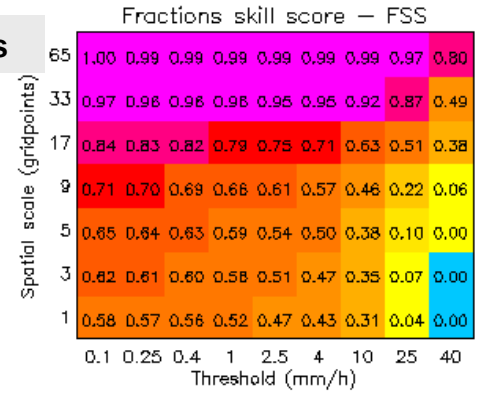
Example:
Fractions skill score (Roberts, N., 2005)

Fraction skill score shows a very reasonable behaviour in case of translations.

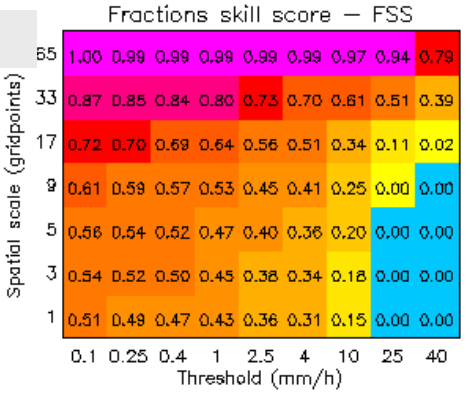
Δx=7.5 points



Δx=15 points

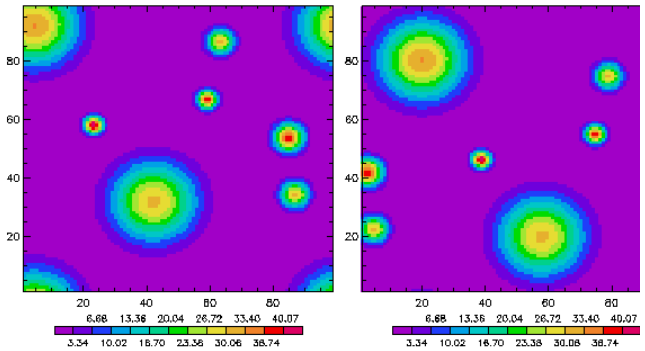


Δx30 points

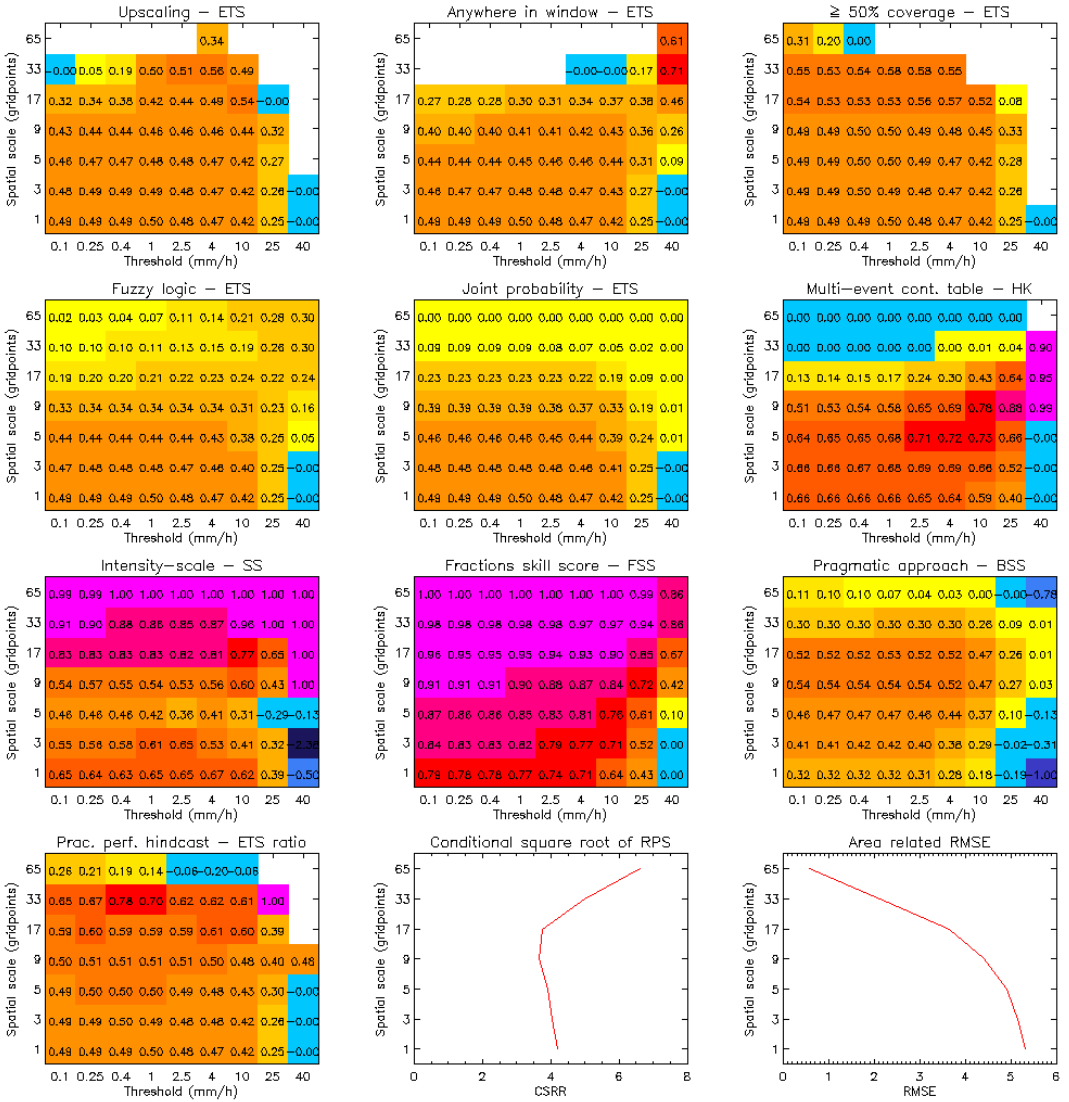


F. Ament & T. Bähler, MeteoSwiss

Testbed: Spatial Translation



Δx=7.5 points

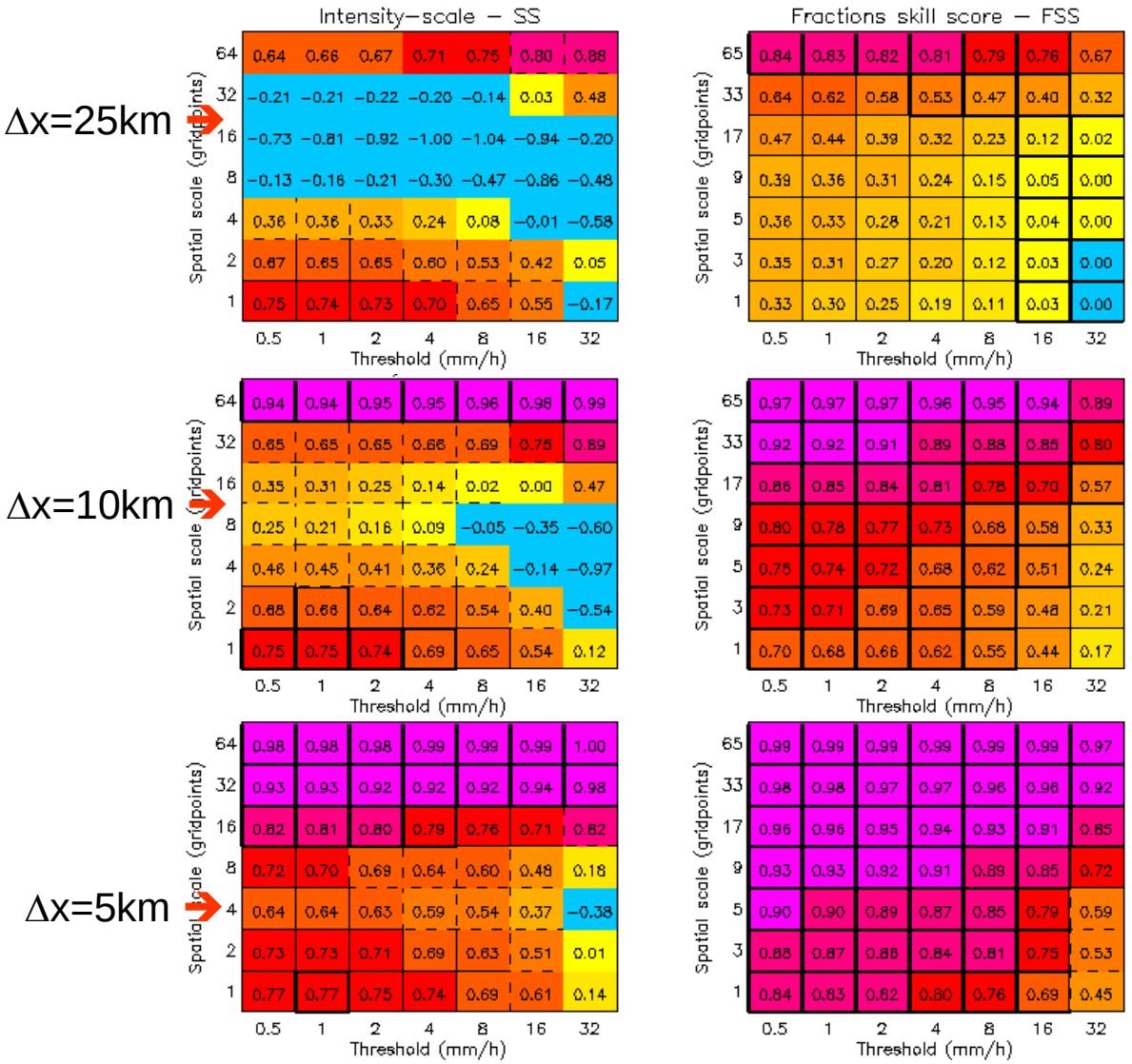


F. Ament & T. Bähler, MeteoSwiss

Spatial detection versus filtering

Horizontal translation (XSHIFT) with variable displacement Δx

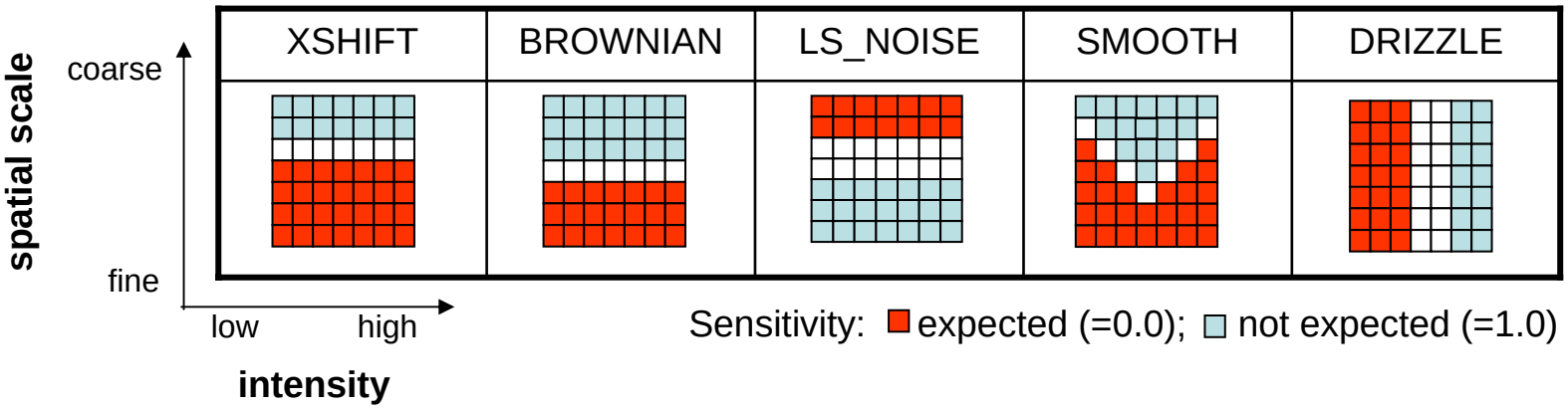
- "Intensity scale" method can detect spatial scale of perturbation
- All other methods like the "Fraction Skill score" just filter small scale errors



