COSMO Verification Activities

Flora Gofa
COSMO WG5 Coordinator – Verification and Case studies

with contributions by WG5 people

Work Group 5 Task List

1. Common Verification Framework
1.1 Operational Verification
1.2 Responsibility for Common Plots Reports
1.3 Verification of vertical profiles using TEMP observations, aircraft data (AMDAR) and wind-profiler data
1.4 Dissemination of daily Grib model output Files

2. Exploitation of observational dataset for operational and scientific purposes
2.1 High density verification of precipitation over Italy
2.2 Exchange of a common data set of non-GTS data DWD
2.3 Evaluation of COSMO models in the lower PBL

3. Evaluation of convection permitting models performance
3.1 Long Term Trend Verification
3.2 Conditional Verification
3.3 Weather Dependant Verification (WDV)
3.4 Severe and High Impact Weather

Work Group 5 Task List

4. Neighborhood method techniques

4.1 Verification of COSMO-7 precipitation forecast using Radar composite network

4.2 Precipitation verification using radar composite network with neighborhood methods

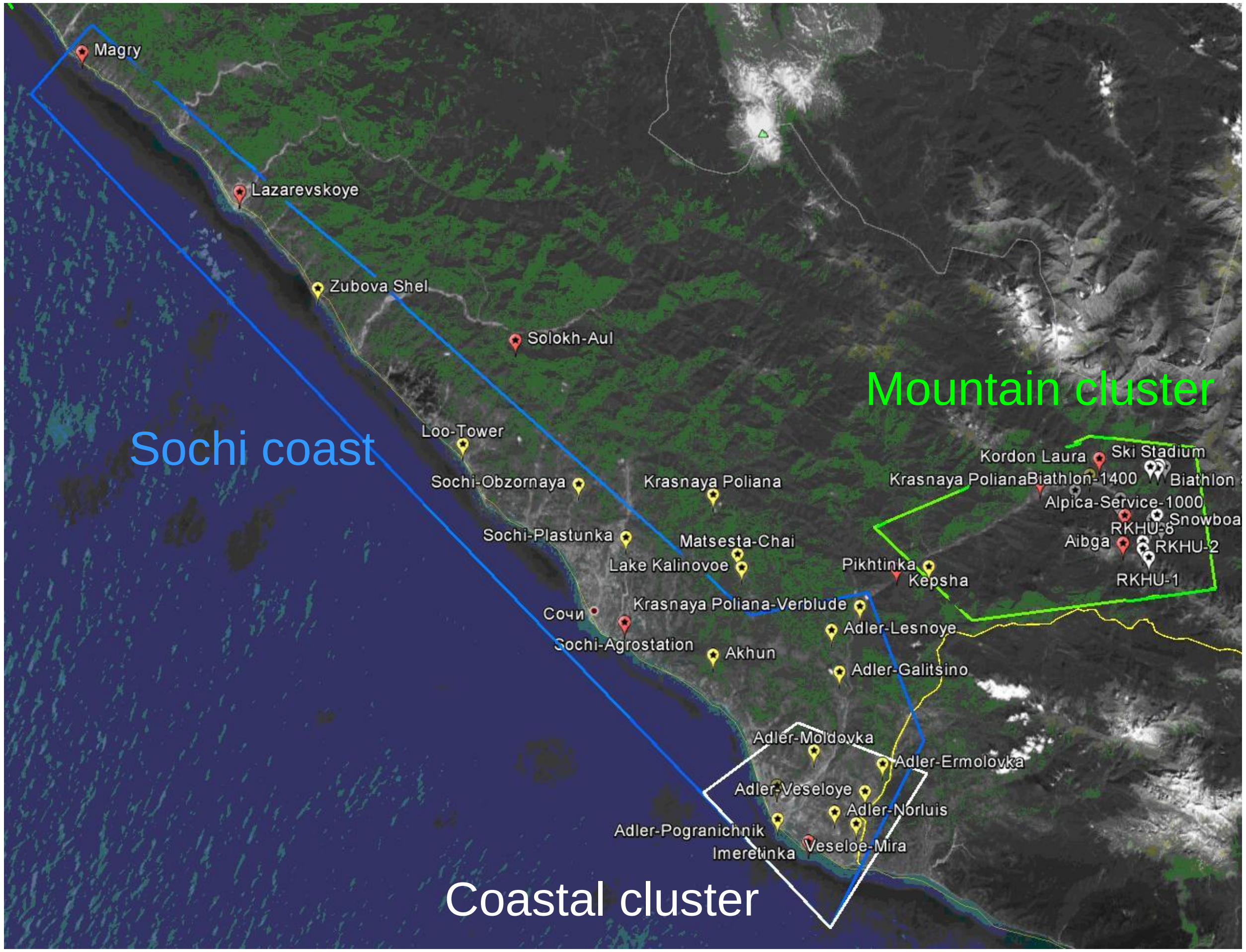
5. Verification of EPS products (Cooperation with WG7)

6. Other

6.1 Annual Workshop/Tutorial on VERSUS2 & WG5

Sochi Olympics Verification Experience

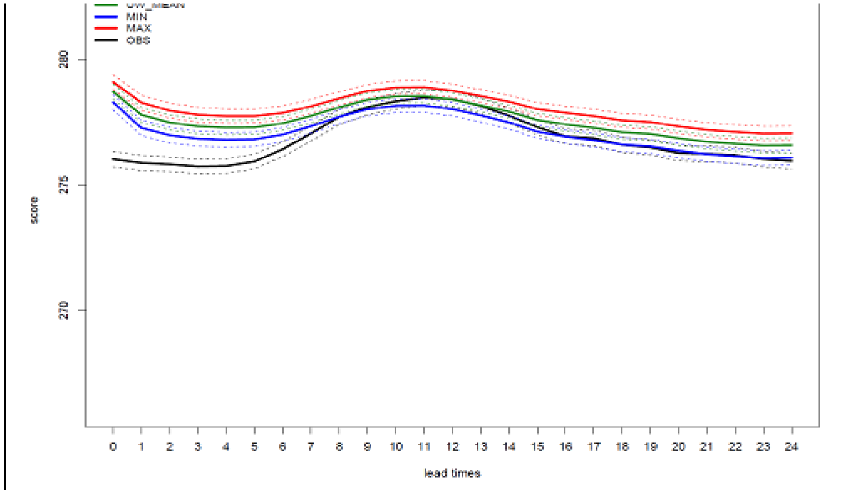
Polygons of verification

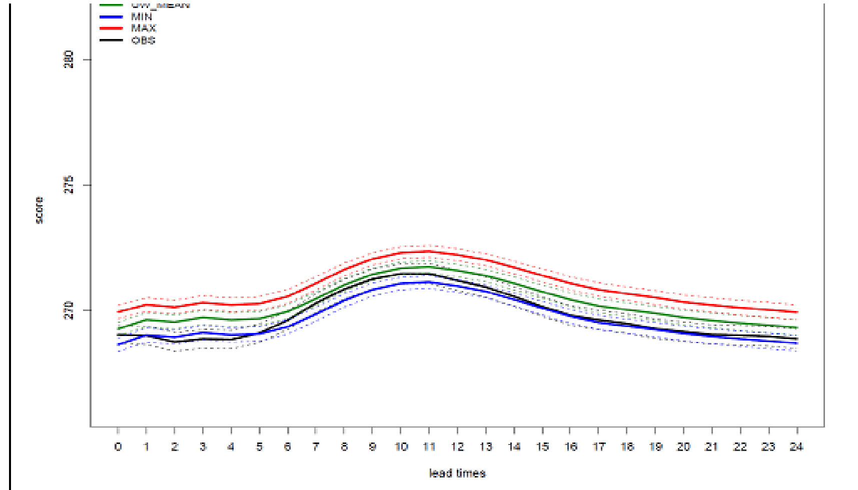


Forecasts for the Mountain cluster are the most important!

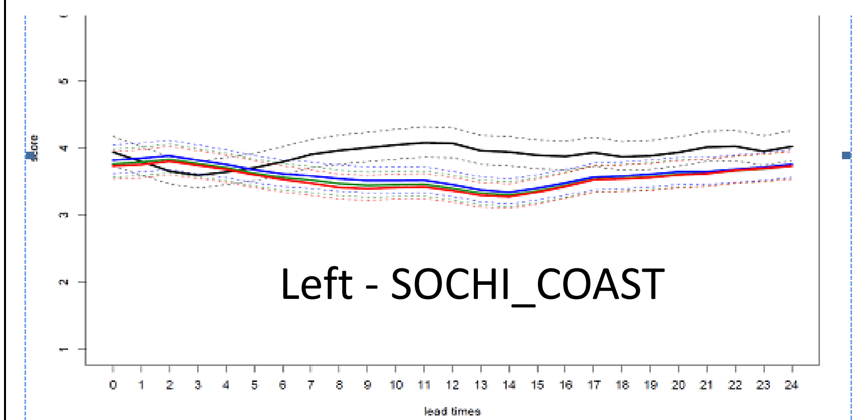
Different interpolation methods give similar results, on average

COSMO mean forecasts from three methods (nearest point, MIN, MAX) and mean observations (black) (first test period)

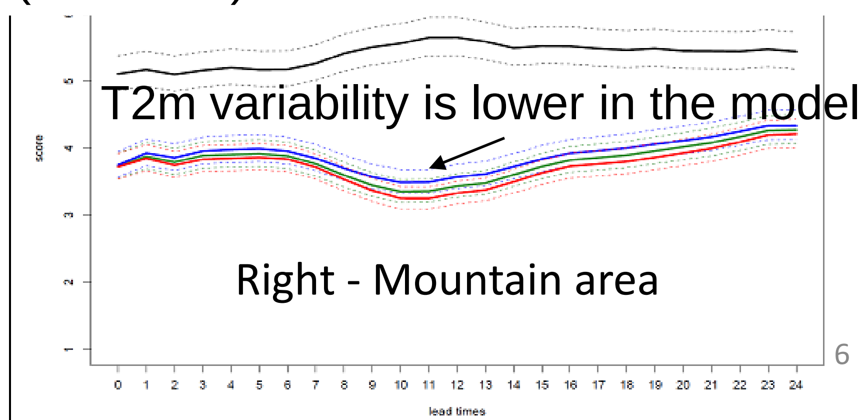




Standard deviations of forecasts from three methods (nearest point, MIN, MAX) and standard deviation of observations (black line)



