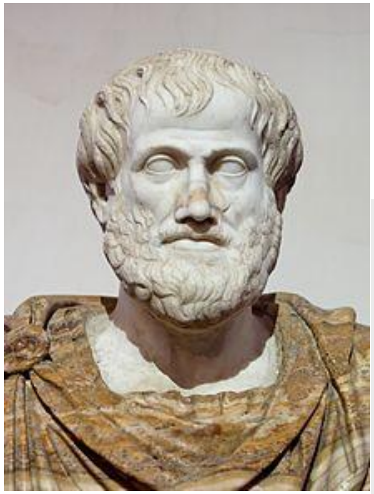
Intercomparison of Spatial Verification Methods for COSMO Terrain (INSPECT): Preliminary Results

Dmitry Alferov (1), Elena Astakhova (1), Dimitra Boukouvala (2), Anastasia Bundel (1), Ulrich Damrath (3), Pierre Eckert (4), Flora Gofa (2), Alexander Kirsanov (1), Xavier Lapillonne (4), Joanna Linkowska (5), Chiara Marsigli (6), Andrea Montani (6), Anatoly Muraviev (1), Elena Oberto (7), Maria Stefania Tesini (6), Naima Vela (7), Andrzej Wyszogrodzki (5), and Mikhail Zaichenko (1)

(1) RHM (a.bundel@gmail.com), (2) HNMS, (3) DWD, (4) MCH, (5) IMGW-PIB, (6) ARPA-SIMC, (7) ARPA-PT

EMS/ECAM, 07 – 11 September 2015, Sofia, Bulgaria

From Aristotle's Meteorologica
Book II, Chap IV, p. 167-169



Sometimes ***drought or rain*** is ***widespread and covers a large area*** of country, sometimes it is only ***local***; for often in the country at large the seasonal rainfall is normal or even above the normal, while in borne districts of it there is a draught ;
At other times, on the other hand, the rainfall in the country at large is meagre, or there is even a tendency to draught, while in a single district the rainfall is abundant in quantity.
The reason is that as a rule a considerable area may be expected to be similarly affected, because ***neighboring*** places lie in a similar relation to the sun, unless they have some local peculiarity ; …
And the reason for this again is the movement of either of the two exhalations across to join that of the neighboring district; the dry, for instance, may circulate in its own, the moist follow to a neighboring district or ***be driven by winds*** still farther afield.

COSMO Priority Project INSPECT

- summarizes the COSMO experience of applying spatial verification methods to high and very-high-resolution systems
- runs in parallel to MesoVICT (several INSPECT tasks involve reruns of COSMO models for MesoVICT test cases and analysis of MesoVICT cases)
- Same as MesoVICT, INSPECT focuses on the *ensembles* and *variables besides precipitation*
- In addition to targeting the goals of MesoVICT, INSPECT provides COSMO users more choice of verification domains and reference data - newer and longer periods, two complex terrains (the Alps and the Caucasus)
- Finally, INSPECT will try to provide criteria for deciding which methods are best suited to particular application
- Share the software

Tasks involving reruns of MesoVICT test cases

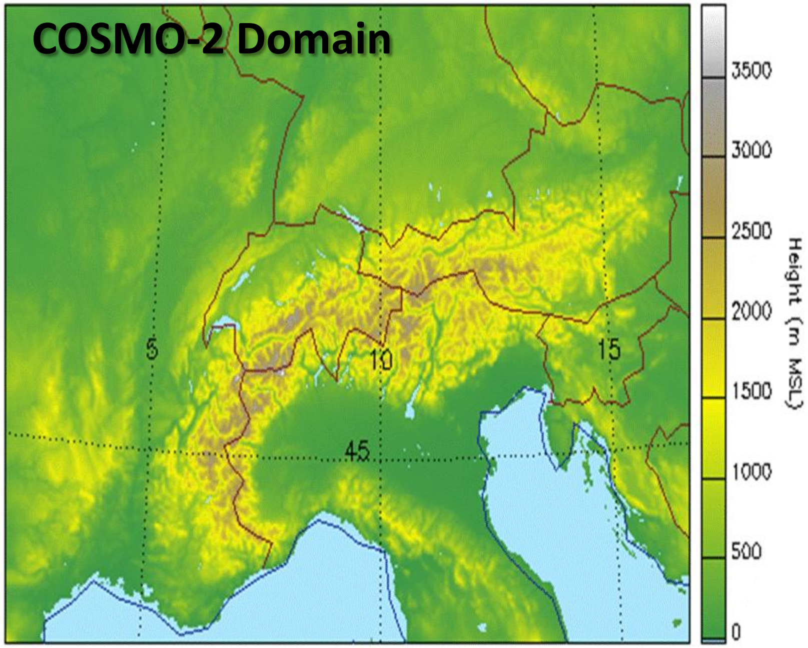
- MeteoSuiss:
  - Reruns of COSMO (COSMO-7, COSMO-2, COSMO-1) models for MesoVICT cases with more recent model versions, at least for case 1
  - Reruns of COSMO-E ensemble (?)
  - Done: The recalculations with COSMO-1 for the first MesoVICT period (20-22 June 2007)
- ARPA-SIMC: Reruns of global model ECMWF-EPS to provide boundary conditions for COSMO-LEPS. Reruns of COSMO-LEPS (Ongoing)
- Roshydromet: To run COSMO-Ru2-EPS from COSMO-LEPS IC&BC

Tasks involving analysis of MesoVICT cases

- ARPA-SIMC: DIST method (possibly also for wind speed)
- ARPA-PT: SAL for the core MesoVICT case (?)
- HNMS, Roshydromet, IMGW: Application of traditional categorical scores and spatial verification methods to analyze extreme precipitation events based on MesoVICT cases
- Verification study of COSMO-Ru-EPS (2.2 km) and, possibly, COSMO-E ensembles for MesoVICT cases
- Follow-up of the MesoVICT activities

MesoVICT core case

by Flora Gofa using COSMO VAST package for the neighborhood methods



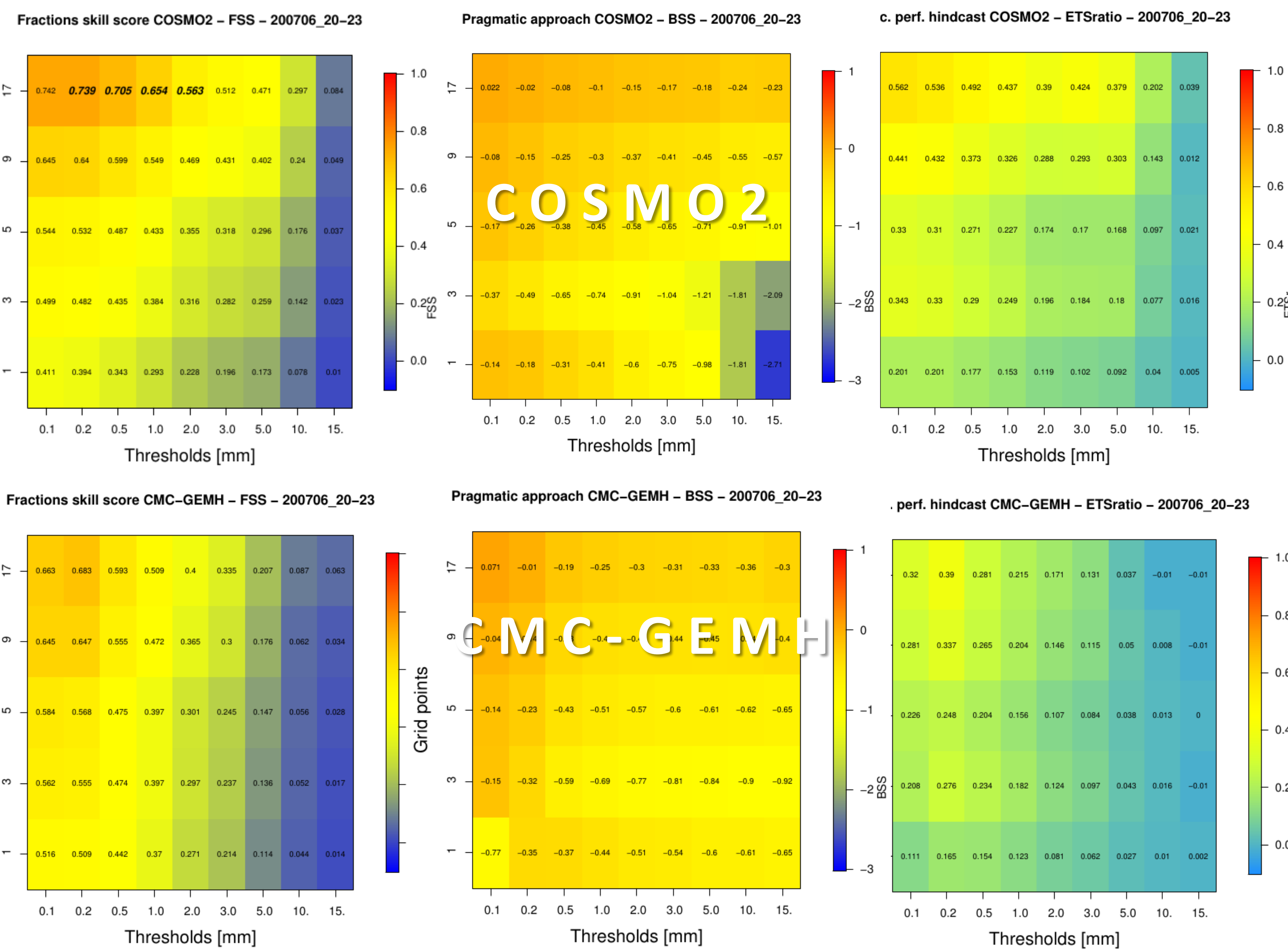
MesoVict Core case

Forecast model used:
1. COSMO-2 extrapolated to ~7km resolution
Data:20, 21,22.06.07:00-24UTC
Precipitation, 1h accumulation