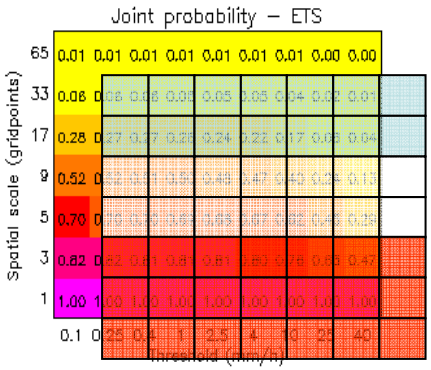
Results of fuzzy verification methods with COSMO over Switzerland and Germany
30th EWGLAM &15th SRNWP meeting, 7 October 2008, Madrid

F. Ament, MeteoSwiss

Expected response to perturbations

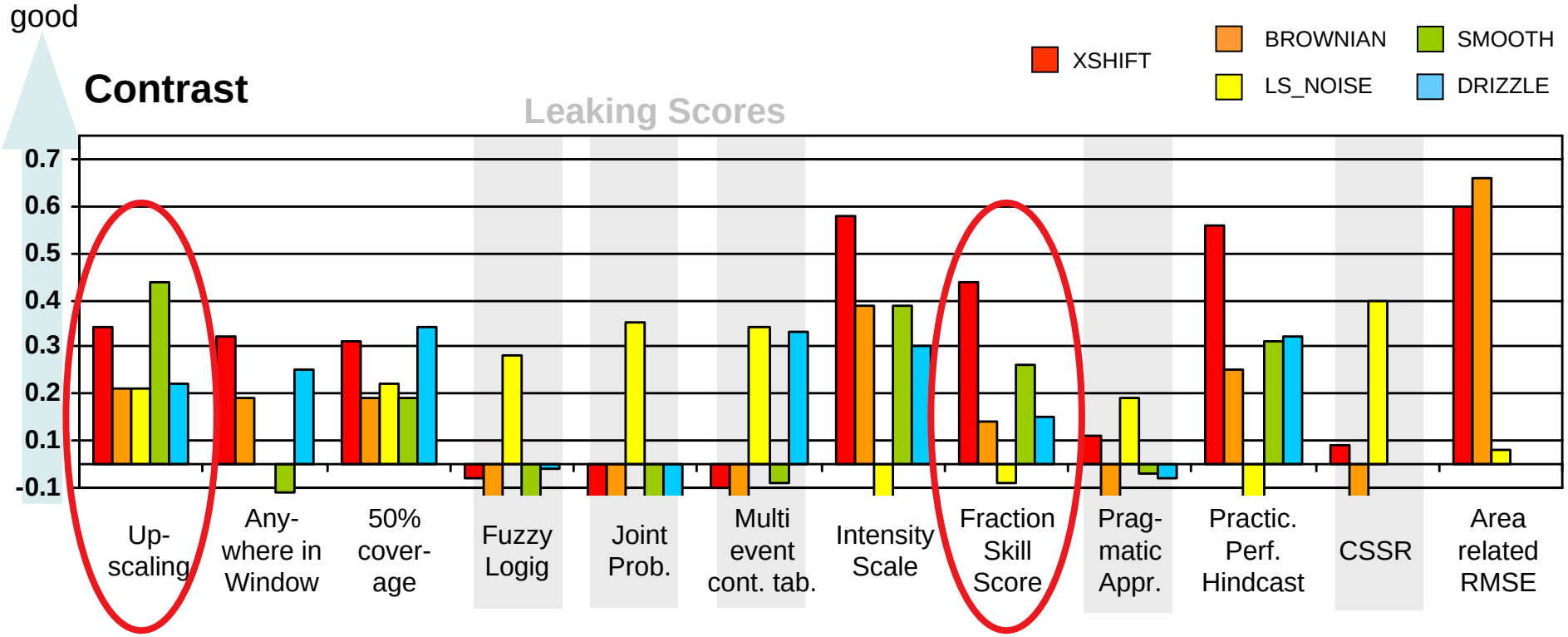


Summary in terms of contrast:



Contrast := mean(□) – mean(■)

Contrast in testbed experiments



- Leaking scores show an overall poor performance
- "Intensity scale" and "Practically Perfect Hindcast" perform in general well, but …
- Many scores have problems to detect large scale noise (LS_NOISE); "Upscaling" and "50% coverage" are beneficial in this respect

Redundancy of scores

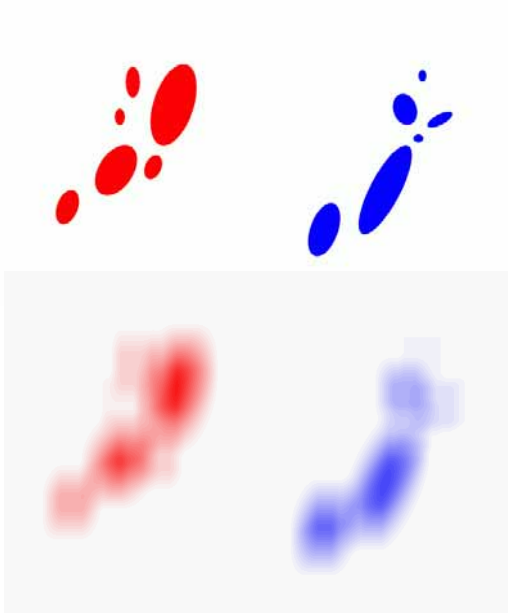
Correlation (%) of resulting scores between all score for all thresholds, window sizes – averaged over all types of perturbation:

	UP	YN	MC	FZ	JP	ME	IS	FB	PG	PP	CS	RM
UP	100	70	91	51	52	26	56	79	62	83	5	29
YN	70	100	71	43	32	35	51	84	17	79	5	27
MC	91	71	100	58	59	34	55	79	66	82	15	25
FZ	51	43	58	100	94	63	13	41	20	34	45	−28
JP	52	32	59	94	100	52	4	30	24	24	48	−32
ME	26	35	34	63	52	100	15	32	38	24	58	−20
IS	56	51	55	13	4	15	100	63	3	63	−12	34
FB	79	84	79	41	30	32	63	100	16	90	1	34
PG	62	17	66	20	24	38	3	16	100	45	7	−2
PP	83	79	82	34	24	24	63	90	45	100	−3	37
CS	5	5	15	45	48	58	−12	1	7	−3	100	−22
RM	29	27	25	−28	−32	−20	34	34	−2	37	−22	100

➔ Groups of scores:

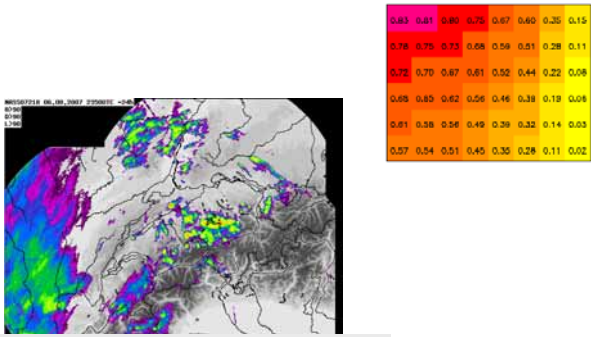
- UP, YN, MC, FB, PP
- FZ, JP
- FB, PP, (IS)

Conclusions



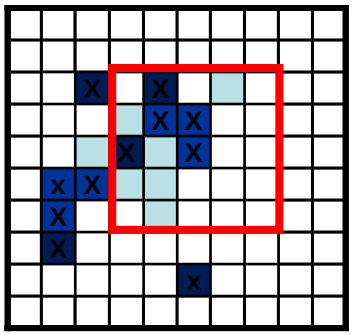
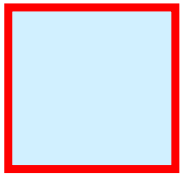
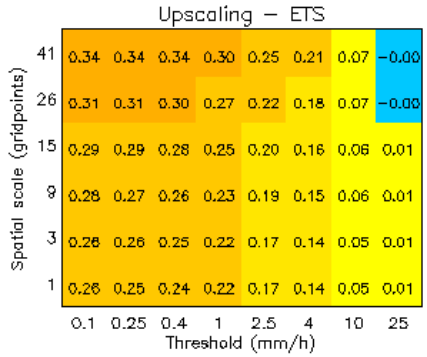
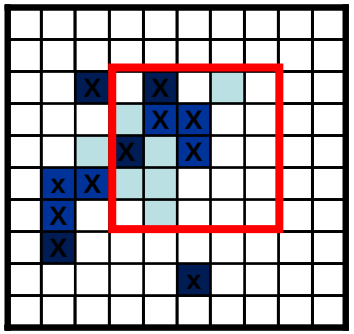
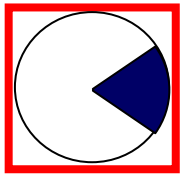
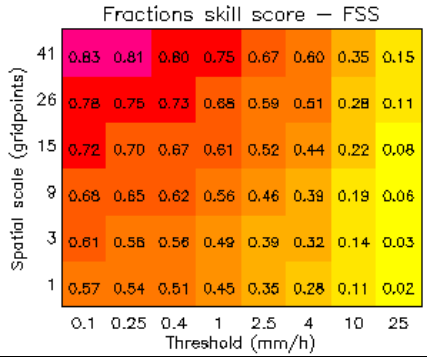
- Intensity scale (IS) is a very promising technique – fast and able to detect a specific scale of an spatial error.
- The Fraction Skill (FS) and Practically perfect hindcast (PP) show also very good result – FS is very popular.
- Set should be completed by Upscaling (UP) to be aware of large scale error patterns.
- Area related RMSE (RM) shows good performance too, but has no intensity component and requires a lot of computational time.
- Leaking scores (FZ, JP, ME, PG, CS) should not be considered for COSMO purposes!
- Reliability (low STD) is good for all scores. Best performance shows Area related RMSE.

