Neighborhood vs Filtering methods

Flora Gofa

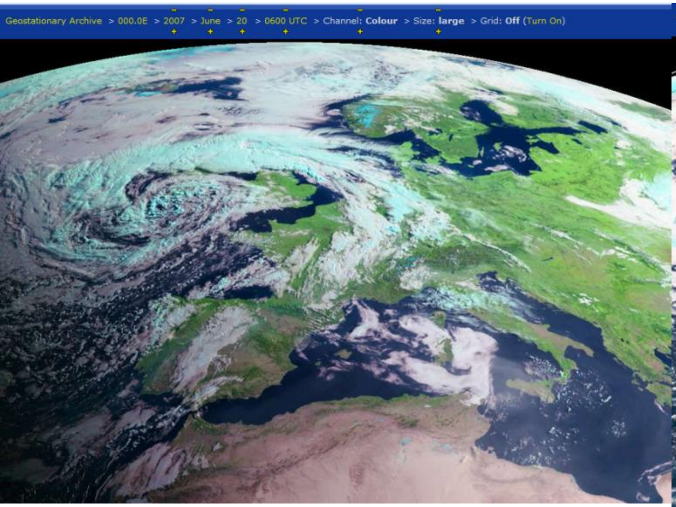
ΕΘΝΙΚΗ
ΜΕΤΕΩΡΟΛΟΓΙΚΗ
ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

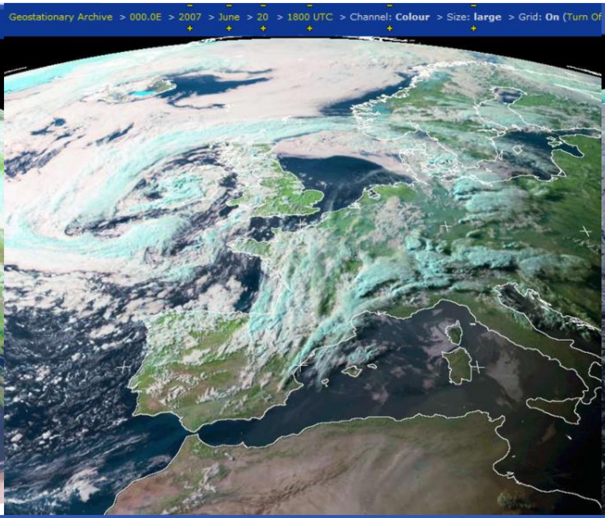
Approach

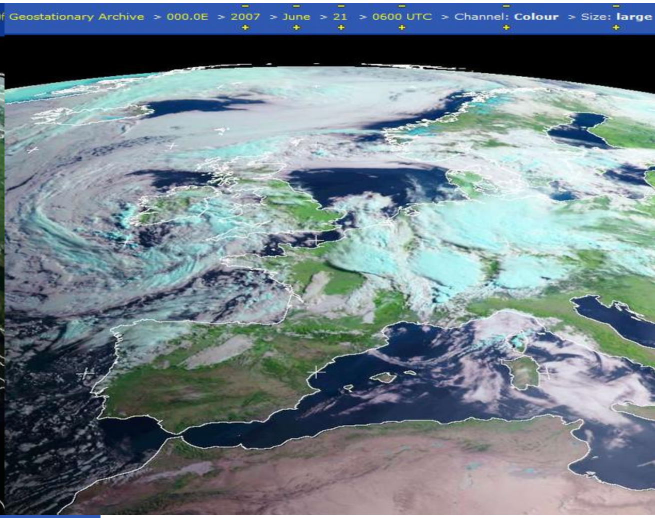
- Attempt to introduce alternative methods for verification of spatial precipitation forecasts
- Techniques that allow to diagnose the skill of a system as a function of the spatial scale of the forecast error and intensity of the precipitation events
- Aim is to compare the performance of quantitative precipitation forecasts for two different resolutions of COSMO model over the core MesoVICT case
- Precipitation events on different spatial scales are caused by different physical processes and the evaluation of forecast skill on various scales becomes important
- Necessity for suitable methods for operational use of any for precipitation forecast evaluation, more comprehensive and easy to apply methods are preferable

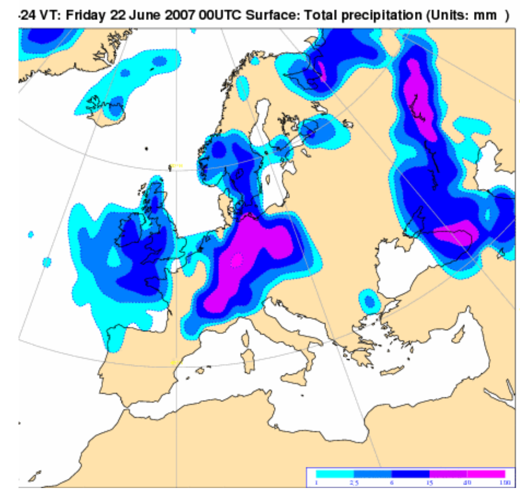
Case 1 (core case): 20-22 June 2007

Ahead of a trough, located over the British Isles, warm moist air is advected towards the Alpine Region. This leads to strong convective events in the evening of 20 June, in the area north of the main mountain range. On the next day (21^{st}) a cold front is reaching the Alps from the west and moves to the east rather quickly. Ahead of the front again convective events are observed. With the passage of the front strong westerly winds occurred.