Left - SOCHI_COAST



Right - Mountain area

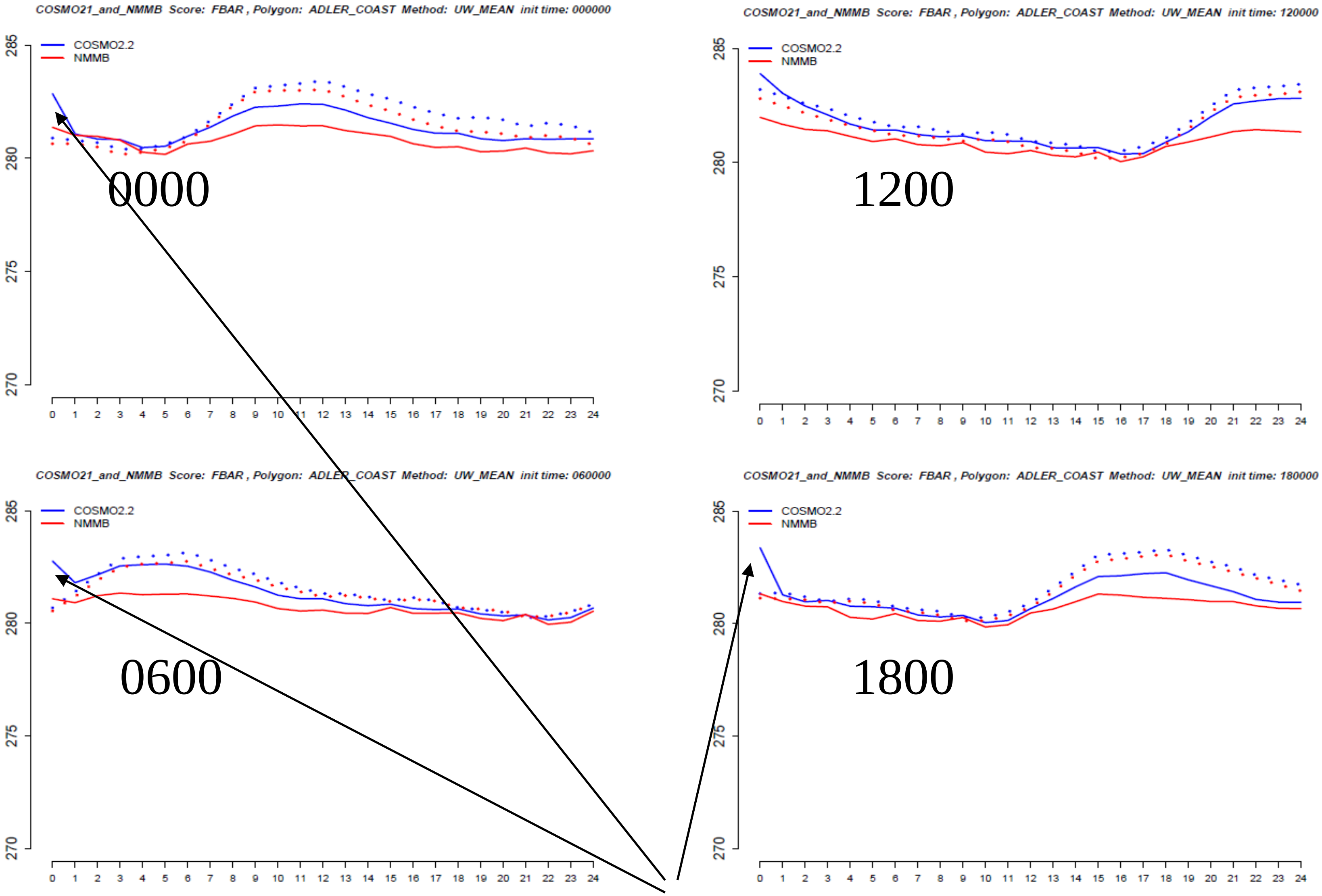
Models

- **2.2-km South region COSMO version** with 40 levels and explicit deep convection calculation (initial and boundary fields from 7-km COSMO-RU) ***interpolated to 1*1-km regular grid using FieldExtra***
- **American 1-km NMMB model**
- Forecast period 24 h, 1-h lead-time step
- 4 initial times (00, 06, 12, 18)

T2m (°K) forecast and observation (dotted) means, COSMO blue, NMMB red

Sochi coast

2nd test period



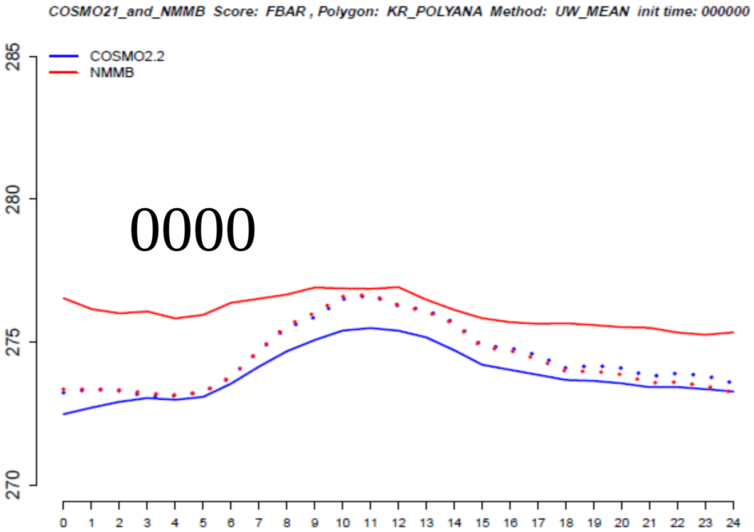
In the coastal polygons, there is a **systematic COSMO error at the initial time** that is likely due to the initial field. It is not detected in the mountain cluster.

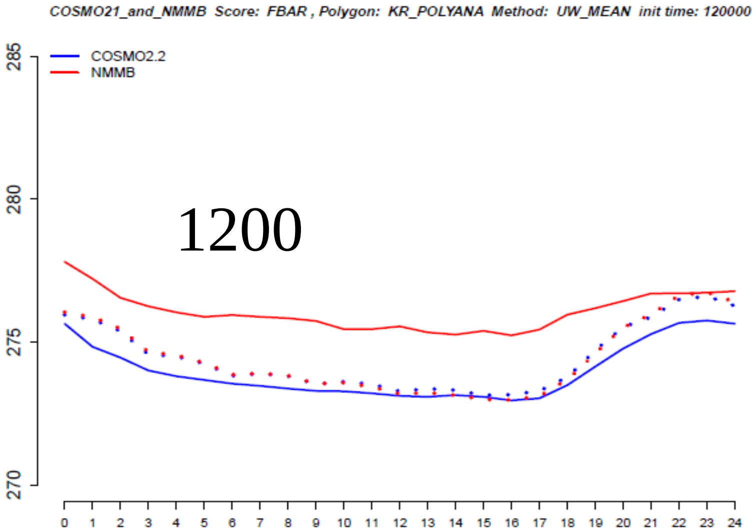
T2m (°K) forecast and observation (dotted) means

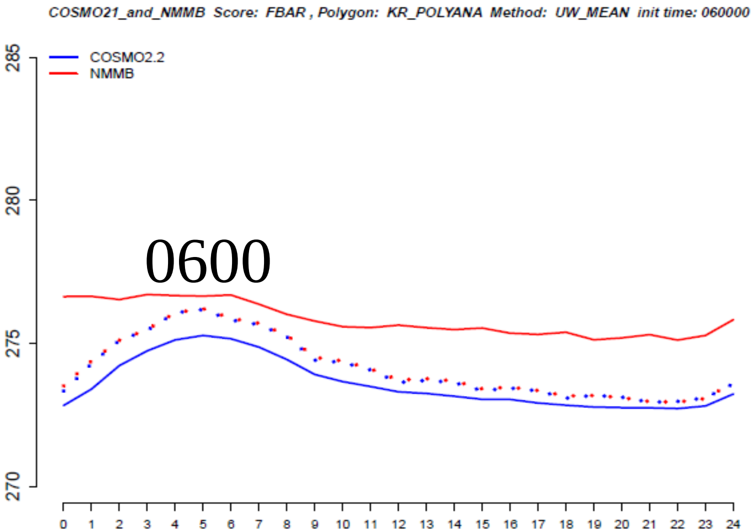
COSMO blue, NMMB red

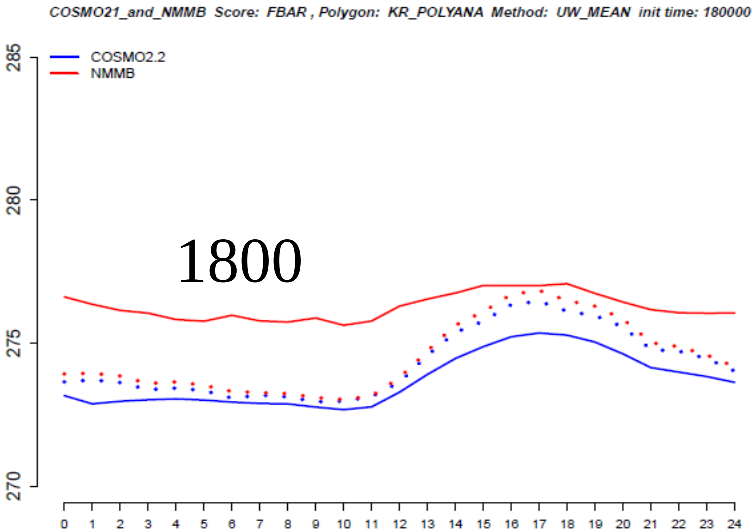
Mountain cluster

2nd test period







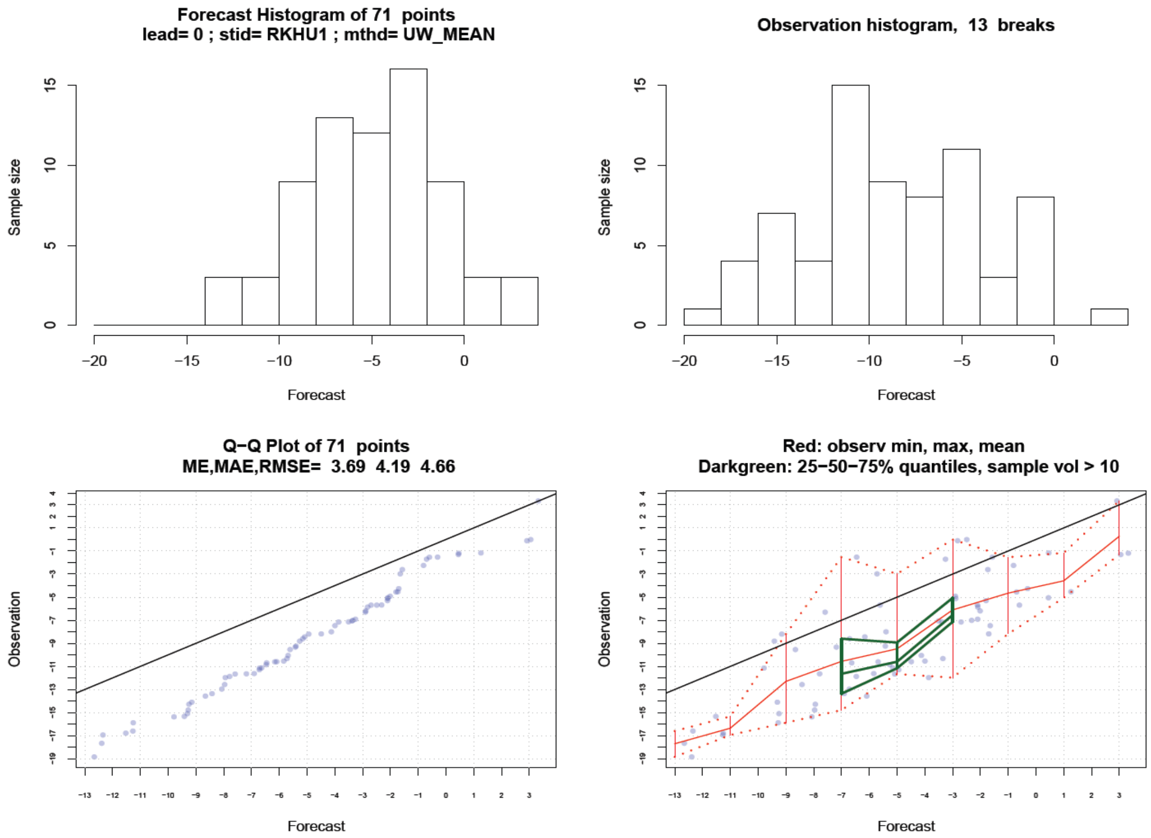


COSMO yields better T2m means and the diurnal cycle, **especially in the mountain cluster**

Diagnostic station-based verification

- “Diagnostic” in the sense that it focuses on the fundamental characteristics of the forecasts, the corresponding observations, and their relationships (A.Murphy,B.Brown,Y.Chen, 1989).
- “Station portraits” are made for each variable, station, lead time, and method (only for COSMO yet).
- They give the possibility to <u>calibrate the forecasts in the whole variable range including the distribution tails</u>, that is, extreme values important for decision making about the competitions;
- show the sample size in different categories.
- The interquartile range values are inversely related to forecast accuracy.