Fuzzy Verification



Radar composite

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

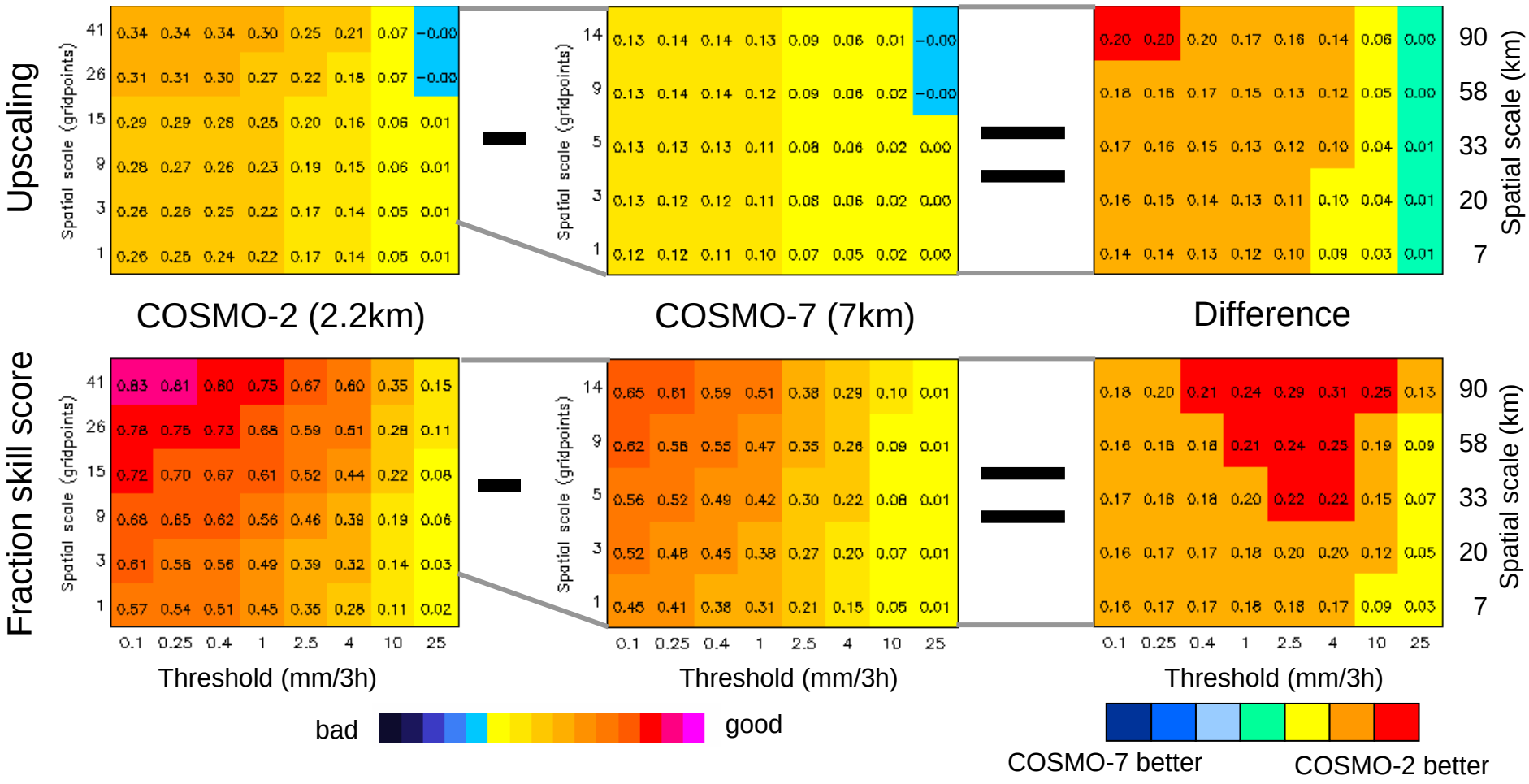
Method	Raw Data	Fuzzyfication	Score	Example result
Upscaling		Average 	Equitable threat score	
Fraction Skill Score (Roberts and Lean, 2005)		Fractional coverage 	Skill score with reference to worst forecast	

Settings

- Nacc = 3h
- Thresh = [0.1, 0.2, 0.5, 1, 2, 5, 10, 20] (mm / 3h)
- $\text{Windows}_{\text{COSMO-CH7}}$ = [1, 3, 5, 9, 15]
- $\text{Windows}_{\text{COSMO-CH2}}$ = [1, 3, 9, 15, 27, 45]
- Methods = Upscaling (UP) and Fraction Skill Score (FB)
- Scores = ETS (for UP) and FSS (for FB)
- Fuzzy-package: Version April 2008

Fuzzy Verification COSMO-2 – COSMO-7

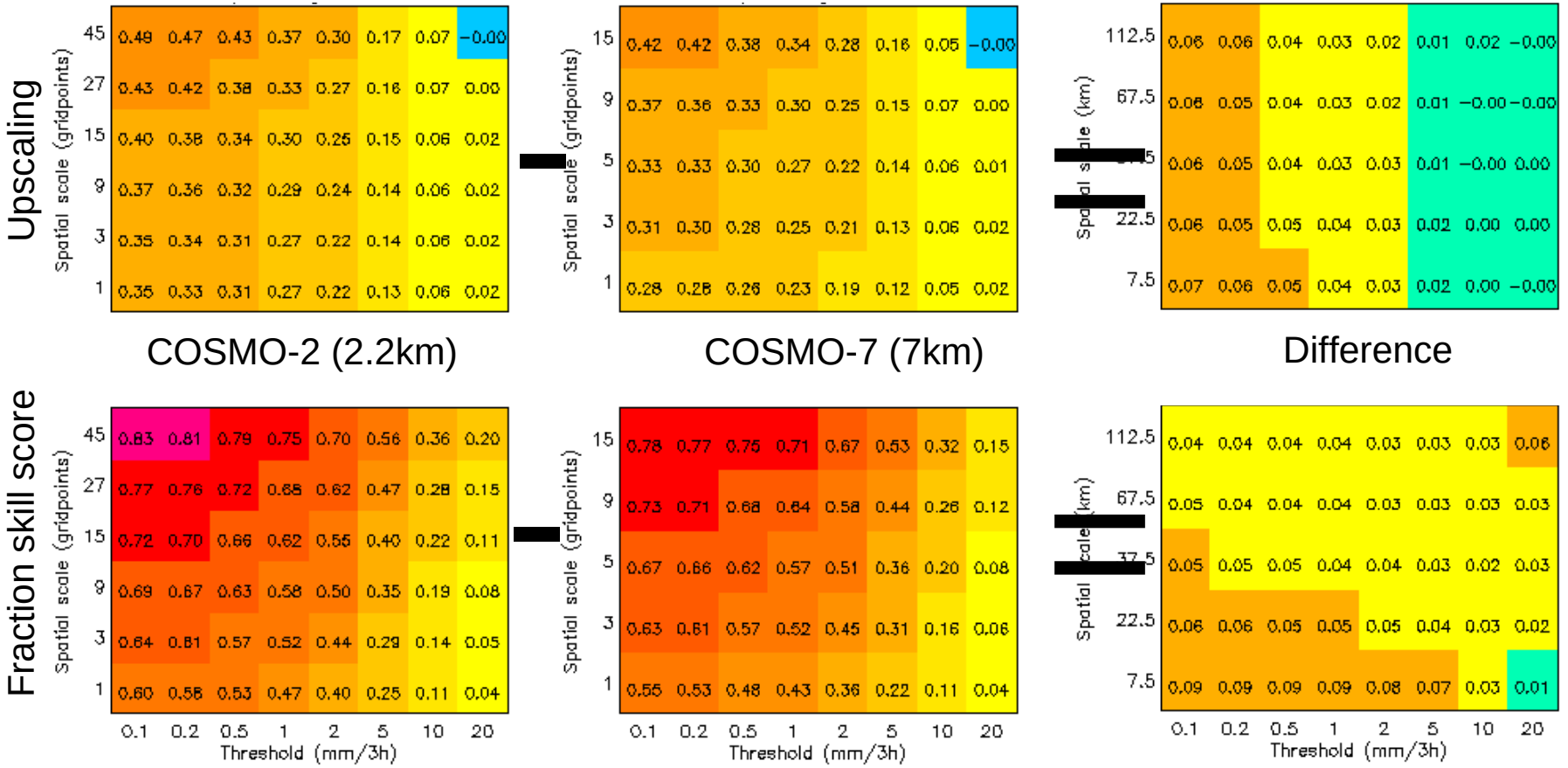
JJA 2007, Verification against Swiss Radar Composite, 3 hourly accumulations, rain events