2. CMC GEMH: in VERA resulution
Originally 2.5 km (0.0225 X 0.0327)
Data:20, 21,22.06.07: 06-18UTC
Precipitation, 6h accumulation

Observation data used: VERA analysis in ~7km resolution resolution

Data adapted by N.Vela and M.S.Tesini


Fractions skill score COSMO2 – FSS – 200706_20–23
0.742 0.739 0.705 0.654 0.563 0.512 0.471 0.297 0.084
0.645 0.64 0.599 0.549 0.469 0.431 0.402 0.24 0.049
0.544 0.532 0.487 0.433 0.355 0.318 0.296 0.176 0.037
0.499 0.482 0.435 0.384 0.316 0.282 0.259 0.142 0.023
0.411 0.394 0.343 0.293 0.228 0.196 0.173 0.078 0.01
17 9 5 3 1
0.1 0.2 0.5 1.0 2.0 3.0 5.0 10. 15.
Thresholds [mm]
FSS
1.0 0.8 0.6 0.4 0.2 0.0
Pragmatic approach COSMO2 – BSS – 200706_20–23
0.022 −0.02 −0.08 −0.1 −0.15 −0.17 −0.18 −0.24 −0.23
−0.08 −0.15 −0.25 −0.3 −0.37 −0.41 −0.45 −0.55 −0.57
−0.17 −0.26 −0.38 −0.45 −0.58 −0.65 −0.71 −0.91 −1.01
−0.37 −0.49 −0.65 −0.74 −0.91 −1.04 −1.21 −1.81 −2.09
−0.14 −0.18 −0.31 −0.41 −0.6 −0.75 −0.98 −1.81 −2.71
COSMO2
17 9 5 3 1
0.1 0.2 0.5 1.0 2.0 3.0 5.0 10. 15.
Thresholds [mm]
BSS
1 0 −1 −2 −3
c. perf. hindcast COSMO2 – ETSratio – 200706_20–23
0.562 0.536 0.492 0.437 0.39 0.424 0.379 0.202 0.039
0.441 0.432 0.373 0.326 0.288 0.293 0.303 0.143 0.012
0.33 0.31 0.271 0.227 0.174 0.17 0.168 0.097 0.021
0.343 0.33 0.29 0.249 0.196 0.184 0.18 0.077 0.016
0.201 0.201 0.177 0.153 0.119 0.102 0.092 0.04 0.005
0.1 0.2 0.5 1.0 2.0 3.0 5.0 10. 15.
Thresholds [mm]
1.0 0.8 0.6 0.4 0.2 0.0
Fractions skill score CMC–GEMH – FSS – 200706_20–23
0.663 0.683 0.593 0.509 0.4 0.335 0.207 0.087 0.063
0.645 0.647 0.555 0.472 0.365 0.3 0.176 0.062 0.034
0.584 0.568 0.475 0.397 0.301 0.245 0.147 0.056 0.028
0.562 0.555 0.474 0.397 0.297 0.237 0.136 0.052 0.017
0.516 0.509 0.442 0.37 0.271 0.214 0.114 0.044 0.014
17 9 5 3 1
0.1 0.2 0.5 1.0 2.0 3.0 5.0 10. 15.
Thresholds [mm]
Grid points
Pragmatic approach CMC–GEMH – BSS – 200706_20–23
0.071 −0.01 −0.19 −0.25 −0.3 −0.31 −0.33 −0.36 −0.3
CMC-GEMH
−0.14 −0.23 −0.43 −0.51 −0.57 −0.6 −0.61 −0.62 −0.65
−0.15 −0.32 −0.59 −0.69 −0.77 −0.81 −0.84 −0.9 −0.92
−0.77 −0.35 −0.37 −0.44 −0.51 −0.54 −0.6 −0.61 −0.65
17 9 5 3 1
0.1 0.2 0.5 1.0 2.0 3.0 5.0 10. 15.
Thresholds [mm]
BSS
1 0 −1 −2 −3
. perf. hindcast CMC–GEMH – ETSratio – 200706_20–23
0.32 0.39 0.281 0.215 0.171 0.131 0.037 −0.01 −0.01
0.281 0.337 0.265 0.204 0.146 0.115 0.05 0.008 −0.01
0.226 0.248 0.204 0.156 0.107 0.084 0.038 0.013 0
0.208 0.276 0.234 0.182 0.124 0.097 0.043 0.016 −0.01
0.111 0.165 0.154 0.123 0.081 0.062 0.027 0.01 0.002
0.1 0.2 0.5 1.0 2.0 3.0 5.0 10. 15.
Thresholds [mm]

COMPACT REPRESENTATION OF LONG-TIME SERIES OF SCORES

By Uli Damrath,
DWD

Comparison of COSMO-EU to COSMO-DE – upscaling ETS (differences)

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Threshold 0.1 0.2 0.5 1 2 5 10 20 50

1.625
0.825
0.425
0.225
0.125
0.075
0.025

Window size

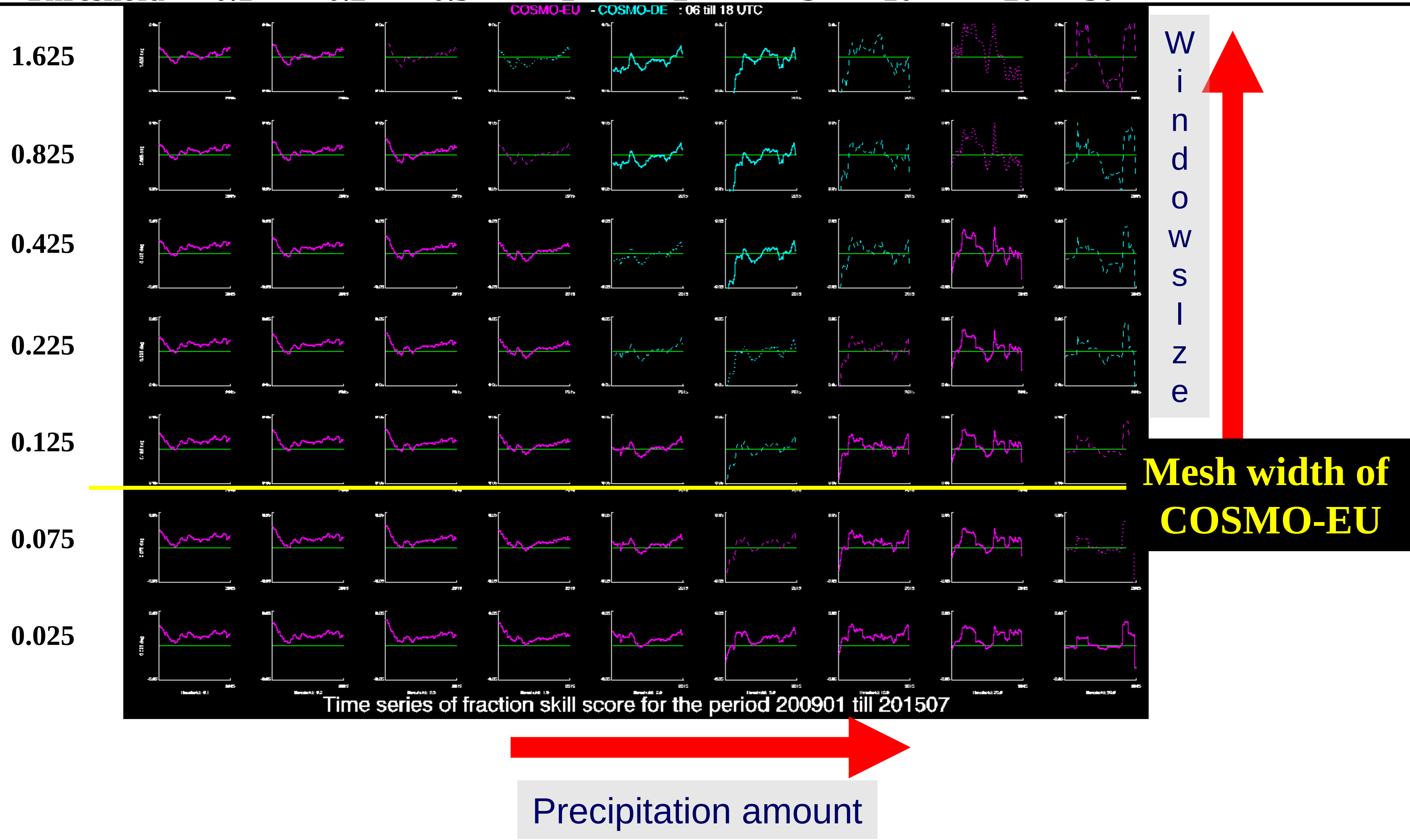
Mesh width of COSMO-EU

Precipitation amount

Comparison of COSMO-EU to COSMO-DE – FSS (differences)

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Threshold	0.1	0.2	0.5	1	2	5	10	20	50



A study on verification of FROST-2014 precipitation forecast fields using neighborhood and CRA methods