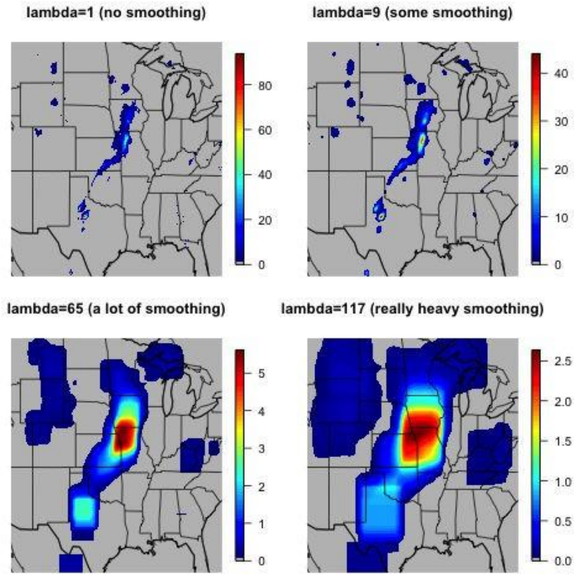




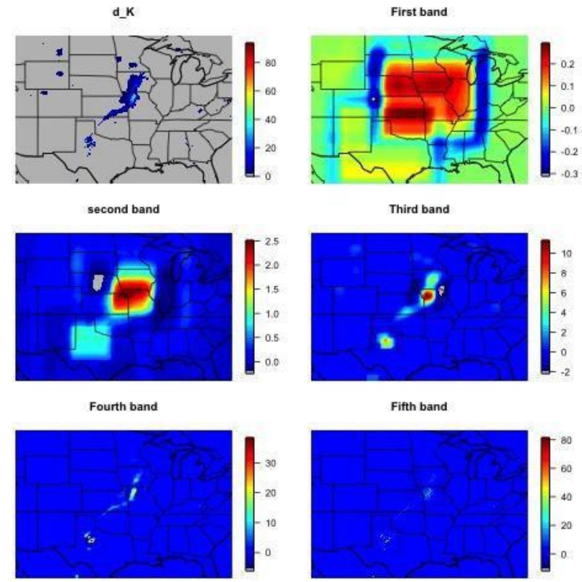




smoothing / neighborhood

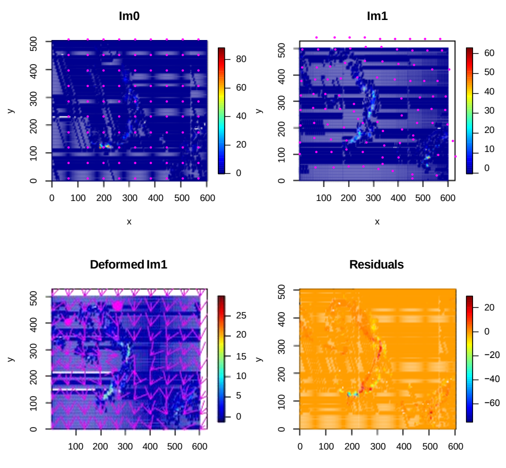


Filter

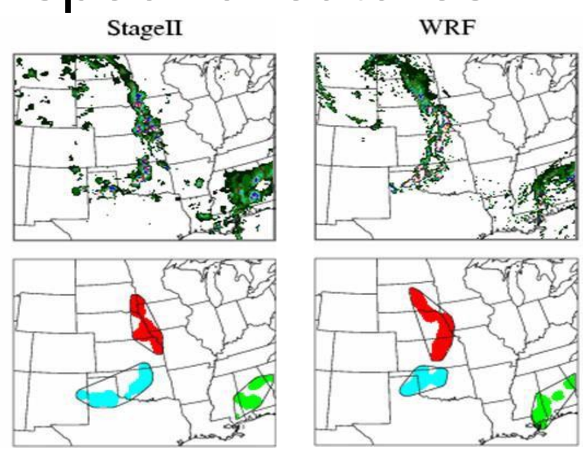


Scale Separation

Entire field



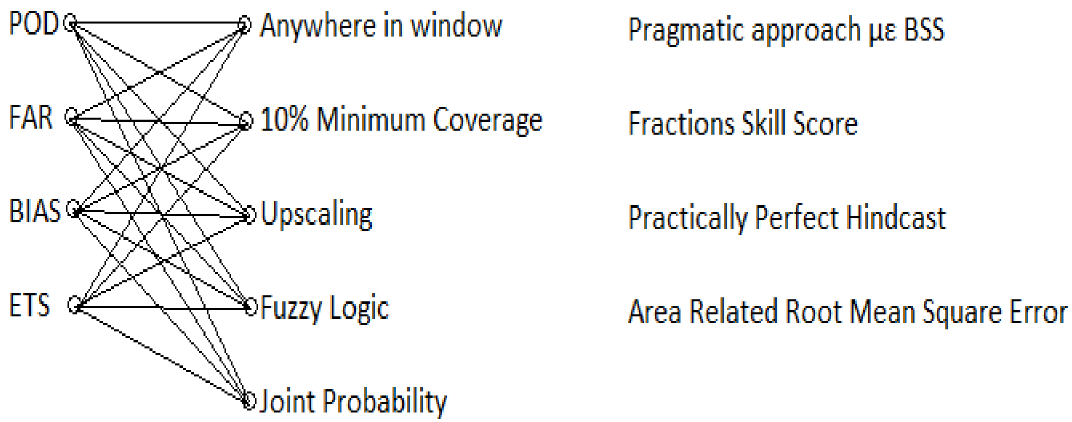
specific features



Spatial Displacement

Some examples on applications with MesoVICT datasets at HNMS

- **Neighborhood (in space) methods/scores :VAST**



Applications on COSMO1, COSMO2 and CMH

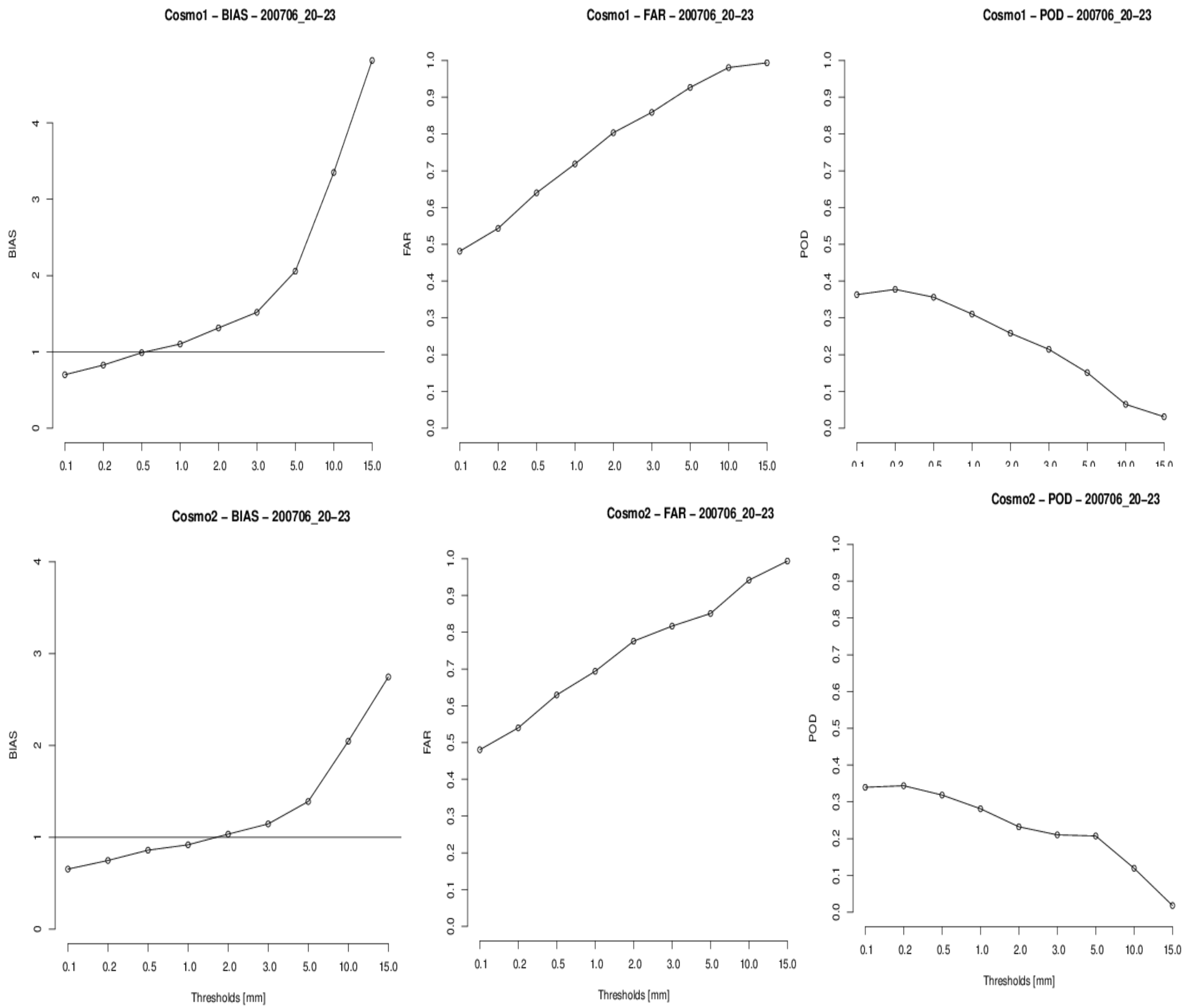
Models upscaled on 8km VERA grid

(can the skill of models with higher resolution than the obs be captured once upscaled?)

Large number of method/score combinations

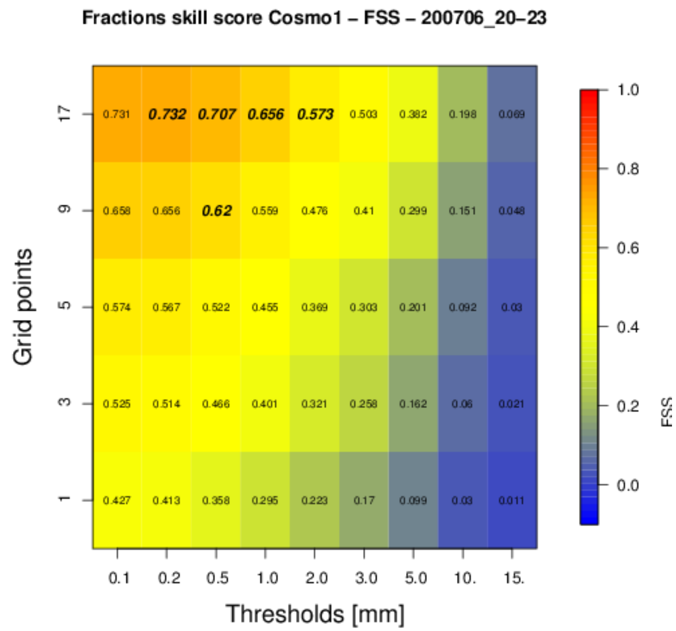
Importance of decision model

Categorical scores: COSMO-1 vs. COSMO-2

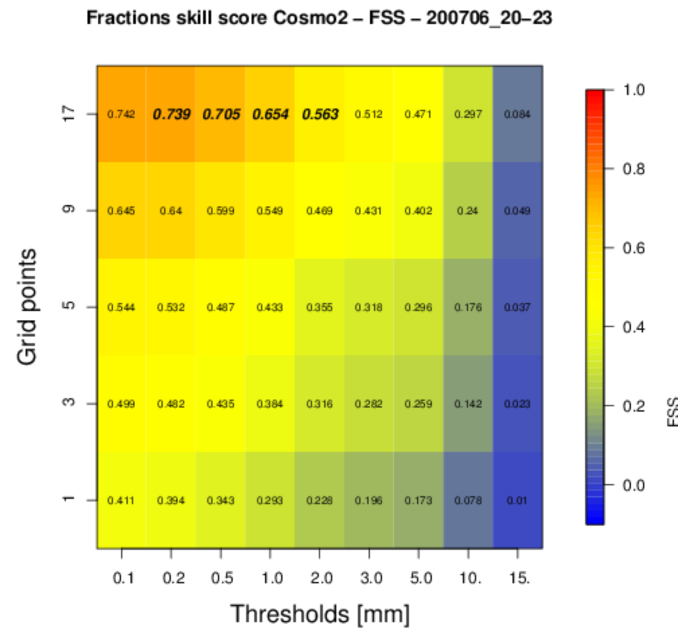


Fraction skill score: Core case

COSMO-1



COSMO-2

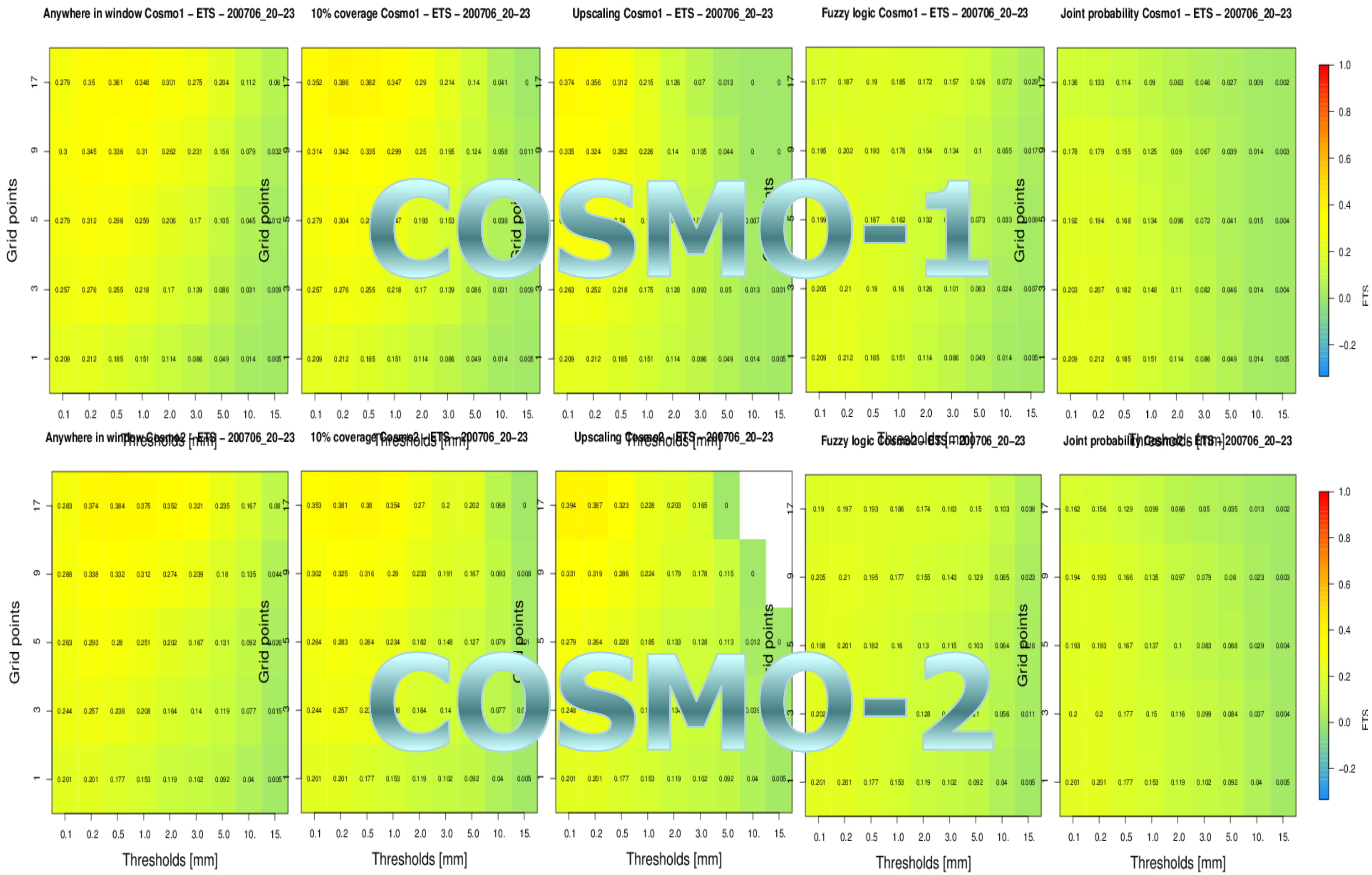


Compare forecast fractions with observed fractions in a *probabilistic* way over different sized neighbourhoods

$$FSS = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(P_{fcst} - P_{obs})^2}{\frac{1}{N}\sum_{i=1}^{N}P_{fcst}^{\ 2} + \frac{1}{N}\sum_{i=1}^{N}P_{obs}^{\ 2}}$$