F. Ament, MeteoSwiss

Fuzzy Verification COSMO-2 – COSMO-7

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

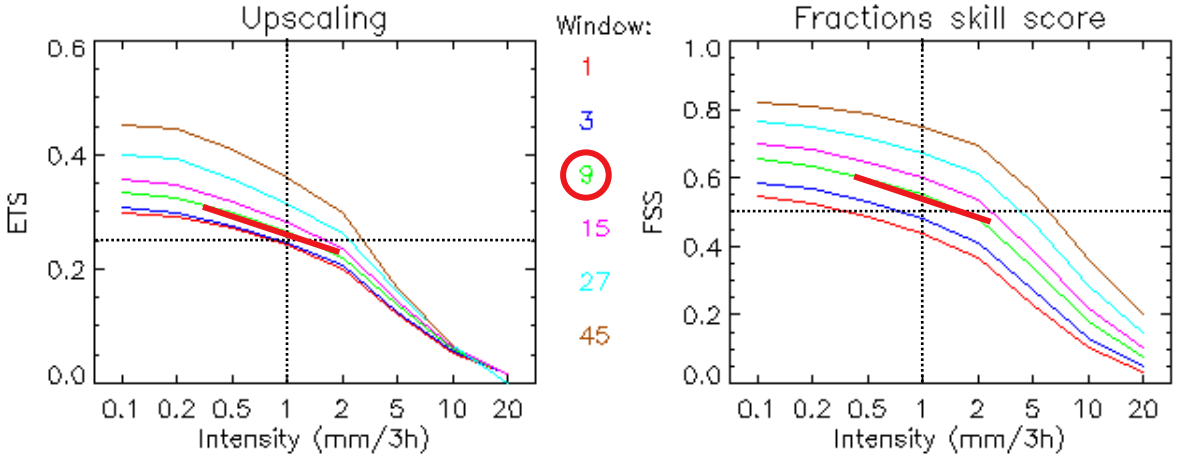


T. Weusthoff, MeteoSwiss

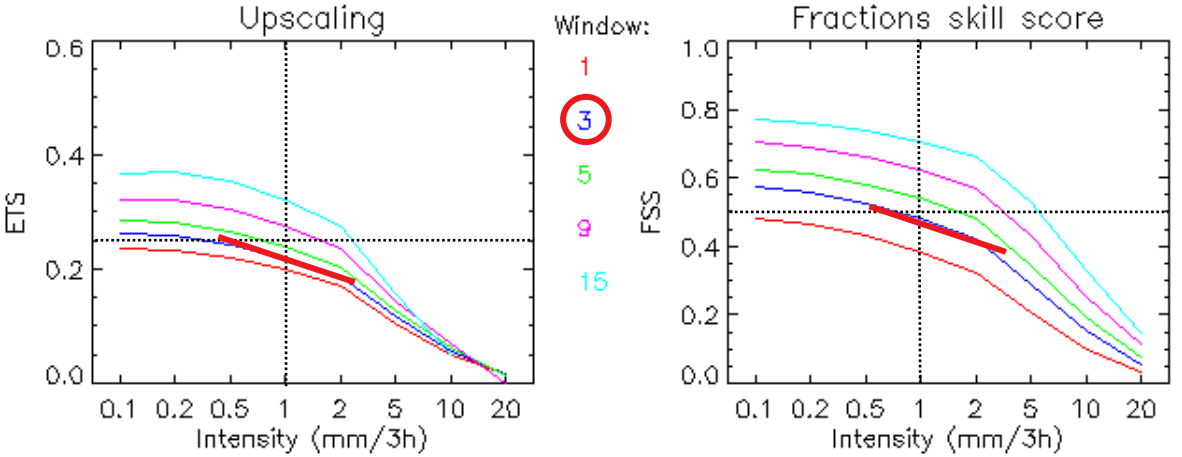
Score vs intensity, entire DOP

JJASON 2007

COSMO-2



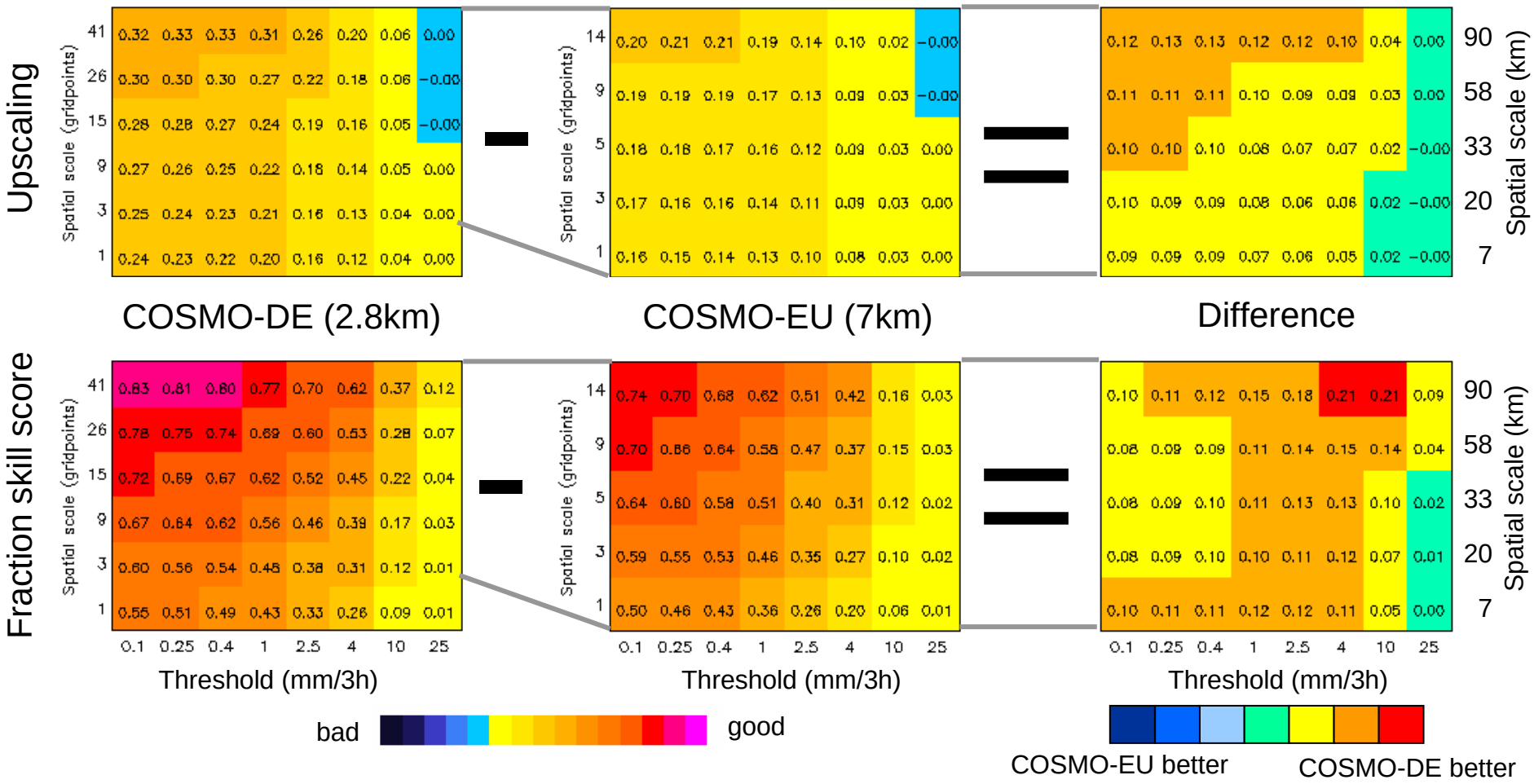
COSMO-7



T. Weusthoff, MeteoSwiss

Fuzzy Verification COSMO-DE – COSMO-EU

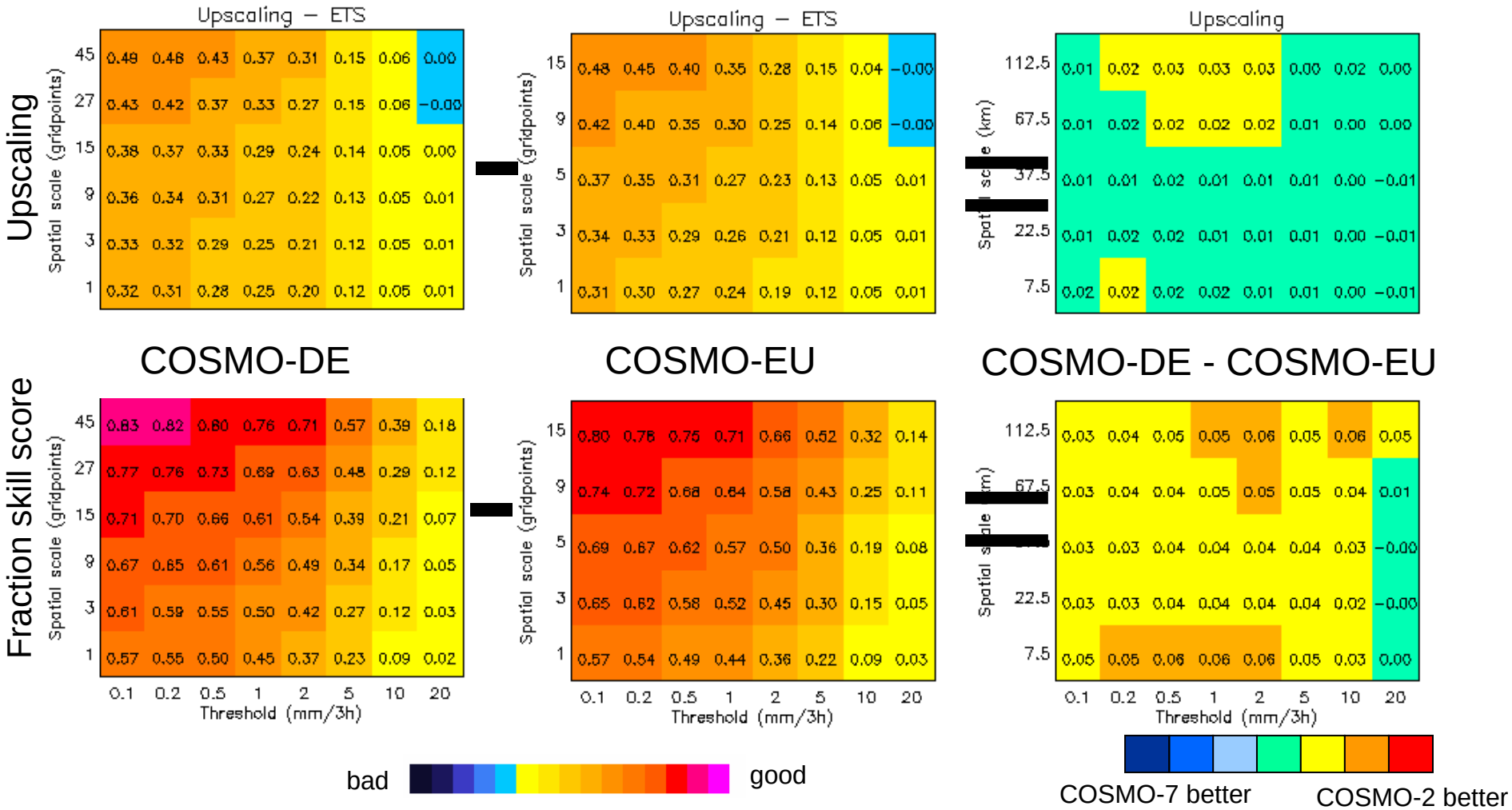
JJA 2007, Verification against Swiss Radar Composite, 3 hourly accumulations, rain events



F. Ament, MeteoSwiss

Fuzzy Verification COSMO-DE – COSMO-EU

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



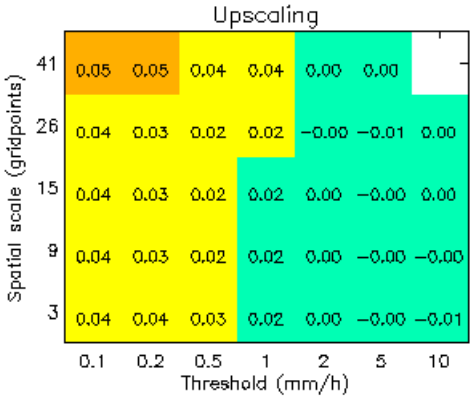
Results of fuzzy verification methods with COSMO over Switzerland and Germany
30th EWGLAM &15th SRNWP meeting, 7 October 2008, Madrid

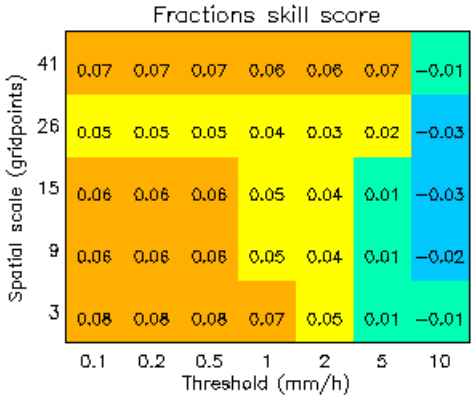
T. Weusthoff, MeteoSwiss

Test of accumulation time

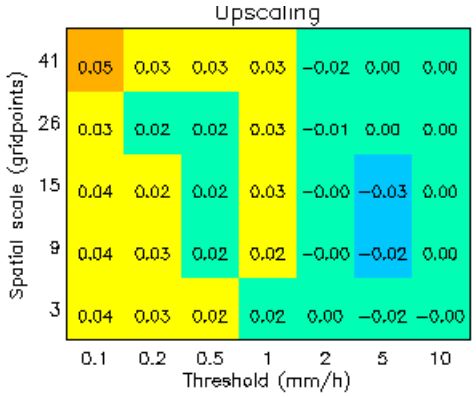
Difference COSMO-DE – COSMO-EU, JJA 07, cut-off 03h