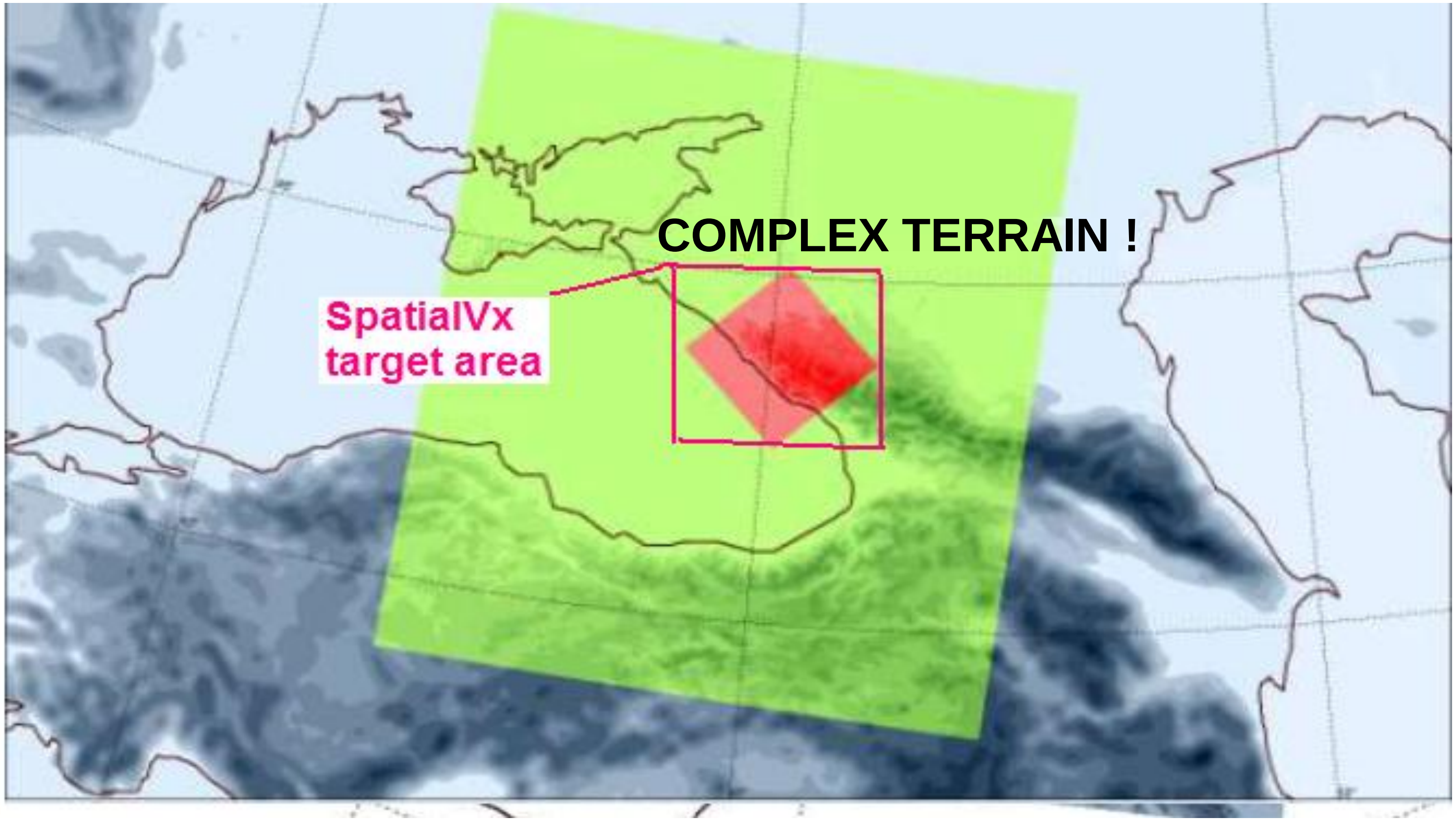
Anatoly Muraviev (1), Anastasia Bundel (1), Dmitry Kiktev (1), Nikolay Bocharnikov (2), and Tatiana Bazlova (2)

(1) Hydrometcentre of Russia/Roshydromet, Moscow,

(2) Institute of Radar Meteorology, Saint-Petersburg, Russia

Area of the study

349 lon points * 481 lat points with **0.00833** lat-lon increments.
1 grid size by **longitude** = 111*0.00833 = **930 m**,
1 grid size by **latitude** = cos(43°35')*930 m = 0.72*930 = **~ 670 m**



COSMO-Ru2 domain

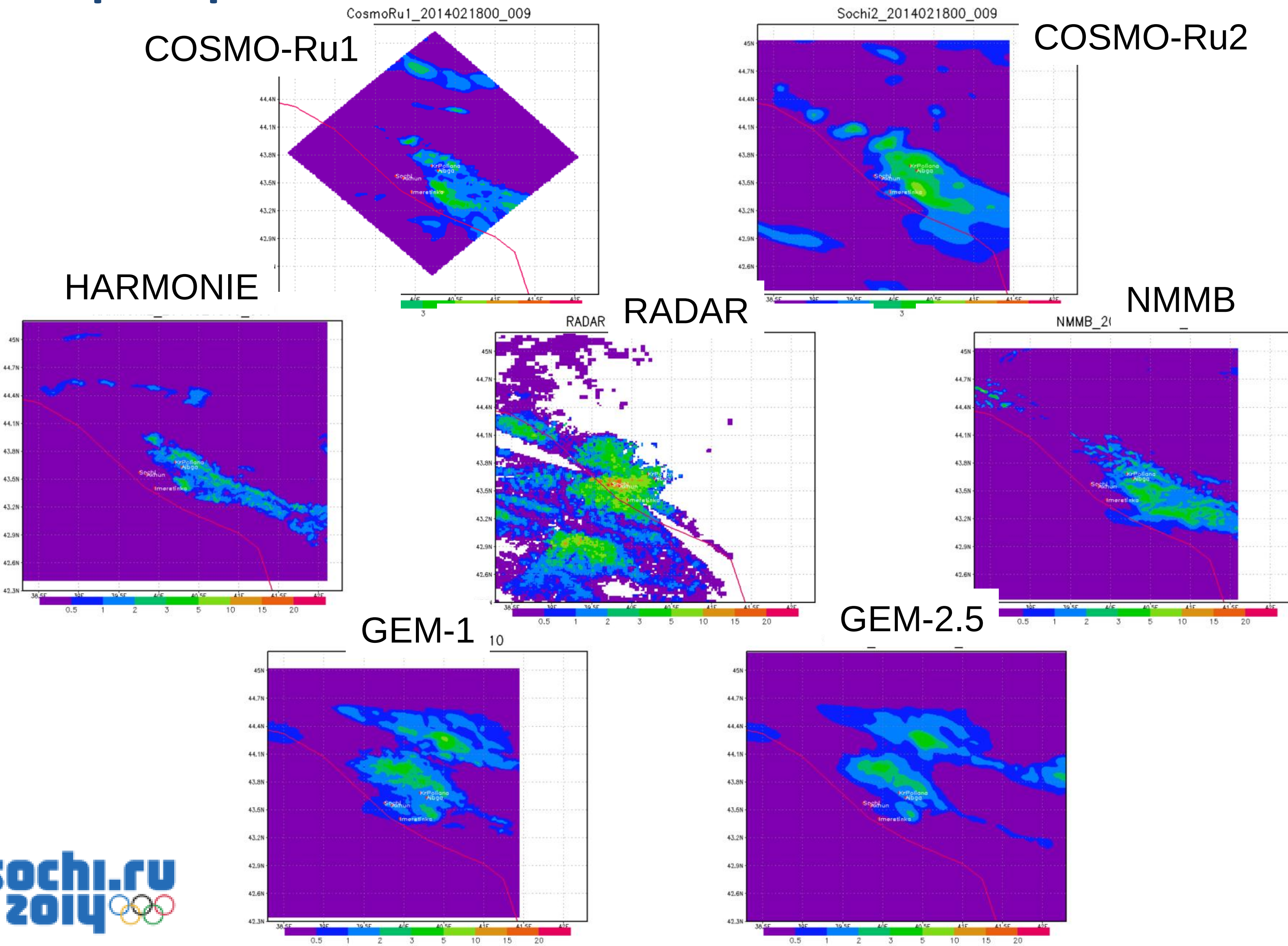
COSMO-Ru1 domain

All the models were interpolated into the radar grid using GRADS (function *Iterp*)

- COSMO-Ru1 (1 km)
- COSMO-Ru2 (2 km)
- NMMB (1 km)
- HARMONIE (1 km)
- GEM-1 (1 km)
- GEM-2.5 (2.5 km)

GEM-0.25: too small domain!

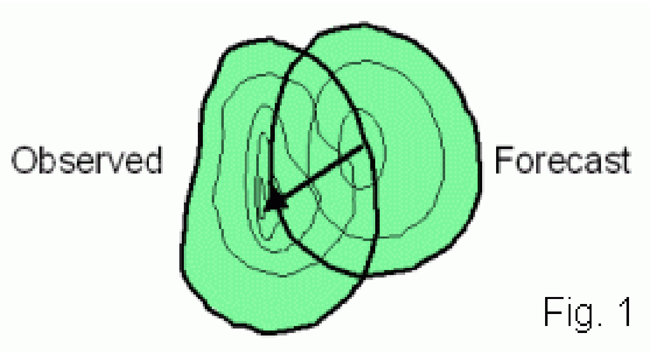
18 Feb 2014, 09 UTC, cold front: All models underestimated max precip and didn’t give precip over the sea.



