QPF verification over catchment areas: the Italian experience with COSMO models and first test with ICON IT/2I in comparison with IFS-ECMWF

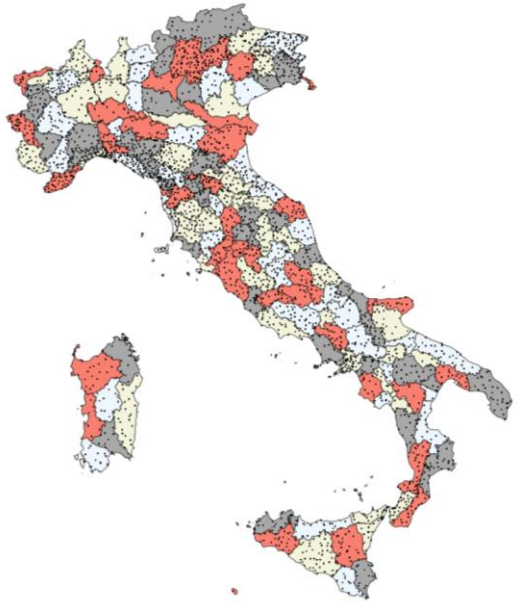
MStefania Tesini (Arpae), Elena Oberto (ArpaP)
Enrico Minguzzi(Arpae), Luca Monaco (PoliTo)

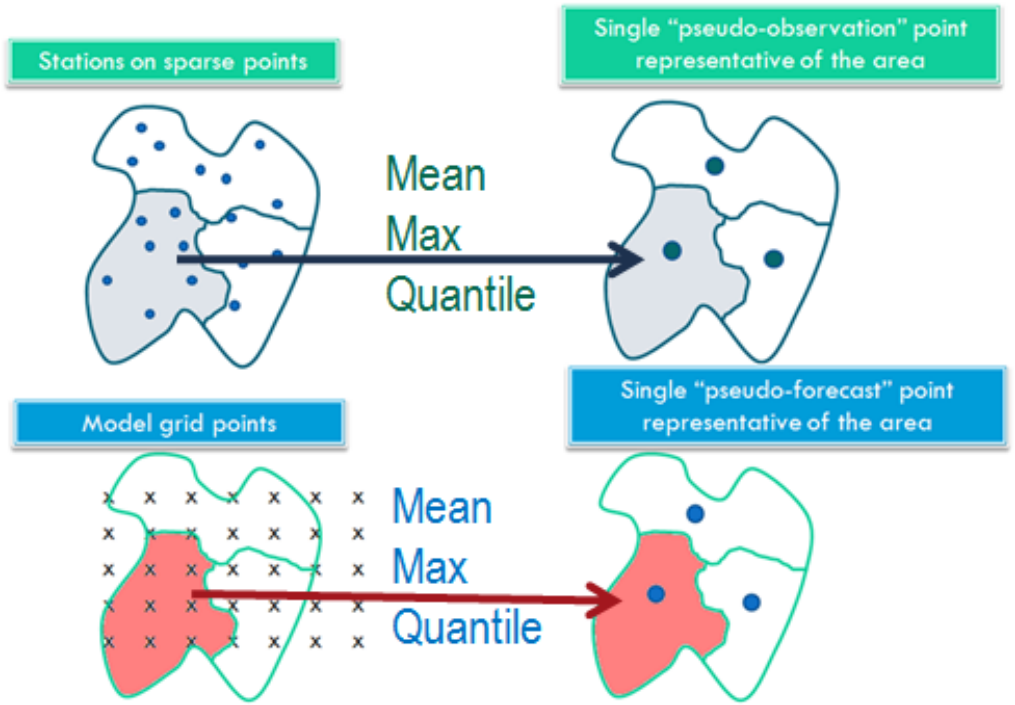
Motivation

The estimation of QPF on catchment areas for purposes related to the issue of Civil Protection alerts for hydro-geological or hydraulic criticality is one of the main activities carried out operationally from the forecasters of Arpae-Emilia Romagna and Arpa-Piemonte.
Many tools were developed to help forecasters and hydrologists to evaluate the mean or the maximum of the precipitation field on the warning areas used by the National Civil Protection Department ,using data from different NWP models (e.g. IFS-ECMWF, COSMO-5M or COSMO-2I) since the exceeding predefined thresholds can give useful indications for situations of intense precipitation possibly leading to floods.

- For these reasons, the verification is, from a long time, carried out using the spatial verification method DIST (Neighborhood obs – Neighborhood fcs)
- The aim of the verification is to provide to the users some information about the performance of the forecast system that can help to decide in which situations one system is better than another

QPF verification over alert areas with DIST methods





$$\overline{OBS} = \frac{1}{K}\sum_{i=1}^{K}(obs)_i$$

- $\overline{OBS}$: Mean value 24h/12h/6h/ cumulated precipitation observed
- K: Number of station points inside alert area
- $(obs)_i$: i-th station precipitation value

$$OBS_{MAX} = MAX(obs)_i$$

$$\overline{FOR} = \frac{1}{N}\sum_{i=1}^{N}(for)_i$$

- $\overline{FOR}$: Mean value 24h/12h/6h/ cumulated precipitation forecasted
- N: Number of grid points inside alert area
- $(for)_i$: i-th grid point precipitation value

$$FOR_{MAX} = MAX(for)_i$$

For each area, several parameters of the distribution of both the observed and forecast values falling in it can be computed (mean, maximum, quantile…).

Verification is then performed using a categorical approach, by comparing for each box the mean or the maximum of forecasts against the corresponding parameter of the observations, using a set of indices.

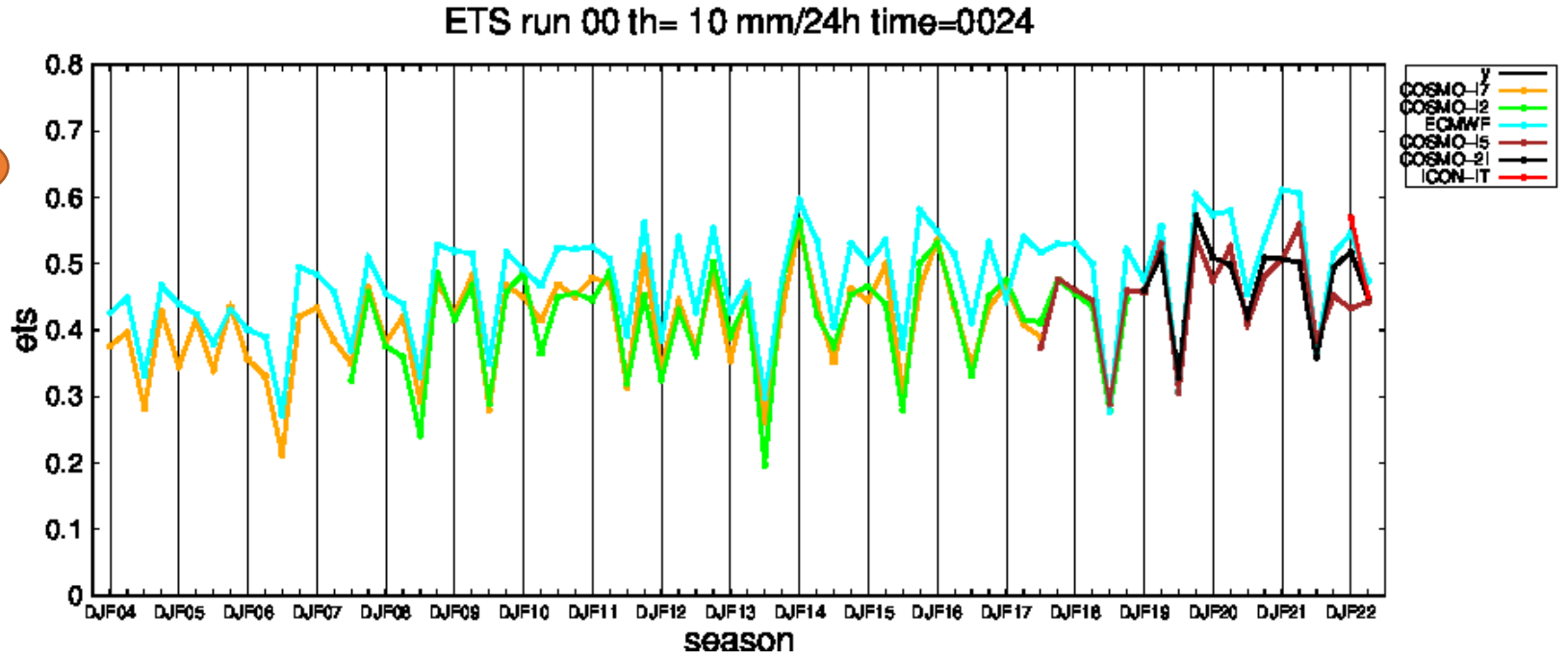
Long Trend: the models' history

To put together the several high-resolution COSMO model runs since 2004 in comparison with ECMWF:

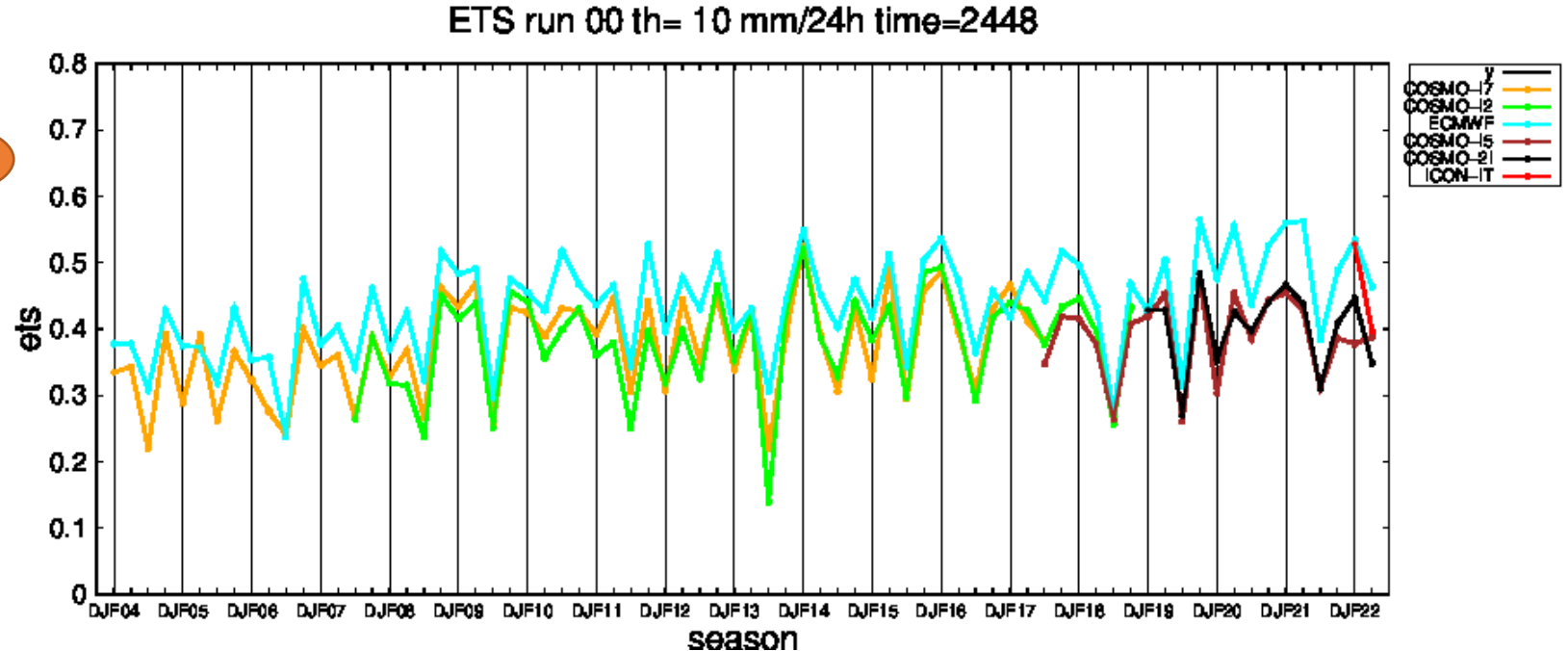
- If we plot an imaginary line starting from DJF 2004 since last season, you can appreciate a general slight models improvement
- If we choose a medium threshold (10 mm/24h) average areal, we cannot lie: the winner is ECMWF, BUT…

Which is the real added value given by our limited area models?

D0



D1



You are a decision maker

C/L ratio model (Richardson)

	Event occurs	Event does not occur
Action taken	***C***	***C***
Action not taken	***L***	***0***

Expense matrix

No forecast info

$E_{always} = C$

$E_{never} = sL$, *s= climatological base rate*

$E_{always} < E_{never}$ → *action*

$E_{always} > E_{never}$ → *no action*

Optimal strategy=mean expense=minimise losses

$E_{climate} = min(C,sL)$

Perfect forecast

$E_{perfect} = sC$

$V \text{ of forecast system} = (E_{climate}-E_{forecast})/(E_{climate}-E_{perfect})$

A maximum value is when the system perfectly forecasts the future. If V >0 the decision maker will gain economic benefit by using forecast info in addition to climatology.