Station “portraits”. Here for T2m RKHU1 station (on the Aibga ridge), nearest point, lead 00 h.



Calibration, p(o|f), defined by the main statistics: conditional means, min-max, quartiles, and medians. Green lines denote the bin sample volume of no less than 10 pairs (sample stability).

Calibration implies a shift of the frc mean-median to the diagonal.
The T2m area outside the green strip indicates sample instability (calibration uncertainty) due to the small data volume.

Importance of the above diagnostic verification for “critical thresholds” that are crucial for decision-making (distribution tails and small samples)

Evaluation of COSMO models in the lower PBL

In the framework of SRNWP data Exchange project available on COSMO web site)) are now available a large set of data from selected station in (Europe for some special parameters like radiation fluxes and soil moisture. It would be interesting and important to use this set of data to verify the PBL surface of our COSMO implementations. VERSUS has been updated to upload some of the information contained in these ASCII files: Radiation (components, budgets) and Fluxes.

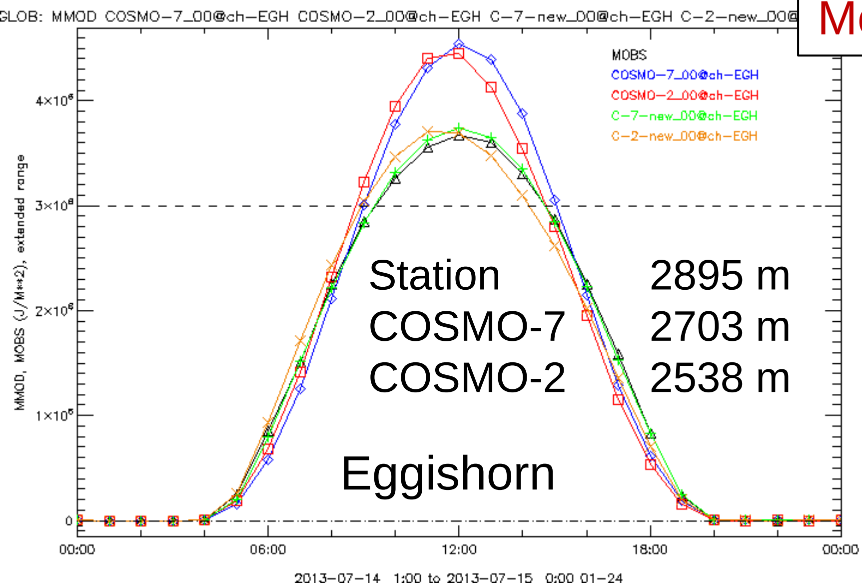
Verification of Global Radiation With Hourly Measurements Over Switzerland

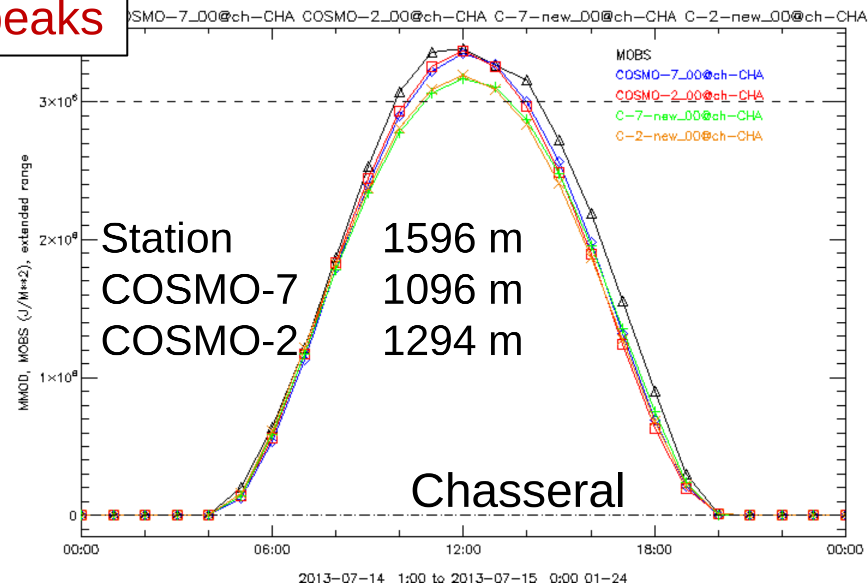
Global Radiation from the Model

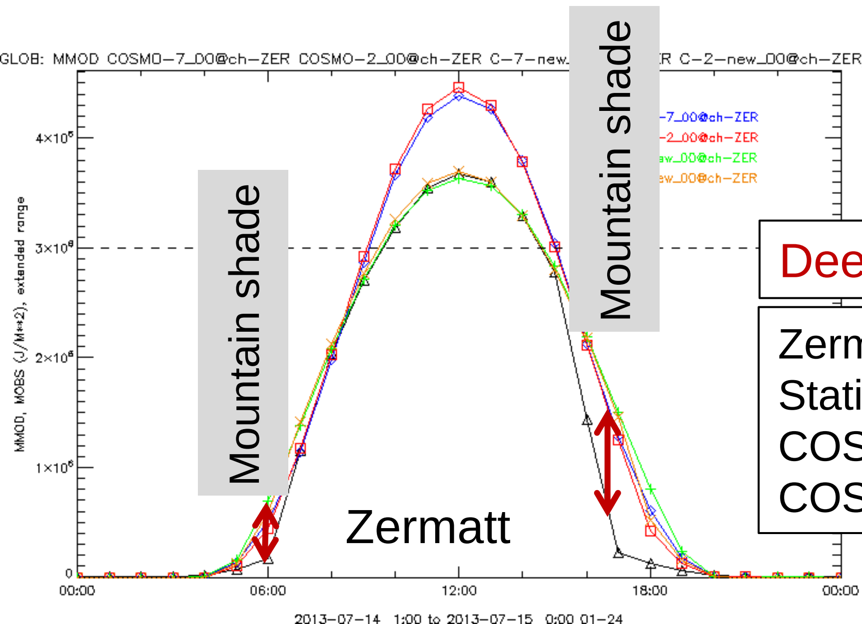
- **Old** approximation (e.g. "Beschreibung des COSMO-DE-EPS und seiner Ausgabe in die Datenbanken des DWD", 2012)
  - GLOB = ASOB_S / (1 - ALB_RAD)
  - Caveats:
    - ALB_RAD is the albedo for the diffuse radiation only
    - ALB_RAD is an instantaneous value, ASOB_S an accumulated value → inconsistency
- **New** output available since about 2 years (but not yet documented): Sum of output parameters
  - GLOB = ASWDIR_S + ASWDIFD_S