CRA – Contiguous Rain Area (E.E. Ebert, J.L. McBride 2000)

http://www.cawcr.gov.au/projects/verification/CRA/CRA_verification.html

MSEtotal = MSEdisplacement + MSEvolume + MSEpattern



MSEdisplacement = MSEtotal – MSEshifted

MSEvolume = (***F*** - ***X***)2
where ***F*** and ***X*** are the CRA mean forecast and observed values after the shift.

MSEpattern = MSEshift – MSEvolume

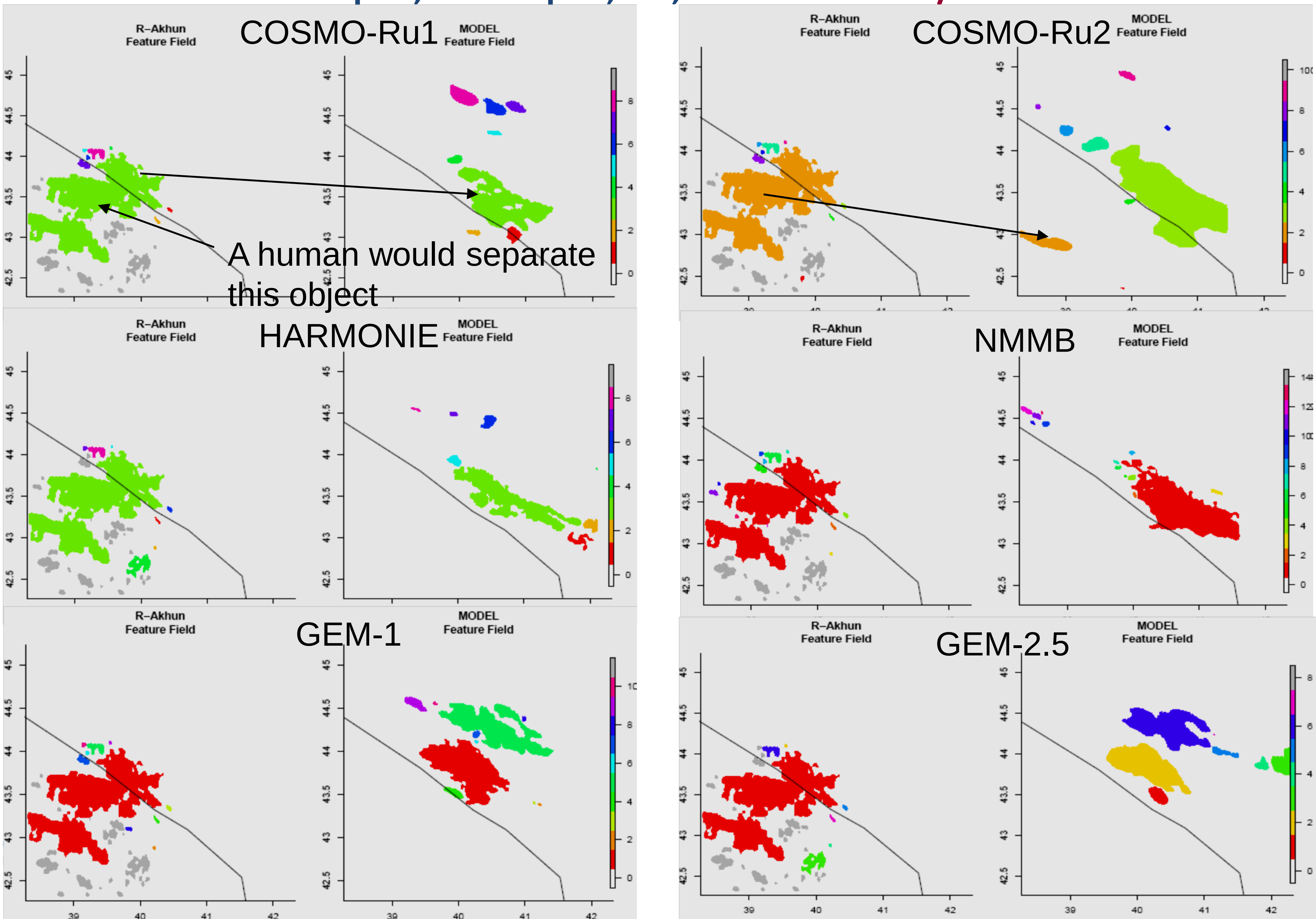
The CRA concept is easy to understand,
but there are many important issues and nuances in application of the CRA

R SpatialVx *craer* function

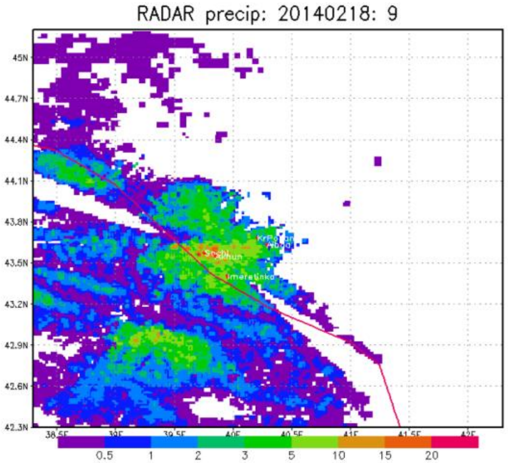
- Convolution threshold technique. First, the field is smoothed using a convolution smoother, and then it is set to a binary image where everything above a given threshold is set to one (Davis et al, 2006)
- *Minboundmatch* function– each object is pared to only one object according to the smallest minimum boundary separation

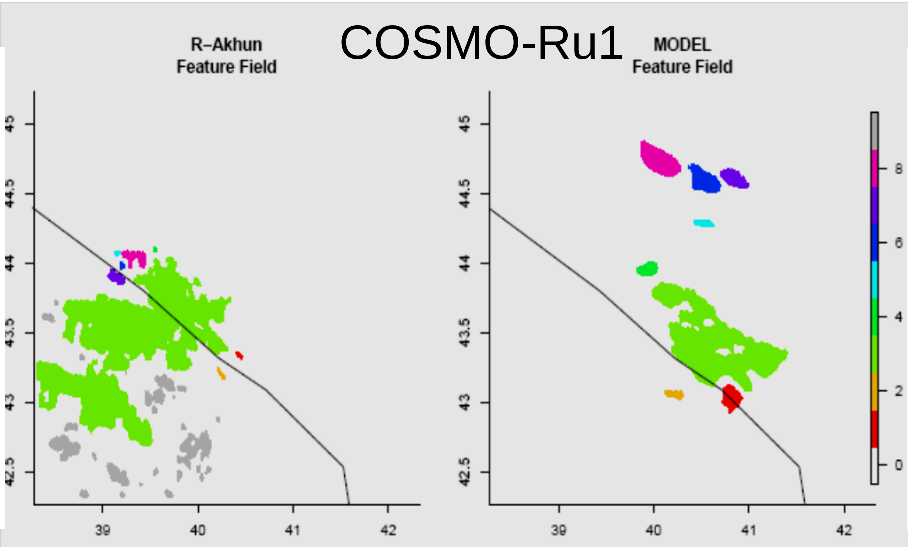
```
hold <- make.SpatialVx(xx, yy, map=TRUE, loc=zz,
field.type="Precipitation", units="mm/h",
data.name=c("Sochi_frcsts", "R-Akhun", "GEM25"))

look <- convthresh(hold, smoothpar=3, thresh=1)

look2 <- minboundmatch( look )

craer( look2, type = "fast", verbose = TRUE)
```

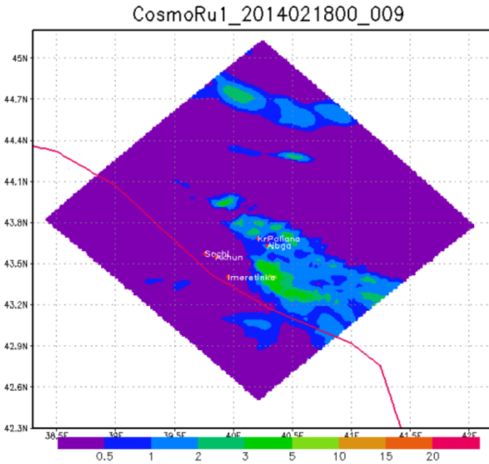
Pairs of matched objects from *craer*, 18 Feb 2014, 09 UTC
Colors indicate the 1st pair, the 2nd pair, etc, threshold: 1mm/h