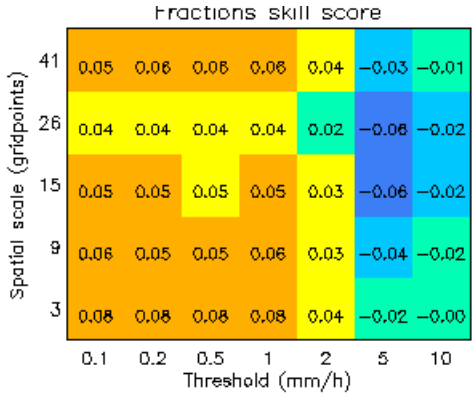
accumulation 03h



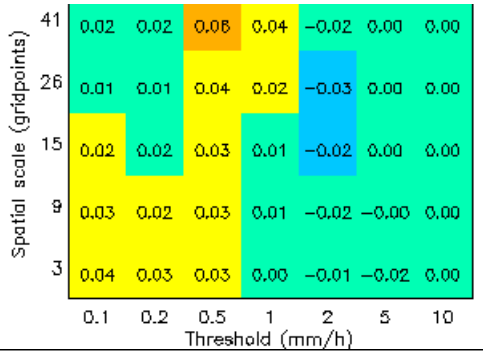


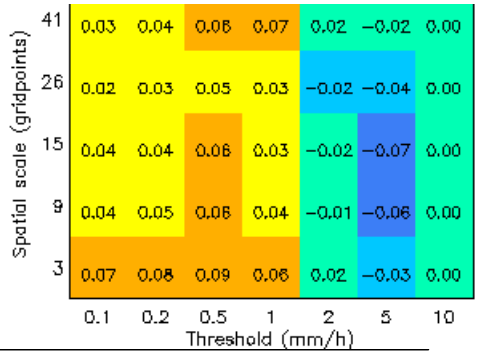
accumulation 06h





accumulation 12h

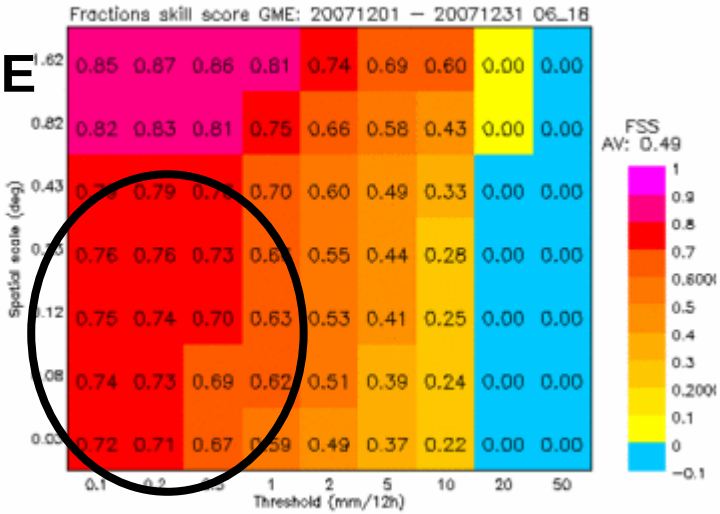




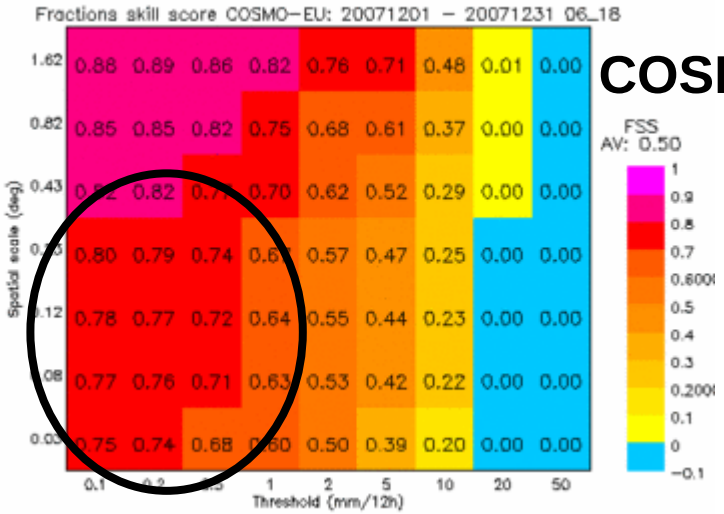
F. Ament, Uni Hamburg

Fraction Skill Score December 2007

GME

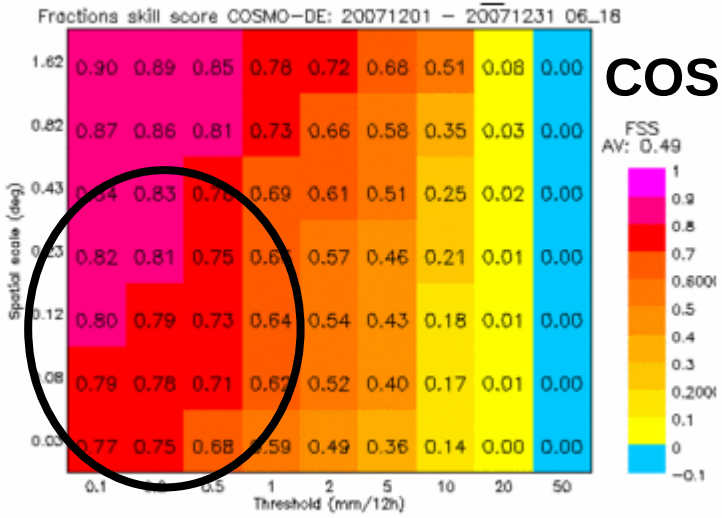


COSMO-EU



whole_area

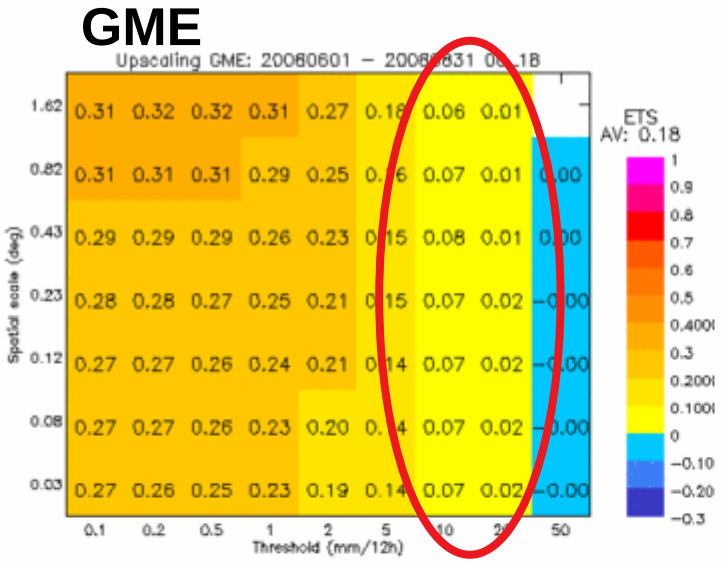
COSMO-DE



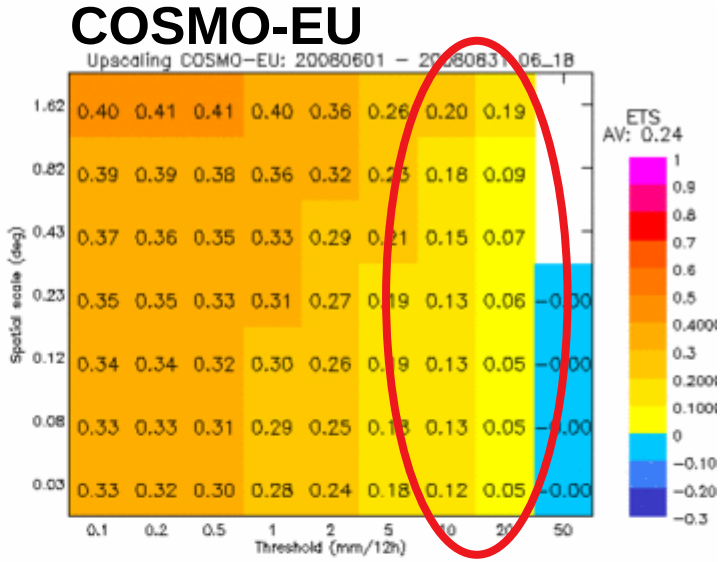
area: Germany

ETS Upscaling Summer 2008

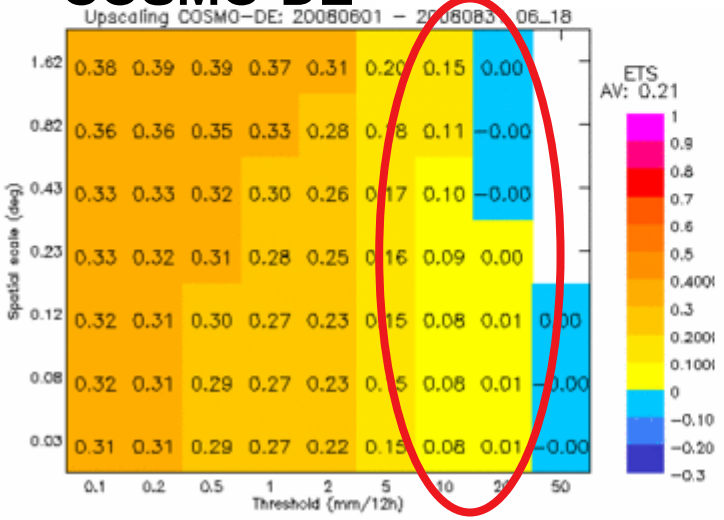
GME



COSMO-EU



COSMO-DE

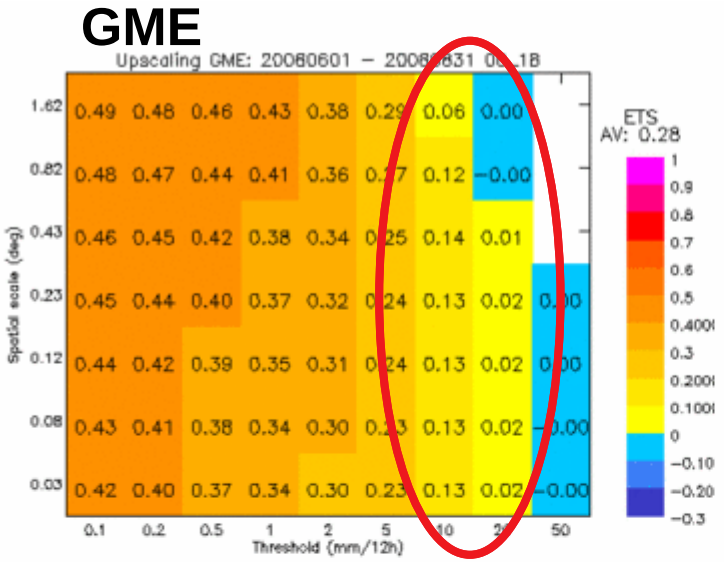


area: Germany

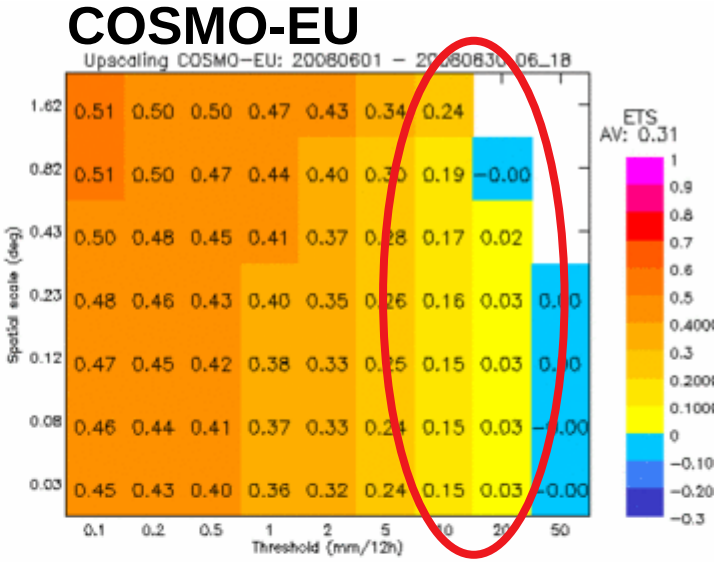
U. Damrath, DWD

ETS Upscaling Summer 2008

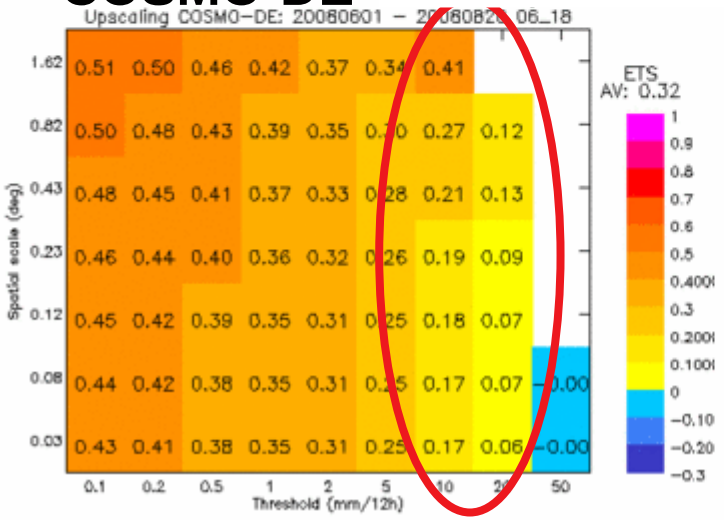
GME



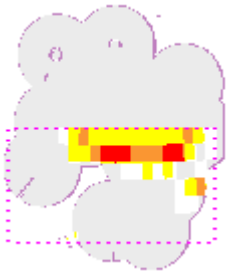
COSMO-EU



COSMO-DE

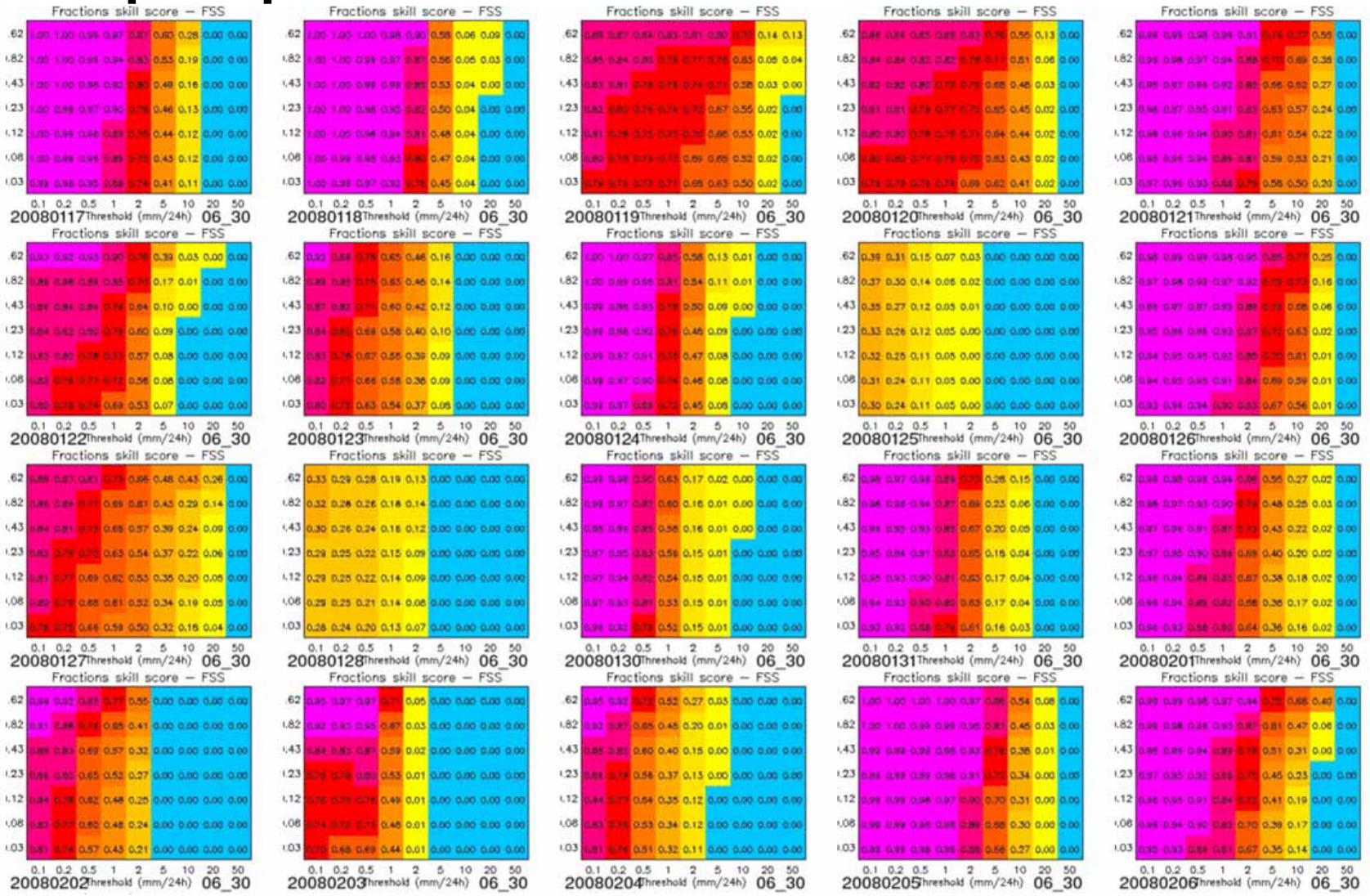


area: Central Part of Germany



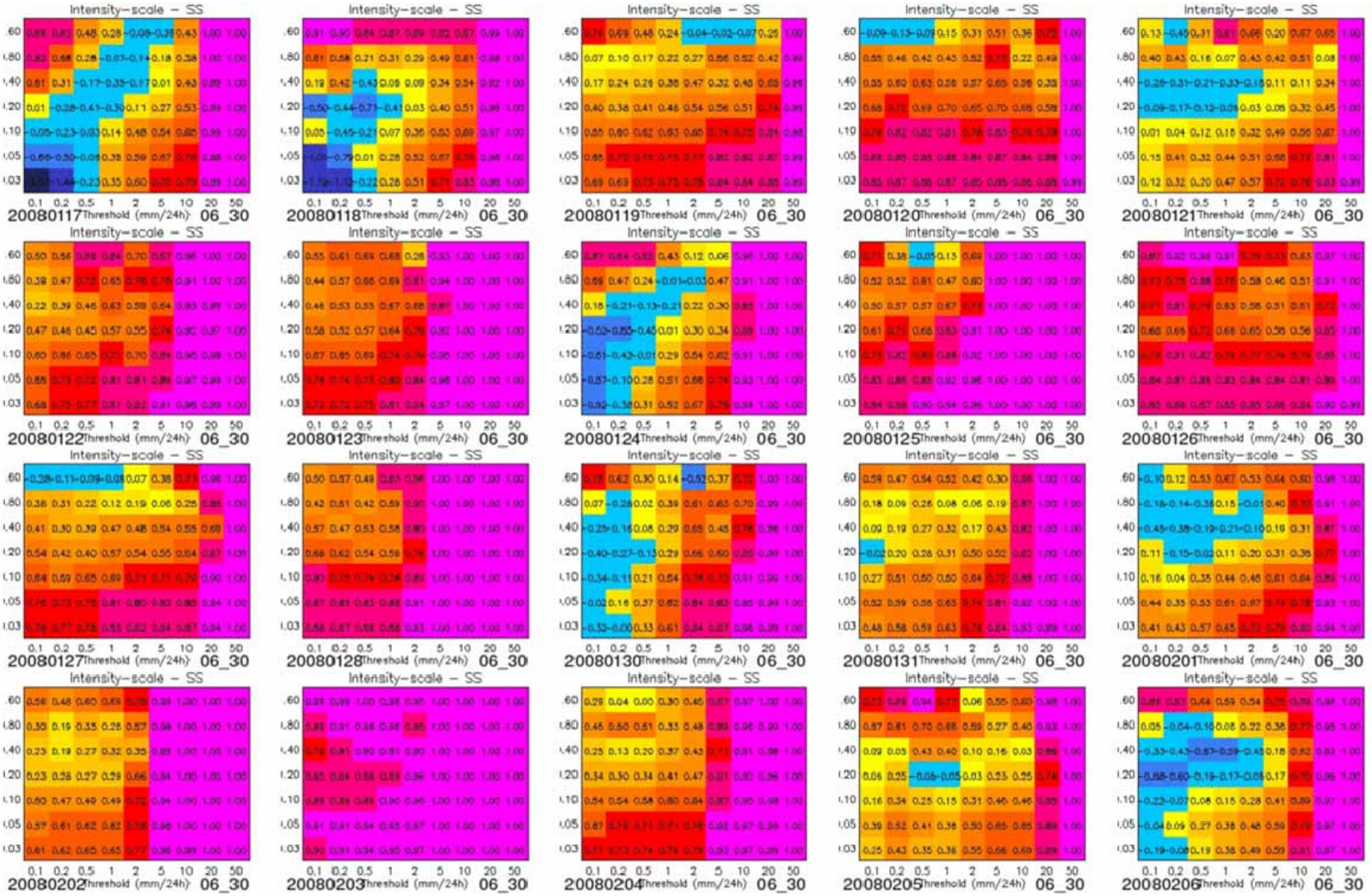