14 July 2013

Mountain peaks





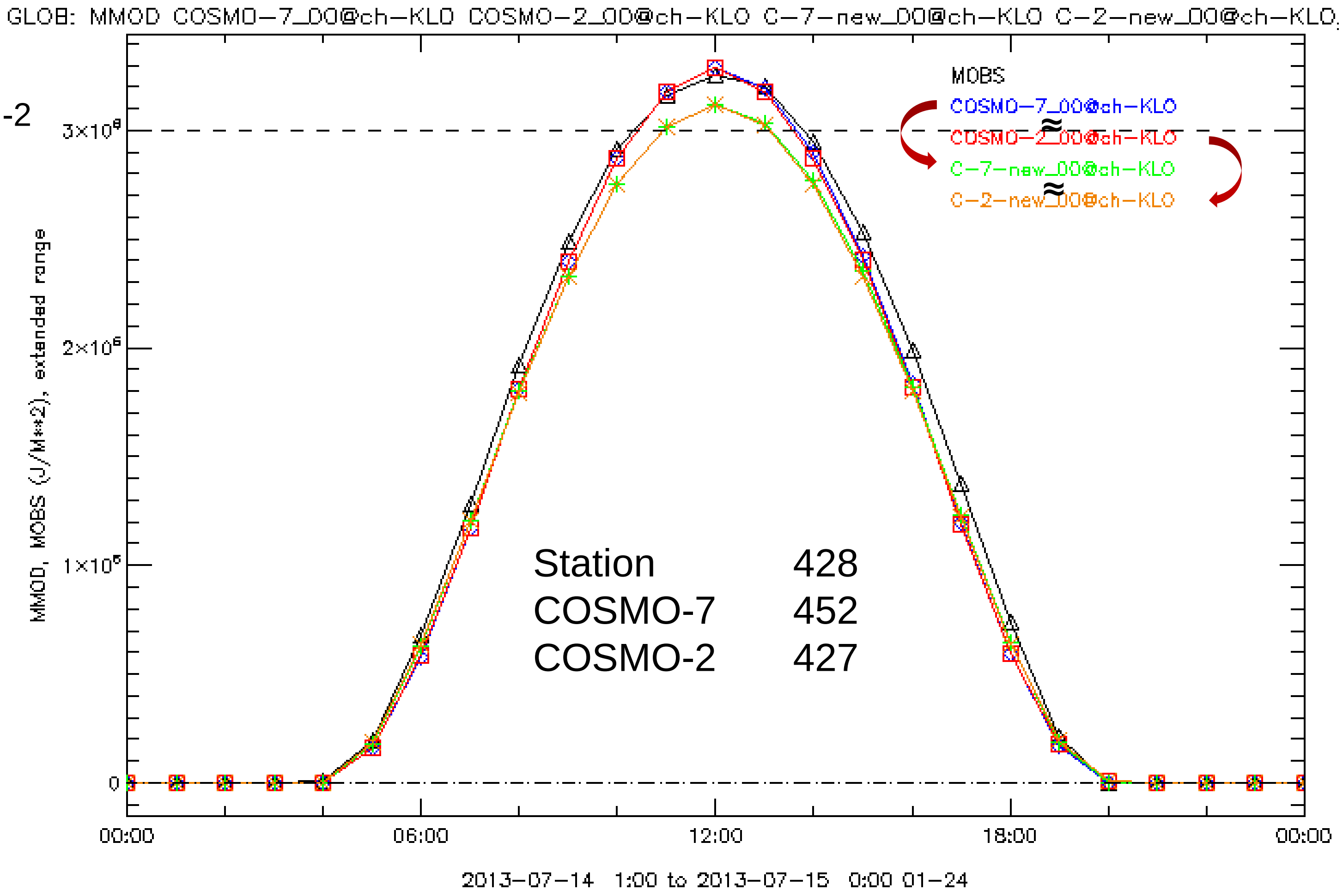


Deep narrow valley

Zermatt	
Station	1640 m
COSMO-7	2719 m
COSMO-2	2077 m

14 July 2013 Zürich-Kloten Representative for Swiss Plateau

833 W m^{-2}

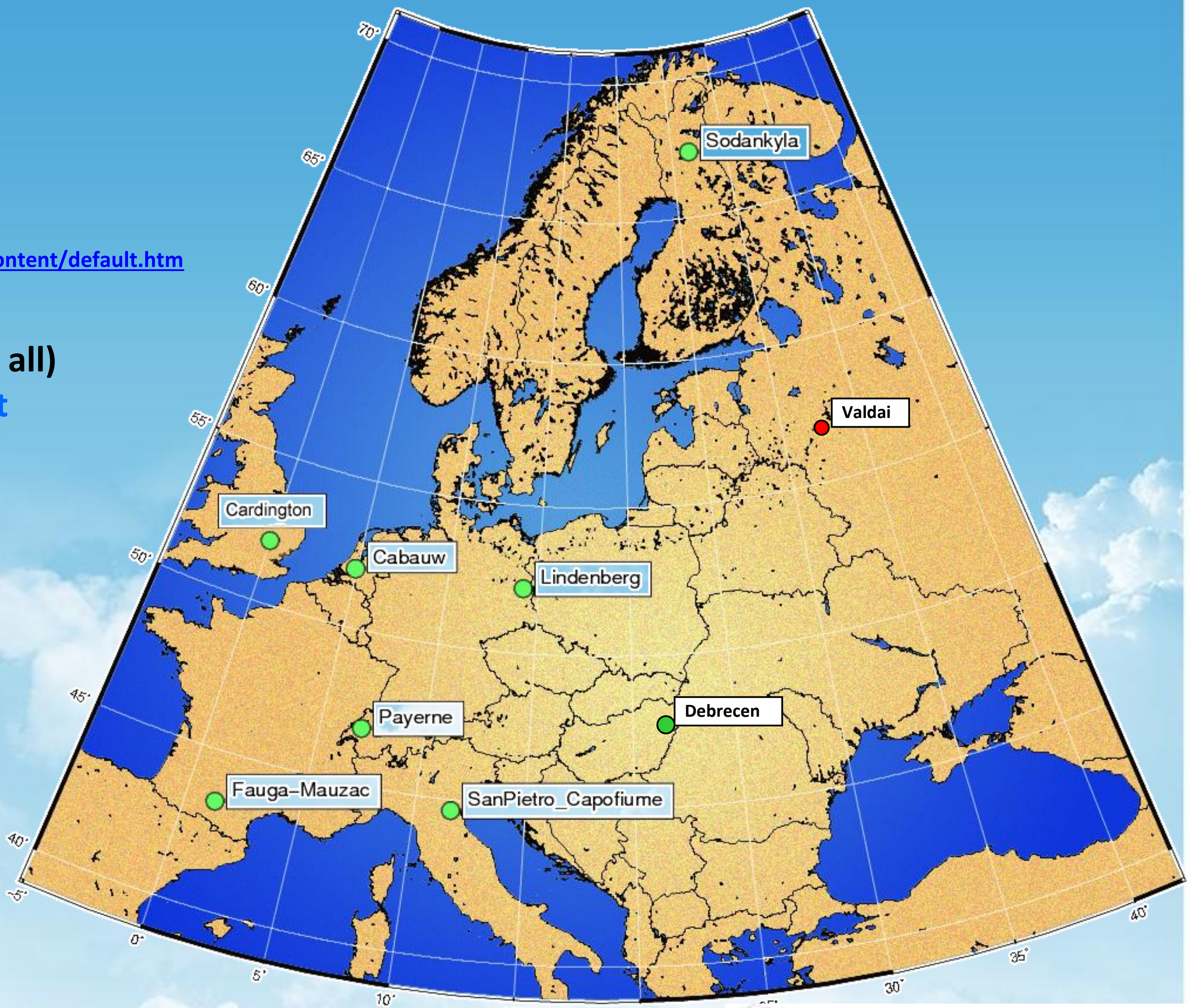


Suggestion

- Proposition to promote the inclusion of hourly accumulated values of global radiation in the international BUFR and SYNOP data exchange over GTS.
- Global radiation is important for photovoltaic solar power plants and a good integral value of the transparency of the atmosphere and occultation by clouds
- Global radiation is an automatic measurement, more widely spread than the manual cloud observation.

Experience with SRNWP data pool PBL data in VERSUS

- Access from COSMO web, password protected
  http://www.cosmo-model.org/srnwp/content/default.htm
- Currently **8 sites**, data from **2006-2012** **(not all)** in a **common ASCII format**
- **Soil**, **surface** and **BL** observations



ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

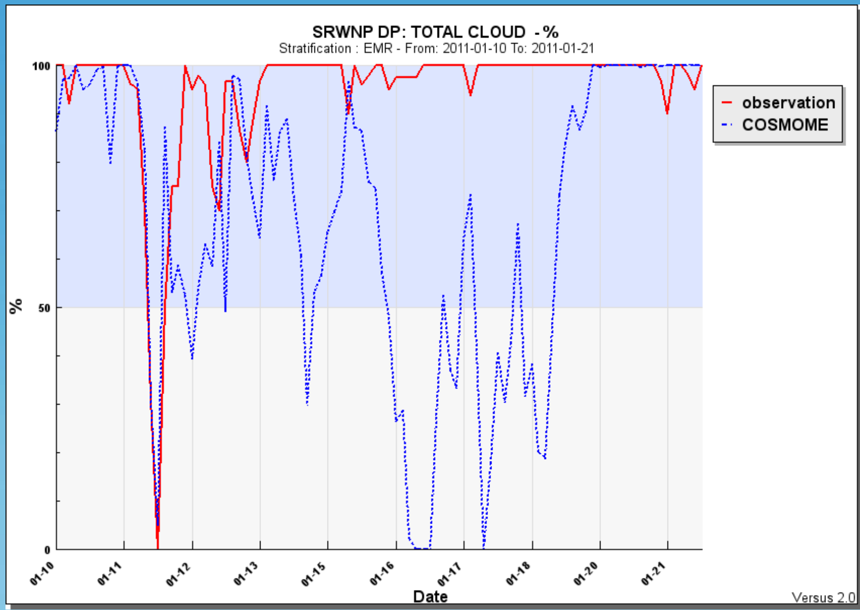
Obs and Fcs data availability

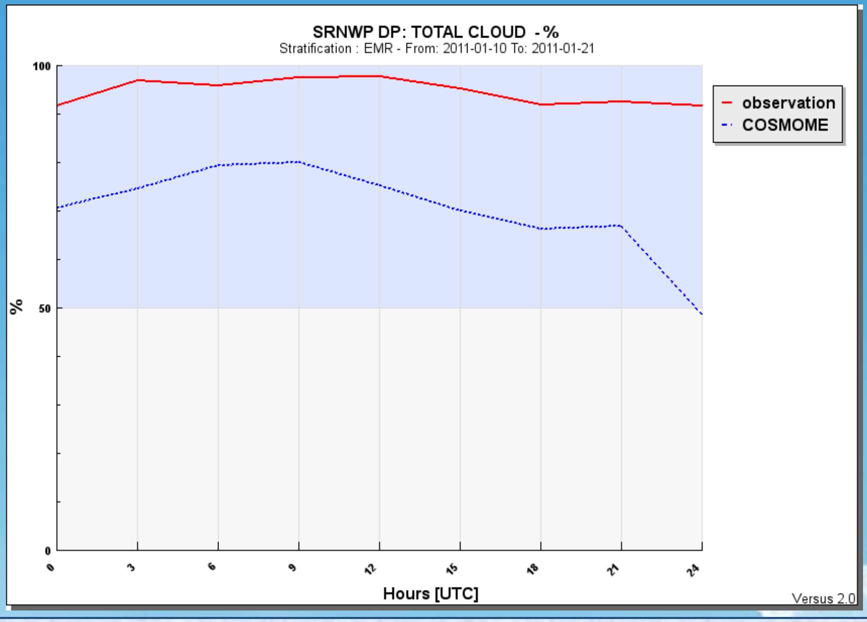
OBS data	FCS data
RSWD: incoming solar radiation	ASWDIR_S aver. direct downward Sw rad. surface
RSWU: reflected solar radiation	ASWDIFD_S aver. diffuse downward Sw rad. Surface
	Avg. Balance of SW
RLWD: incoming thermal radiation	ALWD_S aver. downward Lw radiation at the surface
RLWU: outgoing thermal radiation	ALWU_S averaged upward Lw radiation at the surface
	Avg. Balance of LW
HS: sensible heat flux	Ashfl_s: averaged sensible heat flux
LE: latent heat flux	Alhfl_s: averaged latent heat flux

Balance of SW and LW for obs is internally calculated and stored

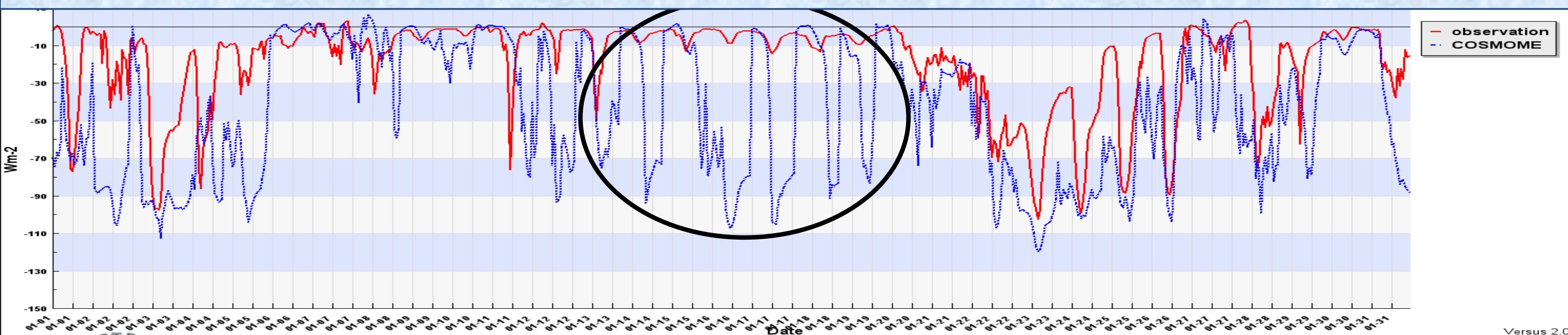
ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

Something interesting





Model predicted wrong values for LW balance but also a completely wrong TCC (much less than reality)