COSMO-Ru1



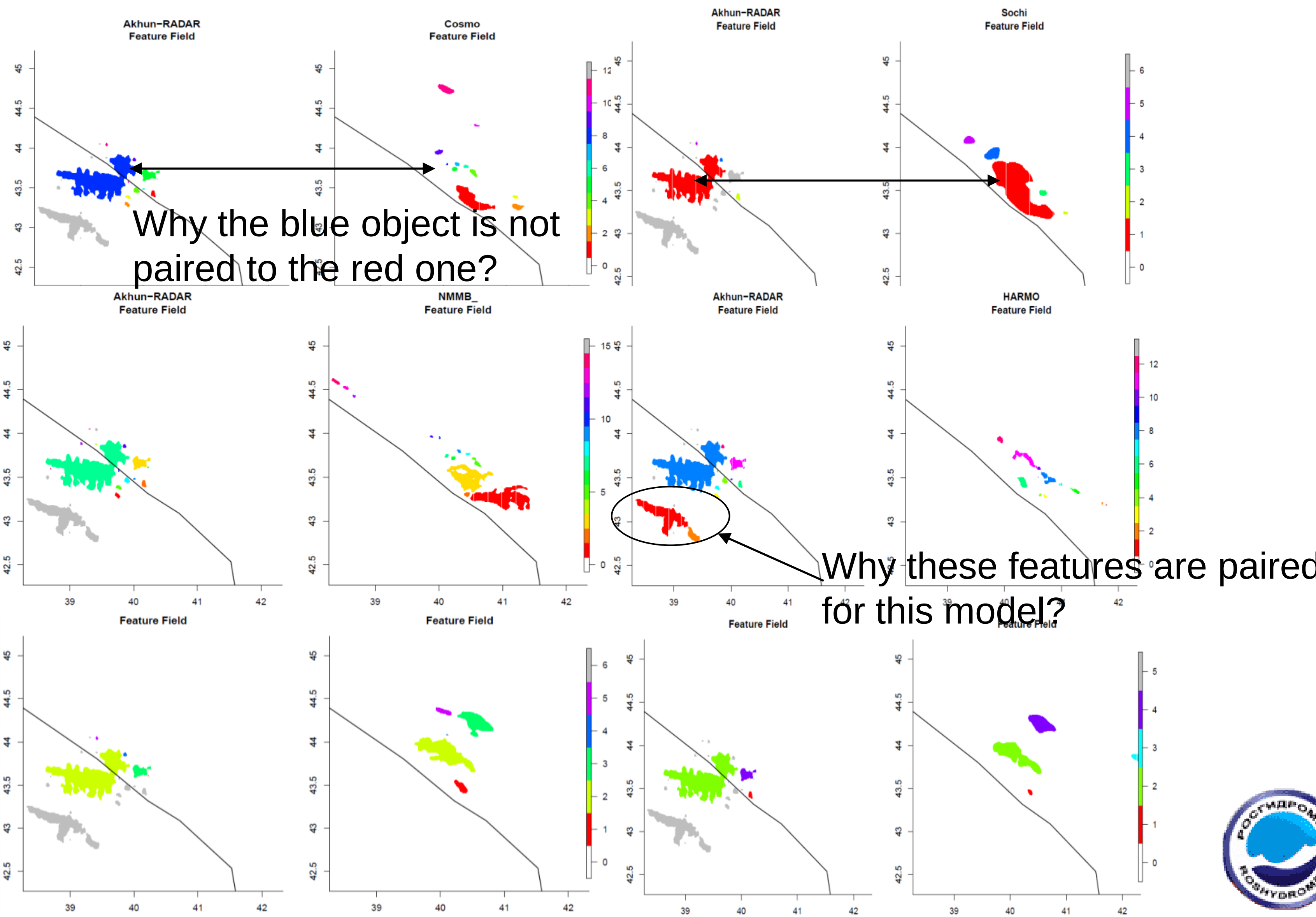




ir	x	y	MSE.total	MSE.shift	MSE.displace	MSE.volume	MSE.pattern
1	45.4021	-36.5179	0.0028	0.0023	0.0005	0.0000	0.0022
2	-2.7630	-17.8333	0.0011	0.0007	0.0004	0.0000	0.0007
3	159.7069	2.3035	0.1246	0.0820	0.0426	0.0027	0.0793
4	45.9893	-16.9170	0.0014	0.0012	0.0002	0.0000	0.0012
5	164.7442	25.5963	0.0011	0.0006	0.0005	0.0000	0.0006
6	159.3112	74.1525	0.0037	0.0033	0.0004	0.0000	0.0033
7	204.6084	85.0732	0.0032	0.0013	0.0019	0.0000	0.0013
8	85.7776	83.9482	0.0068	0.0036	0.0031	0.0000	0.0036

According to these scores, most of the total MSE error comes from the small-scale pattern errors for most object pairs

CRA threshold: 2 mm/h
(3mm/h gives too many little objects!)



Questions:

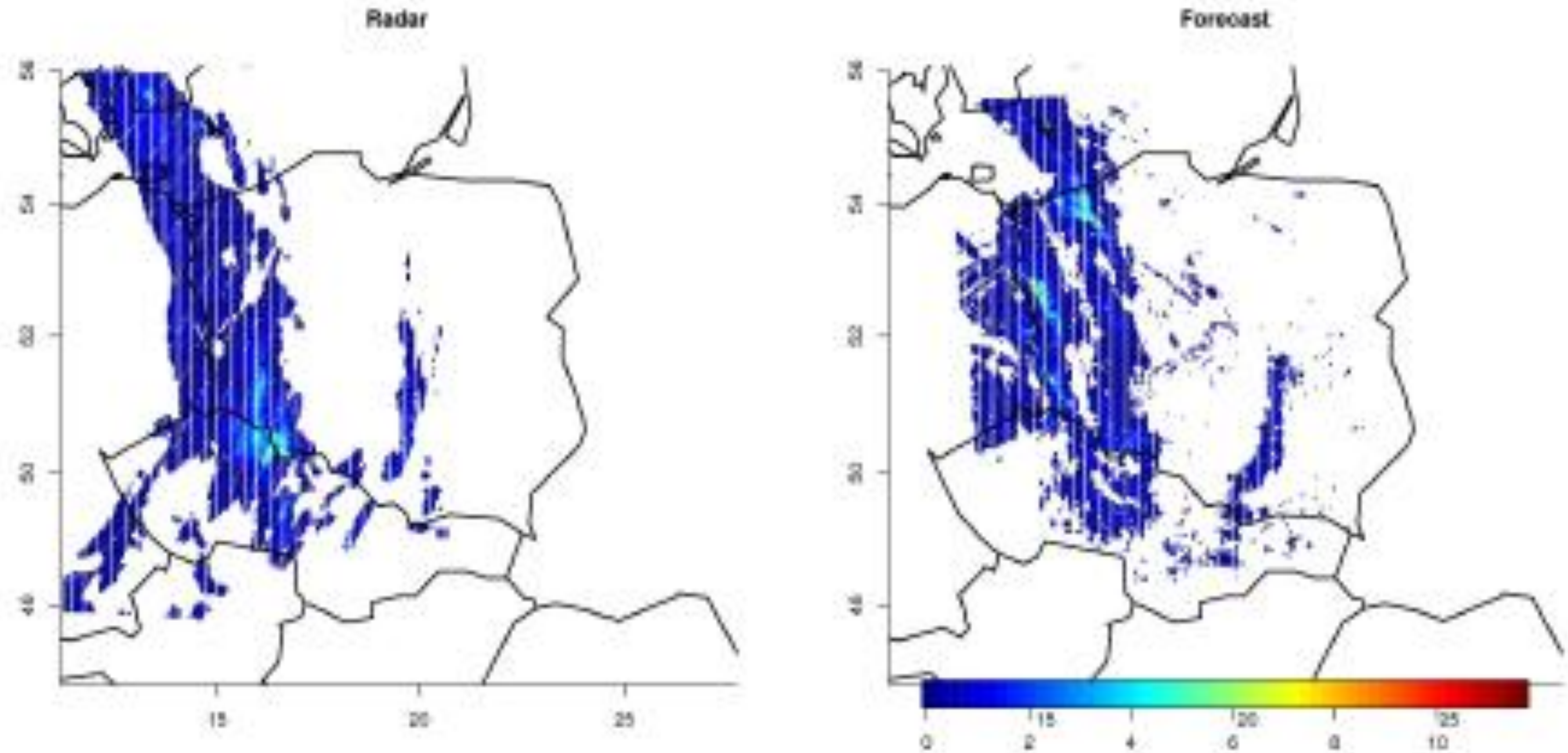
- There are many little objects. Can we set up a limitation on the maximum number of objects?
- Two apparently similar GEM fields: Different model objects are paired with the same radar object.
- Should there be a condition on the area size when pairing the objects? (the largest is paired to the largest) *centmatch*?
- Try another pairing methods (*deltamm, e.g.*)?

This study shows that we are not yet able to give general CRA statistics about the location, volume, and fine-scale structure neither can we yet range the models according to these statistics