U. Damrath, DWD

Fraction skill score 17.01.2008 - 06.02.2008
24h precipitation sums



U. Damrath, DWD

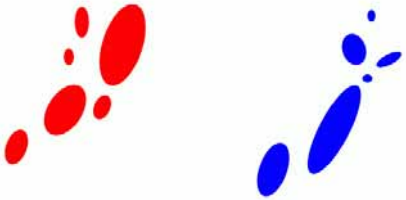
Intensity scale skill score 17.01.2008 - 06.02.2008
24h precipitation sums



Results of fuzzy verification methods with COSMO over Switzerland and Germany
30th EWGLAM &15th SRNWP meeting, 7 October 2008, Madrid

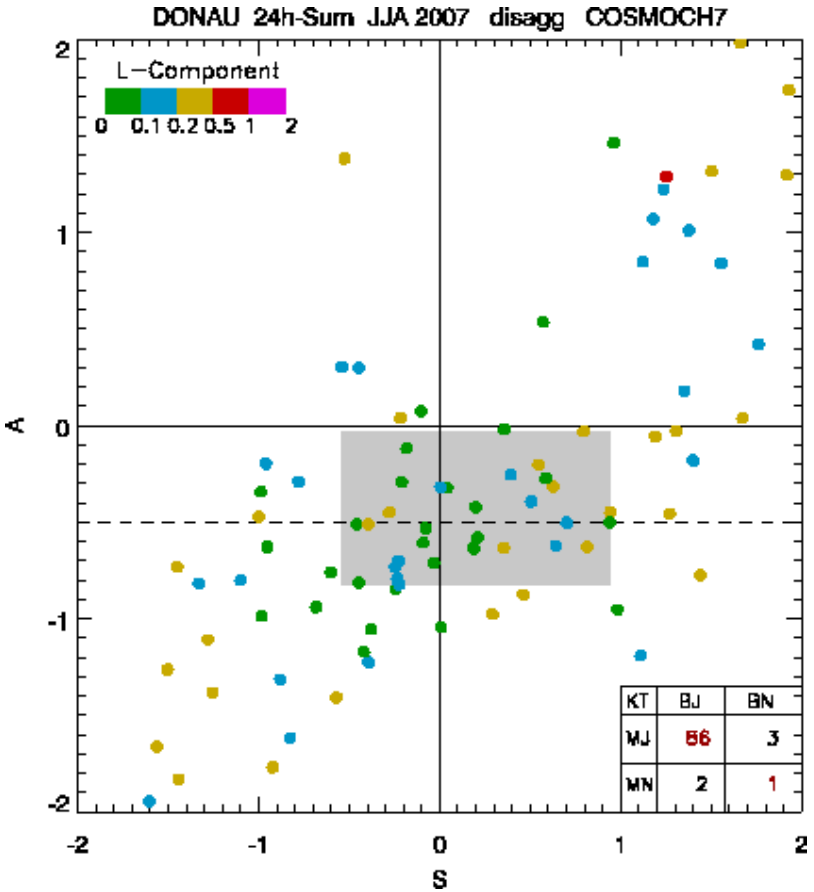
U. Damrath, DWD

Conclusions (so far …)

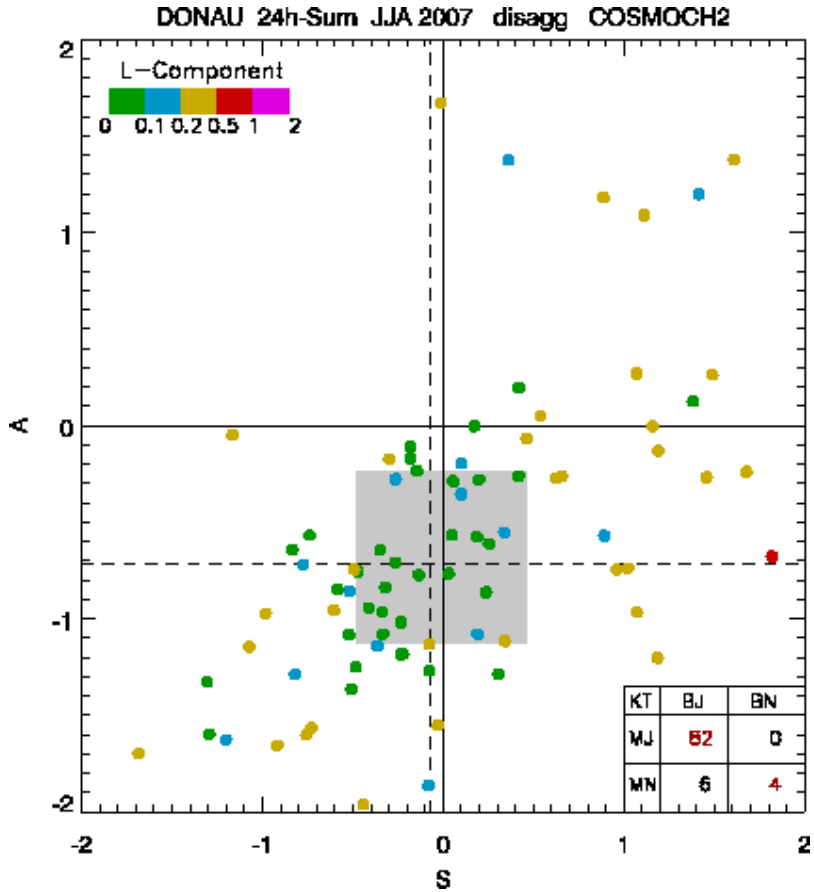


- Fraction skill score and Upscaling are the two fuzzy verification methods chosen inside COSMO, although Intensity-scale is also very promising.
- First results regarding COSMO 2.2/2.8 km vs COSMO 7km show:
  - some advantages for 2.2/2.8 km especially in regions where topography plays a major role and for situations with mesoscale character
  - 2.2/2.8 km has advantages for shorter accumulated periods.
  - 2.2/2.8km shows better scores for low thresholds and for small to medium space scales.