$V_{relative} = [min(C/L,s)-F(1-s)C/L+Hs(1—C/L)-s]/[min(C/L,s)-sC/L]$, *s=a+c (base rate)*

V relative depends on quality of system, observed base rate and user's C/L

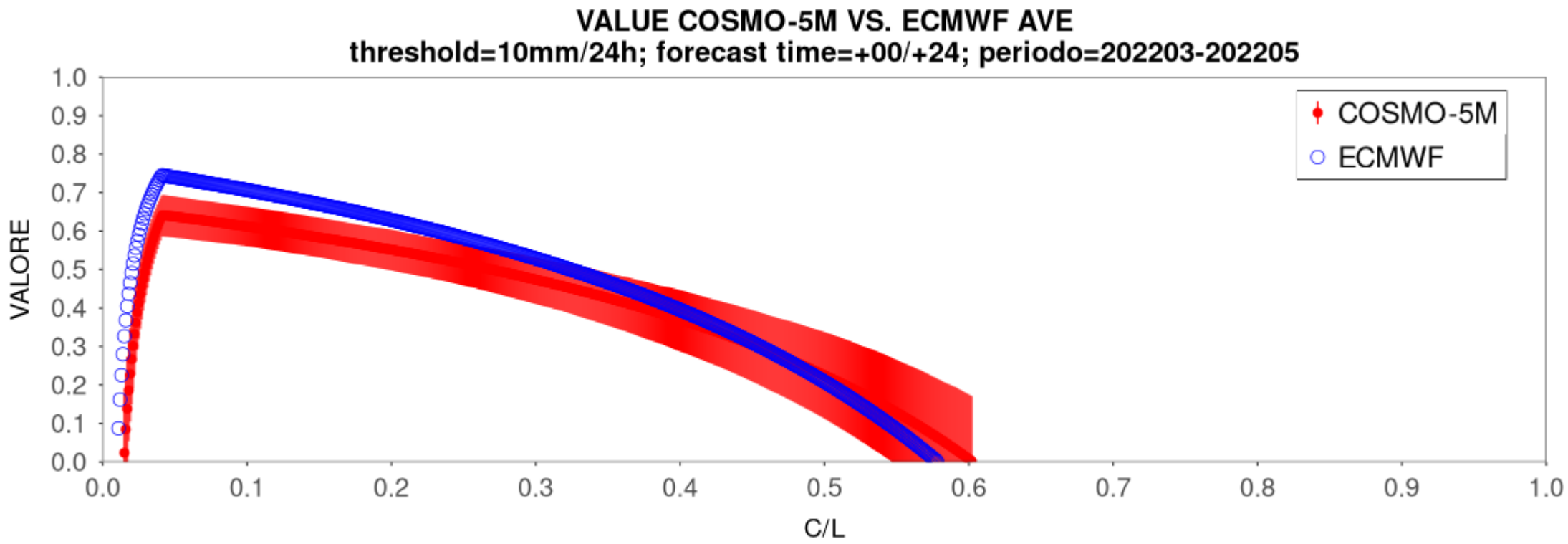
The RV is a useful index to investigate the relative improvement of your forecast system

Relative Value: MAM 2022, averaged values, 10mm/24h

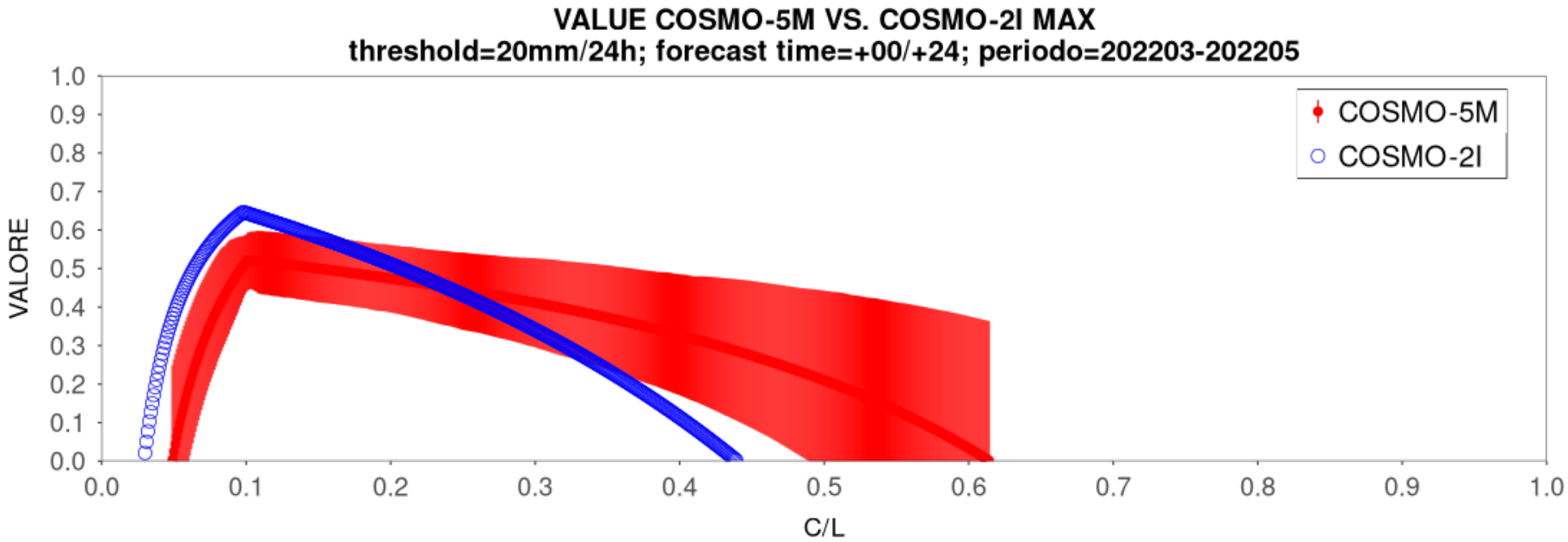
Which is the real added value given by our limited area models?

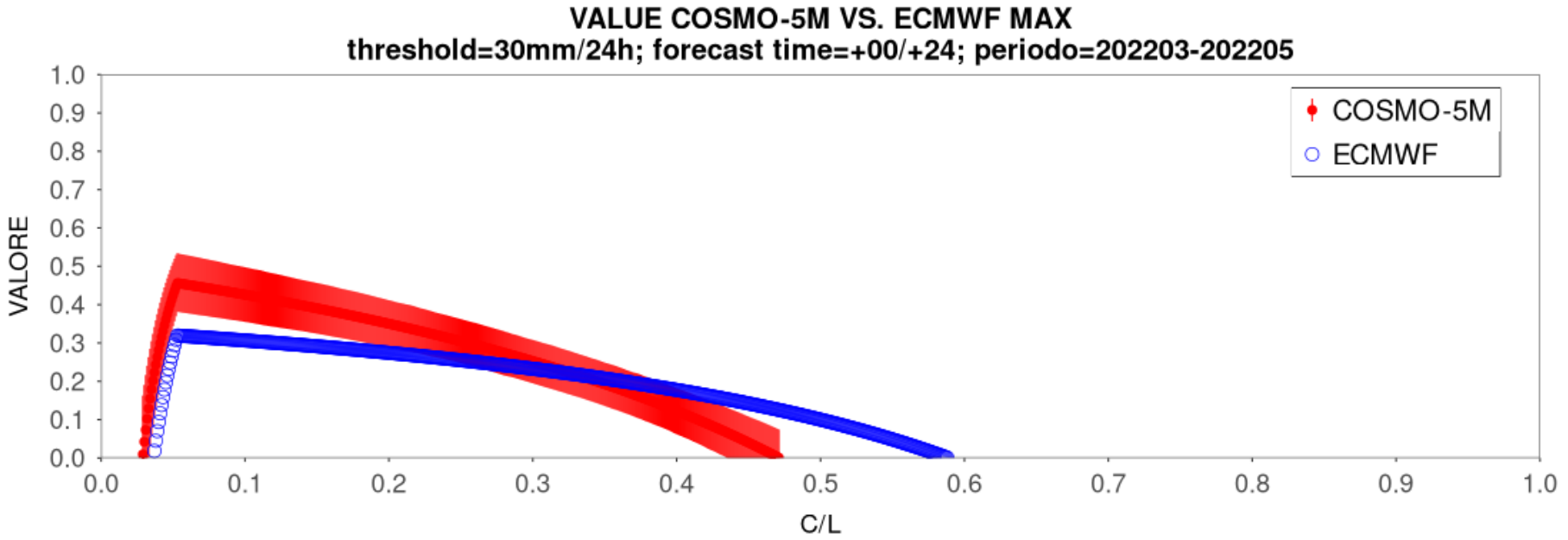
It depends on the user: if the user is the Civil Protection and the decision-making process is aimed at issuing an alert for exceeding the threshold, we look at C/L values that are very low and close to zero, as very high losses (L) can happen.

In this case, for 10mm/24 average areal, it is more convenient to use ECMWF, but if we investigate for higher thresholds (20 mm or 30 mm /24h , maximum values) we get real added value from COSMO models.



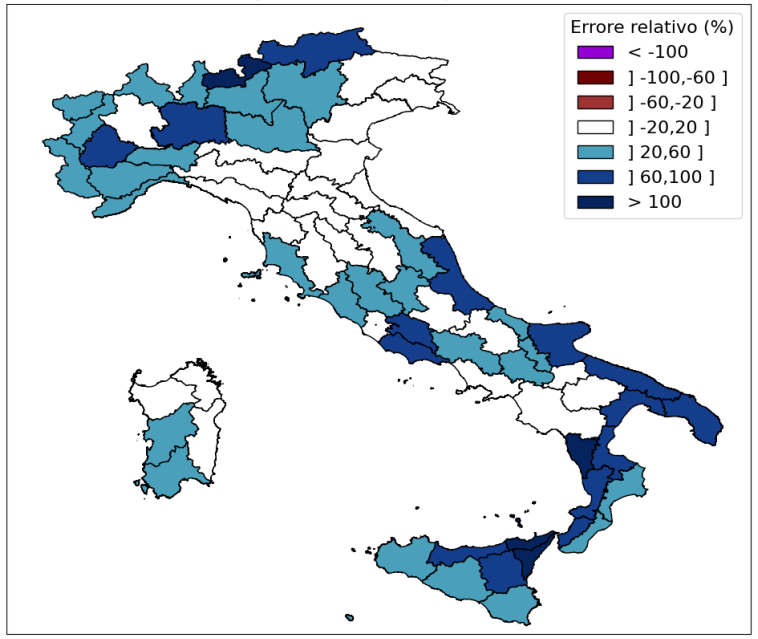
Relative Value: MAM 2022, maximum values, 20mm/24h 30mm/24h



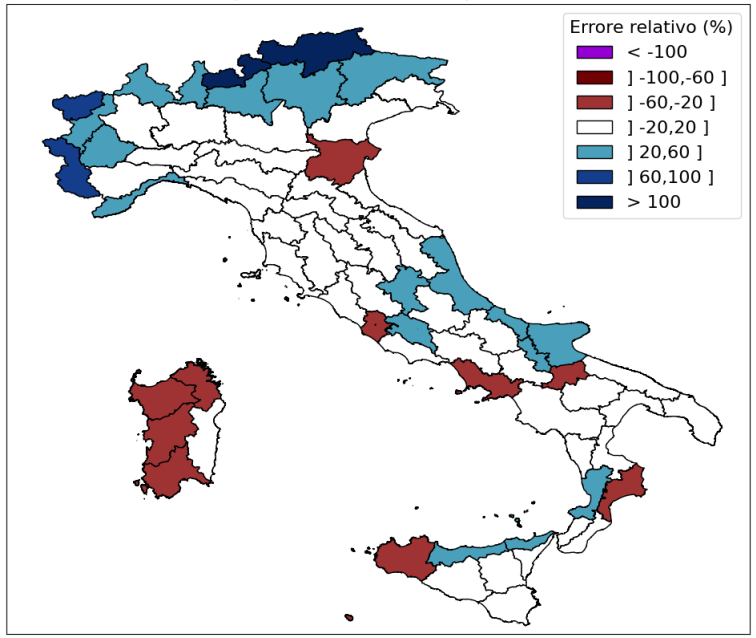


Relative error %: MAM 2022, D0

2022 MAM, ECMWF RUN 00, SCADENZA D0



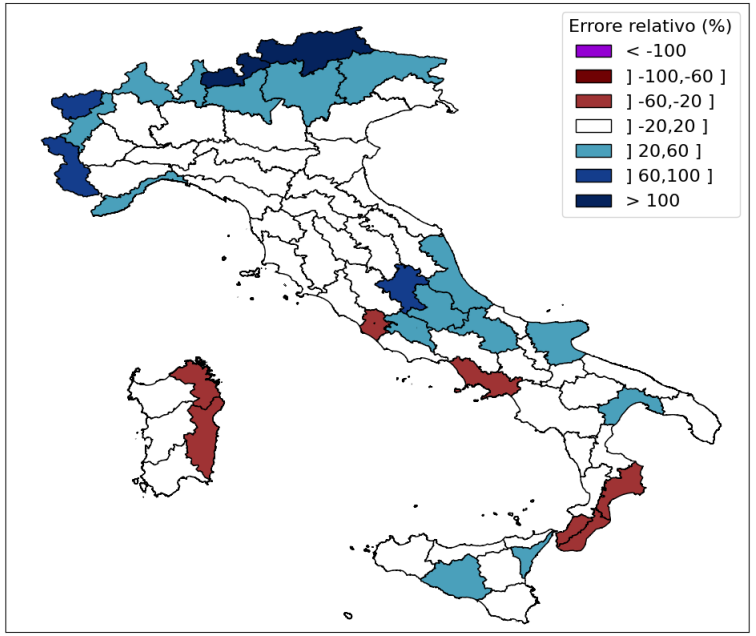
2022 MAM, COSMO-5M RUN 00, SCADENZA D0



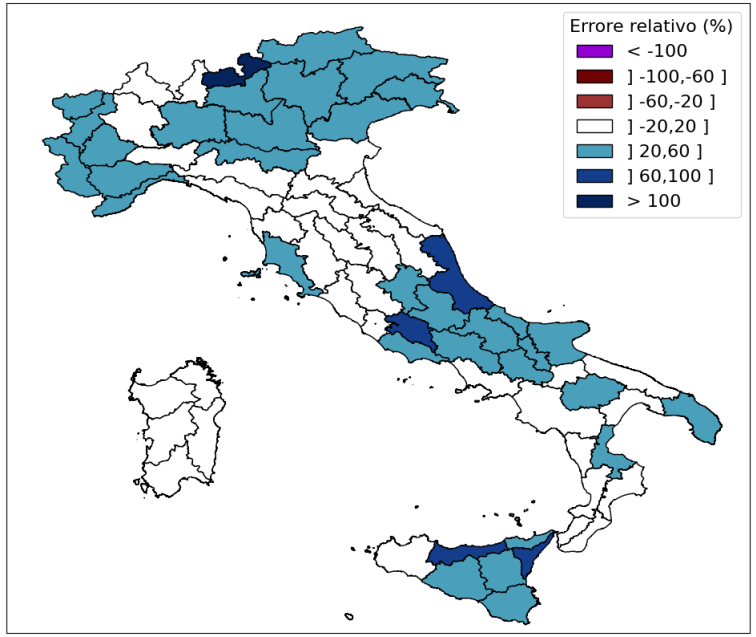
In general, it is interesting for Civil Protection purposes to plot the seasonal relative error (accompanied by the other statistical indices) for each alert area in order to have a quantitative idea regarding to the error spatial distribution.

In this case, for the last spring the marked ECMWF overestimation is evident as well as for ICON-IT.