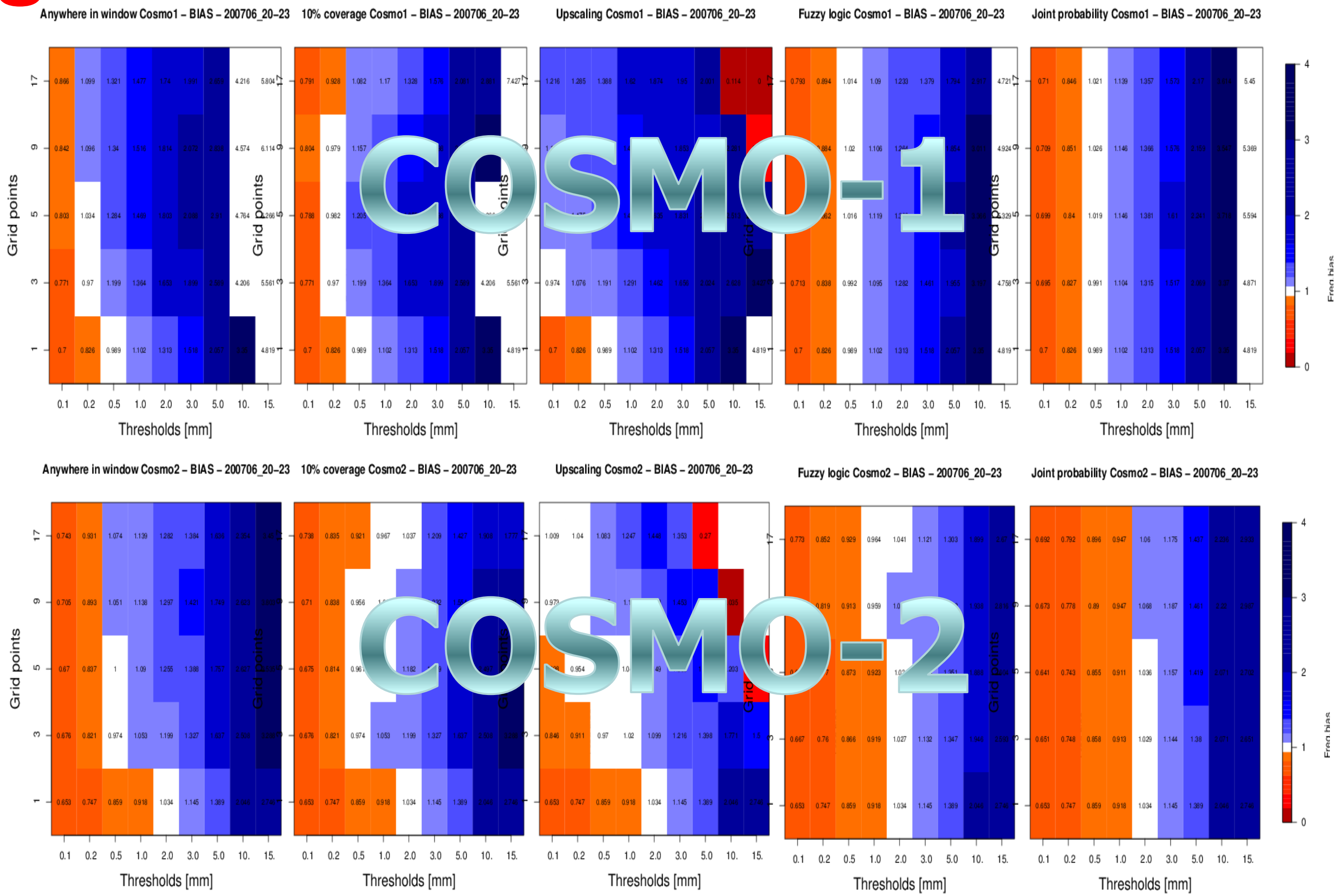
Neighborhood method scores COSMO-1 vs COSMO-2

- **ETS**



Neighborhood method scores COSMO-1 vs. COSMO-2

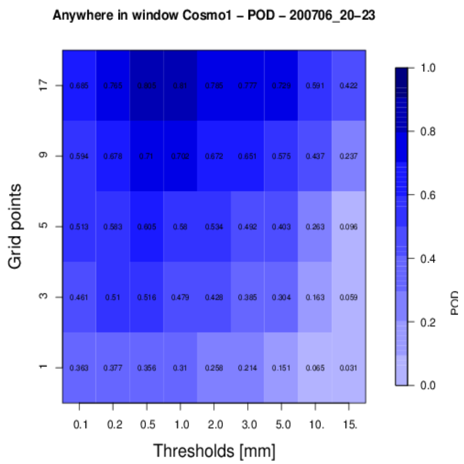
- **BIAS**

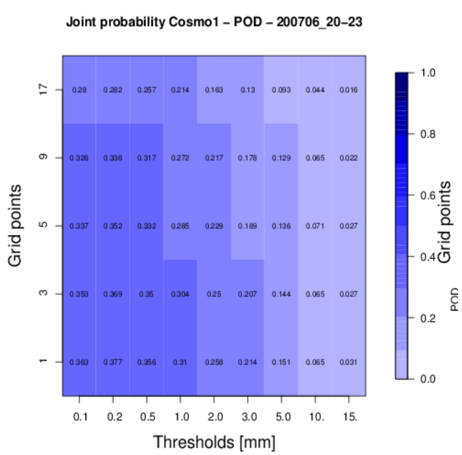


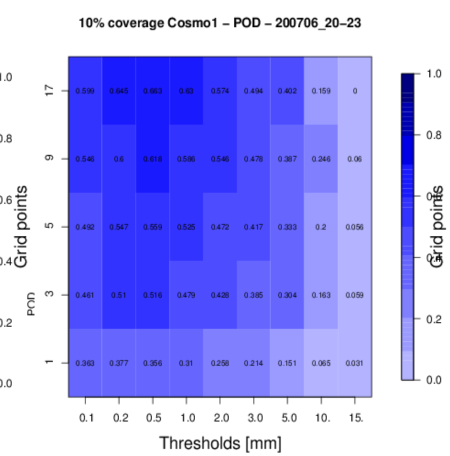
COSMO-1

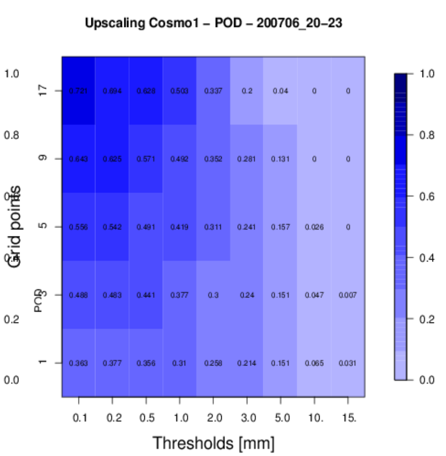
- **POD**

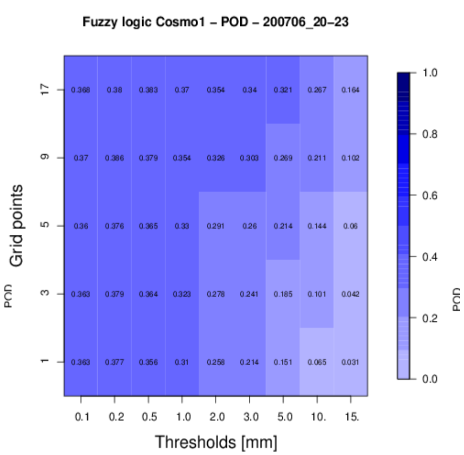
Neighborhood method	Decision Model
Upscaling	In a good forecast it is predicted the same average value of precipitation with the predicted
Fraction Skill Score	A forecast is useful when the frequency of forecasted events is similar with the frequency of the observed events
Minimum coverage	A forecast is useful if the event is forecasted in a minimum fraction of the area of interest
Anywhere in the window	A forecast is useful if it is more true than untrue, characterizing the events as probabilities
Pragmatic approach	A useful forecast has higher probability to discriminate the events and the non events.(in reference with the climatological value derived from the observations
Area related RMSE	A useful forecast has the same distribution of intensity as the observations