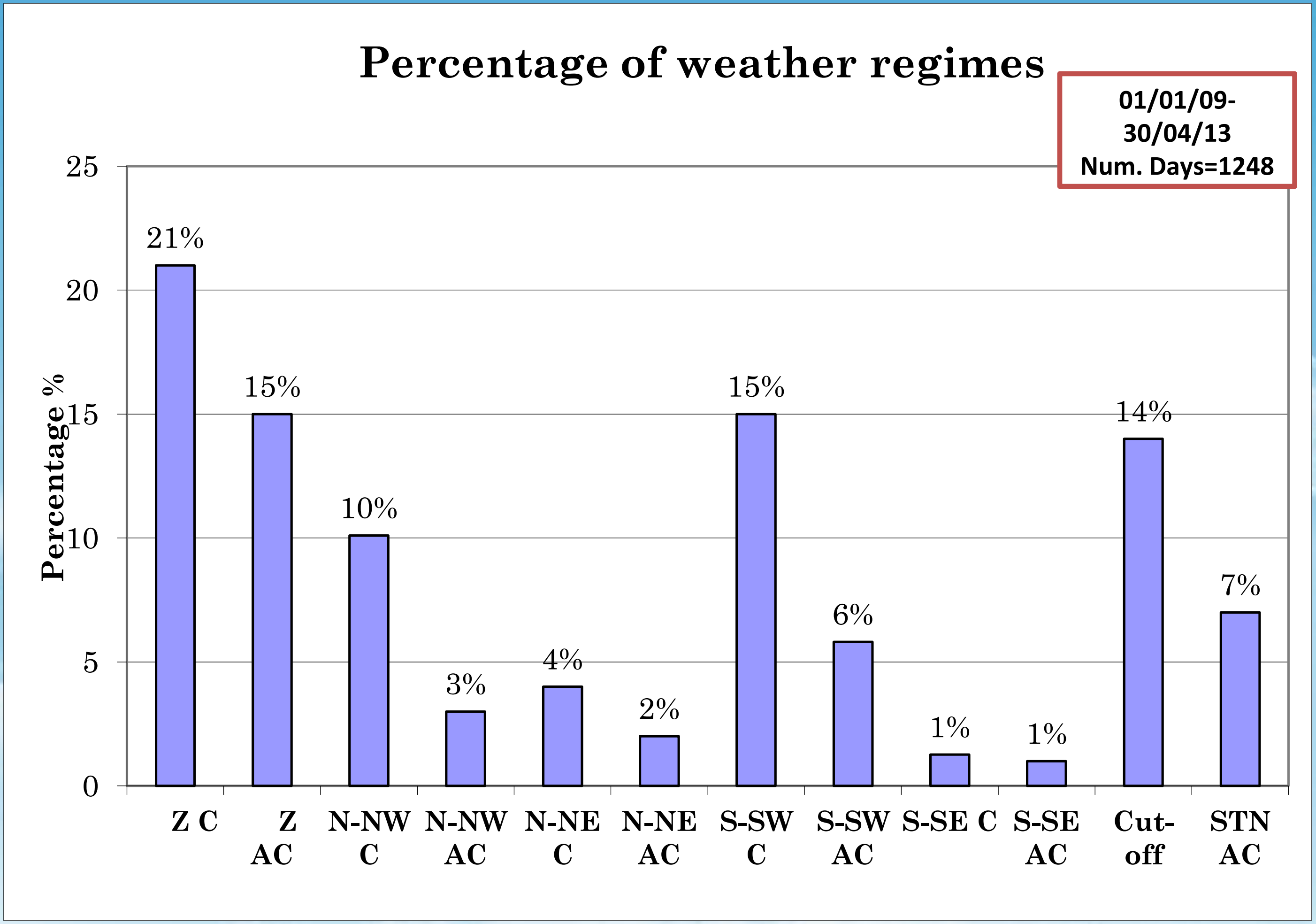
SERVIZIO METEOROLOGICO DELL'AERONAUTICA

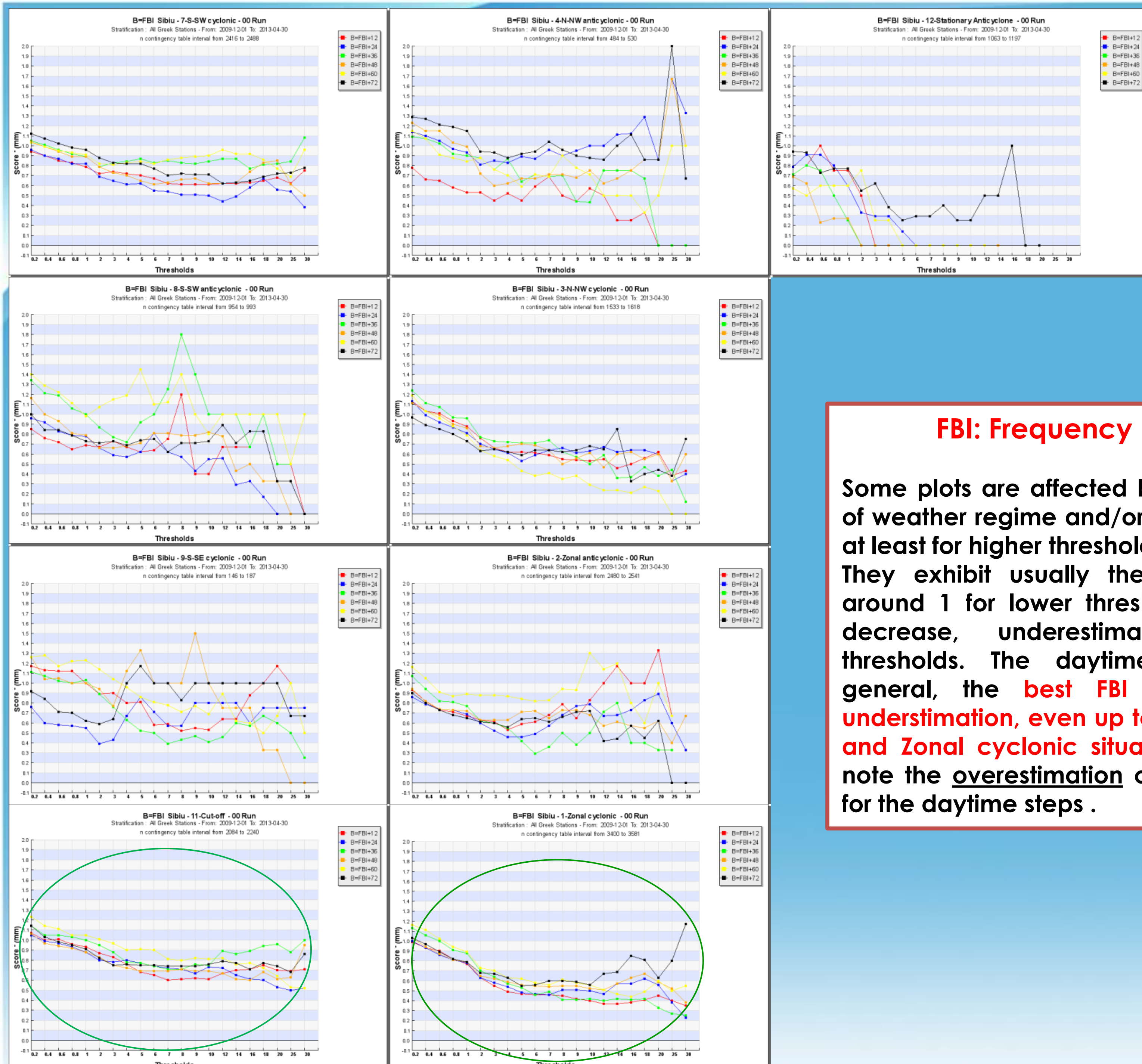
ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

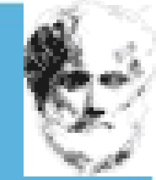
Weather Defined Verification

Weather Classification: 01/09/2009-30/04/2013=1248days

1	Zonal cyclonic
2	Zonal anticyclonic
3	N-NW cyclonic
4	N-NW anticyclonic
5	N-NE cyclonic
6	N-NE anticyclonic
7	S-SW cyclonic
8	S-SW anticyclonic
9	S-SE cyclonic
10	S-SE anticyclonic
11	Cut-off
12	Stationary Anticyclone