2022 MAM, COSMO-2I RUN 00, SCADENZA D0

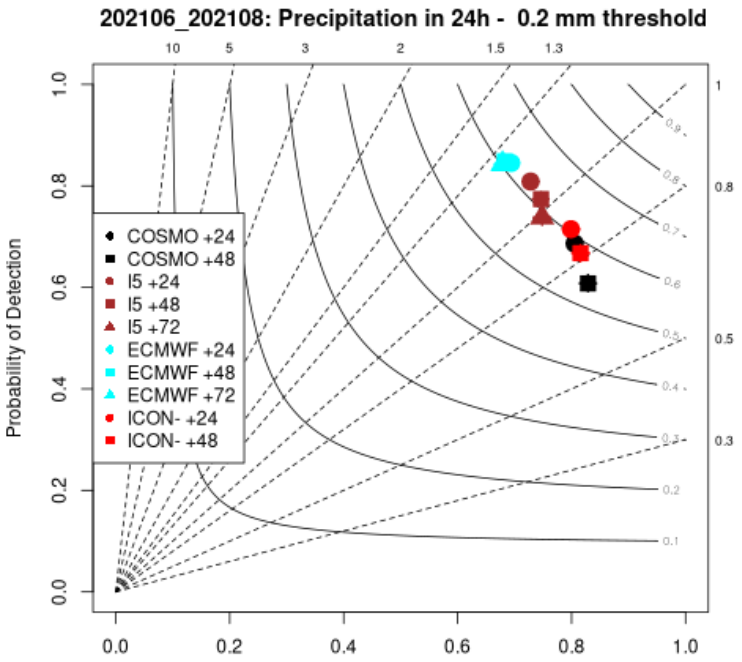


2022 MAM, ICON-IT RUN 00, SCADENZA D0

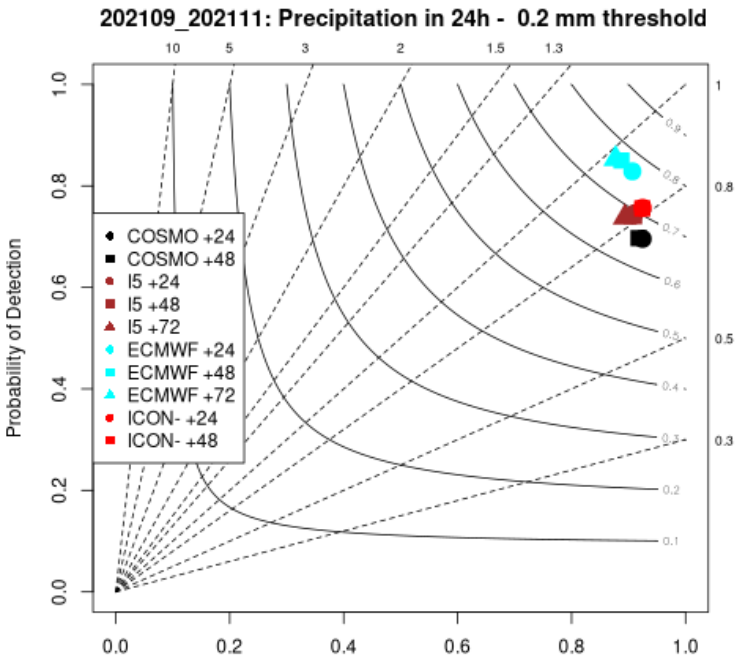


Performance diagram
Maximum Values
0.2mm/24h

JJA 2021

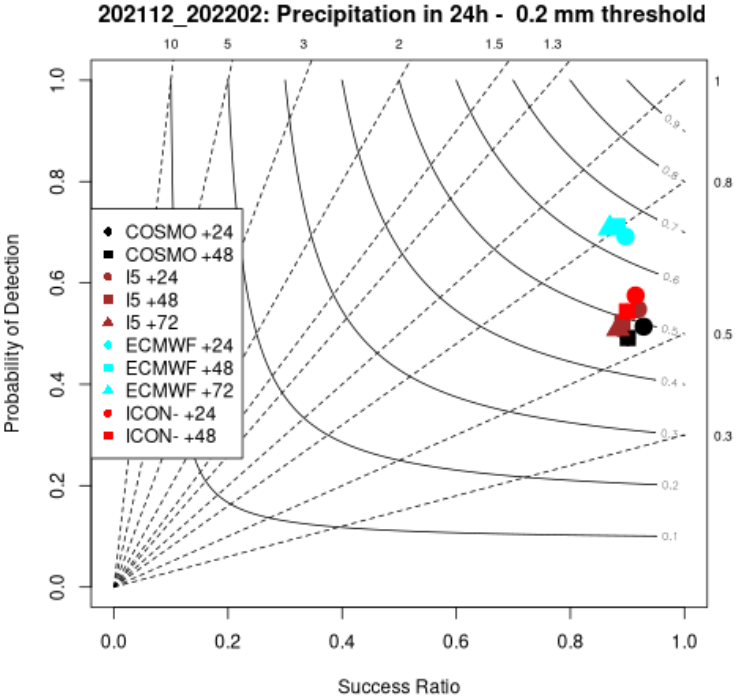


SON 2021

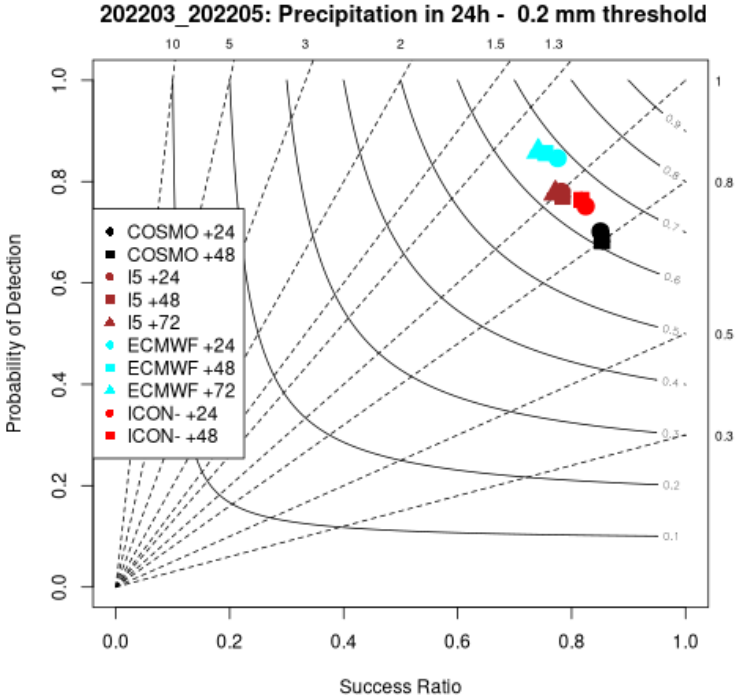


Rain/No Rain: general tendency to underestimate (the number of events) for COSMO models