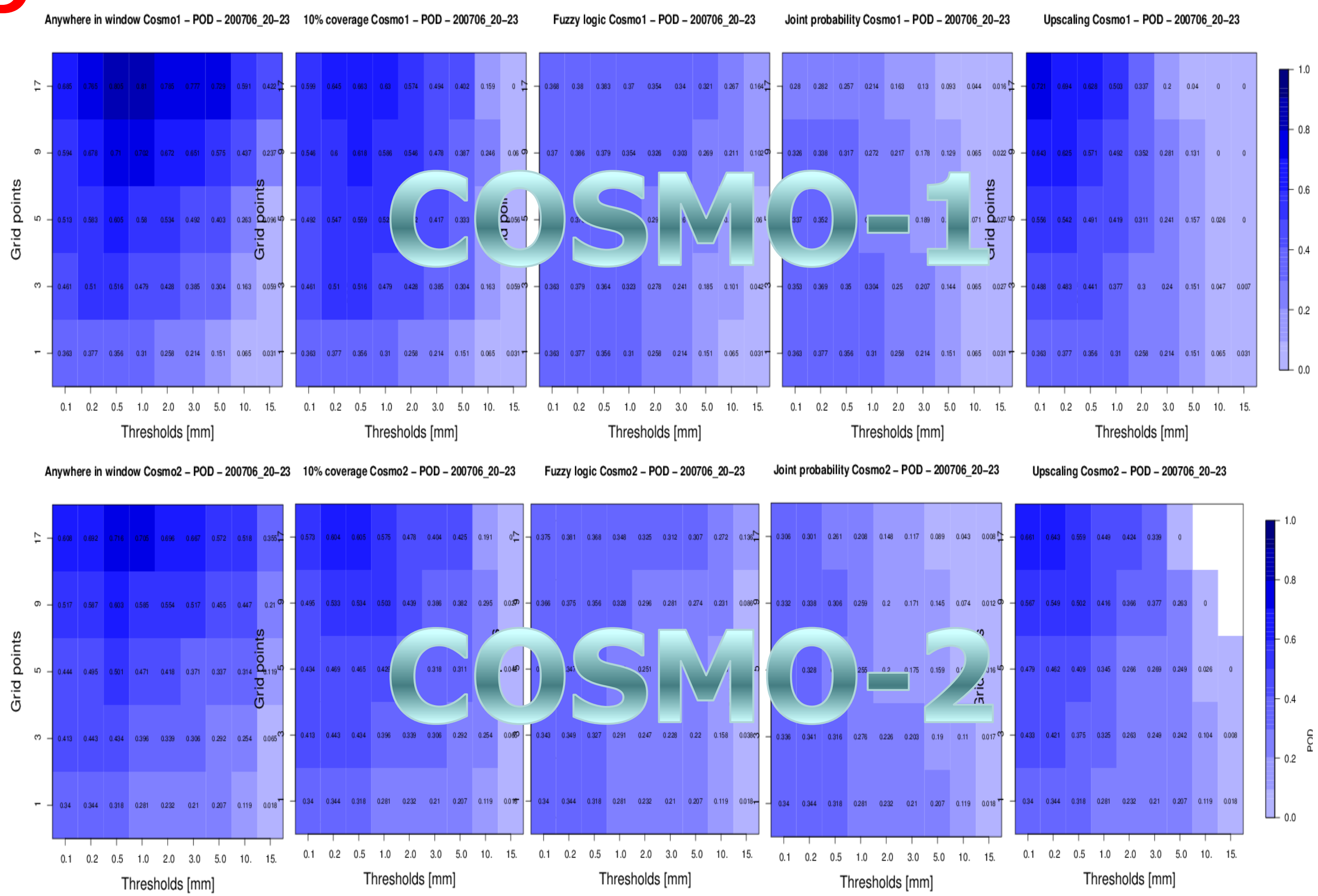






Neighborhood method scores
COSMO-1 vs COSMO-2

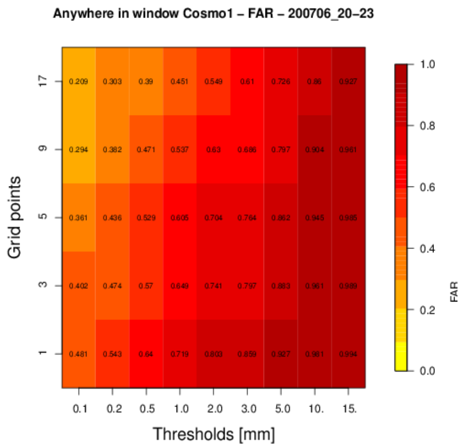
- **POD**

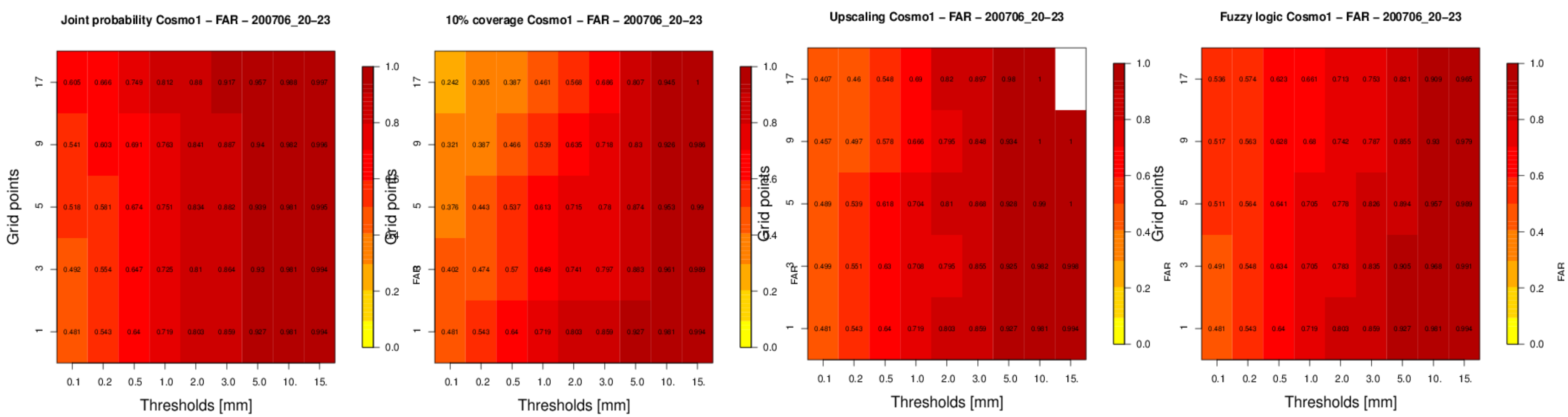


COSMO-1

- **FAR**

Neighborhood method	Decision Model
Upscaling	In a good forecast it is predicted the same average value of precipitation with the predicted
Fraction Skill Score	A forecast is useful when the frequency of forecasted events is similar with the frequency of the observed events
Minimum coverage	A forecast is useful if the event is forecasted in a minimum fraction of the area of interest
Anywhere in the window	A forecast is useful if it is more true than untrue, characterizing the events as probabilities
Pragmatic approach	A useful forecast has higher probability to discriminate the events and the non events.(in reference with the climatological value derived from the observations
Area related RMSE	A useful forecast has the same distribution of intensity as the observations





Comments on neighborhood methods use

- Neighbourhood methods offer a wide range of averaging techniques and scores, but for the operational verification requirements only a few of the methods and scores are to be selected.

- Once the goal is to evaluate and compare the relative performance of two model configurations, the decision model is the main factor determining which neighbourhood methods to be used for this comparison. . Again one has to decide on the criteria that make his forecast trustworthy and then apply and focus on few methods and scores (avoid vast amount of information with conflicting).

- With the right choice of decision model and aggregation on several timesteps/runs can provide a more **operationally** “useful” type of information.

- Suitable for other parameters than precipitation.

- Easiness on the application and good graphical representation

Intensity-scale verification
(filtering method)

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

B. Casati, G. Ross and D.B. Stephenson (2004) "A New intensity-scale verification approach for the verification of spatial precipitation forecasts", Meteorol Appl, vol 11, 141-154 pp

Evaluate the forecast skill as a function of the precipitation intensity and the spatial scale of the error

NOTE: **scale** = single band spatial filter → features of different scales → feedback on different physical processes and model parameterizations

In the neighborhood based (fuzzy) verification, the **scale** is the neighborhood size (low band pass filter): as the scale increases the exact positioning requirements are more and more relaxed

Steps in IS verification

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

- Binary error decomposition: Thresholding is used to convert the forecast and analysis into binary images for each of the thresholds. Binary error is the difference of this $Z=I_{y'}-I_x$
- Binary error image is then expressed as the sum of components on different spatial scales by performing a 2-dimentional discrete Haaer wavelet decomposition $Z=\sum_{l=1}^{L} Z_l$
- Most substantial binary error image of the mother wavelet components are calculated for various spatial scales (ℓ=1,..L=7 that corresponds to XXkm). The spatial scales refer to the spatial scale of the error and not that of the precipitation features or their displacement as it happens in the neighborhood methods
- The MSE of the binary error image is calculated from while for each threshold the skill score can be calculated fror $\mathrm{MSE}=\sum_{i=1}^{L}\mathrm{MSE}_l$ $\mathrm{MSE}_l=\overline{Z_l^2}$ $\mathrm{SS}=\frac{\mathrm{MSE}-\mathrm{MSE}_{\mathrm{random}}}{\mathrm{MSE}_{\mathrm{best}}-\mathrm{MSE}_{\mathrm{random}}}$ where and $\mathrm{MSE}_{\mathrm{random}}$ is associated with a random forecast calculated from the bias and the base rate at each threshold
- Intensity scale verification technique is a spatial generalization of traditional binary verification (HSS, PSS)
- Application using SpatialVx: waveIS routine

Two-dimensional discrete Haar wavelet filter (Casati et al, 2004)

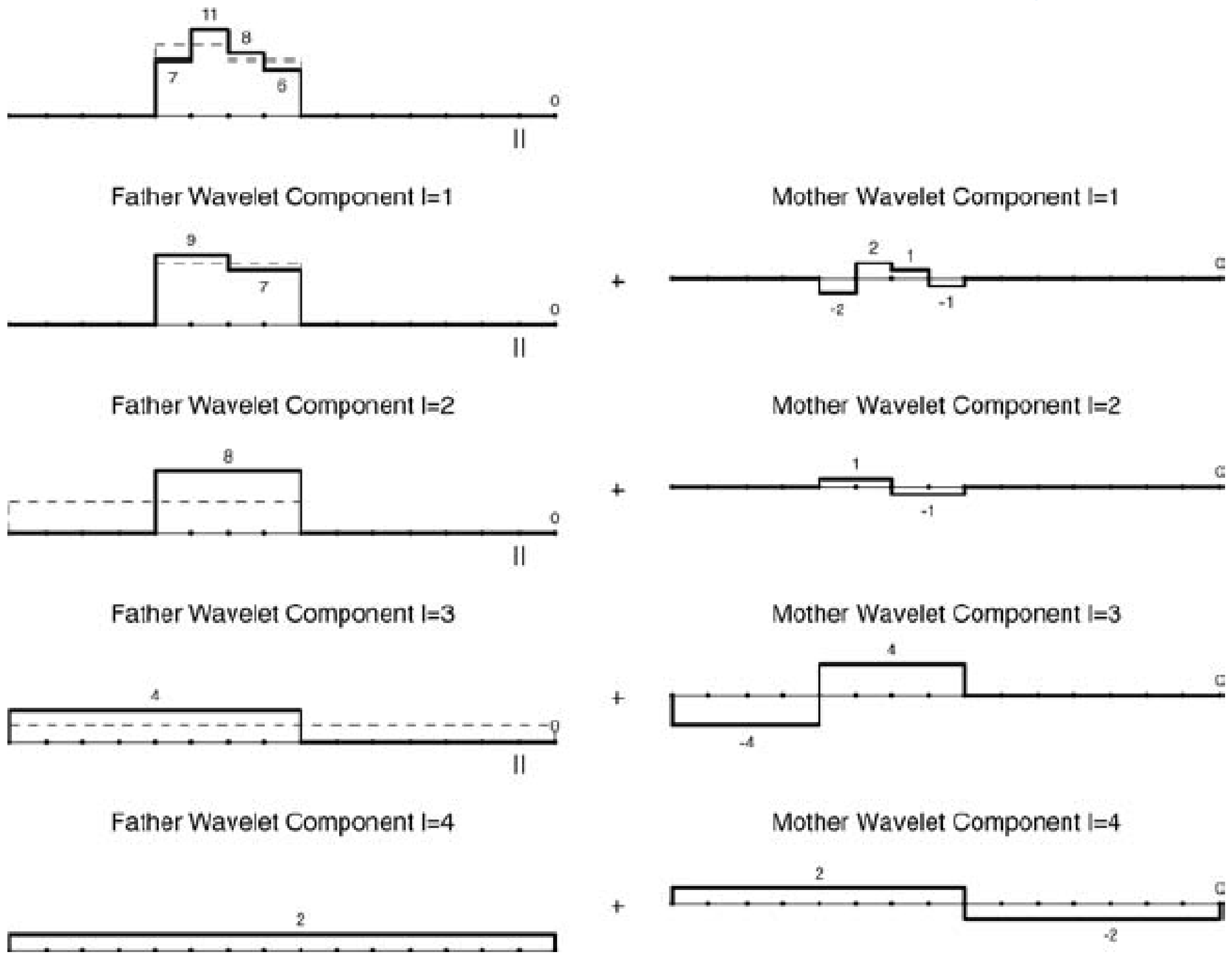
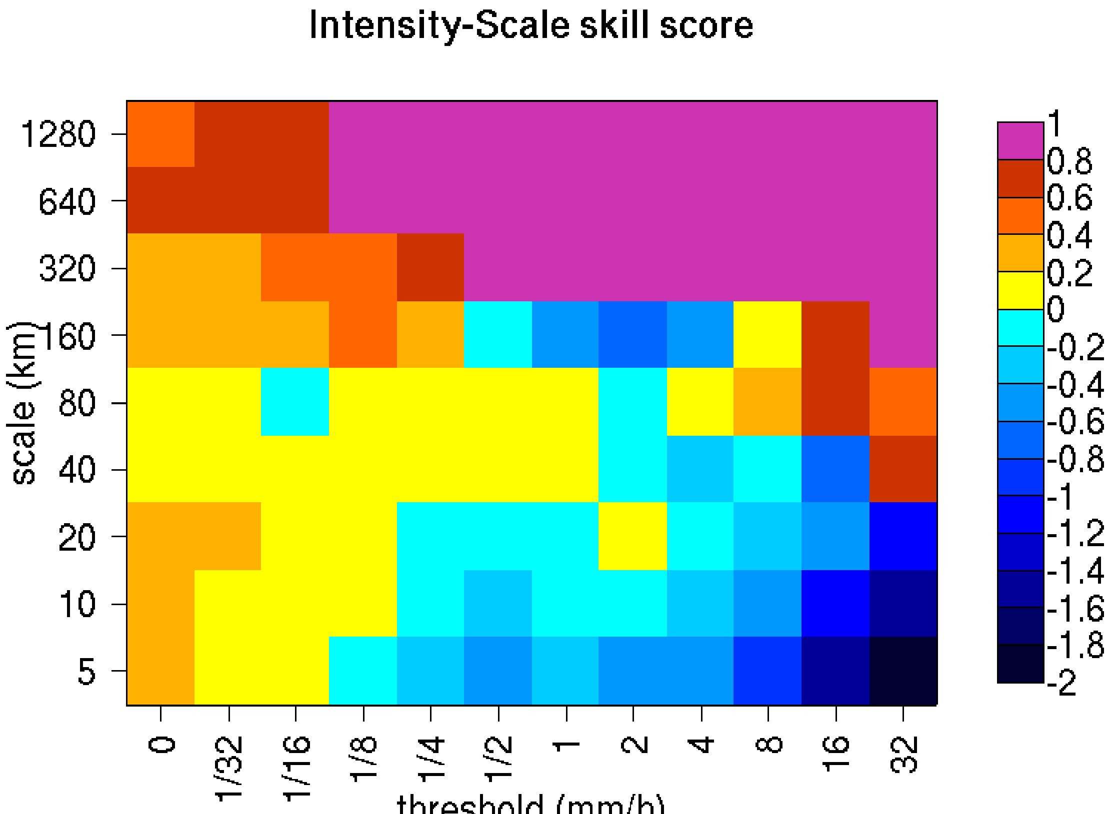