Another fuzzy method: SAL
JJA 2007, catchment Danube, 24h-sums



COSMO-7, 7 km



COSMO-2, 2.2 km

Results of fuzzy verification methods with COSMO over Switzerland and Germany
30th EWGLAM &15th SRNWP meeting, 7 October 2008, Madrid

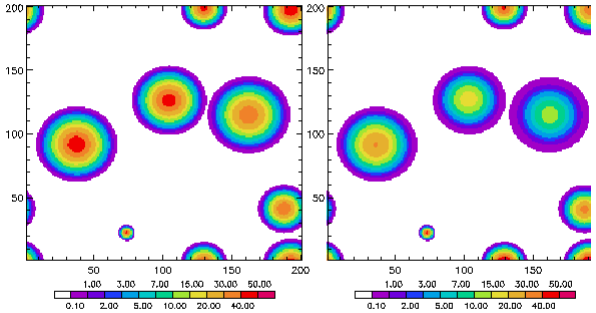
H. Wernli et al., Uni Mainz
to appear in MWR

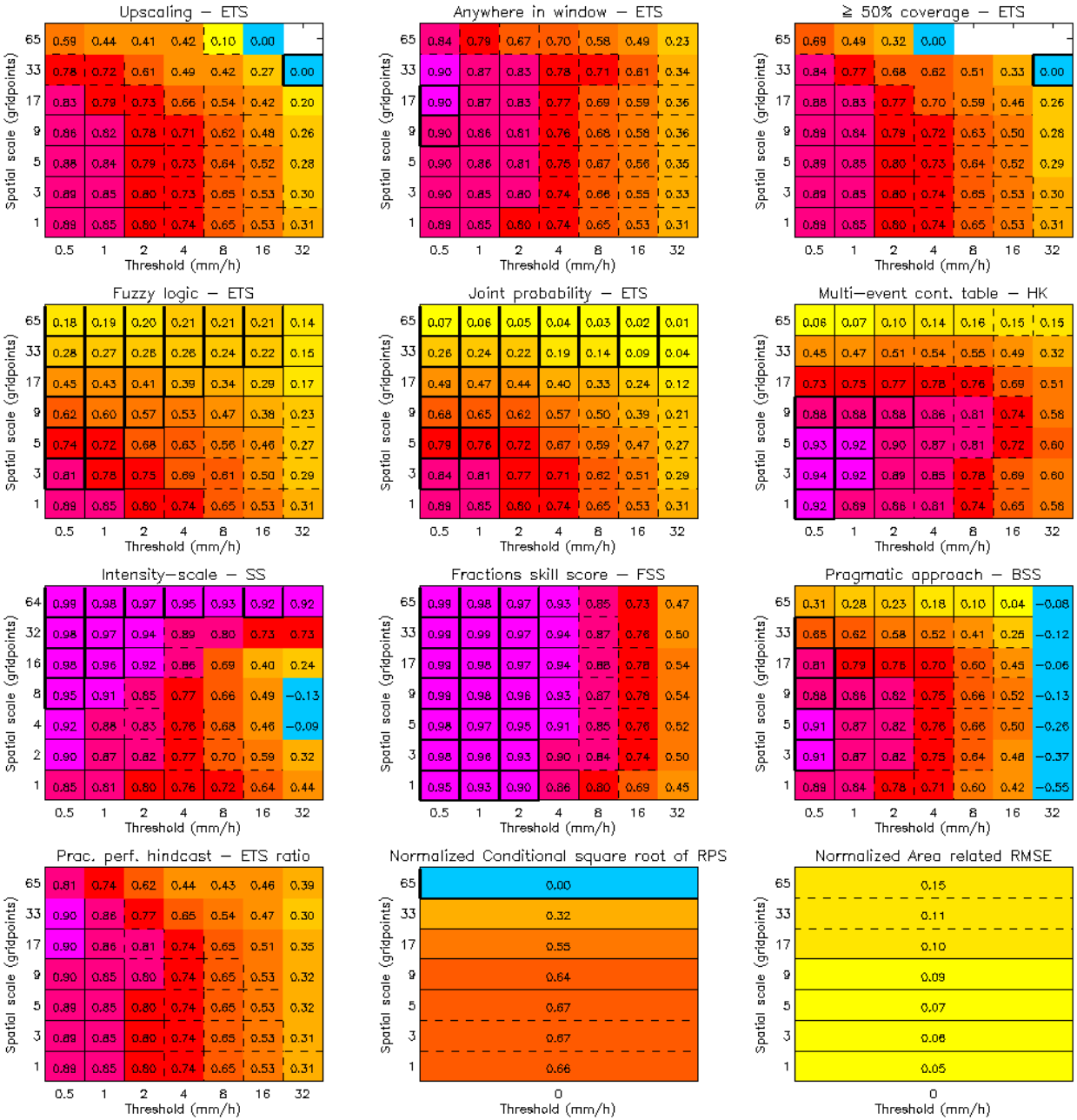
Thank you for your attention!

A (Fuzzy) Verification testbed
Perturbations

Perturbation	Type of forecast error	Algorithm	Example
PERFECT	No error – perfect forecast!	-	
XSHIFT	Phase shift	Horizontal translation (10 grid points	
SCALE	Perfect structure but quantitatively wrong	Multiplication by a constant factor (e.g. 2)	
SMOOTH	High horizontal diffusion (or coarse scale model)	Moving Window arithmetic average	
DRIZZLE	Overestimation of low intensity precipitation	Moving Window filter setting each point below average point to the mean value	
BROWNIAN	No small scale skill	Random exchange of neighboring points (Brownian motion)	
LS_NOISE	Wrong large scale forcing	Multiplication with a disturbance factor generated by large scale 2d Gaussian kernels.	

LNOISEMI (Ideal)

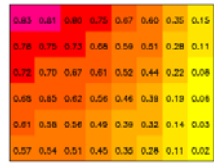




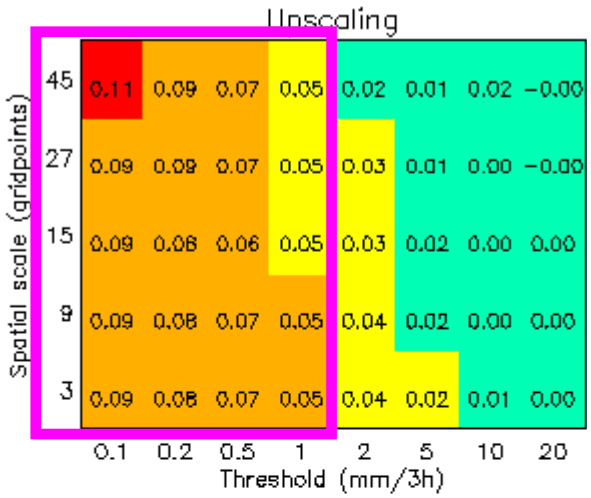
Results of fuzzy verification methods with COSMO over Switzerland and Germany

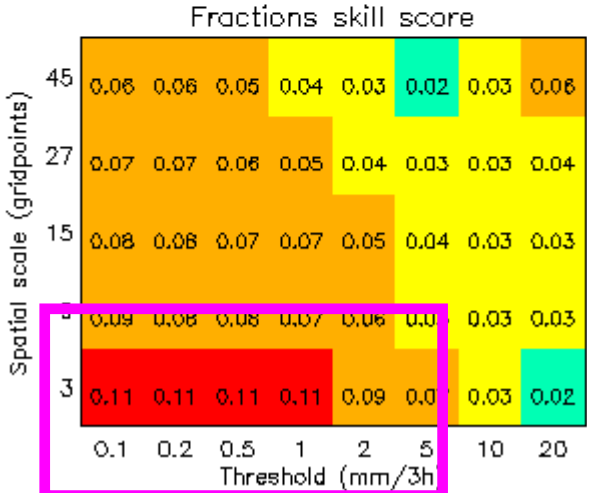
30th EWGLAM &15th SRNWP meeting, 7 October 2008, Madrid

COSMO-2 – COSMO-7 (2007)

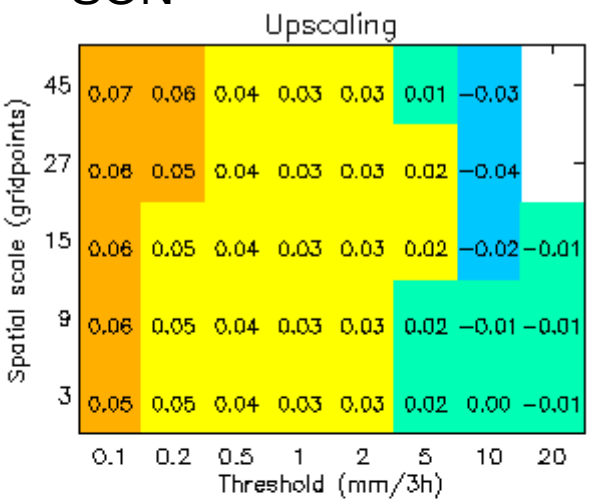


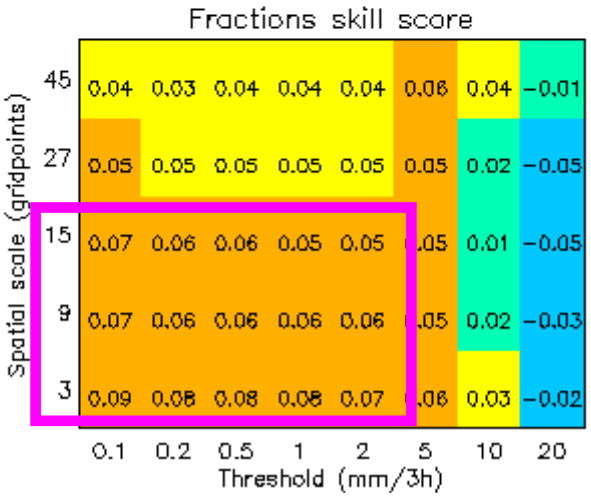
JJA



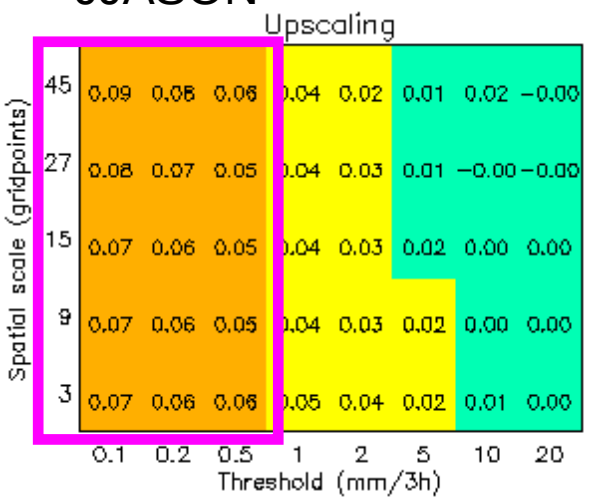


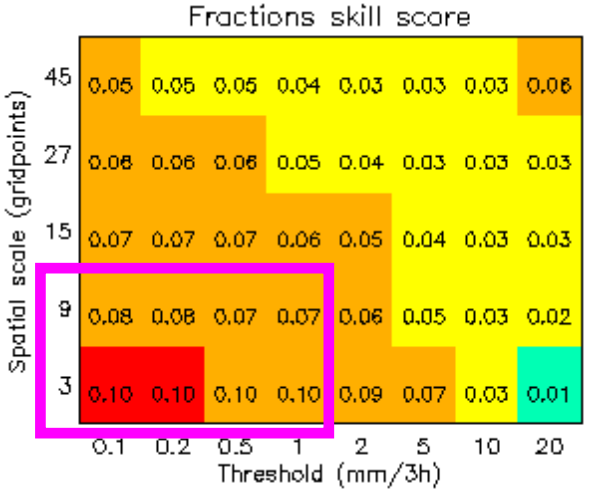
SON



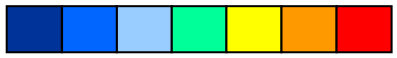


JJASON



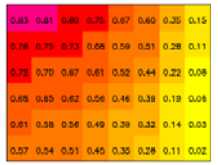


COSMO-7 better

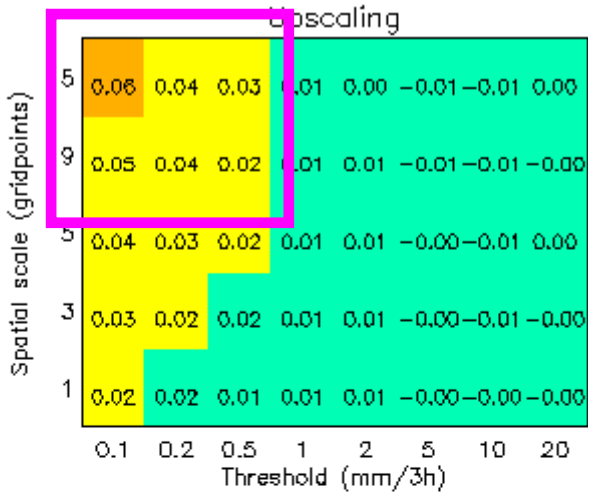


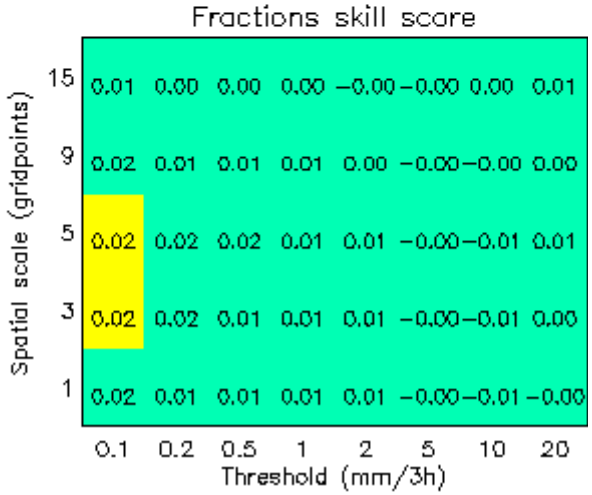
COSMO-2 better

COSMO-EU - COSMO-7 (2007)