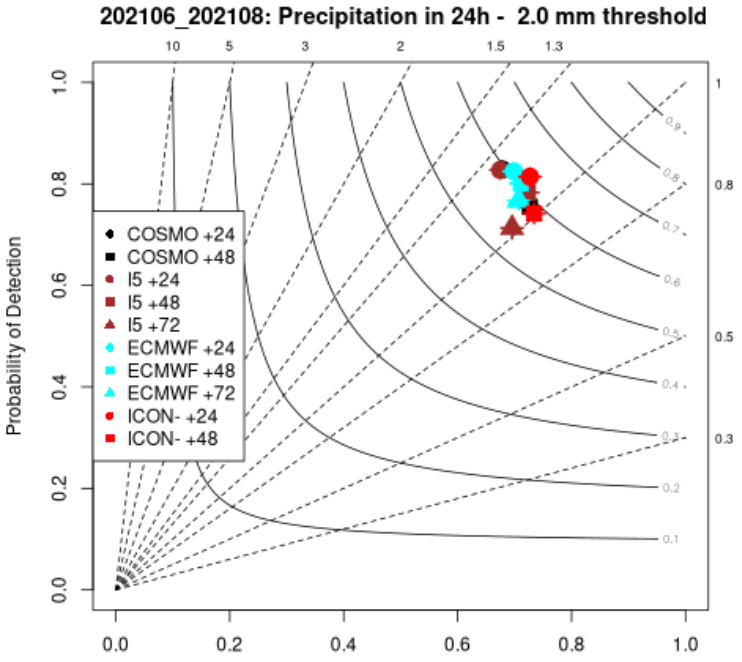
DJF 2022



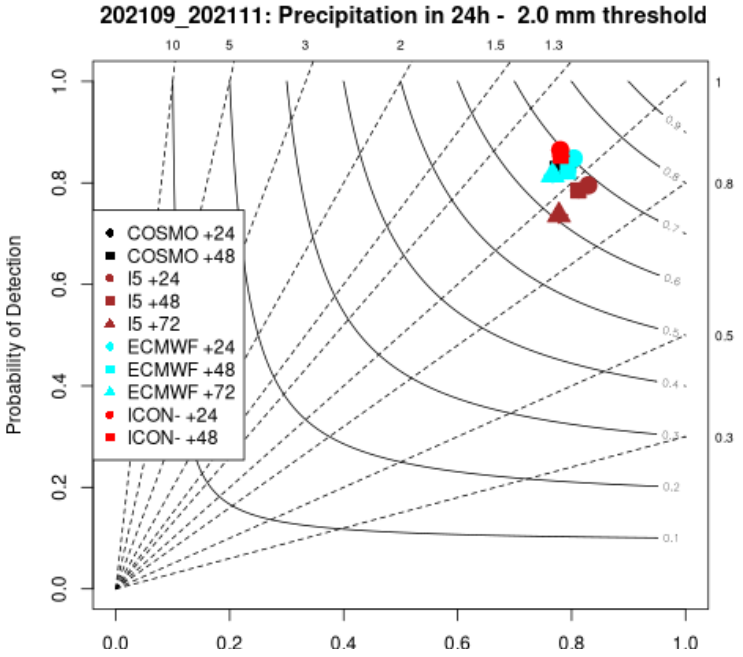
MAM 2022



Performance diagram
Maximum Values
2mm/24h

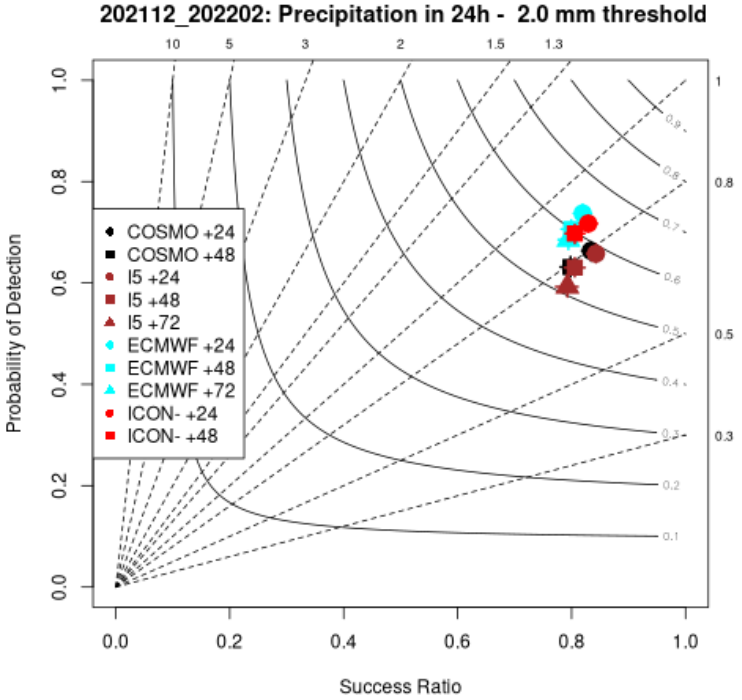


JJA 2021

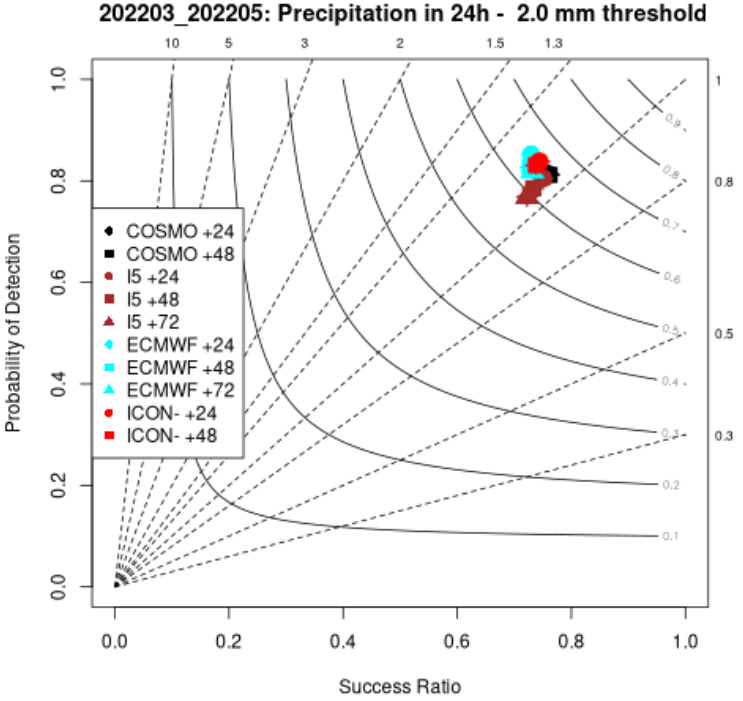


SON 2021

Low threshold: generally good skills for all the models

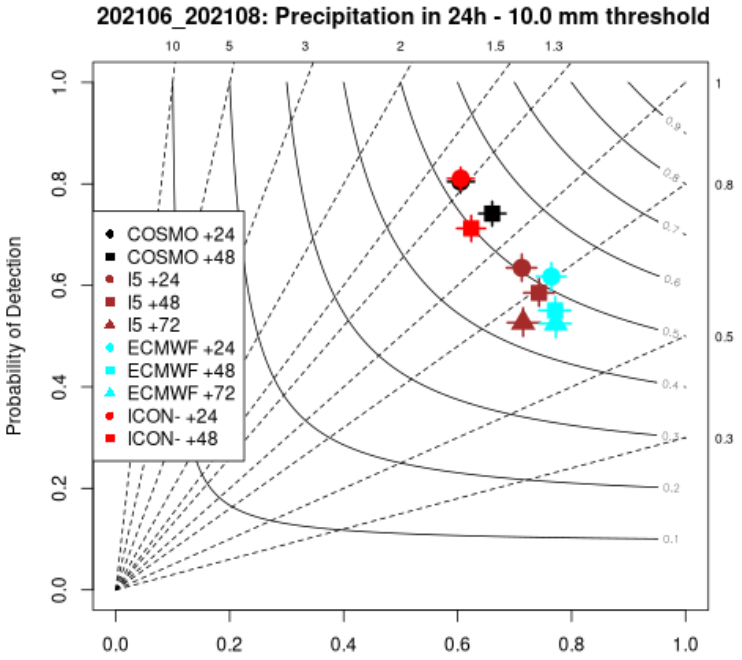


DJF 2022

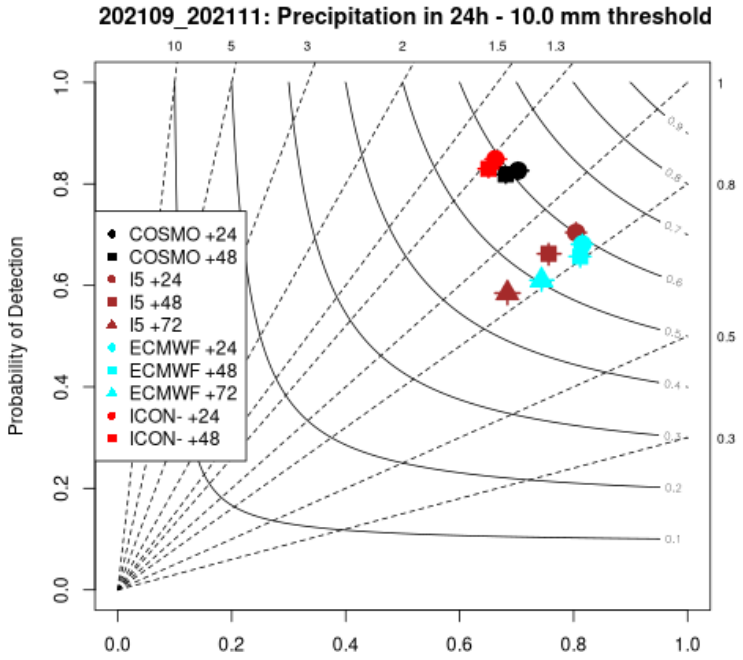


MAM 2022

Performance diagram
Maximum Values
10mm/24h



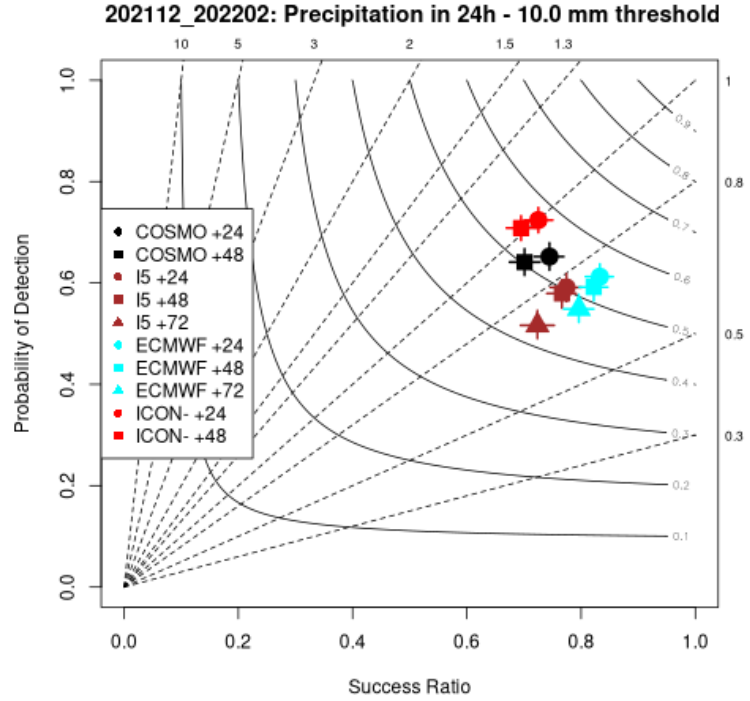
JJA 2021



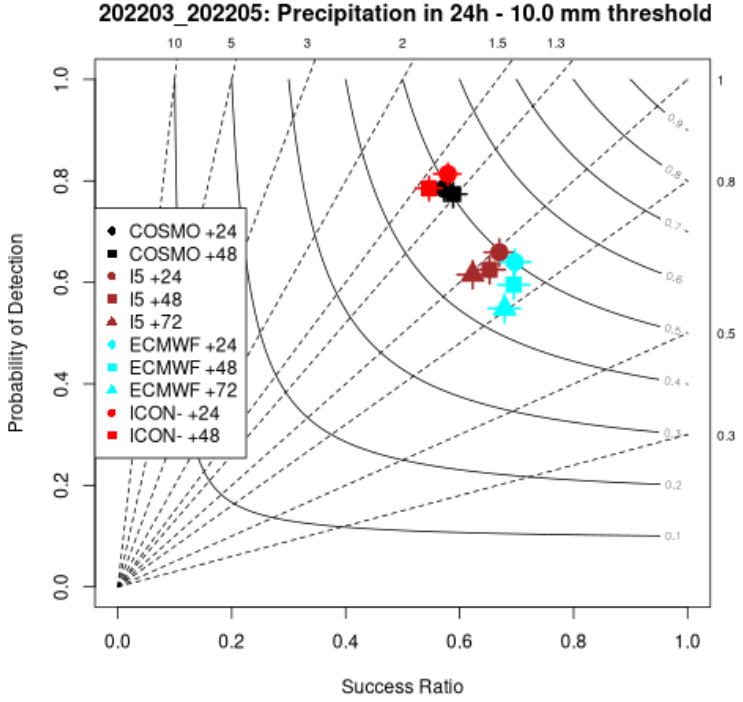
SON 2021

Medium threshold:

- we begin to see the differences in behavior
- A little overestimation (the number of events) for very high resolution COSMO model
- A little underestimation (the number of events) for ECMWF and COSMO 5km



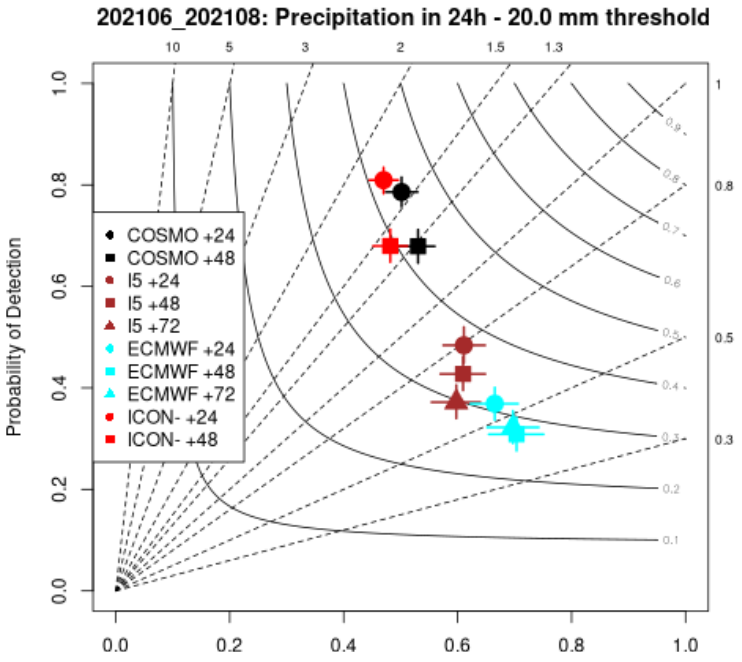
DJF 2022



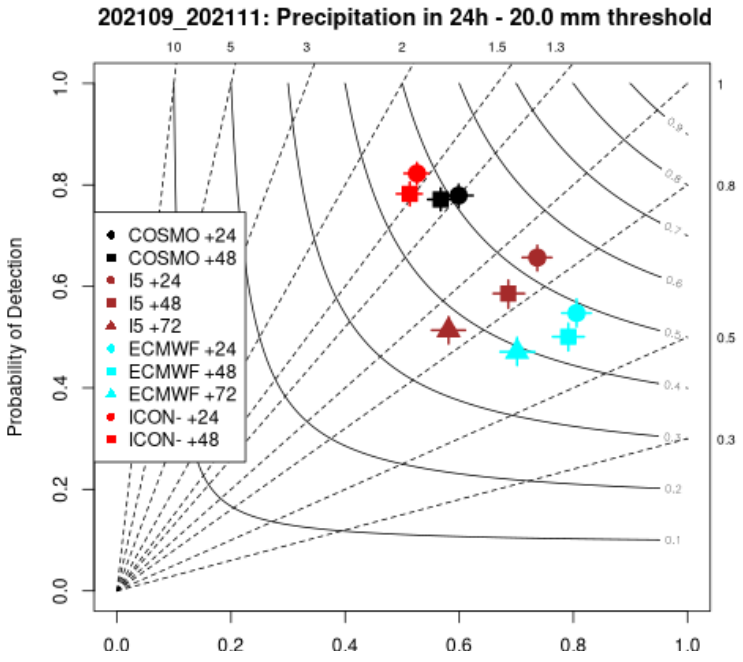
MAM 2022

Performance diagram
Maximum Values
20mm/24h

JJA 2021



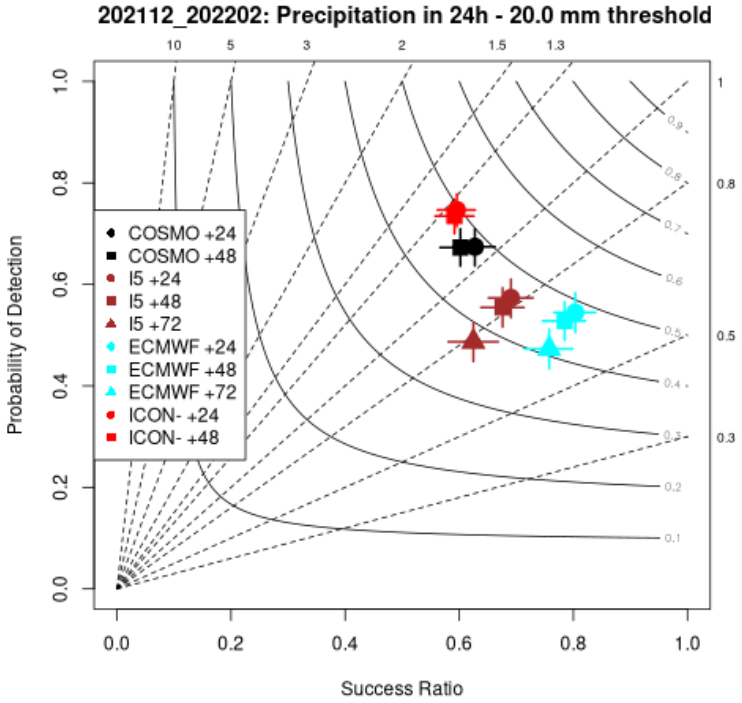
SON 2021



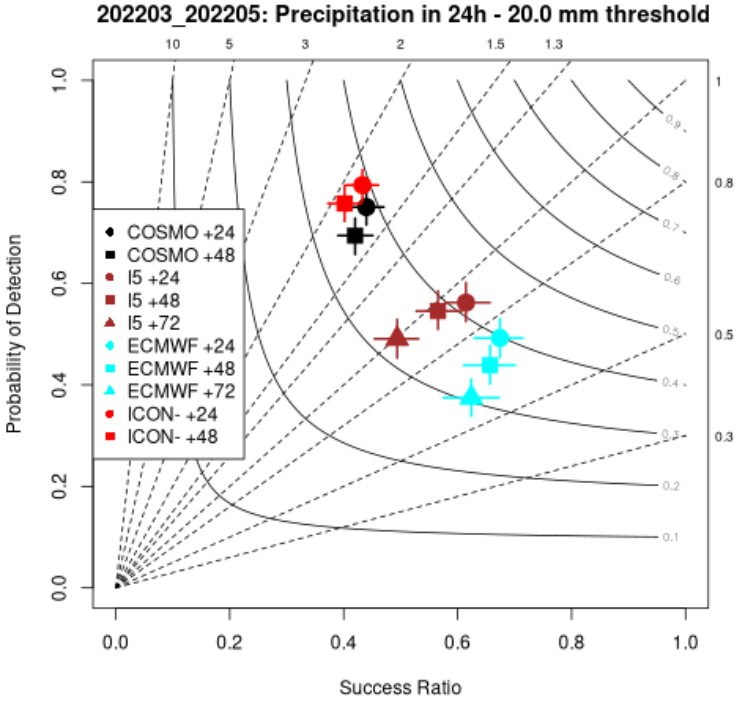
High threshold:

- we see a big difference in behavior, the models are grouped into three different subsets
- Big overestimation (the number of events) for 2km models but quite good TS
- An underestimation (the number of events) for COSMO-5M
- A strong underestimation (the number of events) for ECMWF