Figure 11. *Example of the one-dimensional discrete Haar wavelet filter applied to an example function (top left panel). At the first step the function is decomposed into the sum of a coarser* mean *function (the first father wavelet component) and a* variation-about-the-mean *function (the first mother wavelet component). At each step the Haar wavelet filter decomposes the father wavelet component obtained from the previous step into the sum of a coarser* mean *function (the l^{th} father wavelet component) and a* variation-about-the-mean *function (the l^{th} mother wavelet component). The l^{th} father wavelet component is obtained from the initial function by a spatial averaging over 2^l pixels. The process stops when the largest father wavelet component (mean over the whole domain) is found.*

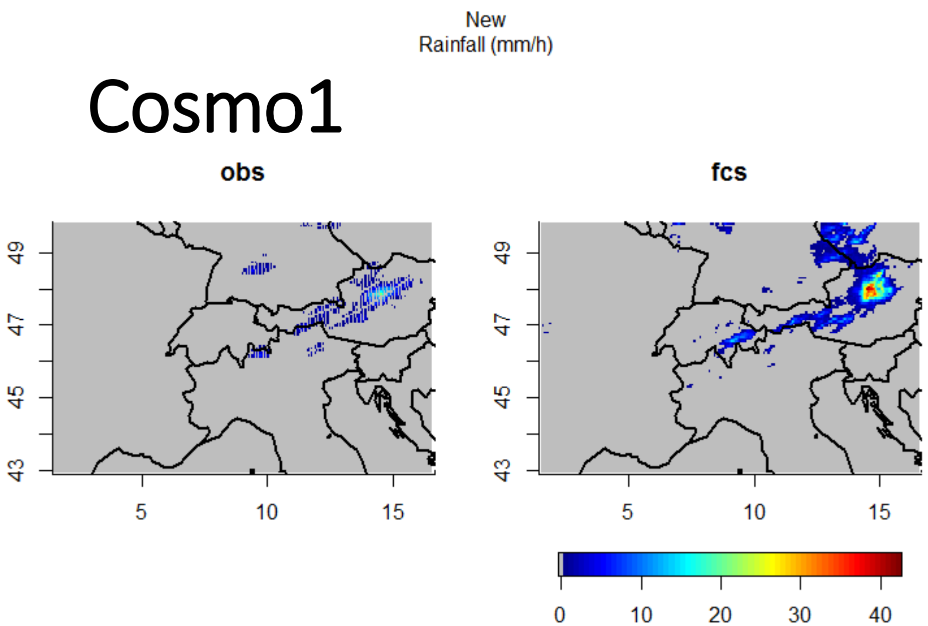
Intensity-scale skill score (SS)

- **For each threshold and scale component: skill score can be calculated associated to the MSE of binary images (= HSS).**
- The IS skill score is capable of isolating specific scale-dependent errors.
- Often, small scales exhibit negative skill or close to zero skill, whereas large scales exhibit positive skill

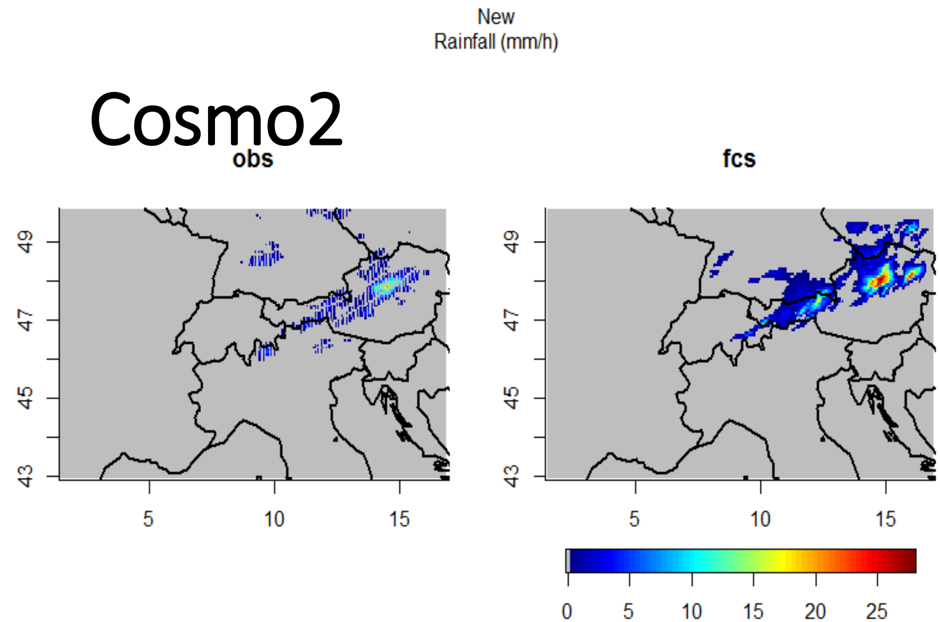


20070621-15:map

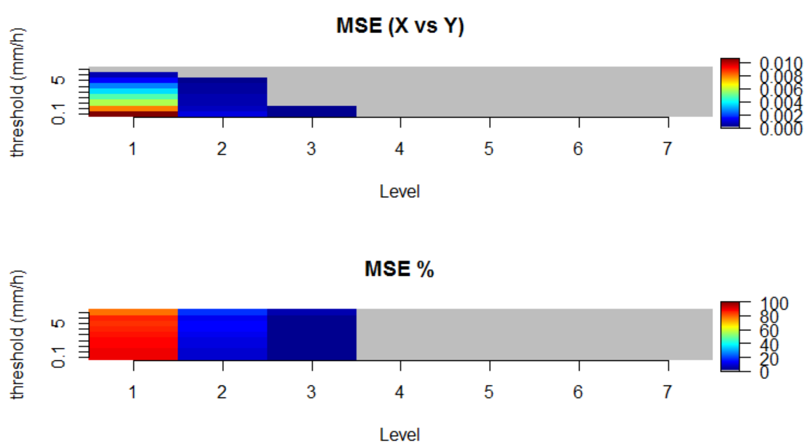
Cosmo1

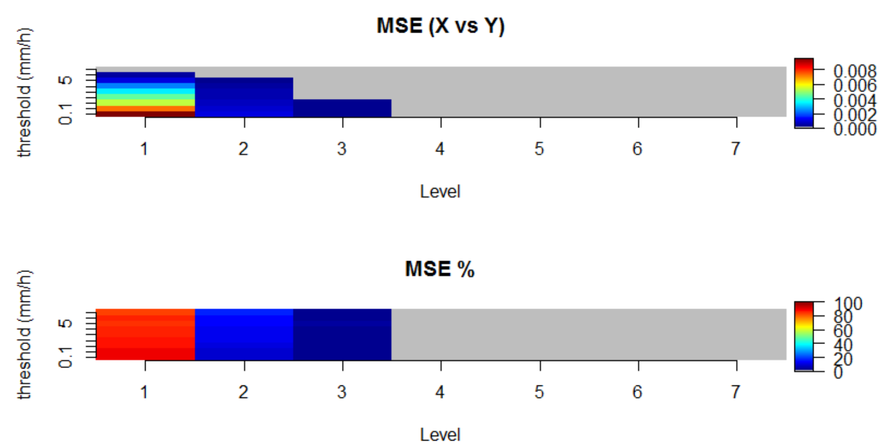


Cosmo2



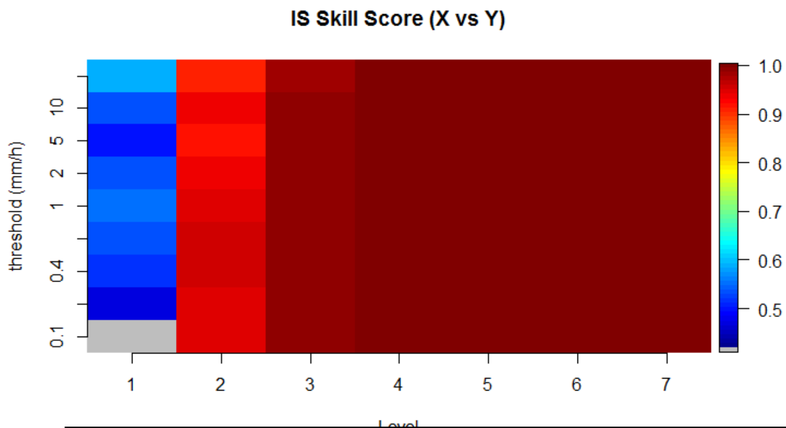
MSE almost disappears for error tiling of the order of 3x3 grid points, COSMO-1 slightly improved behavior