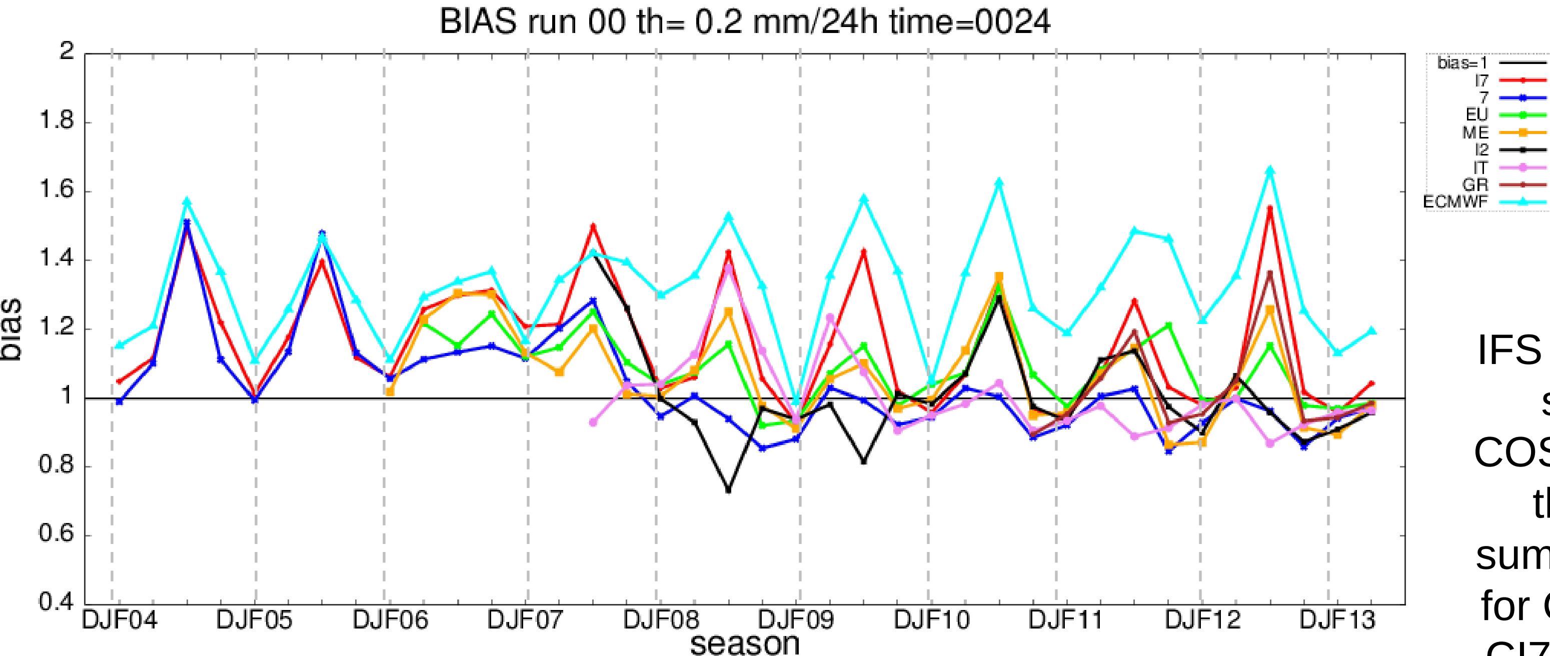


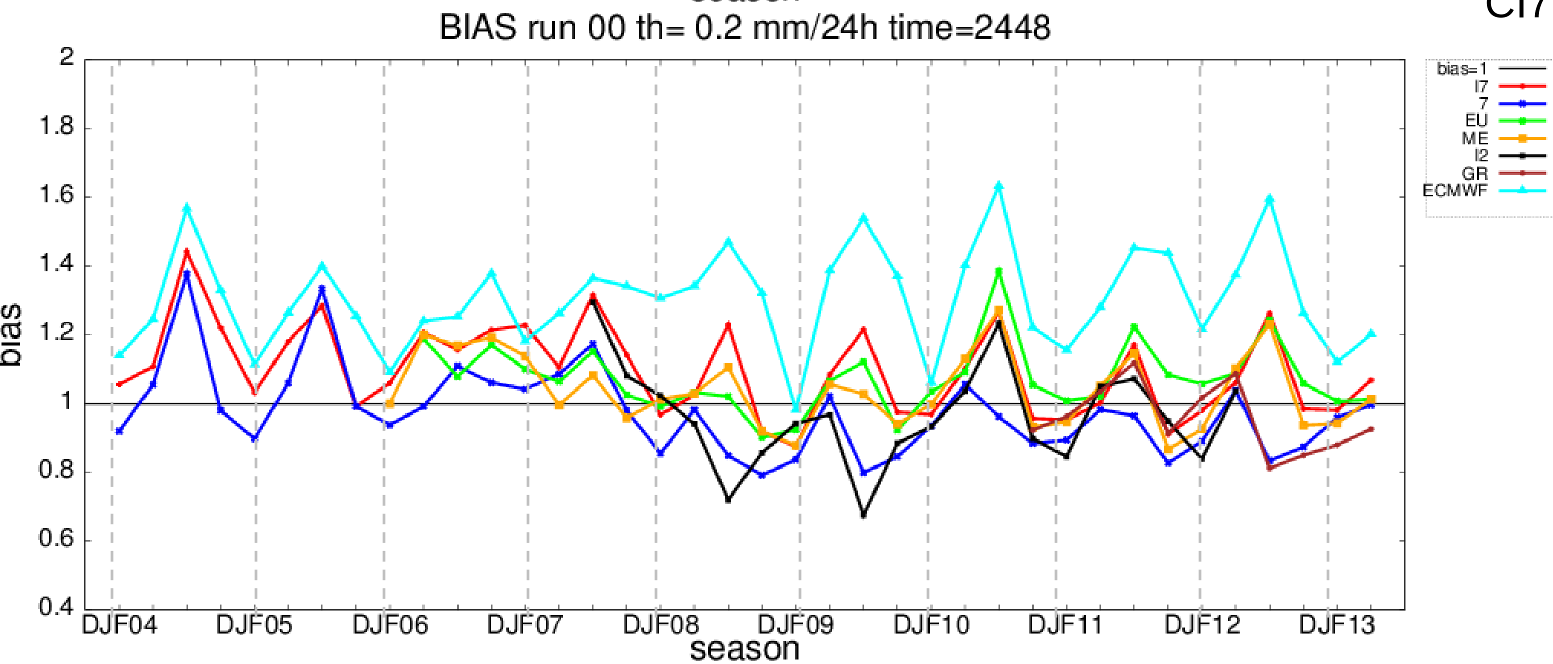


FBI: Frequency Bias Index

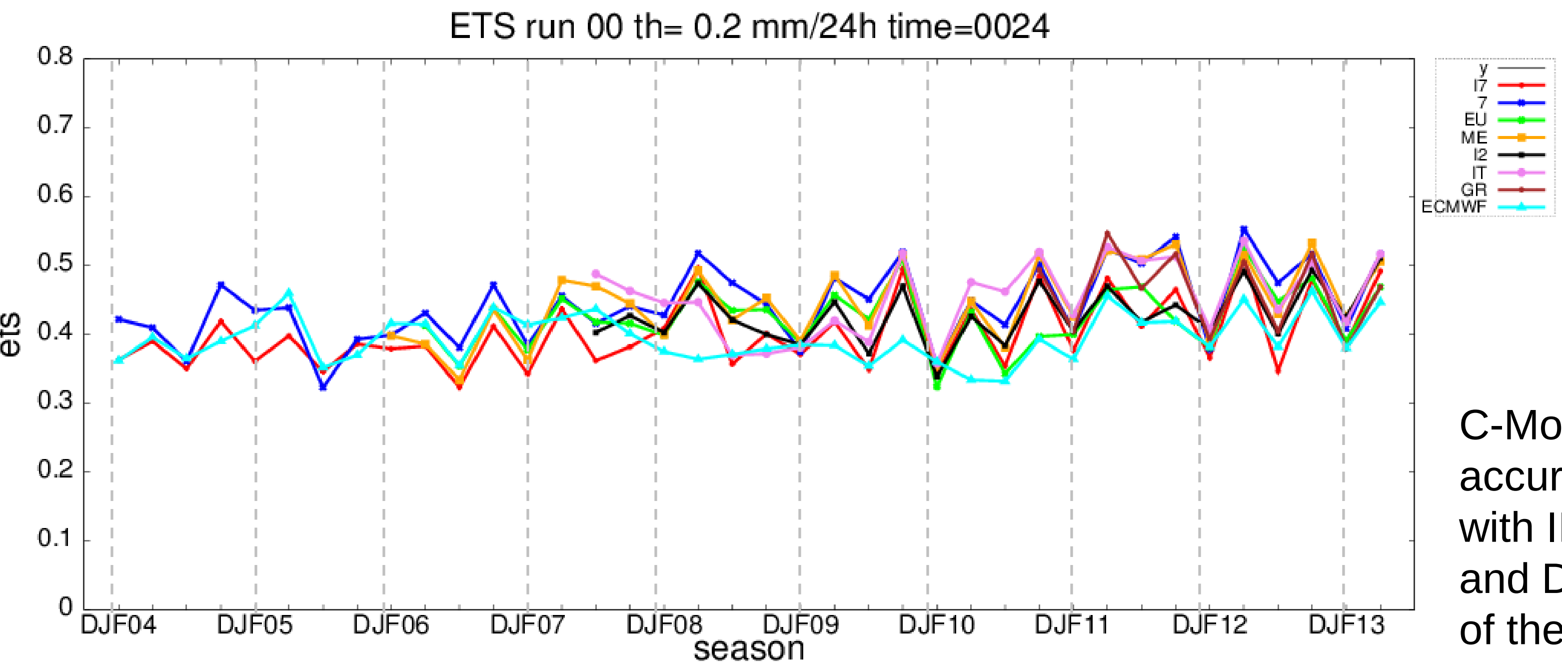
Some plots are affected by the poor sample of weather regime and/or precipitation event at least for higher thresholds.
They exhibit usually the tendency for FBI around 1 for lower thresholds that tends to decrease, underestimating the higher thresholds. The daytime steps show, in general, the best FBI in terms of less understimation, even up to +72h is for Cut-Off and Zonal cyclonic situations. It is worth to note the overestimation of rain/norain cases for the daytime steps .