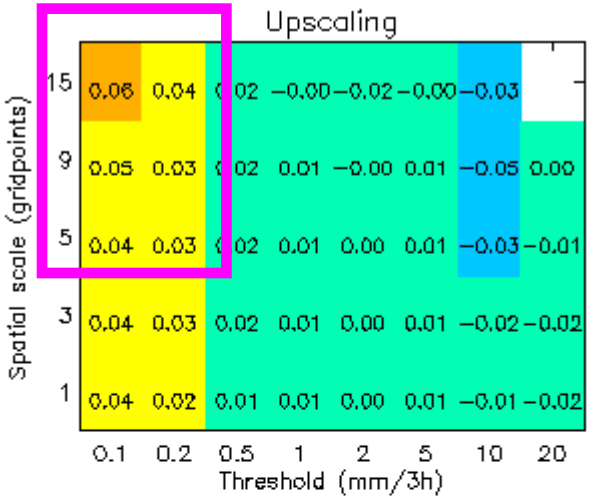


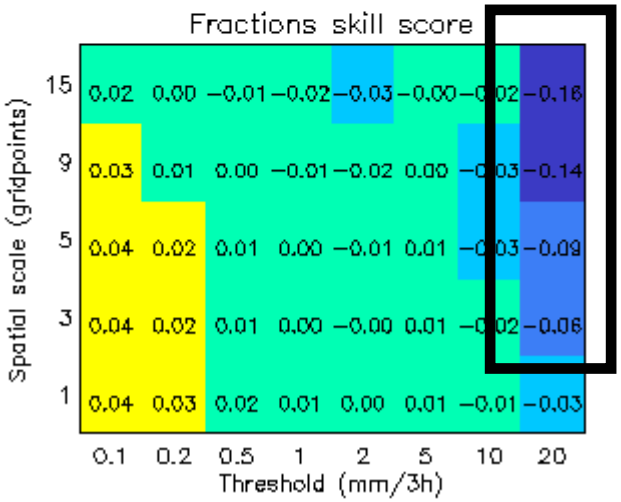
JJA



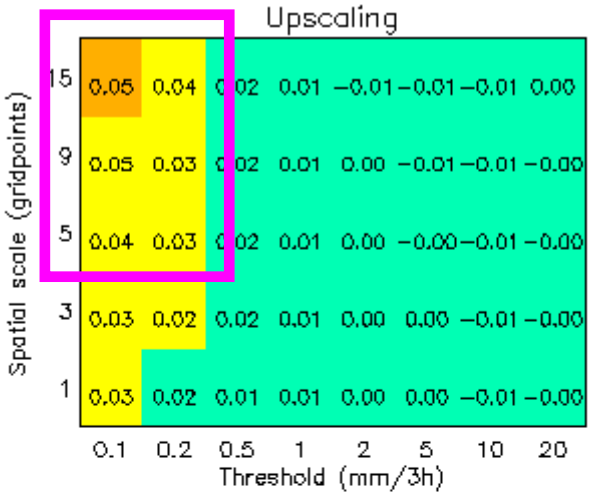


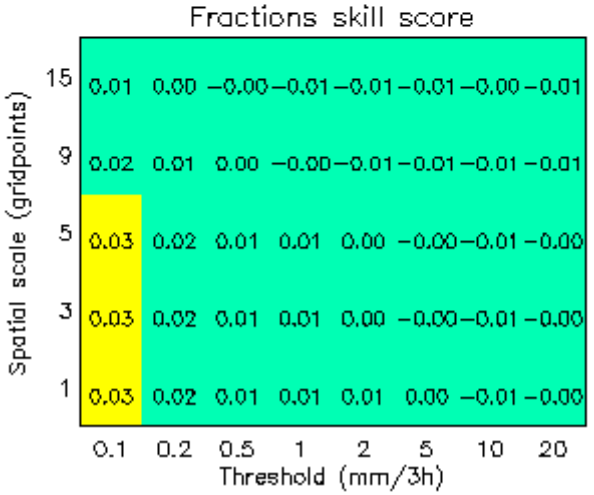
SON

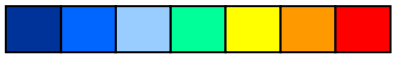




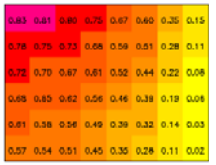
JJASON



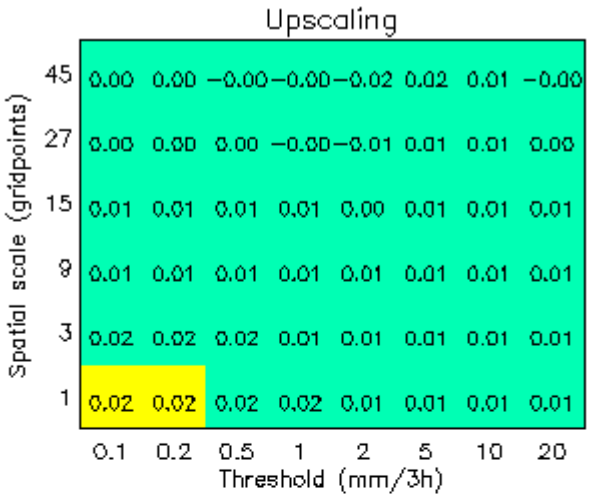


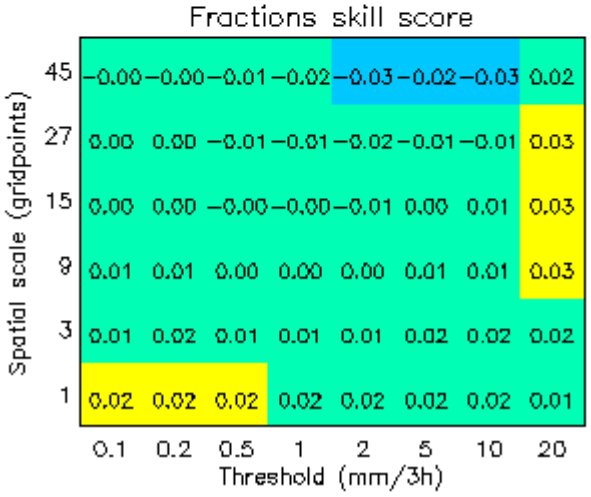
COSMO-7 better  COSMO-EU better

COSMO-DE - COSMO-2 (2007)

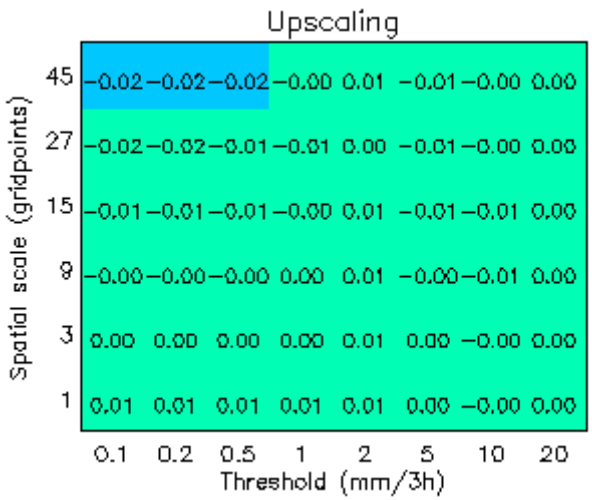


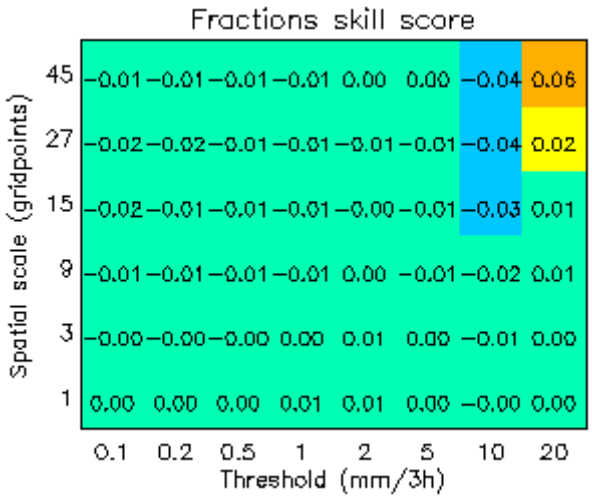
JJA



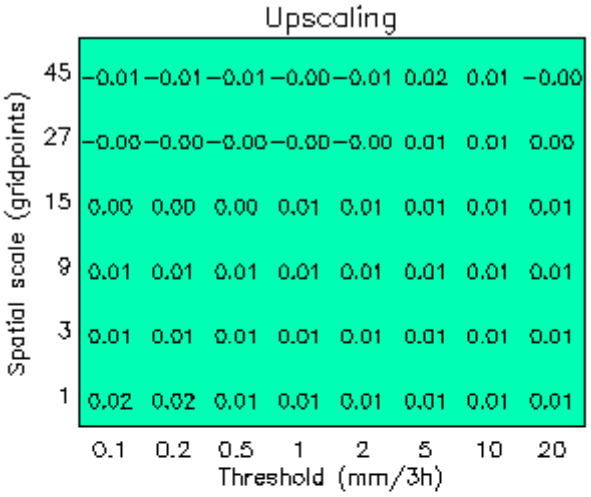


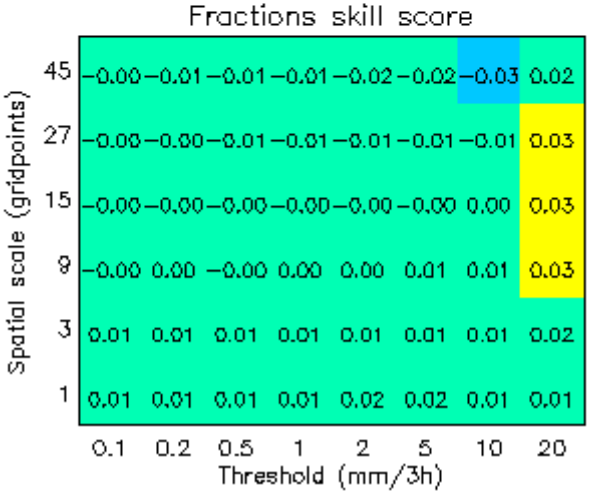
SON

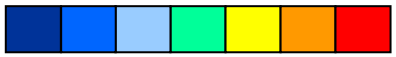




JJASON





COSMO-2 better  COSMO-DE better