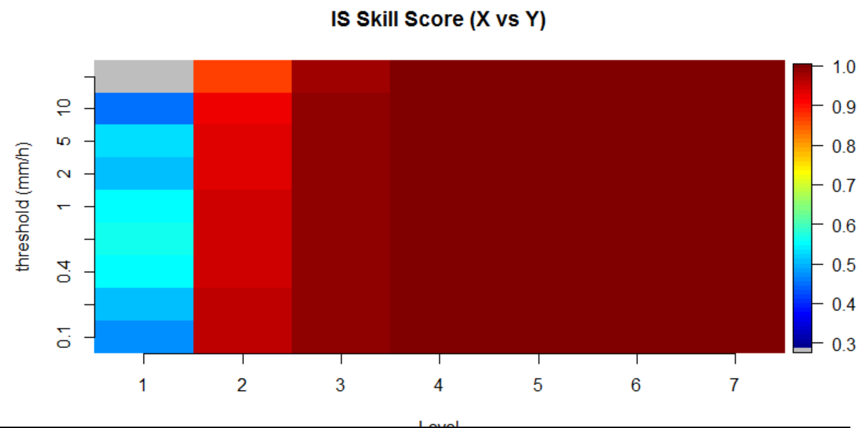


20070621-15: Skill

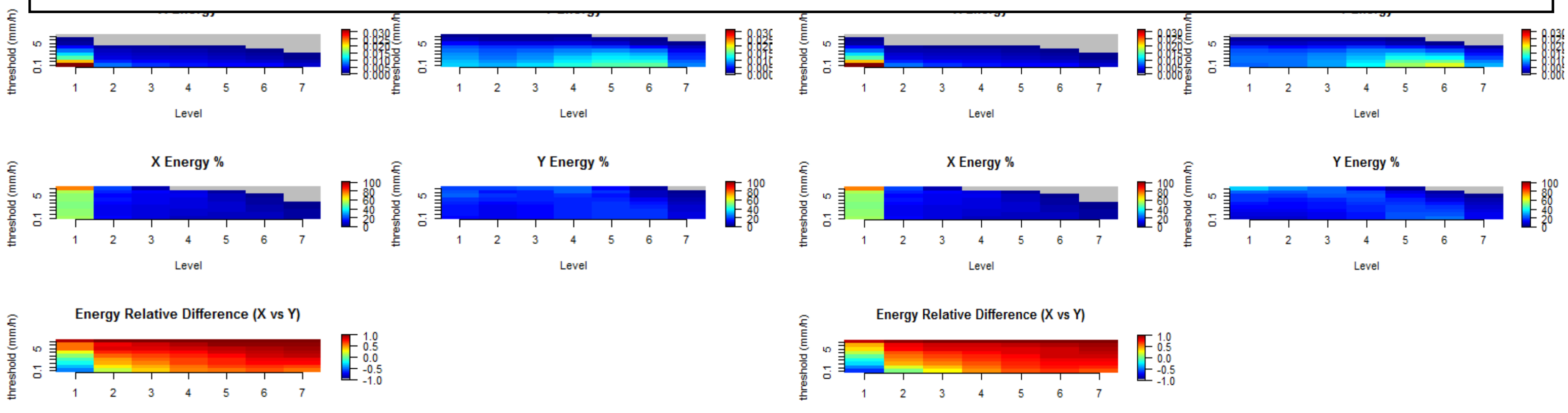
Cosmo1



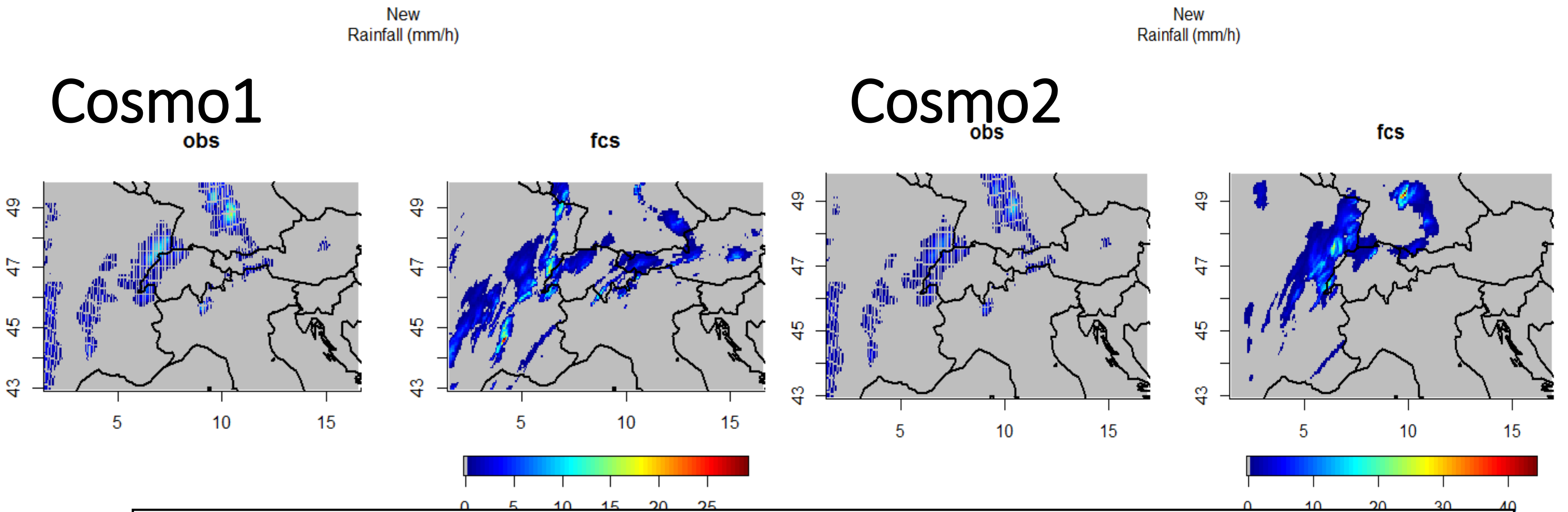
Cosmo2



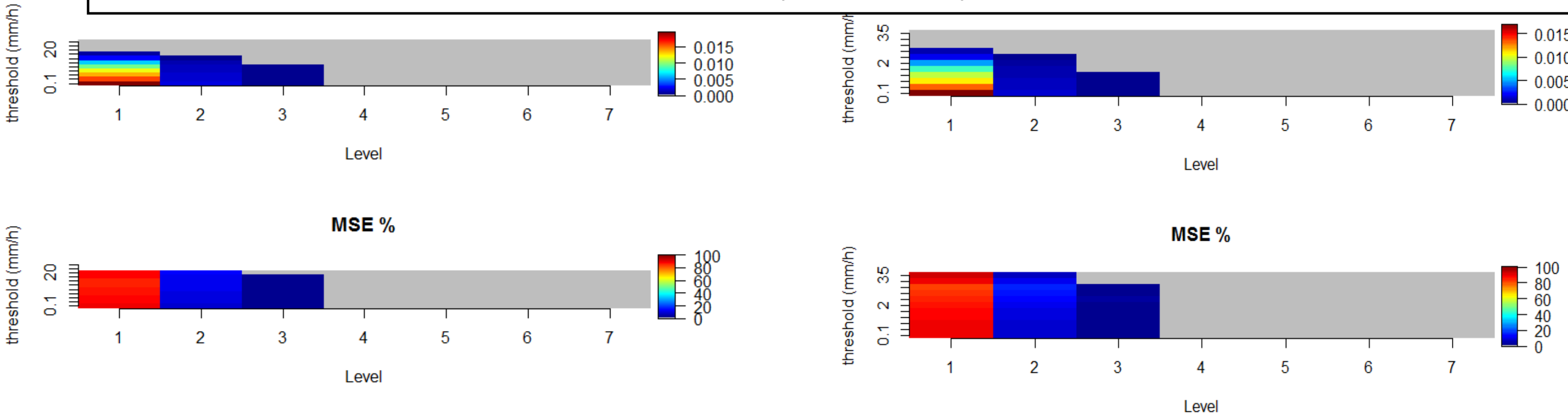
Small scales 1x1 02 2x2 scales have skill close to zero , while large scales exhibit large skill. COSMO1 ISS graphs exhibit that errors due to displacements of small spatial scale features are more important compared to those of COSMO-2



20070621-01:map

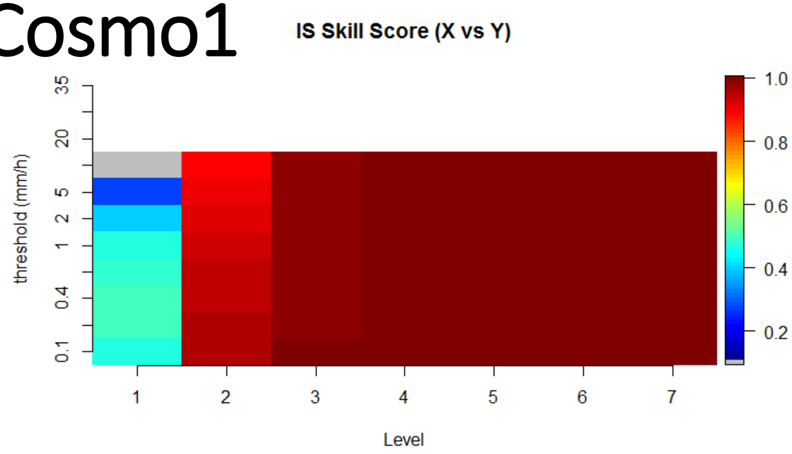


MSE almost disappears for error tiling of the order of 3x3 grid points, higher impact of threshold amount than impact of displacement on both models