DJF 2022

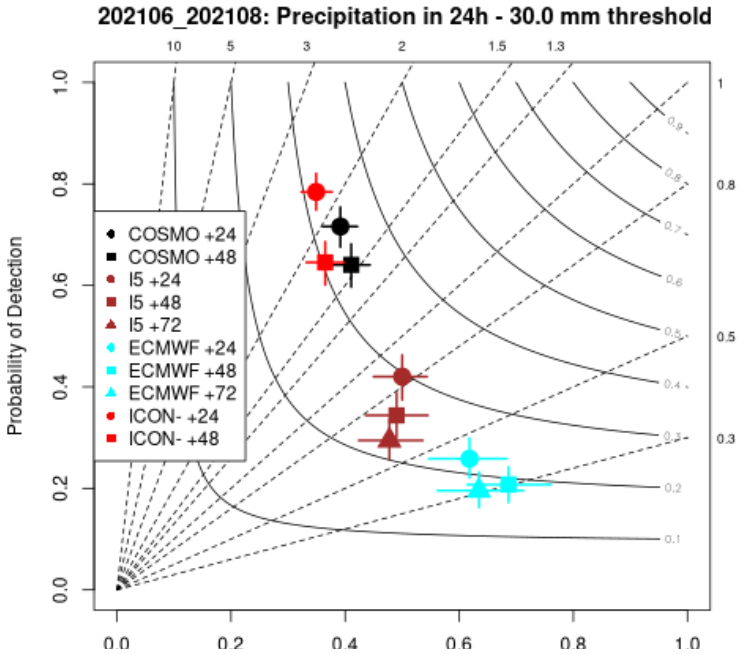


MAM 2022

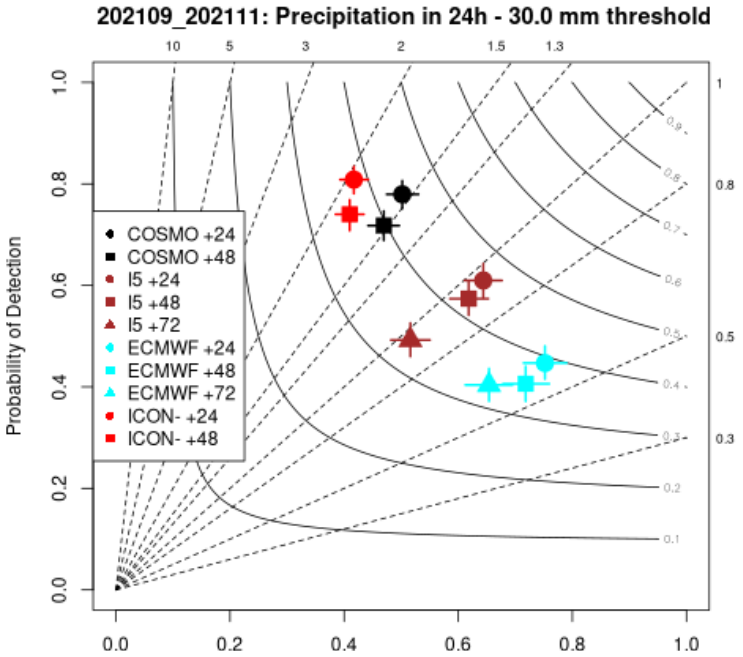


Performance diagram
Maximum Values
30mm/24h

JJA 2021



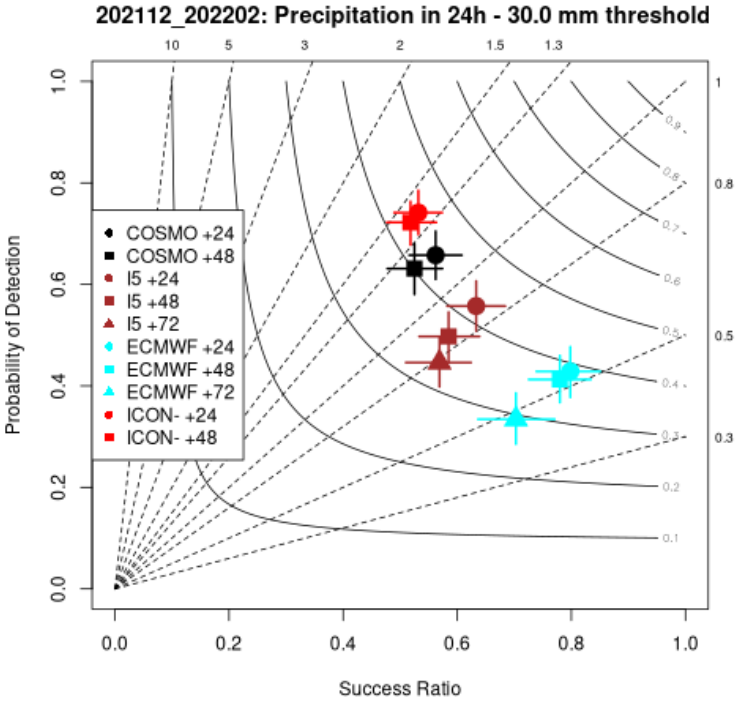
SON 2021



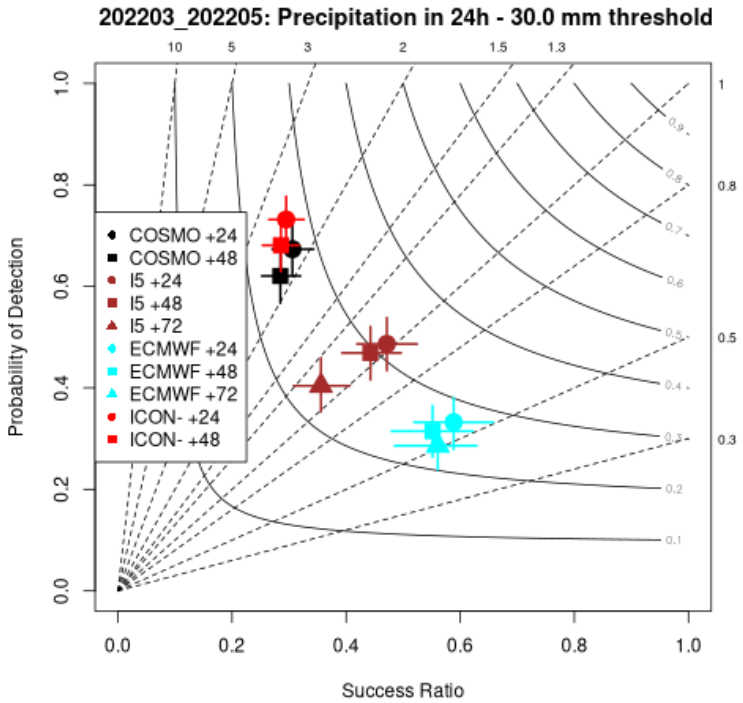
Very high threshold:

- we see a big difference in behavior, the models are grouped into three different subsets
- Big overestimation (number of events) for 2km models but quite good TS
- An underestimation (number of events) for COSMO-5M
- A strong underestimation (number of events) for ECMWF

DJF 2022



MAM 2022

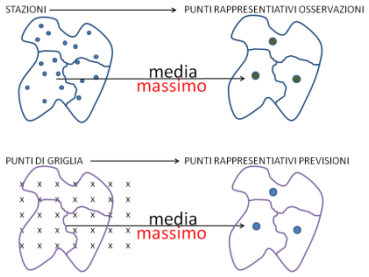


Validation for precipitation classes with "bubble plot diagram"

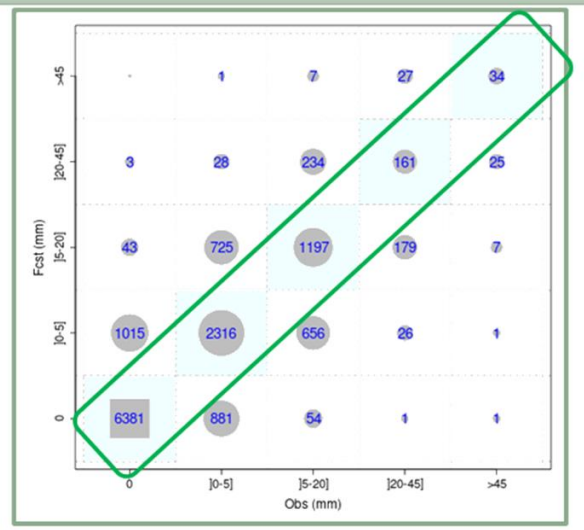
For the evaluation of critical hydrogeological or hydraulic conditions, it is useful to forecast average and maximum precipitation on each alert area categorized into classes:

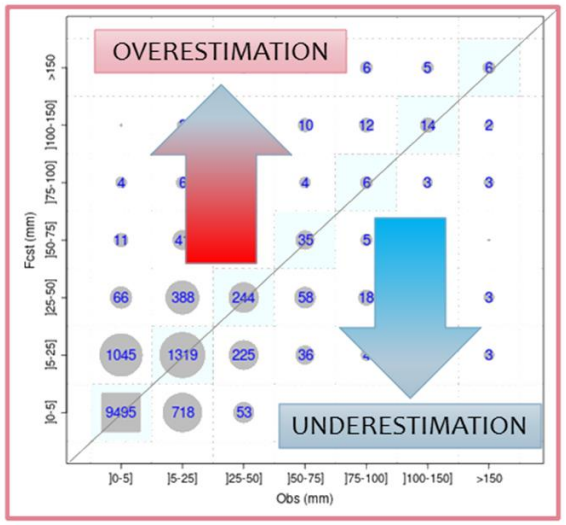
CLASSES FOR MEAN PRECIPITATION	
mm/24h	**mm/3h**
0-0.2	0-0.2
0.2-5	0.2-1
5-20	1-5
20-45	5-10
>45	>10

CLASSES FOR MAXIMUM PRECIPITATION	
mm/24h	**mm/3h**
0.2 -5	0-0.2
5-25	0.2-2
25-50	2-10
50-75	10-30
75-100	>30
100-150	
>150	



✓ CORRECT FORECASTS LIE ON THE DIAGONAL





It is a useful visual verification tool that allows you to diagnose the type of error in terms of overestimation or underestimation of events for each class.

Bubble plot is a sort of the scatter plot, in which the data points are replaced with bubbles.

The sizes of the bubbles are determined by the number of events (or the percentage respect to the total number if the events are too many)
(The square symbol is used for the most populated category to preserve the proportions of the other bubbles)

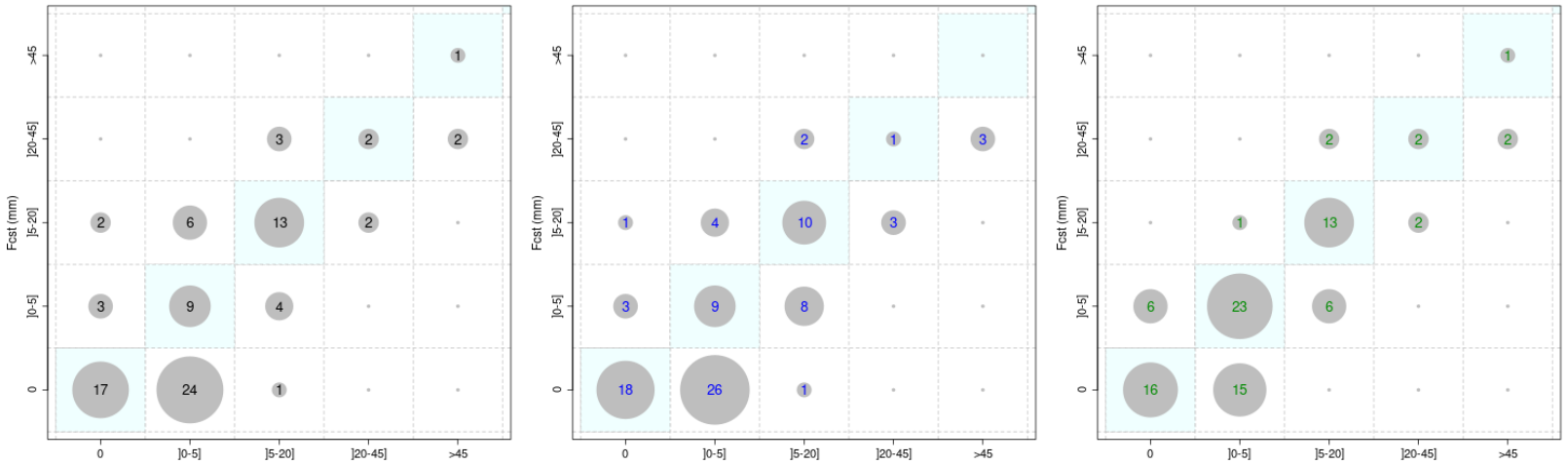
Validation for precipitation classes: some examples

COSMO-2I

COSMO-5M

IFS-ECMWF

mean

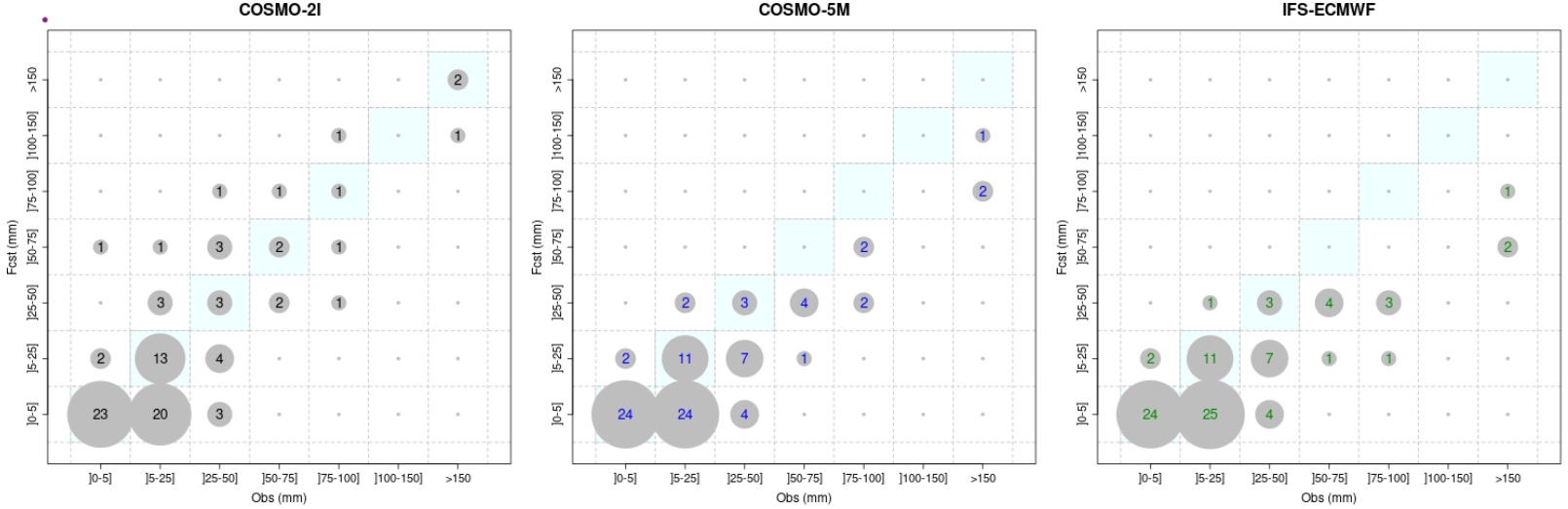


COSMO-2I

COSMO-5M

IFS-ECMWF

max

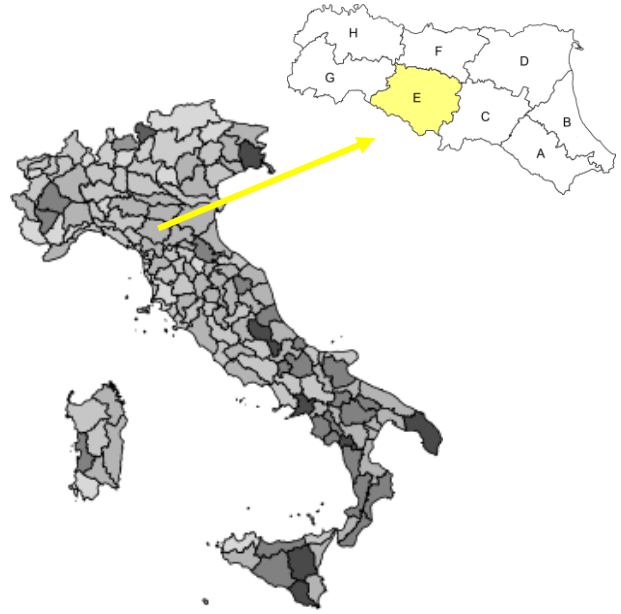


24 hours accumulated precipitation at +48h (D1)