Long Term Trends

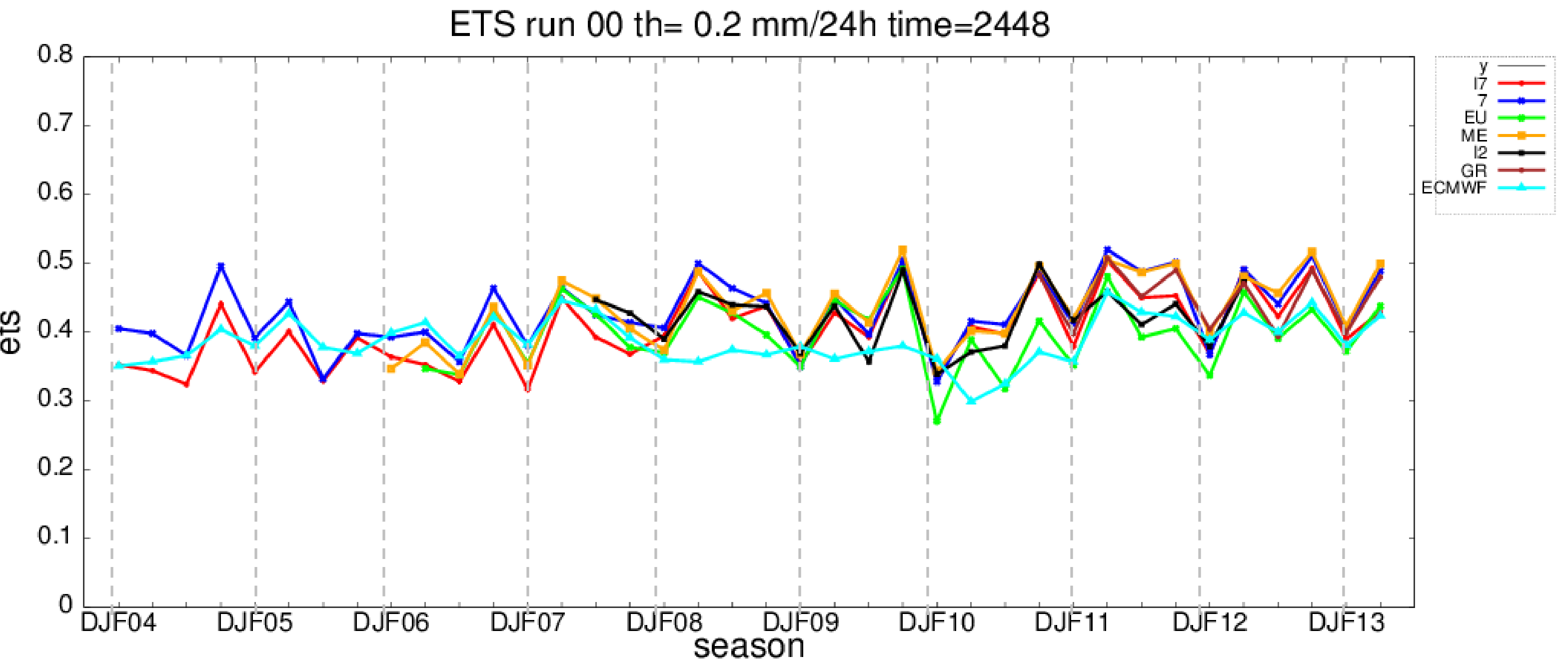


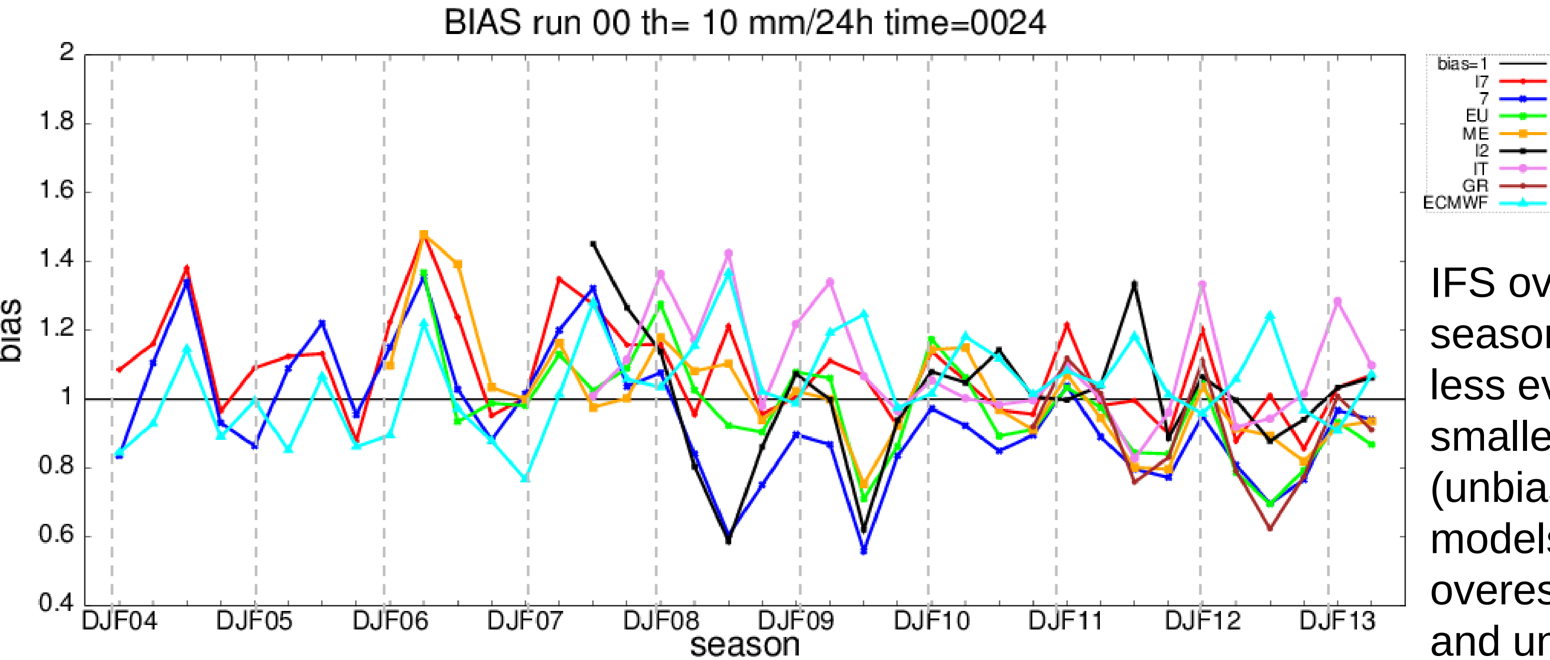


IFS Overestimates all seasons, while COSMO models keep this tendency in summer, eg JJA 2012 for CEU, CME, CGR, CI7. Less evident for D+2

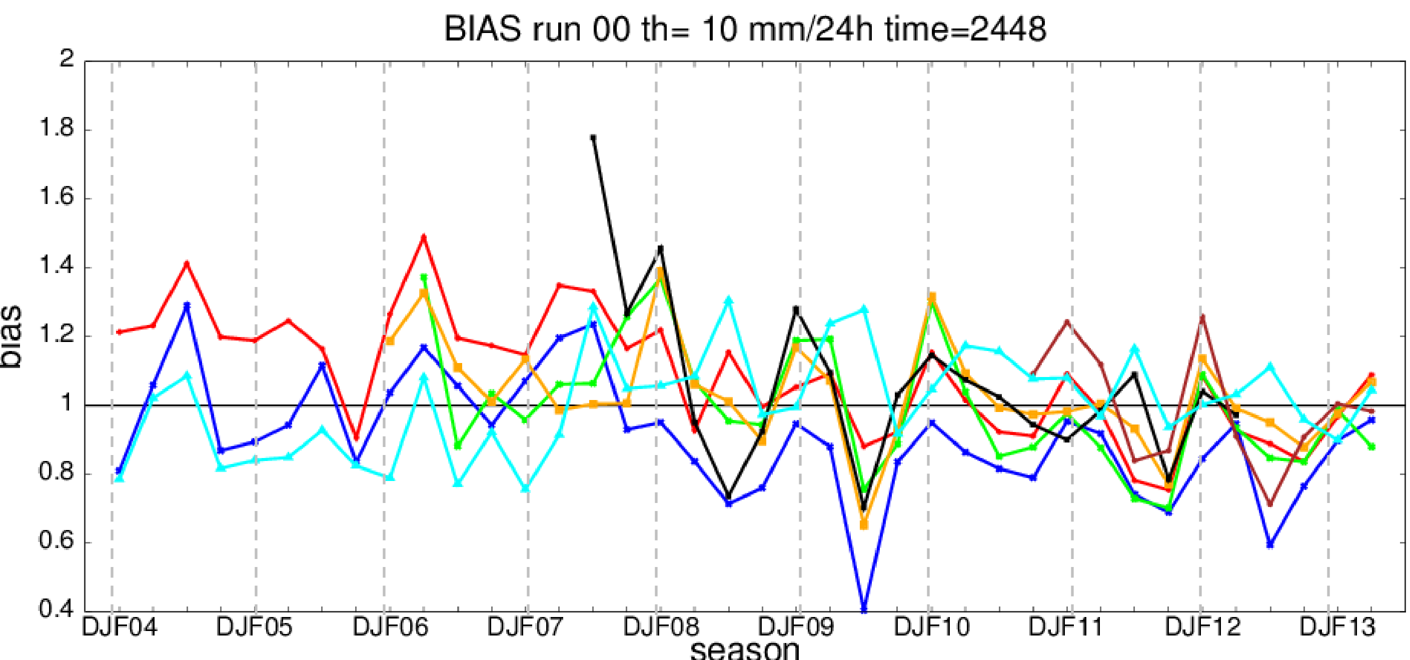


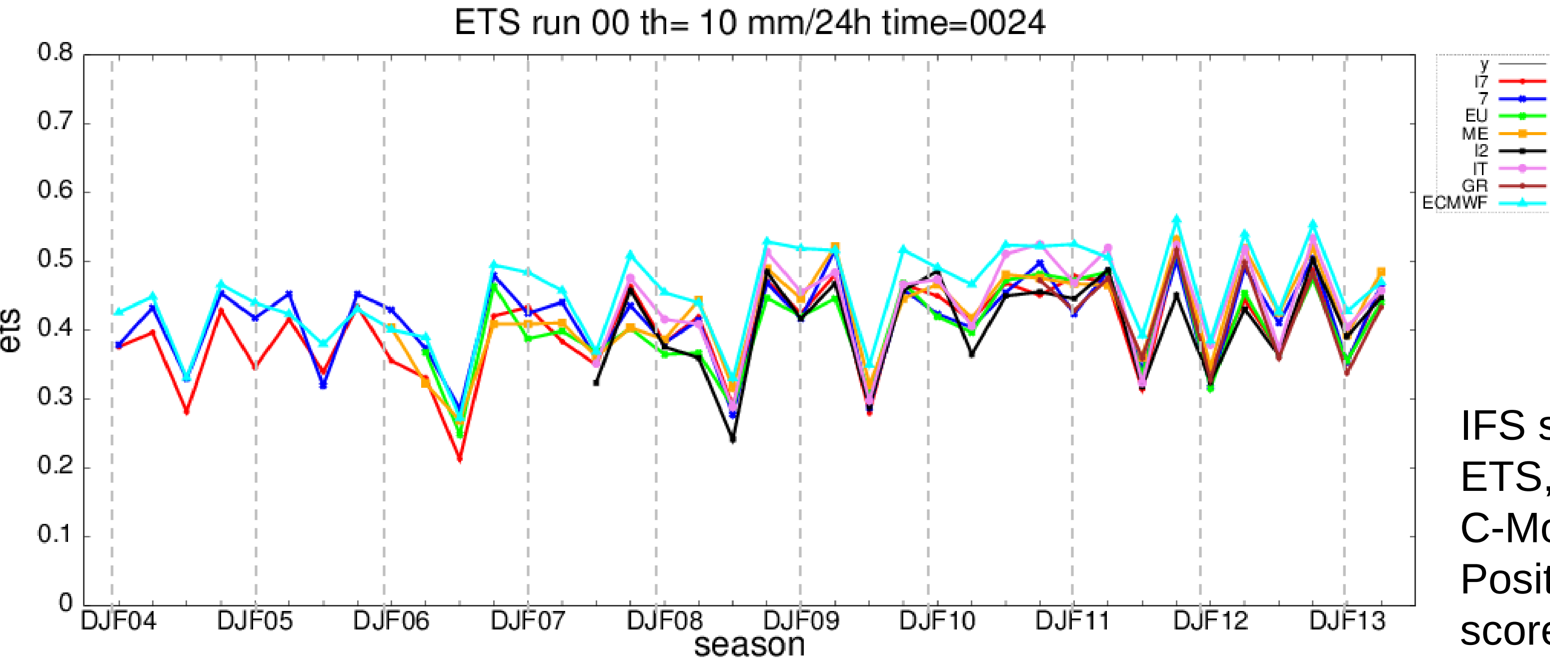
C-Models show higher accuracy compared with IFS for both D1 and D2. Positive trend of the score.



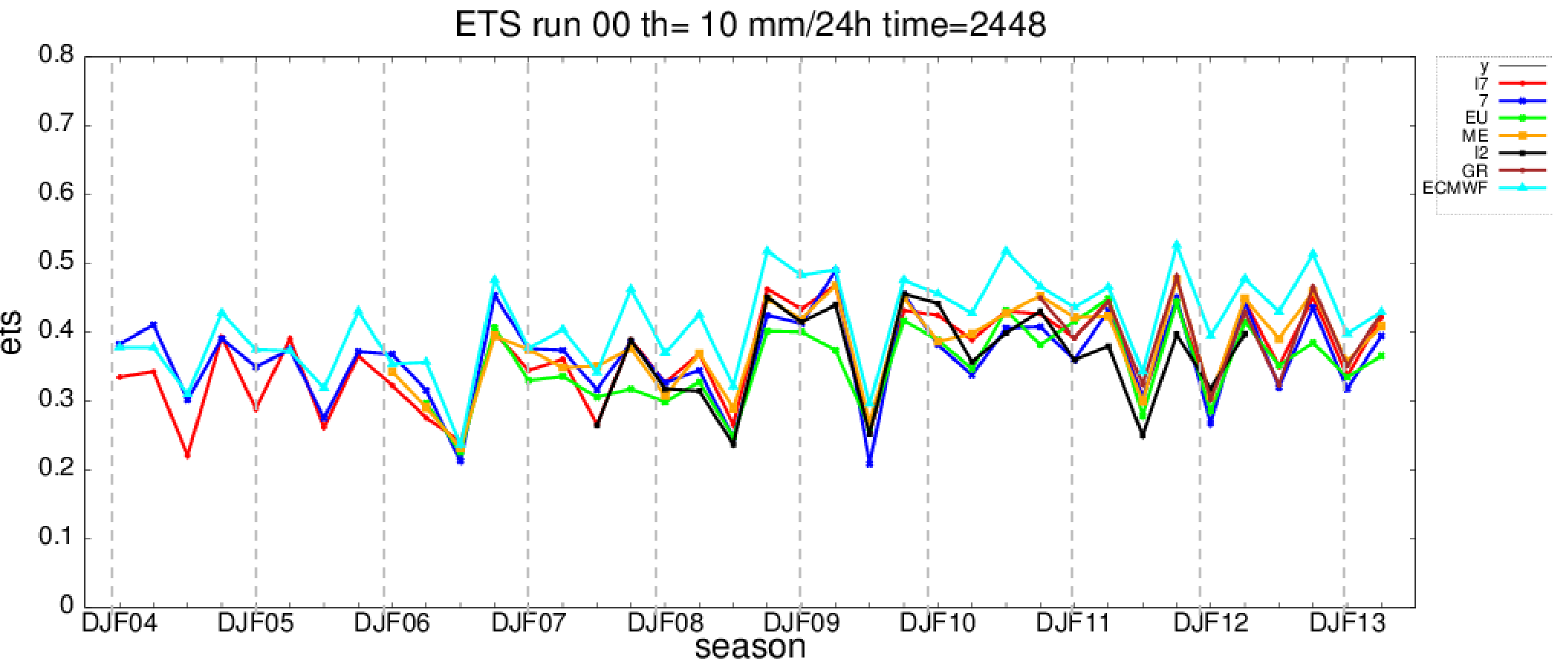


IFS overestimates all seasons in D1, but it is less evident than smaller threshold (unbiased in D2). C-models tend to overestimate in DJF and underestimate in JJA (more), but not all of them. General tendency for C7 to underestimate the event for all the thresholds