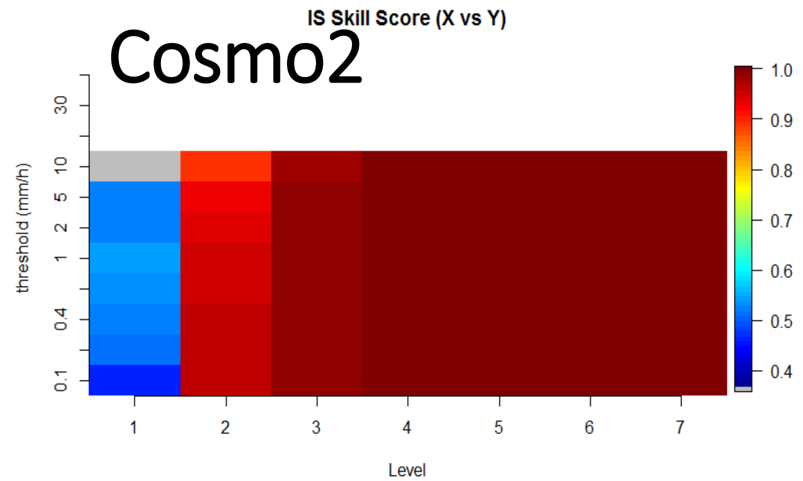


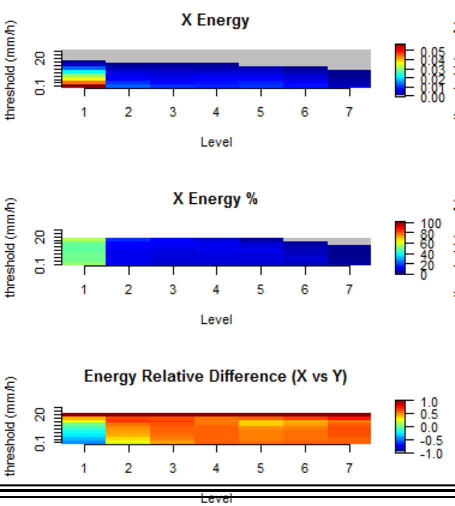
20070621-01: Skill

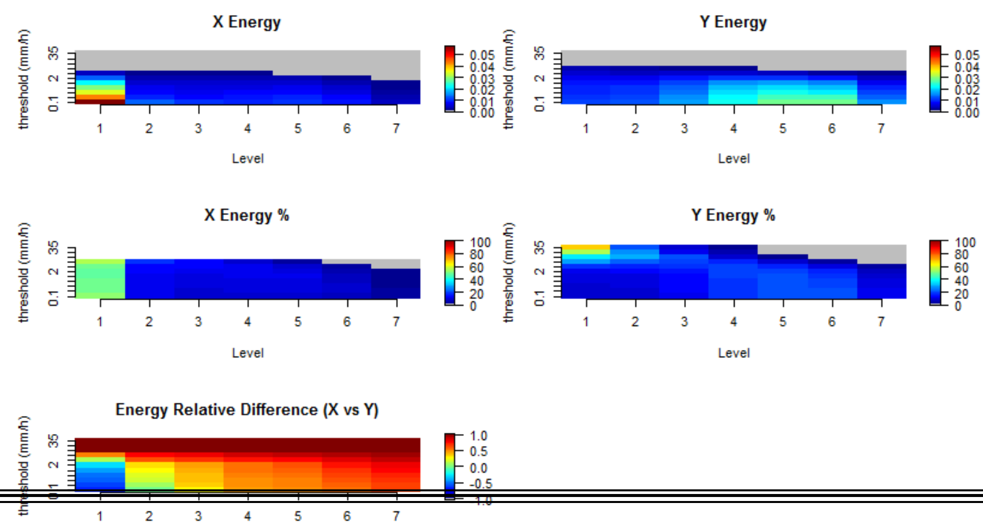
Cosmo1



Cosmo2



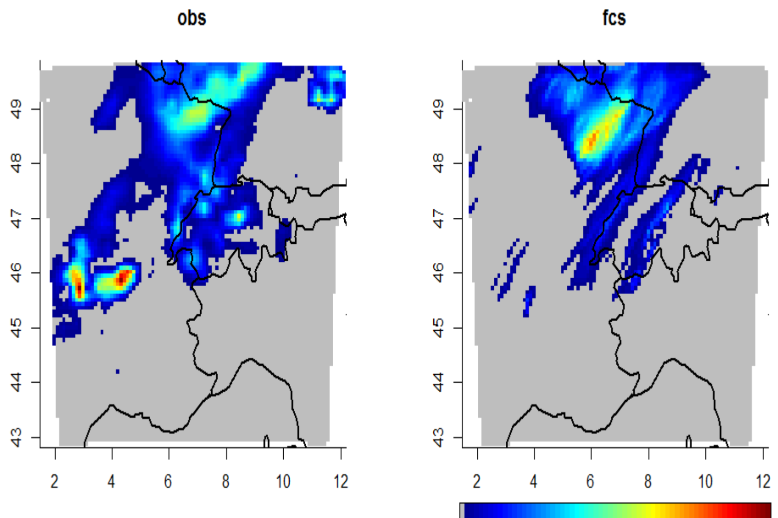




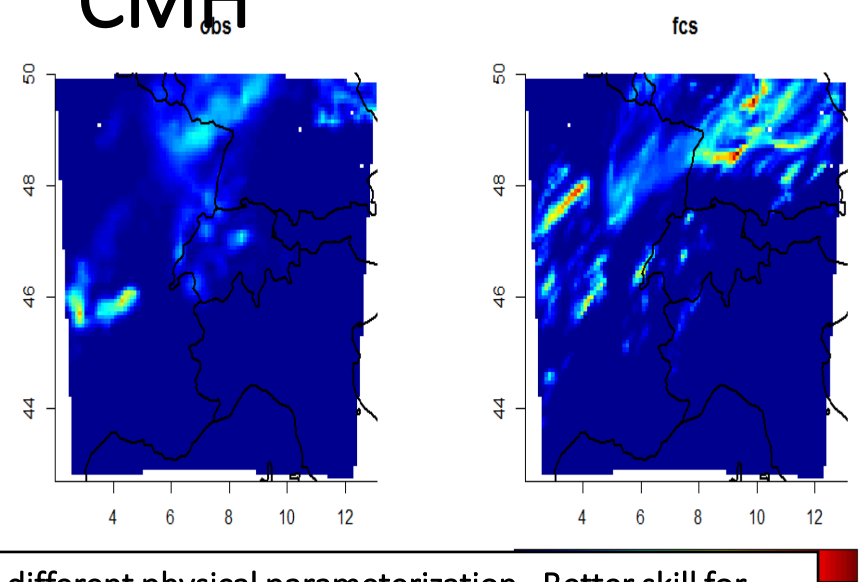
ΕΘΝΙΚΗ
ΜΕΤΕΩΡΟΛΟΓΙΚΗ
ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

20070621-18: 12hpreci

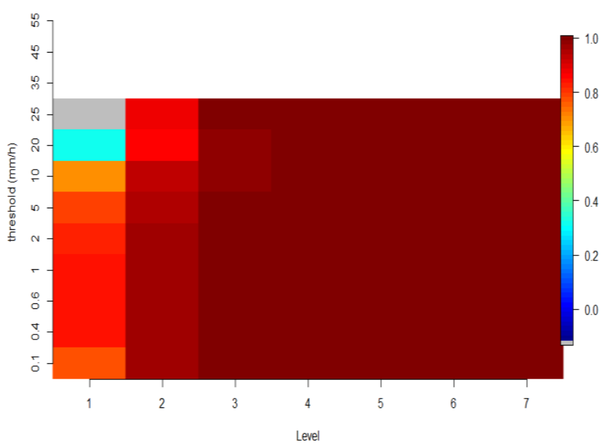
Cosmo2

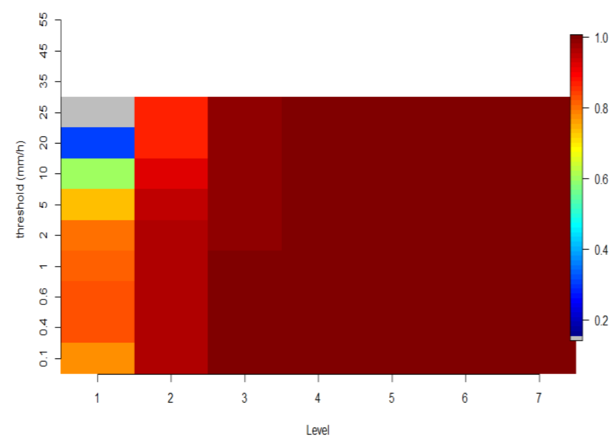


CMH



Comparison of models with similar resolution but different physical parameterization. Better skill for COSMO model, obvious at small scales for all thresholds . Very good skill for both models for scales higher than 2x2 points.