run 00 UTC

Period: DJF 2020-2021

Alert area in Emilia-Romagna region



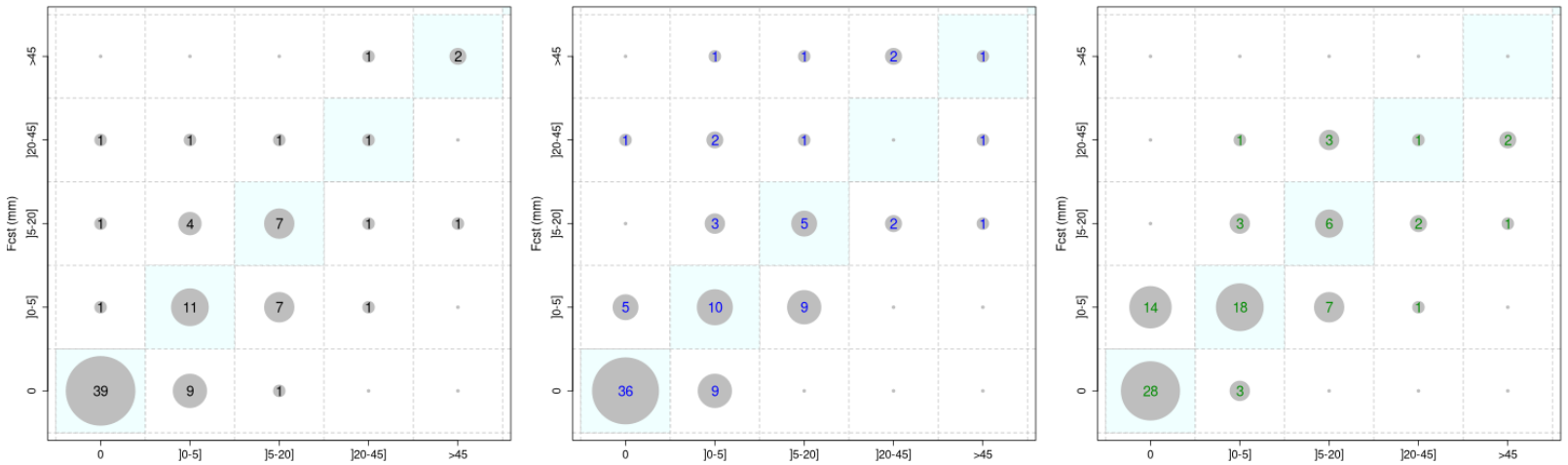
Validation for precipitation classes: some examples

COSMO-2I

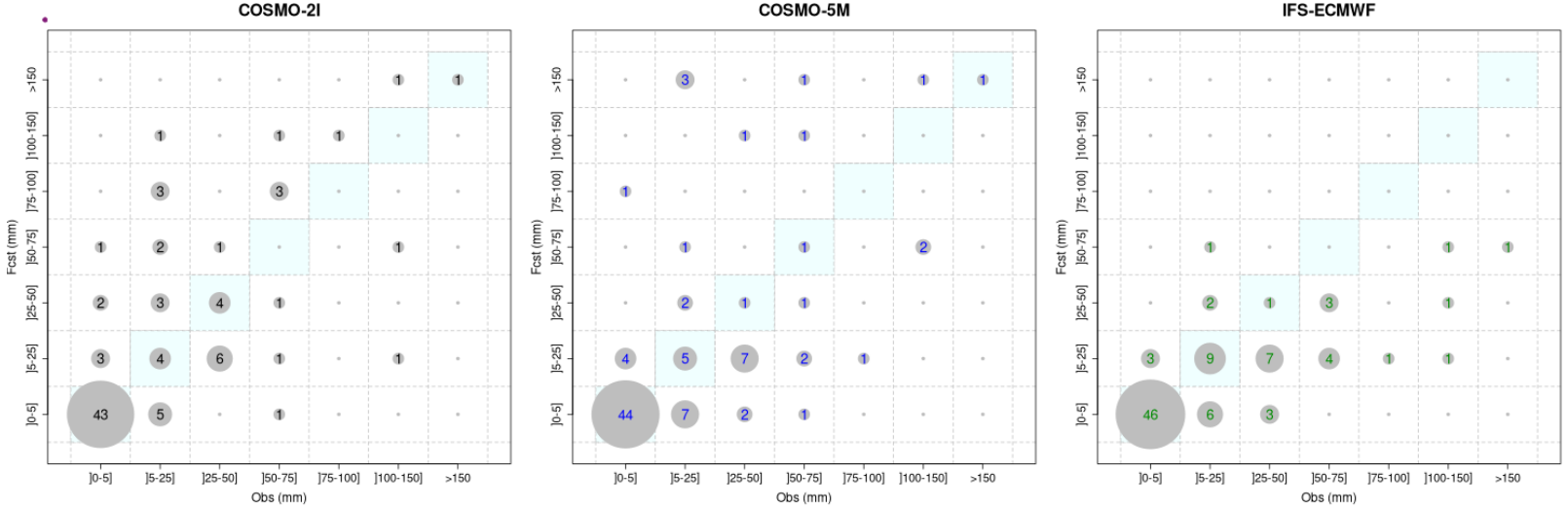
COSMO-5M

IFS-ECMWF

mean



max

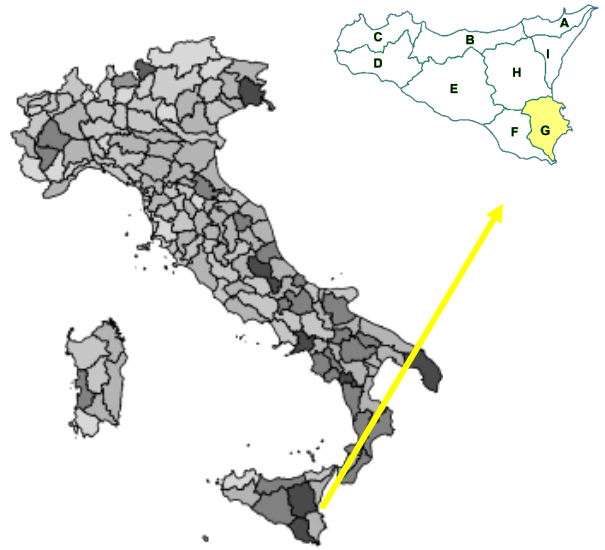


24 hours accumulated precipitation at +48h (D1)

run 00 UTC

Period: SON2021-22

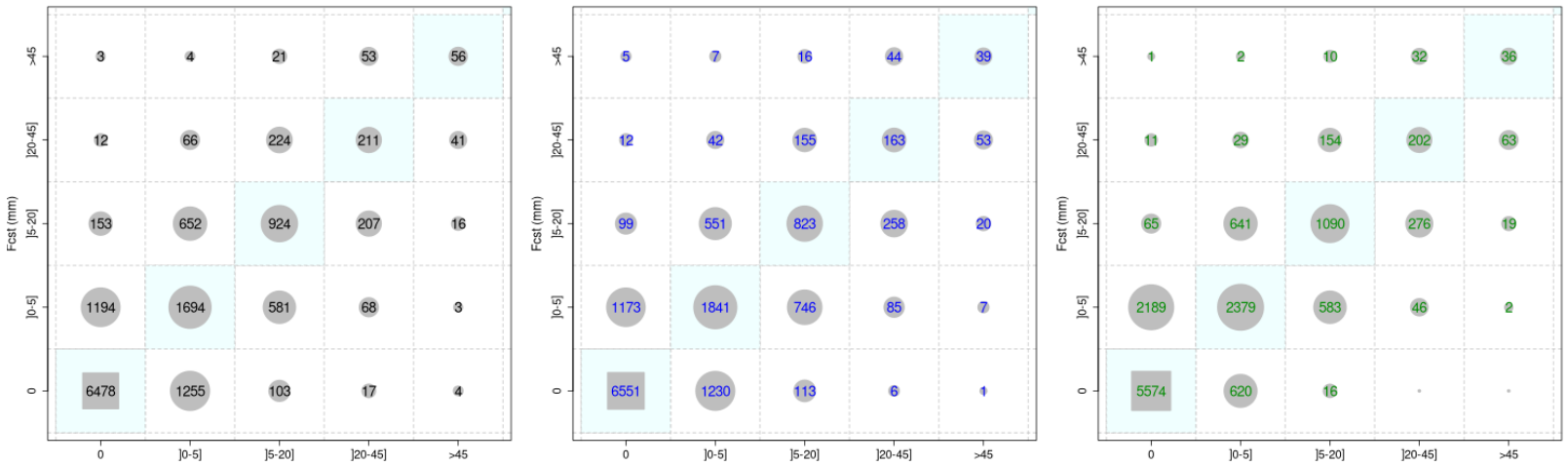
Alert area in Sicilia region



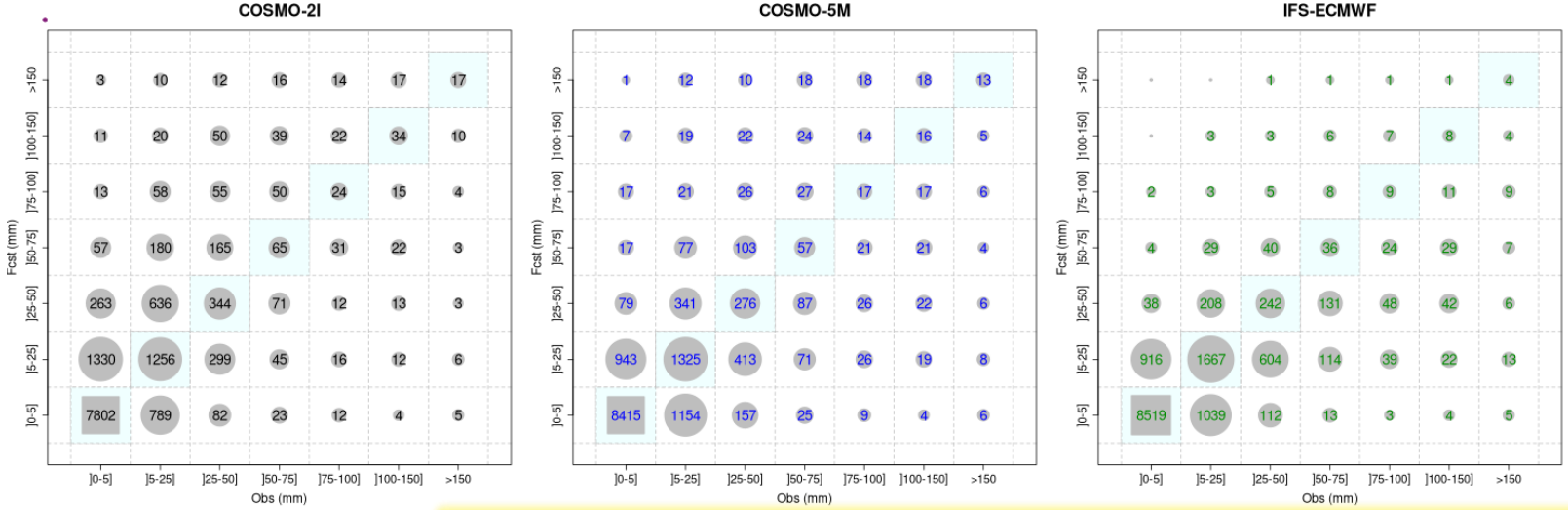
Validation for precipitation classes: some examples

COSMO-2I | COSMO-5M | IFS-ECMWF

mean



max



24 hours accumulated precipitation at +48h (D1)

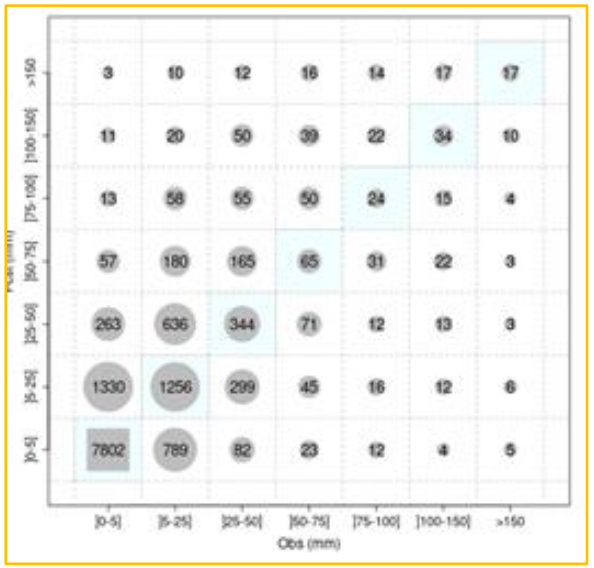
run 00 UTC

Period SON2021-22

All Italian alert areas togheter

TOO MANY DATA FOR A VISUAL VALIDATION → NEED TO HAVE AN OBJECTIVE SUMMARY

Multi-category verification – Gerrity Score



Multi-category Contingency Table

		Observed Category				Total
	i,j	1	2	...	K	
Forecast Category	1	$n(F_1,O_1)$	$n(F_1,O_2)$	...	$n(F_1,O_K)$	$N(F_1)$
	2	$n(F_2,O_1)$	$n(F_2,O_2)$	...	$n(F_2,O_K)$	$N(F_2)$
	...	...	...	...	...	...
	K	$n(F_K,O_1)$	$n(F_K,O_2)$	...	$n(F_K,O_K)$	$N(F_K)$
Total		$N(O_1)$	$N(O_2)$	...	$N(O_K)$	N

In this table $n(F_i,O_j)$ denotes the number of forecasts in category i that had observations in category j, $N(F_i)$ denotes the total number of forecasts in category i, $N(O_j)$ denotes the total number of observations in category j, and N is the total number of forecasts.

The "Gerrity Score" allows to evaluate the ability of the model to correctly separate the various classes/category

Smaller errors are penalized less than larger forecast errors. This is achieved through the use of the scoring matrix

Range: -1 to 1, 0 indicates no skill.