MODE application to COSMO PL

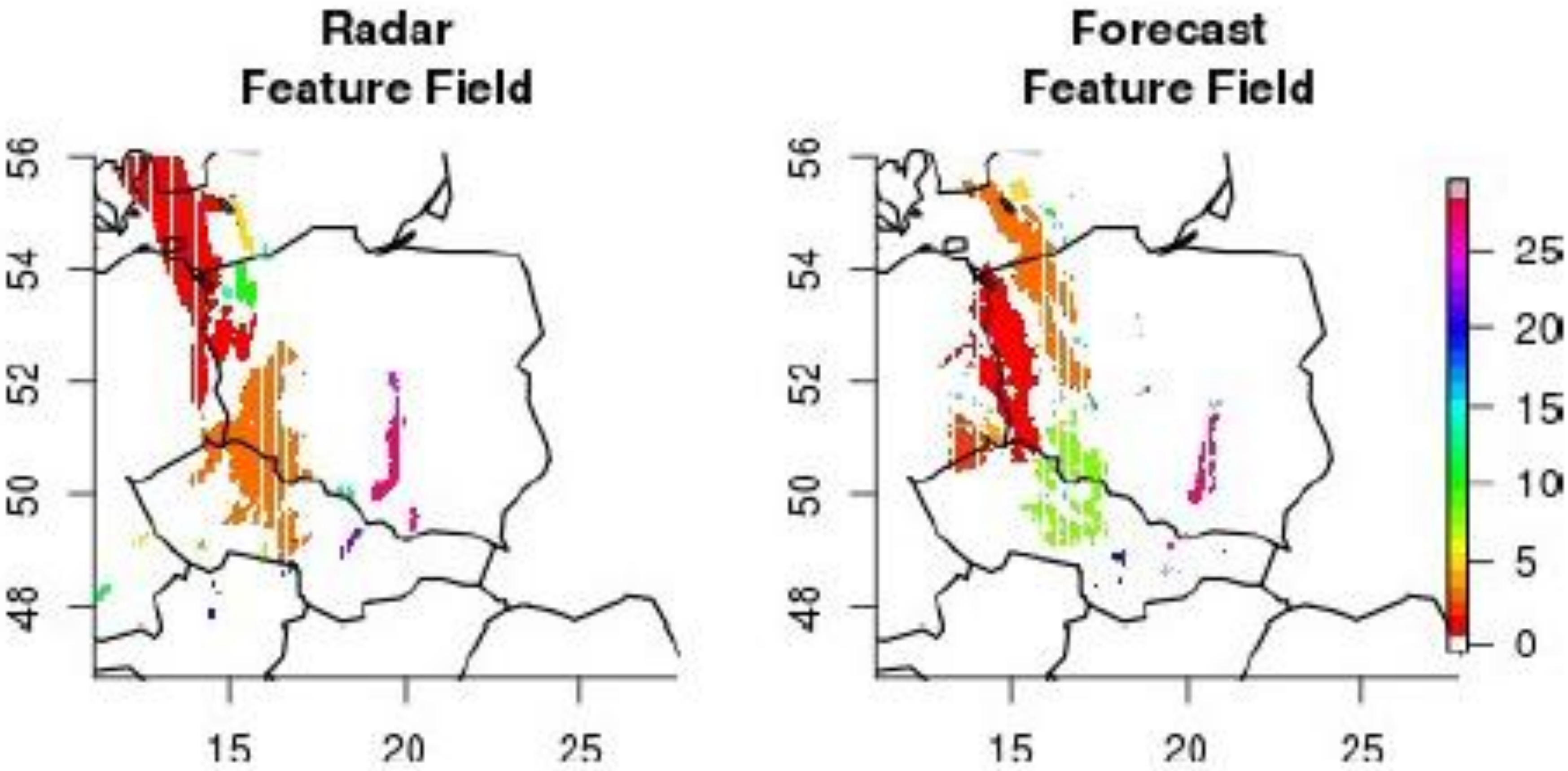
Joanna Linkowska
Andrzej Wyszogrodzki
IMGW-PIB

25.08.2015, 07 UTC
COSMO PL 2.8, 24.08.2015 18 UTC + T13

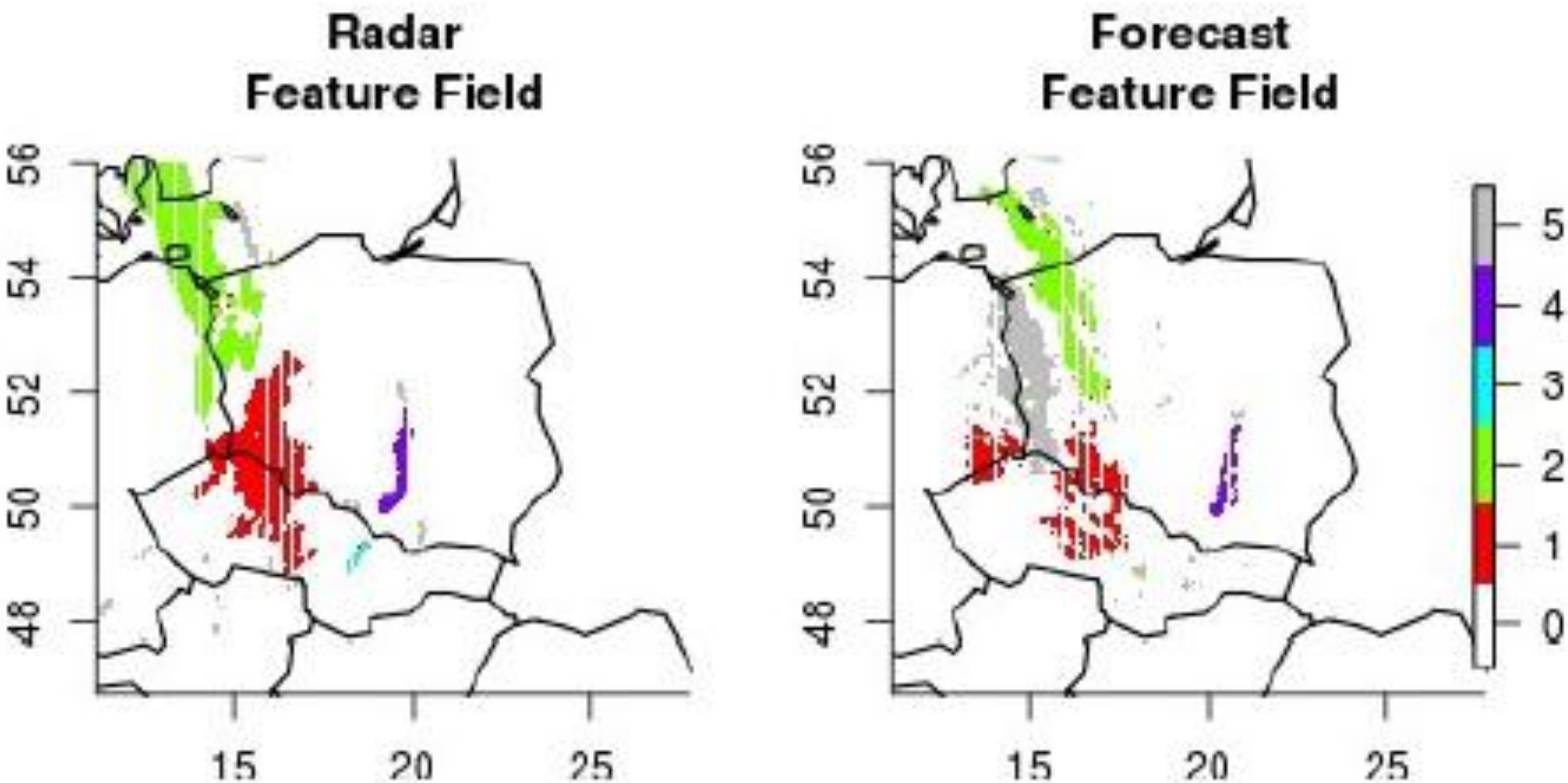


make.SpatialVx

Minboundmatch, smoothpar = 1.5, thresh = 0.15
25.08.2015, 07 UTC
COSMO PL 2.8, 24.08.2015 18 UTC +T13



Centmatch = 2, smoothpar = 1.5, thresh = 0.15
25.08.2015, 17 UTC
COSMO PL 2.8, 24.08.2015 18 UTC +T13



White - Zero values, Grey - Unmatched features

D < **average size of the two features**

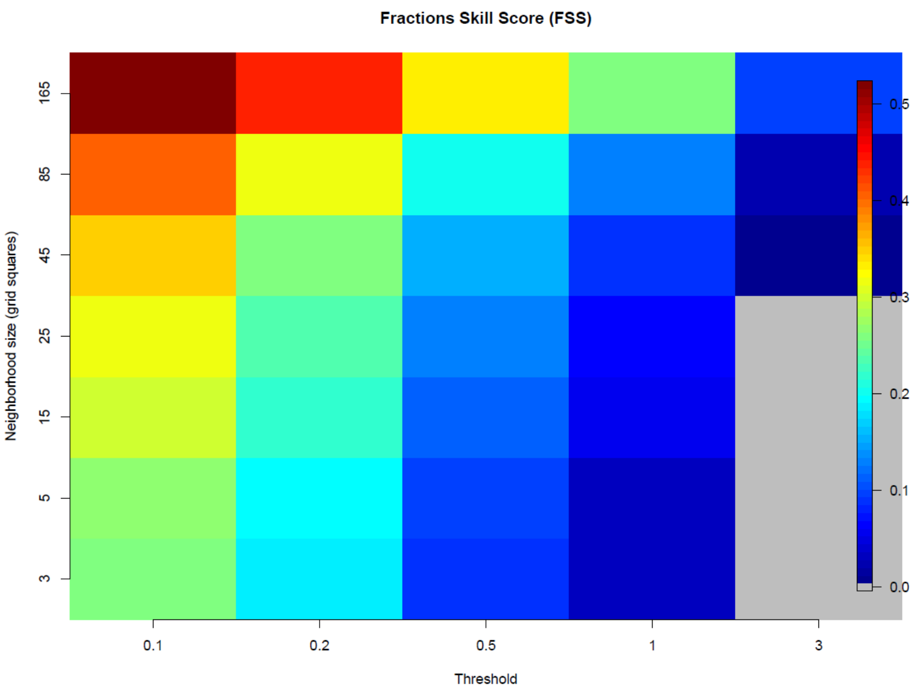
The main benefit of INSPECT

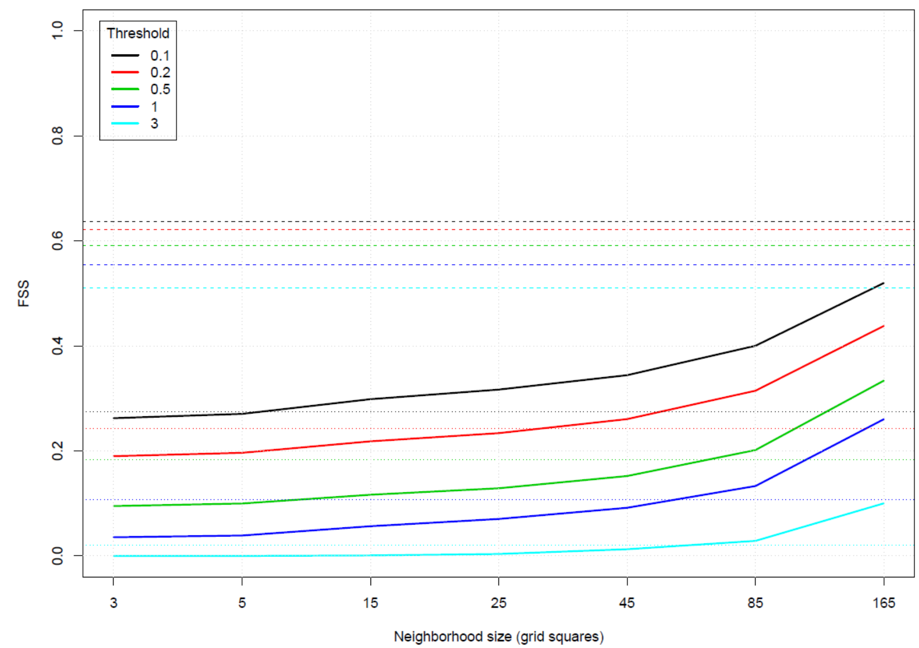
will be that a wide range of spatial verification methods will become commonly used within the COSMO community and COSMO Guidelines will be proposed to ensure the correct interpretation of results of these methods.