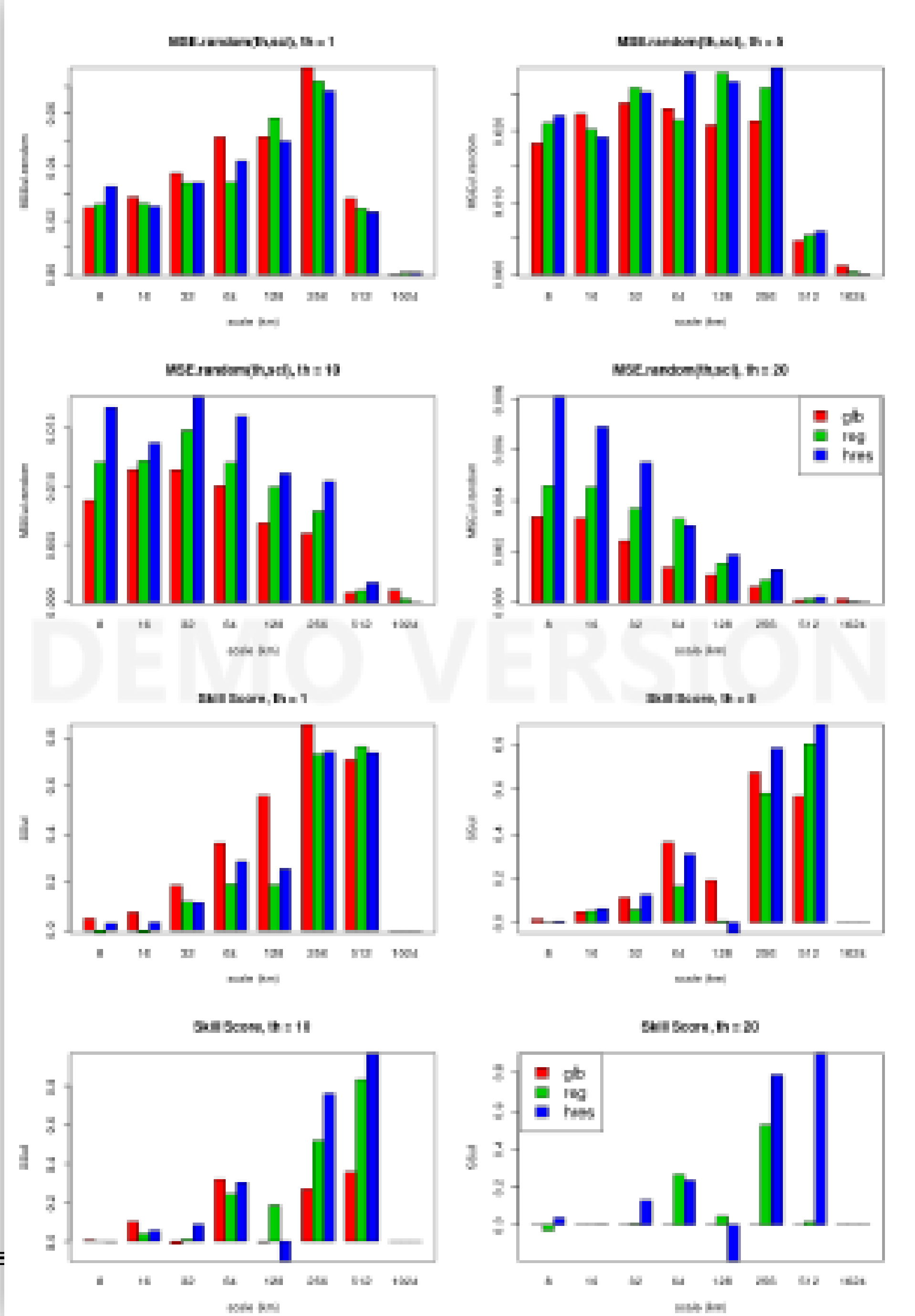
Comments on IS (filtering method) use

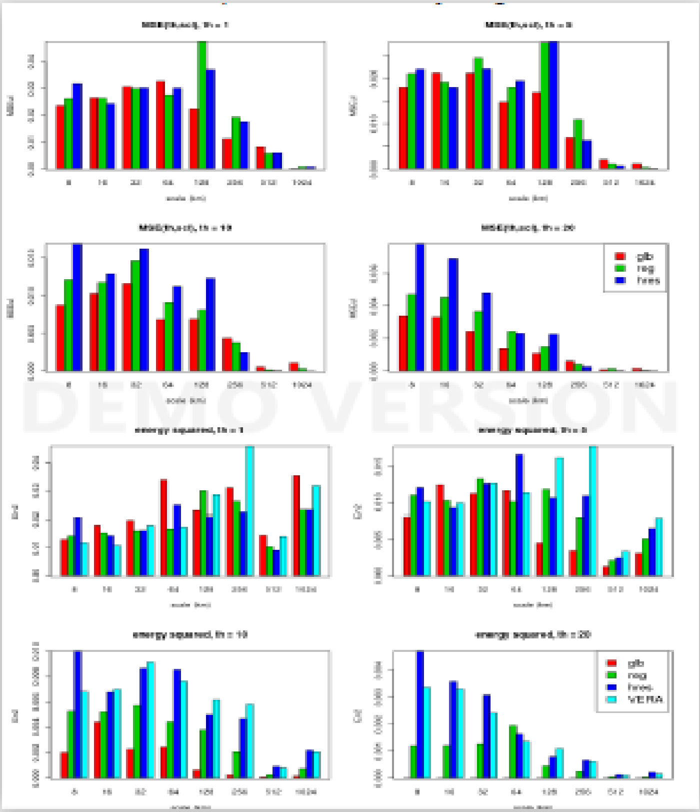
ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

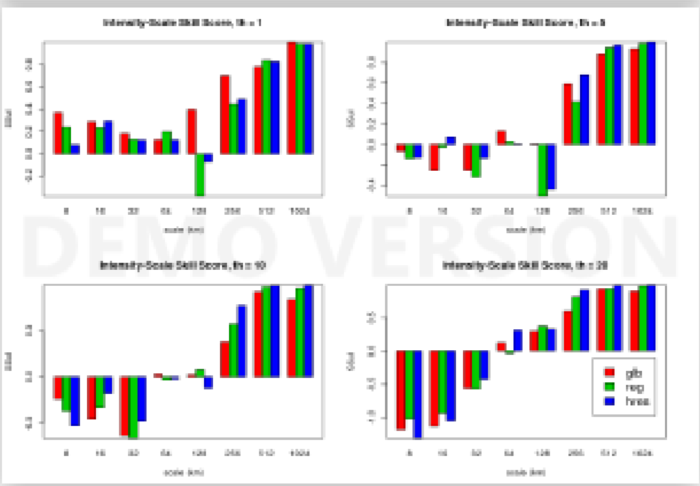
- Application of IS spatial method was attempted over a few instances related to the MesoVICT core case
- Both models have small skills at the smallest scale but skill improves when considering larger spatial scales.
- ISS decreases as the precipitation threshold is increased and this is due to the poor ability of the model to go beyond just the yes/no rain discrimination
- The technique allows the skill to be diagnosed as a function of the scale of the forecast error and intensity of the precipitation events. Results show that reduction of skill is mainly due to the small-scale misplacement of more intense (rare) precipitation events.
- Wavelet-based scale-separation MSE skill score and scale-separation statistics are suitable for comparing models with different resolutions.
- IS method constraints are related to the request to have precipitation analysis available for each grid point of the forecast field, and to the fact that Haar wavelet decomposition is designed for a square domain

- Finally, this kind of analysis and graphical representation of scores is not ultimately useful for the operational verification as it does not provide insight on the average behavior of the model over areas but on the forecast of a single day. Further investigation is required on the adaptation of score as am aggregation of many events and the information such “averaging” could provide

Ευχαριστώ πολύ


glb
reg
hres


glb
reg
hres
VERA


glb
reg
hres

Intensity-scale verification technique summary

- ➢ evaluate the **skill as function of precipitation intensity and spatial scale of the error**

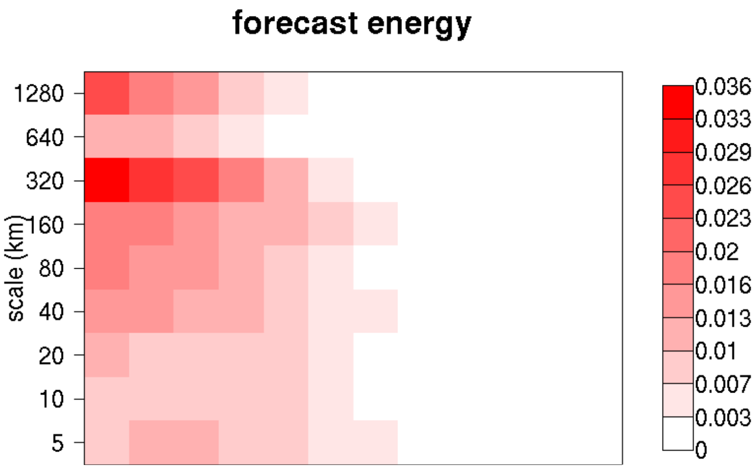
it is capable of isolating specific IS errors (e.g. displaced storm → negative minimum in skill for 160 km scale)

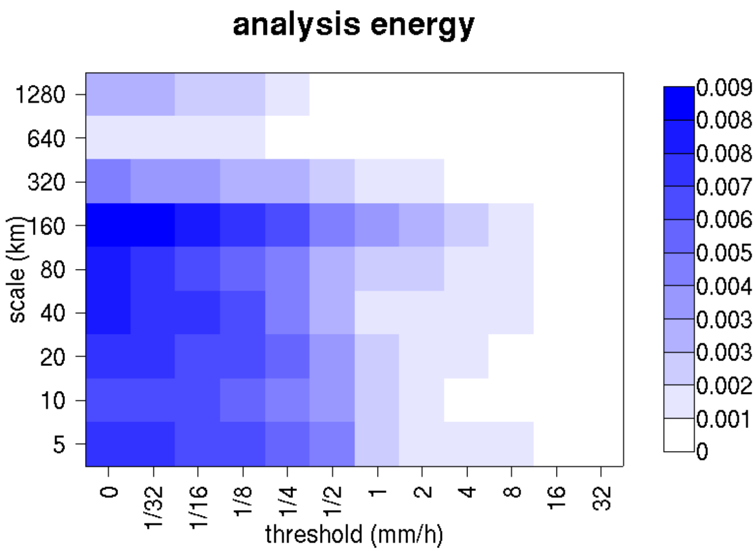
in general small scales have negative skill (small scale displacements) and large scales have positive skill

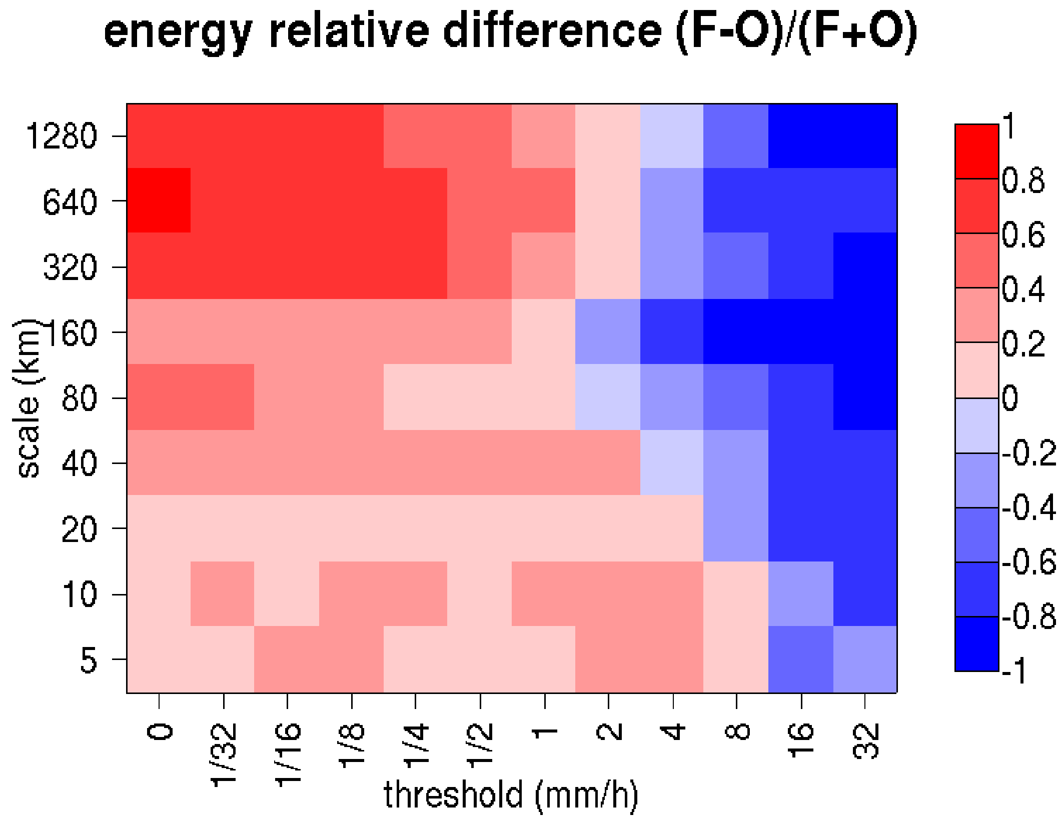
- ➢ bridges **categorical approaches** (joint distribution) and **scale verification** approaches (physical properties)

- ➢ is suitable for **spatial precipitation forecasts**: wavelets (discontinuities and features) + categorical approach (robust and resistant)

Energy and Bias







For each threshold and scale:

the energy informs on the amount of events (energy = mean(X^2))

the energy relative difference measures the bias = (B-1)/

Conclusions

The intensity-scale verification approach evaluates bias and skill for different precipitation intensities and spatial scales.

Precipitation forecasts in general exhibit negative skill on small scales, and positive skill on large scales (predictability).
However the IS skill score is capable of identifying specific scale-dependent errors (e.g. Nimrod displaced storm ◊ negative skill at 160km scale).

The scale of the error is associated with both the features' size and their displacement; the IS statistics are sensitive to displacement and intensity errors.

Tiling smooths the effects of the discrete wavelet support: appropriate for single cases verification. For aggregated precipitation forecasts tiling, cropping, padding or interpolating provide similar results.

THANK YOU!