Perfect score: 1

Gerrity score - $$GS = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{K} n(F_i,O_j)\, s_{ij}$$

where s_{ij} are elements of a scoring matrix given by

$$s_{ii} = \frac{1}{K-1}\left(\sum_{r=1}^{i-1} a_r^{-1} + \sum_{r=i}^{K-1} a_r\right) \quad (i = j, \text{ diagonal}),$$

$$s_{ij} = s_{ji} = \frac{1}{K-1}\left(\sum_{r=1}^{i-1} a_r^{-1} - (j-i) + \sum_{r=i}^{K-1} a_r\right) \quad (i \neq j, \text{ off-diagonal})$$

$$a_i = \left(1 - \sum_{r=1}^{i} p_r\right) \Big/ \sum_{r=1}^{i} p_r$$

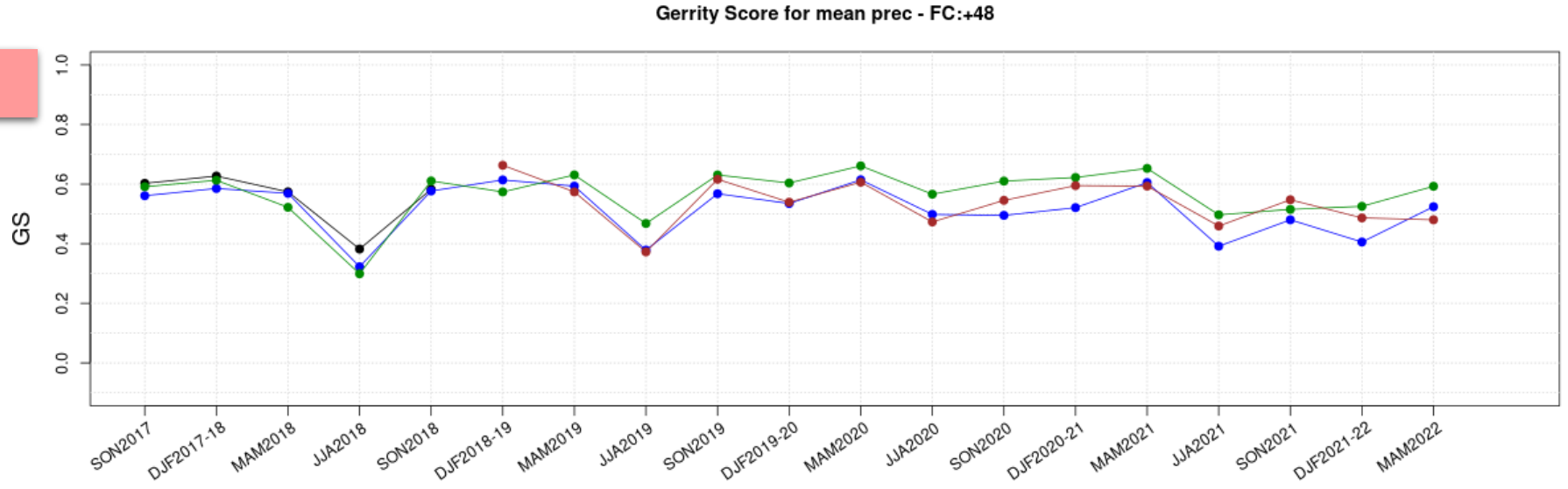
with the sample probabilities (observed frequencies) given by $p_i = N(O_i) / N$

Trend of “Gerrity Score”

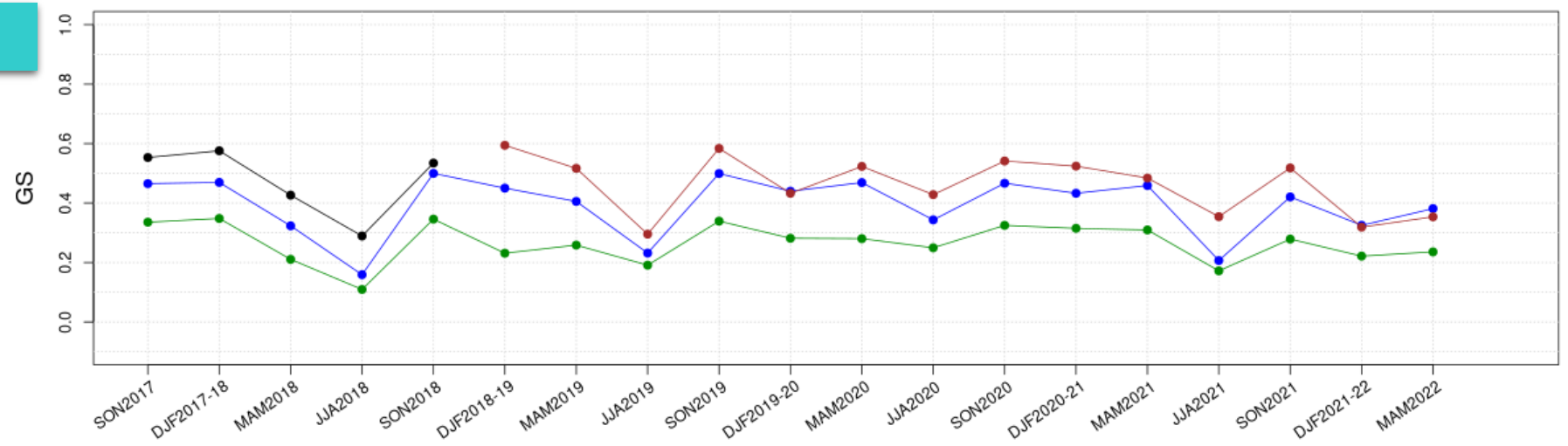
24 accumulated precipitation at +48h (run 00 UTC)

COSMO-I2
COSMO-5M
IFS-ECMWF
COSMO-2I

mean



max



ECMWF seems more accurate in predicting the correct category if we consider the mean value, even if COSMO-2I performed very similarly in the last year (except MAM)

COSMO models (in particular2I) are more accurate in representing the correct category for the maximum precipitation

First results applied to some test periods for ICON-2I

6 periods with different precipitation regimes:

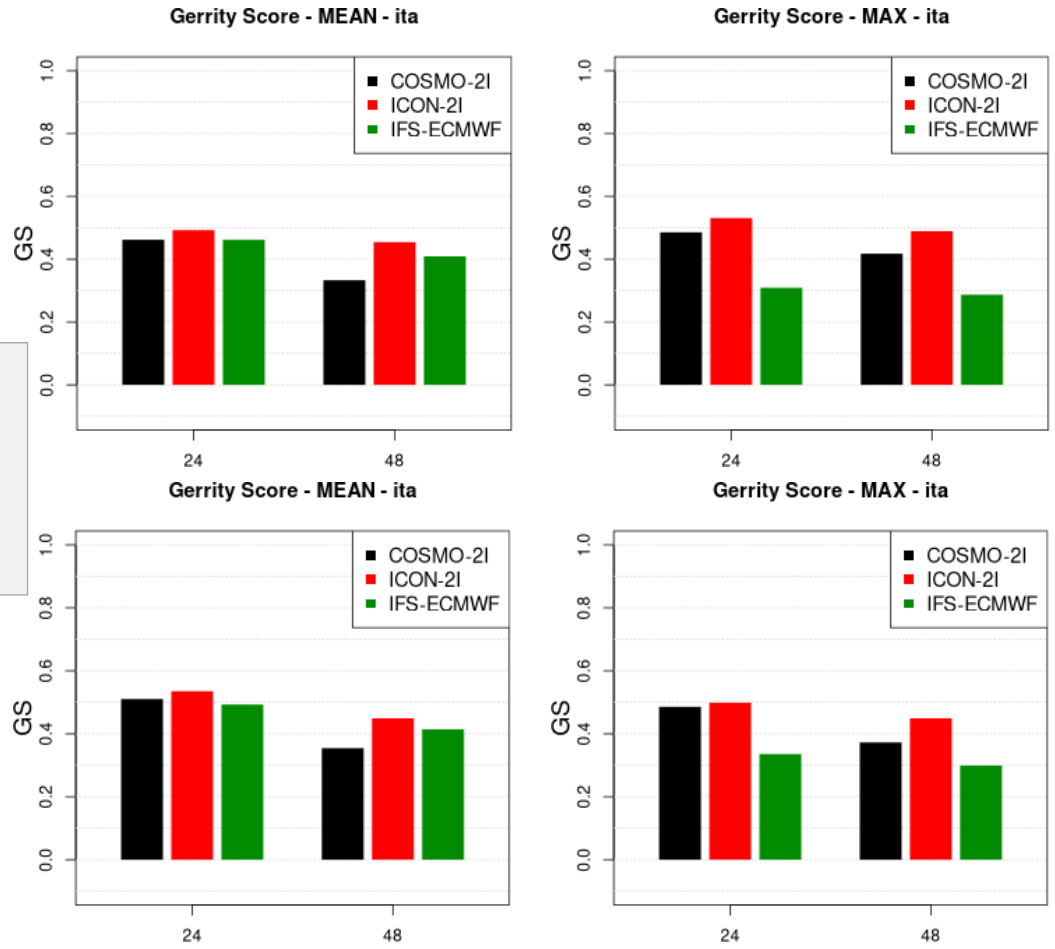
W2: 27 apr - 29 mag 2019 (33 gg)
S3: 01 lug - 02 ago 2019 (33 gg)
S5: 22 ago - 23 set 2019 (33 gg)
W3: 01 nov - 02 dic 2019 (32 gg)
W4: 24 dic 2020 - 25 gen 2021 (33 gg)
W6: 28 oct - 29 nov 2021 (33 gg)

W2-W6 Mixed type large scale/convective

Spring 2019

3 hours accumulated precipitation:
+3h to +24h →24 (D0)
+27h to +48h →48 (D1)

Autumn 2021



Information about ICON-2I
(for more details Enrico Minguzzi eminguzzi@arpae.it)

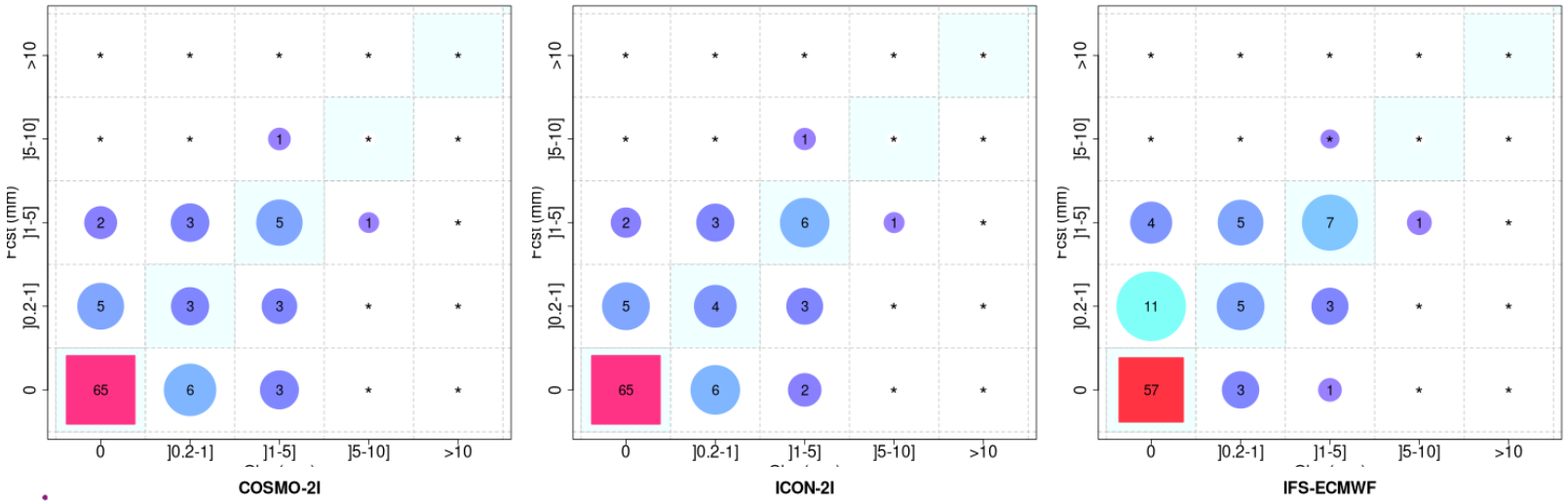
- version 2.6.4
- domain 40% larger than Cosmo-2I; resolution 2.2 km (R9B8)
- IC & BC taken from IFS (no high-resolution assimilation)
- DWD setup

ICON-2I seems to perform better than ECMWF (for max in particular) and COSMO-2I, expecially in D1 probably because COSMO-2I has data assimilation cycle (KENDA)

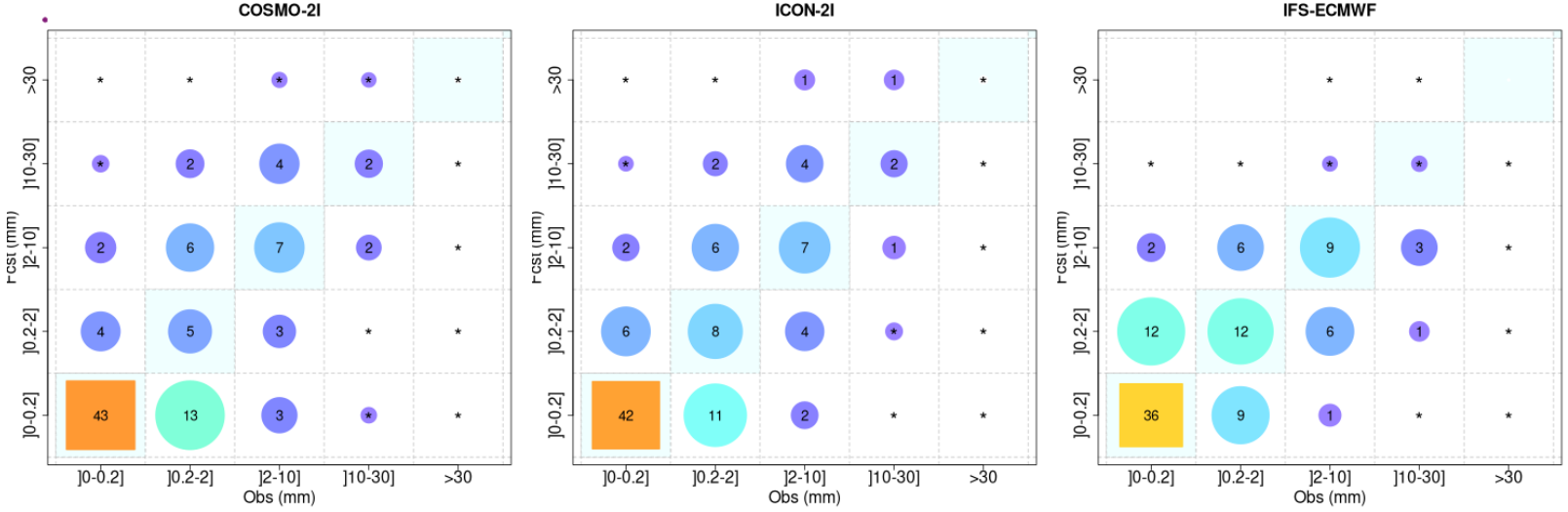
Validation for precipitation classes: some examples