Thank you for your attention!

hoods2d

- Different scores were calculated, but the FSS (Roberts and Lean 2008) is presented as one of the most useful neighborhood statistics (see, e.g., COSMO INTERP project)

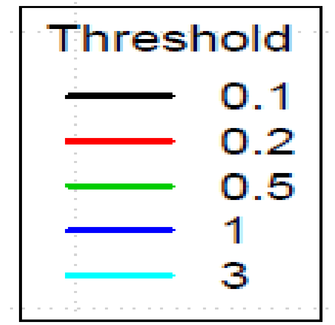


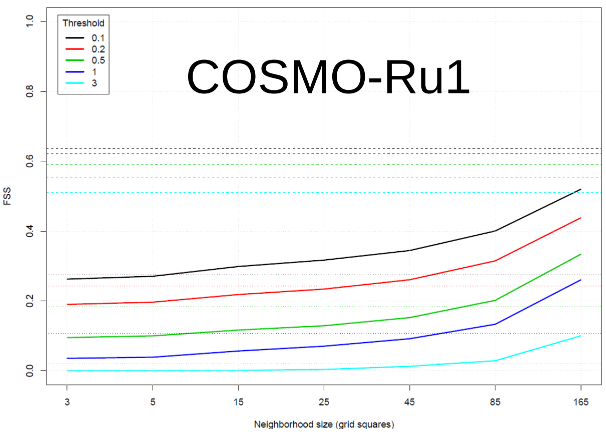


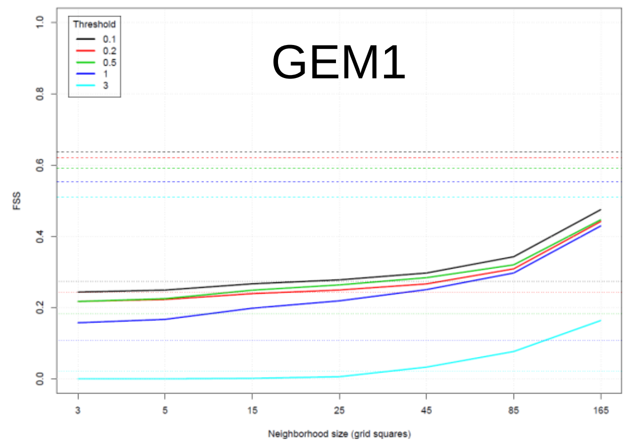
FSS, 18 Feb 2014, 09 UTC

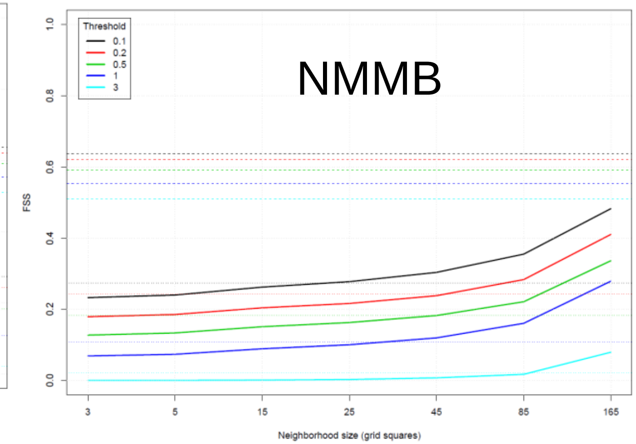
Note: 2-2.5-km models are interpolated onto ~1km grid!

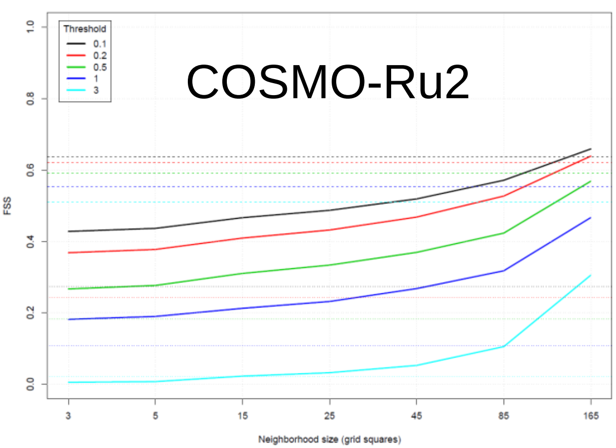
COSMO-Ru2 best here, its FSS is useful at all scales except for the highest threshold (precip ≥ 3mm/h)
GEM-1 is good for middle thresholds (0.5 and 1 mm/h)