COSMO-2I | ICON-2I | IFS-ECMWF

mean



max



3 hours accumulated precipitation
+27h to +48h (D1)

run 00 UTC

Period: **spring 2019**

All Italian alert areas togheter

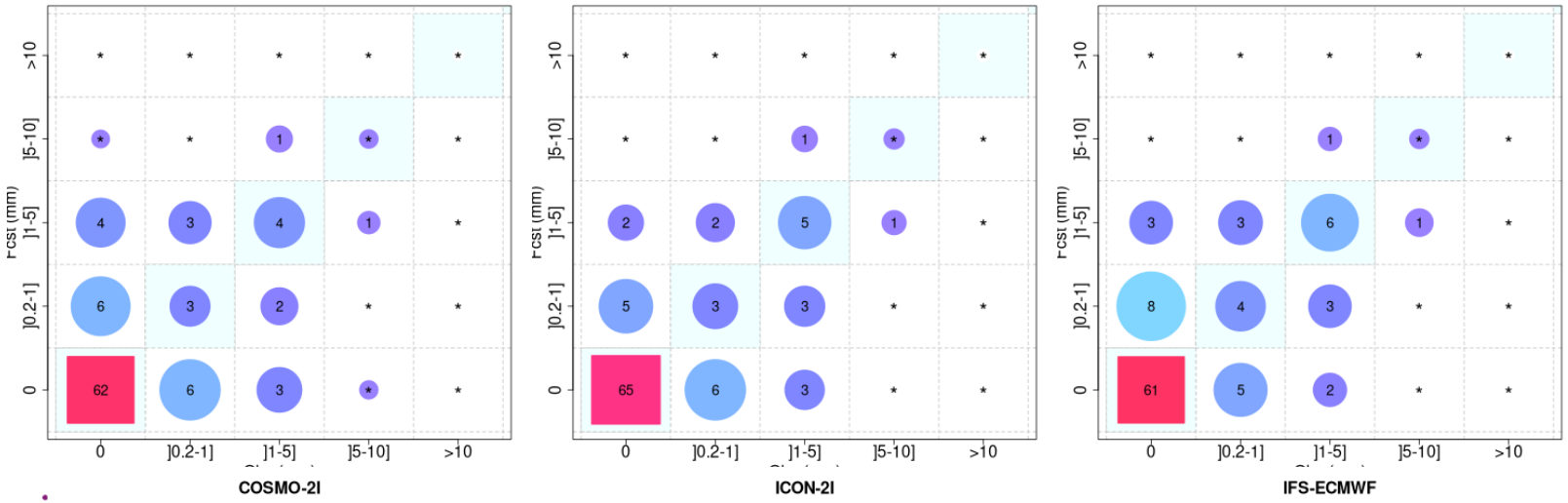
The size of the bubbles are proportional to the fraction of events respect to the total number.

The value in the bubble represents the percentage

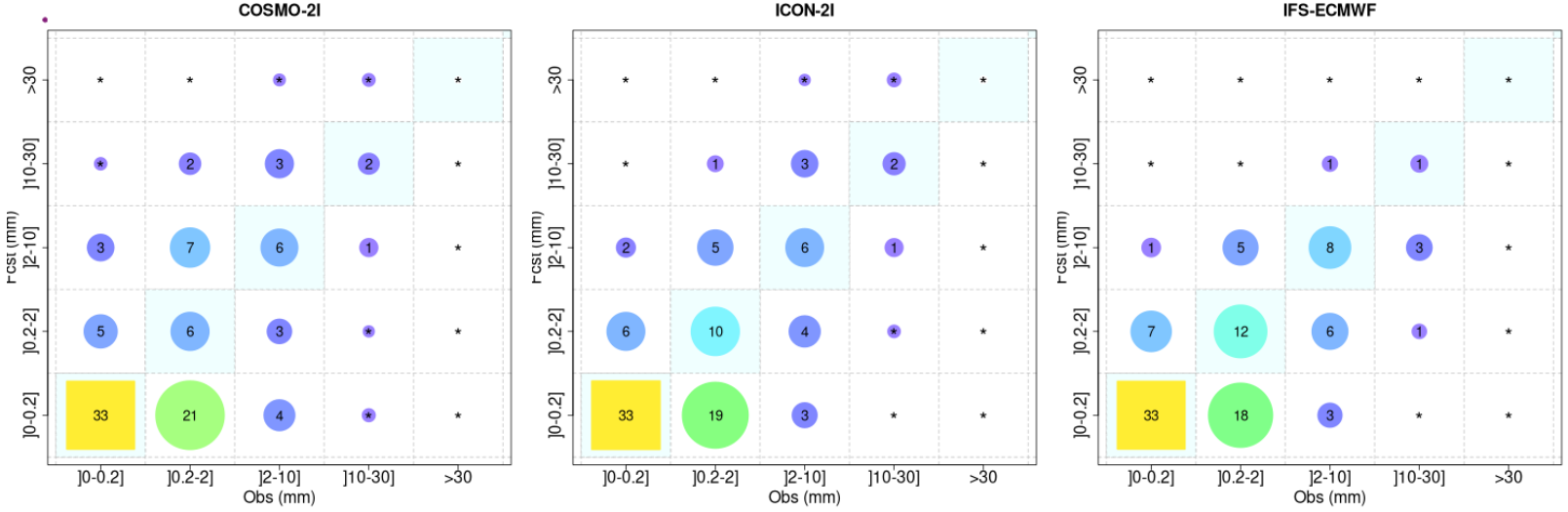
Validation for precipitation classes: some examples

COSMO-2I | ICON-2I | IFS-ECMWF

mean



max



3 hours accumulated precipitation
+27h to +48h (D1)

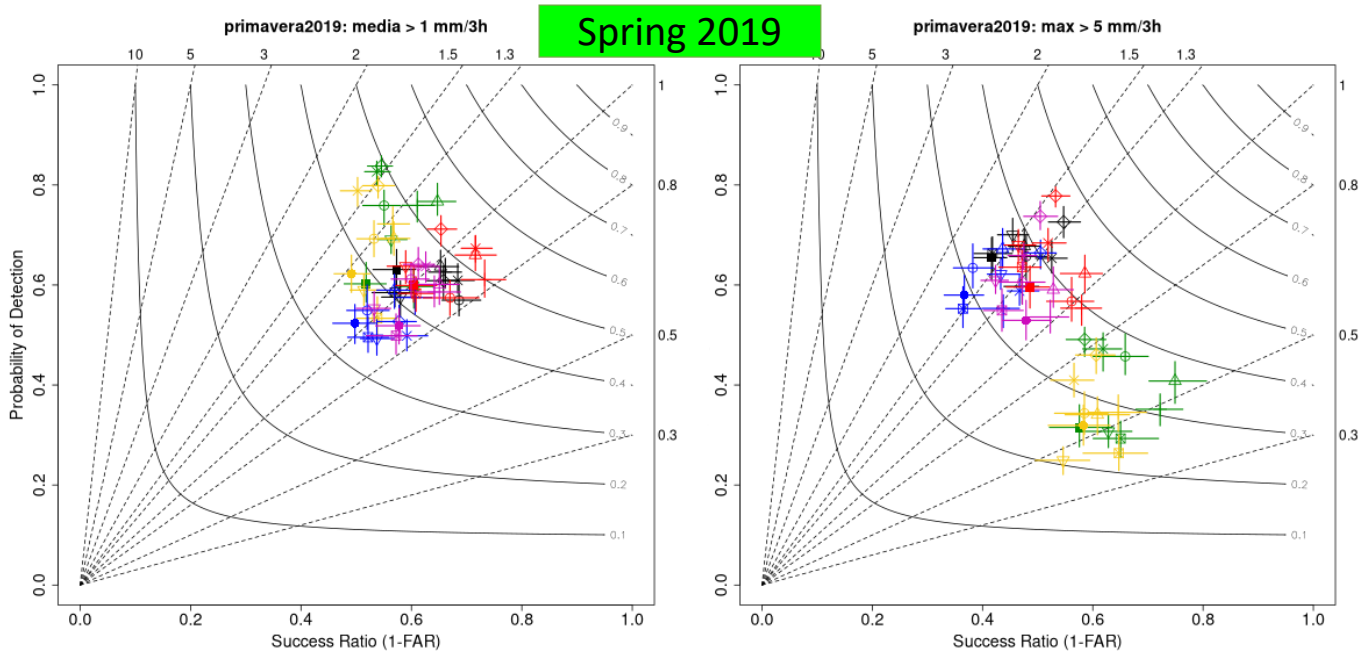
run 00 UTC

Period: **novembre 2021**

All Italian alert areas togheter

The size of the bubbles are proportional to the fraction of events respect to the total number.

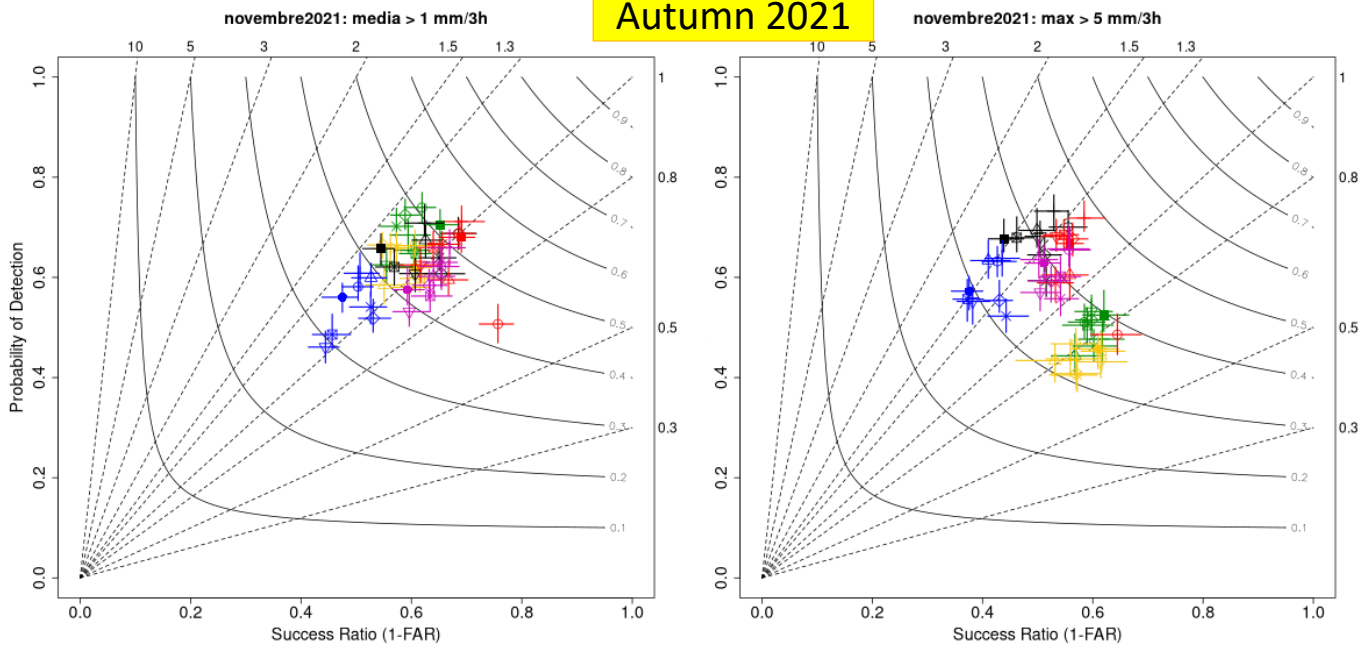
The value in the bubble represents the percentage



Spring 2019

mean

max



Autumn 2021

In addition to the "usual" behavior related to the resolution of the model, for ICON-2I all the forecast step of D1 are much closer to those of D0

In Autumn 2021 the +3h is very different from the others and worst respect to +27h This behavior does not appear in Spring (maybe because of the regime of precipitation), but other tests suggested possible problems in the initial part of the run → to be investigated

- IFS-ECMWF +3h
- IFS-ECMWF +6h
- IFS-ECMWF +9h
- IFS-ECMWF +12h
- IFS-ECMWF +15h
- IFS-ECMWF +18h
- IFS-ECMWF +21h
- IFS-ECMWF +24h
- IFS-ECMWF +27h
- IFS-ECMWF +30h
- IFS-ECMWF +33h
- IFS-ECMWF +36h
- IFS-ECMWF +39h
- IFS-ECMWF +42h
- IFS-ECMWF +45h
- IFS-ECMWF +48h
- COSMO-2I +3h
- COSMO-2I +6h
- COSMO-2I +9h
- COSMO-2I +12h
- COSMO-2I +15h
- COSMO-2I +18h
- COSMO-2I +21h
- COSMO-2I +24h
- COSMO-2I +27h
- COSMO-2I +30h
- COSMO-2I +33h
- COSMO-2I +36h
- COSMO-2I +39h
- COSMO-2I +42h
- COSMO-2I +45h
- COSMO-2I +48h
- ICON-2I +3h
- ICON-2I +6h
- ICON-2I +9h
- ICON-2I +12h
- ICON-2I +15h
- ICON-2I +18h
- ICON-2I +21h
- ICON-2I +24h
- ICON-2I +27h
- ICON-2I +30h
- ICON-2I +33h
- ICON-2I +36h
- ICON-2I +39h
- ICON-2I +42h
- ICON-2I +45h
- ICON-2I +48h

D0

D1

Conclusion

- Using aggregate QPF on alert areas produces good results: COSMO models in general are better performing in predicting maximum precipitation while for the average ECMWF still seems to prevail
- COSMO 2I and ICON-IT/2I seem to best represent the precipitation spectrum within the alert areas, distinguishing well especially the precipitation maxima
  - This aspect is fundamental since the QPF is used for Civil Protection purposes
- The good results of the verification carried out considering the precipitation in steps of 3 hours show that the timing of the phenomena is quite well forecasted

Thank you for your attention!