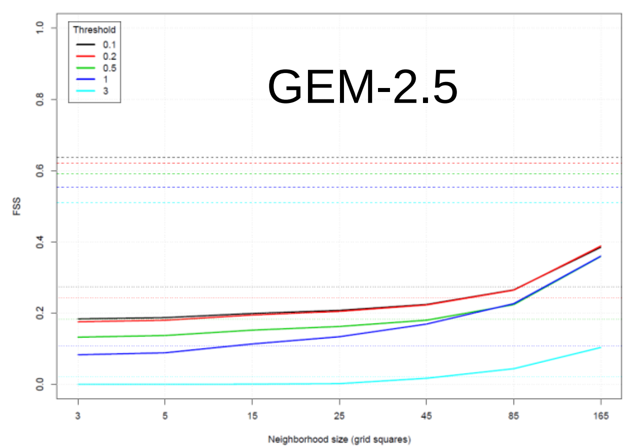


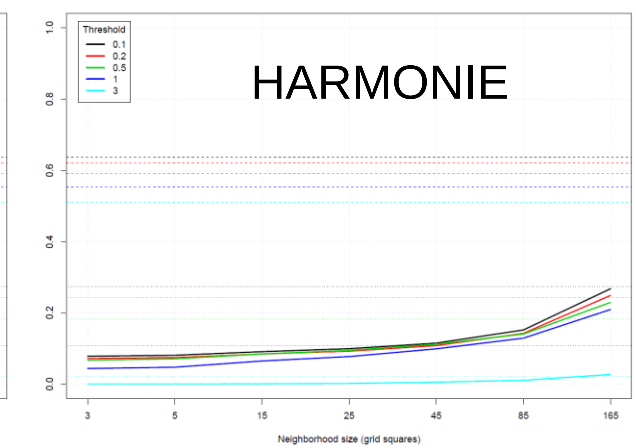




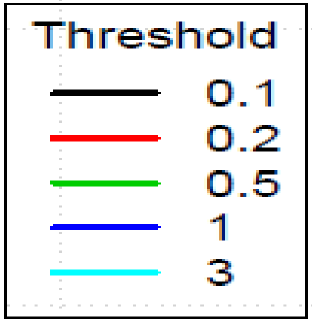




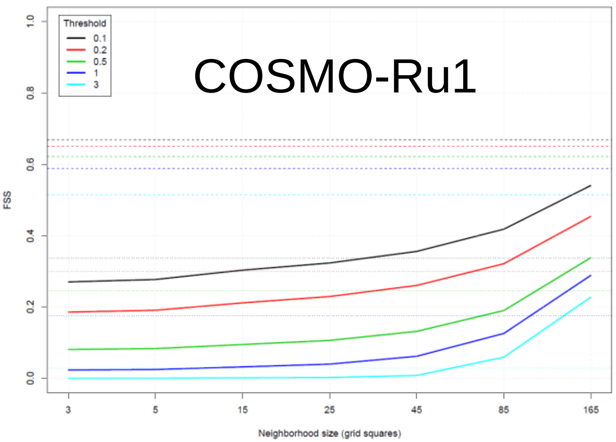


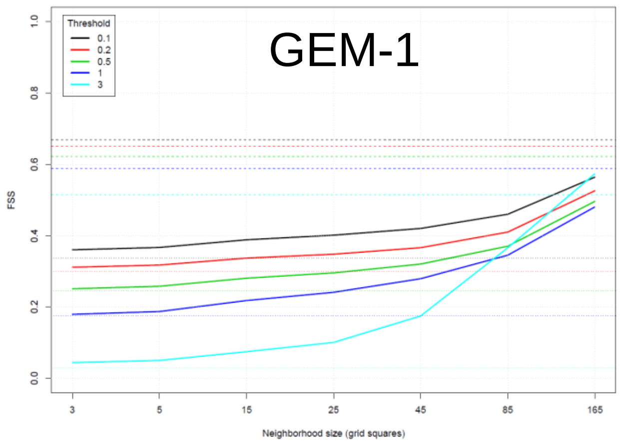


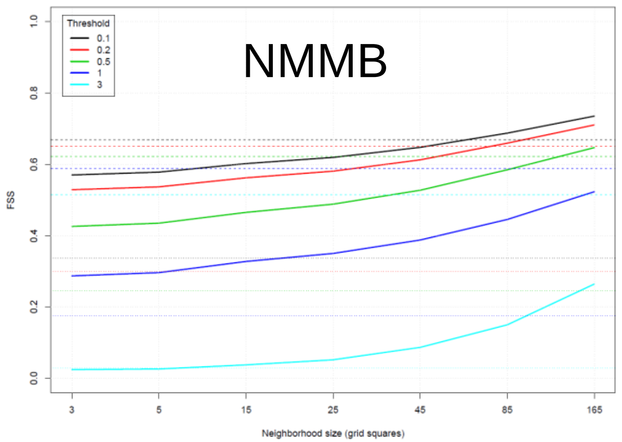
FSS, 18 Feb 2014, 17 UTC

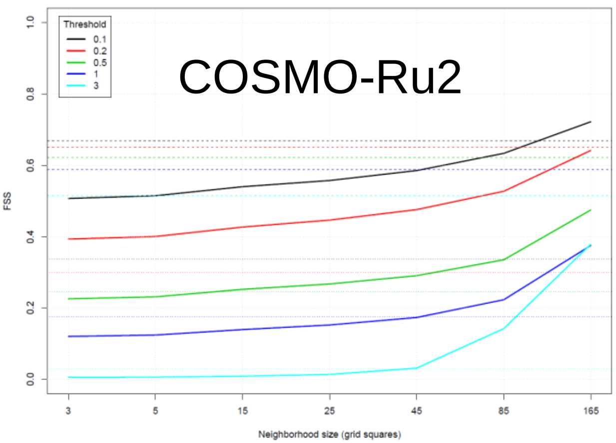


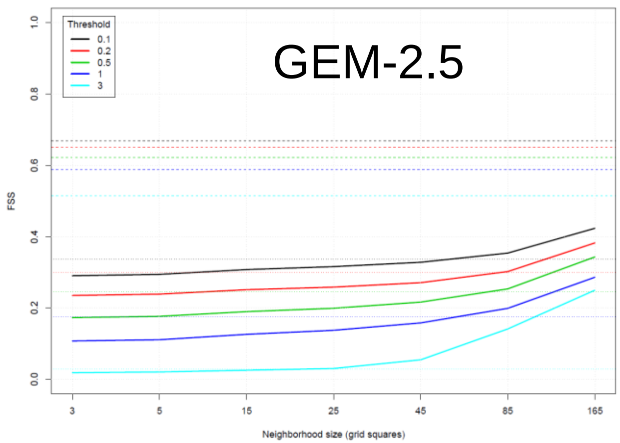
NMMB and HARMONIE have comparable high skill.
COSMO-Ru2 looses its skill for higher thresholds

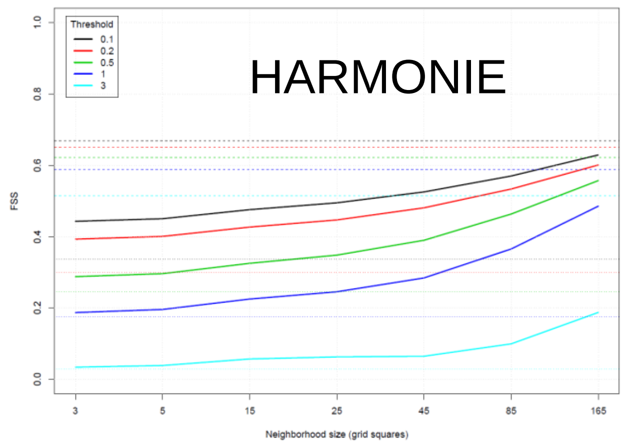




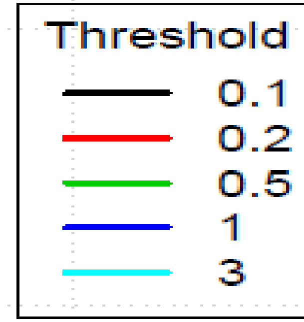




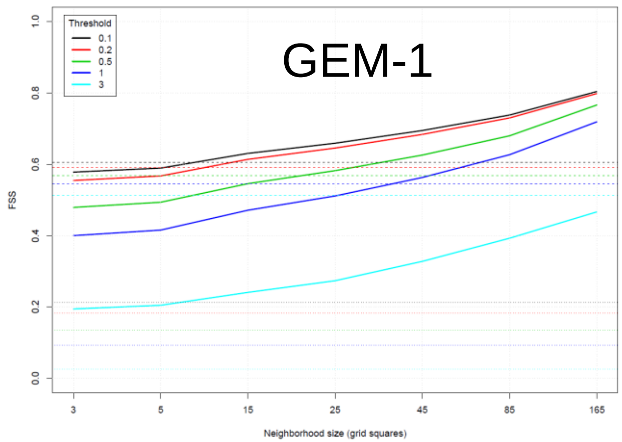


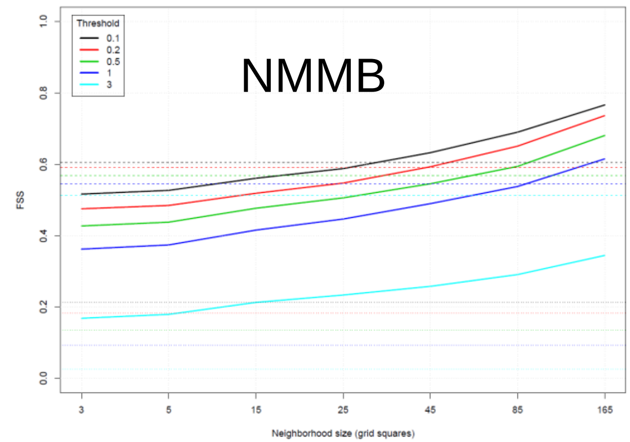


22 Jan 2014, 23 UTC, intense precipitation



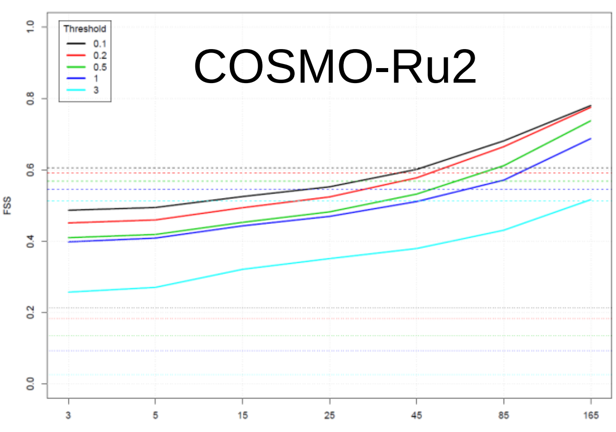
Good forecast by all models.
COSMO-Ru2 and GEM-1 are the leaders

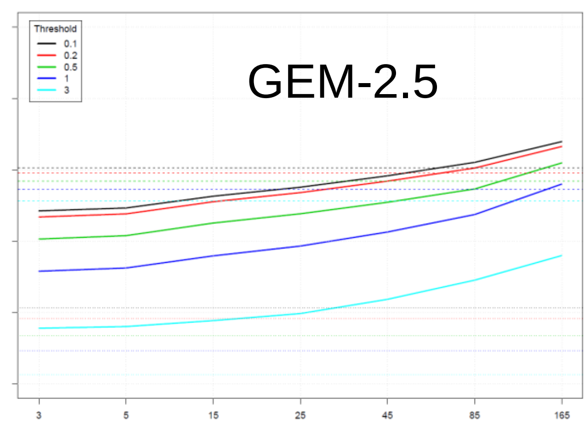


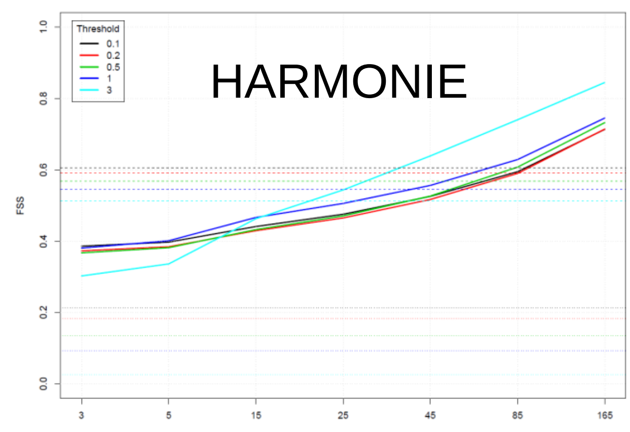


COSMO-Ru1

Not avail. until 29 Jan







Neighborhood: conclusions and plans

- **All the models underestimated the maximum precipitation**
- **According to the FSS, COSMO-Ru2 tends to be better then COSMO-Ru1, GEM-1 is better than GEM-2**
- **All the models (esp. COSMO-Ru2) loose skill for precip ≥ 3mm/1h (the last threshold)**

We need to:

- **aggregate neighborhood scores over all cases to estimate the systematic models’ behavior**
- **include the cases where precipitation was predicted, but not observed**
- **analyze timing errors**

18 Feb 2014, 17 UTC, all models predicted expanding precipitation area, but not the max value