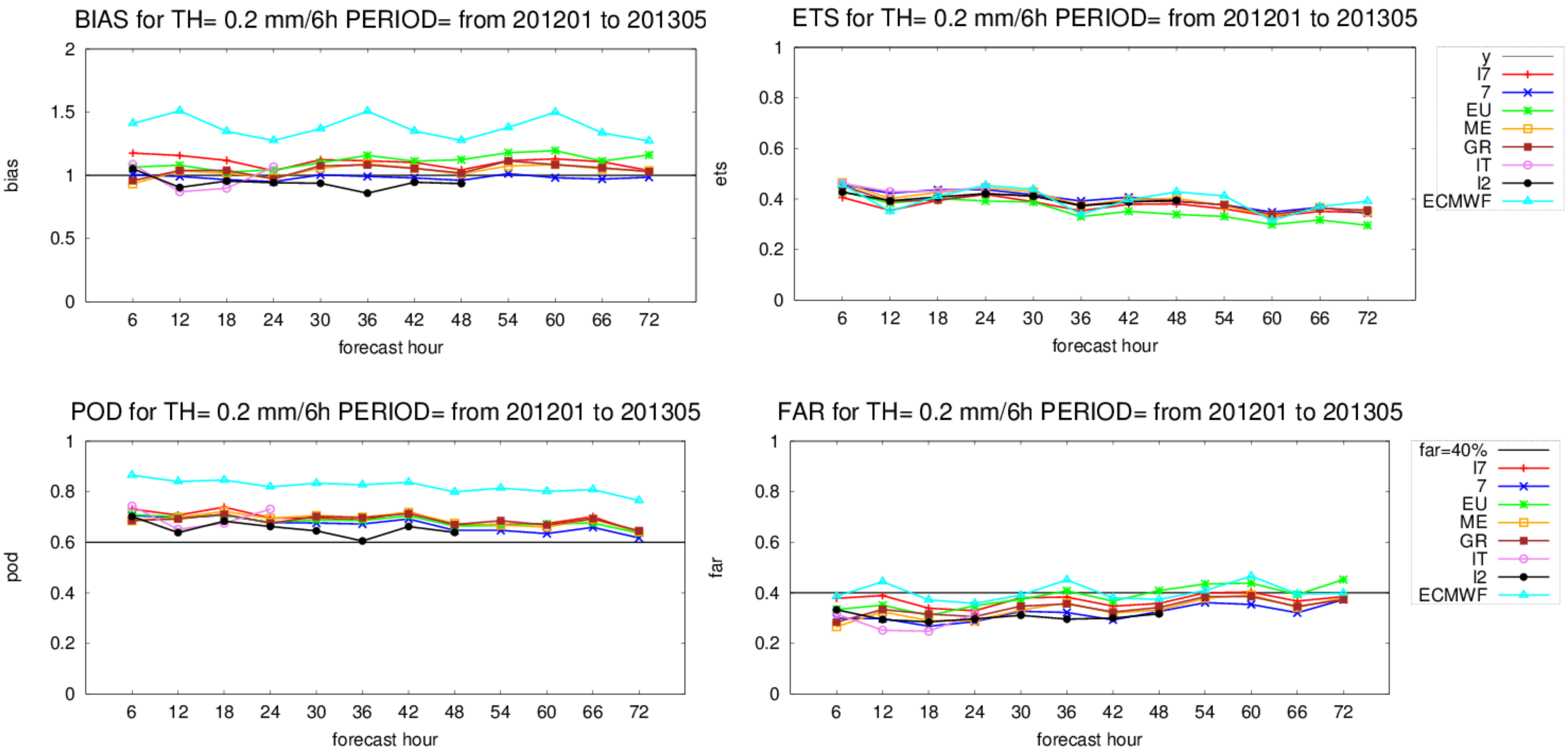
IFS shows higher ETS, but very close to C-Models.
Positive trend of the score.



6h cumulated precipitation average over areas: 201201-201305

Rain/NoRain case 201201 - 201305

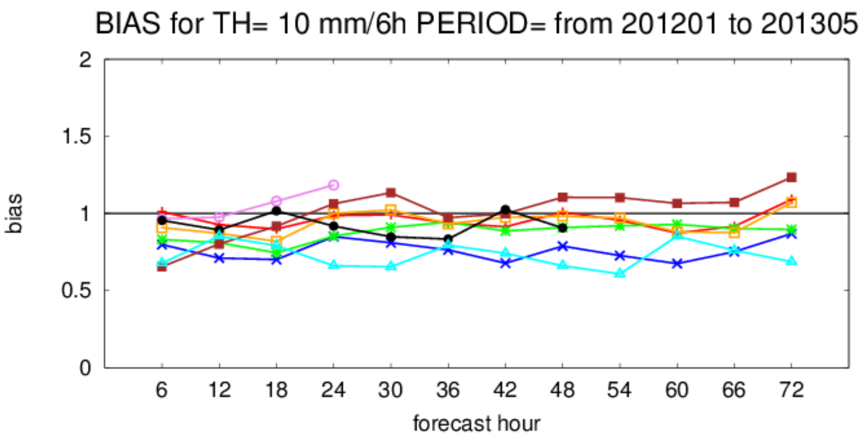
Overestimation for IFS -> higher POD. Low Bias for C-Models

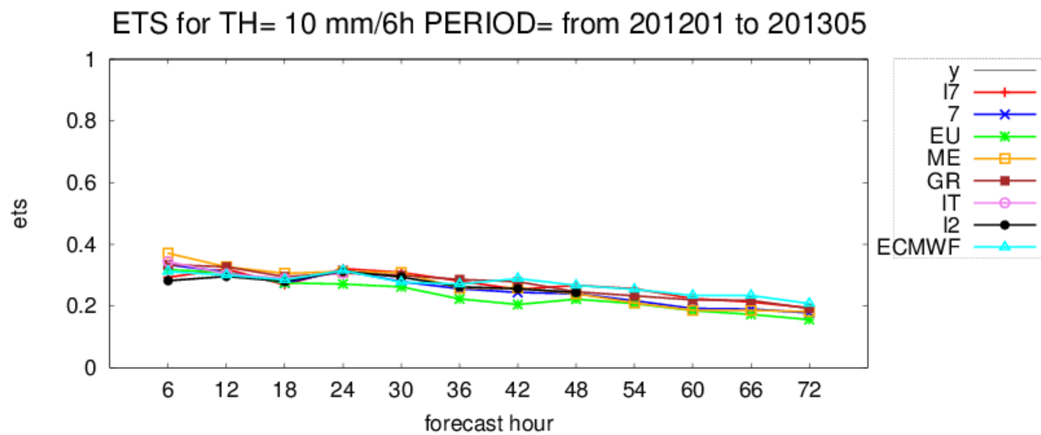


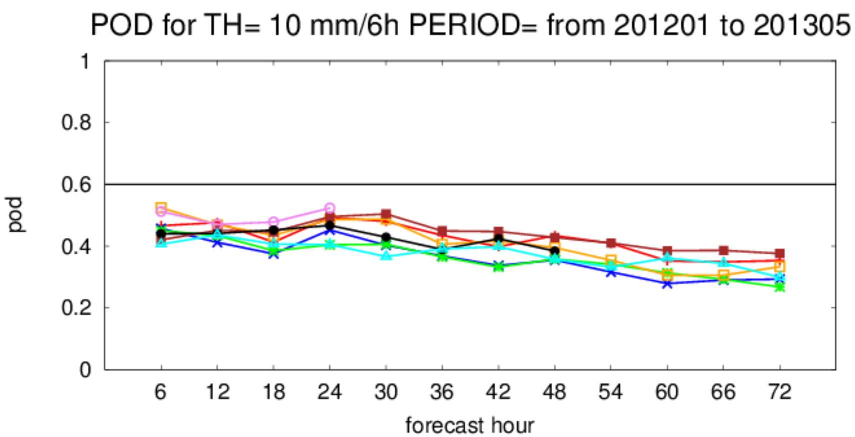
6h cumulated precipitation average over areas: 201201-201305

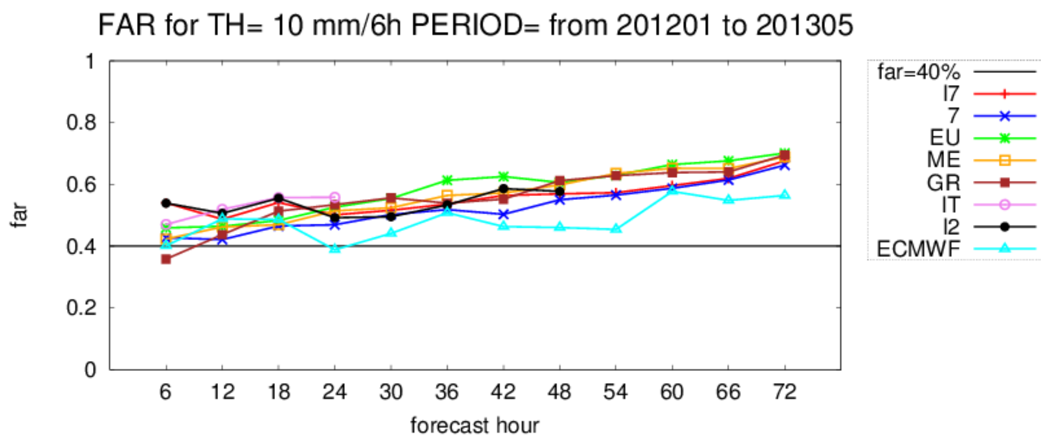
10mm/6h case 201201 - 201305

Underestimation for IFS -> Low POD. ETS now comparable with C-Models, but also low FAR. C-7 (also CEU) underestimates all the fcs steps







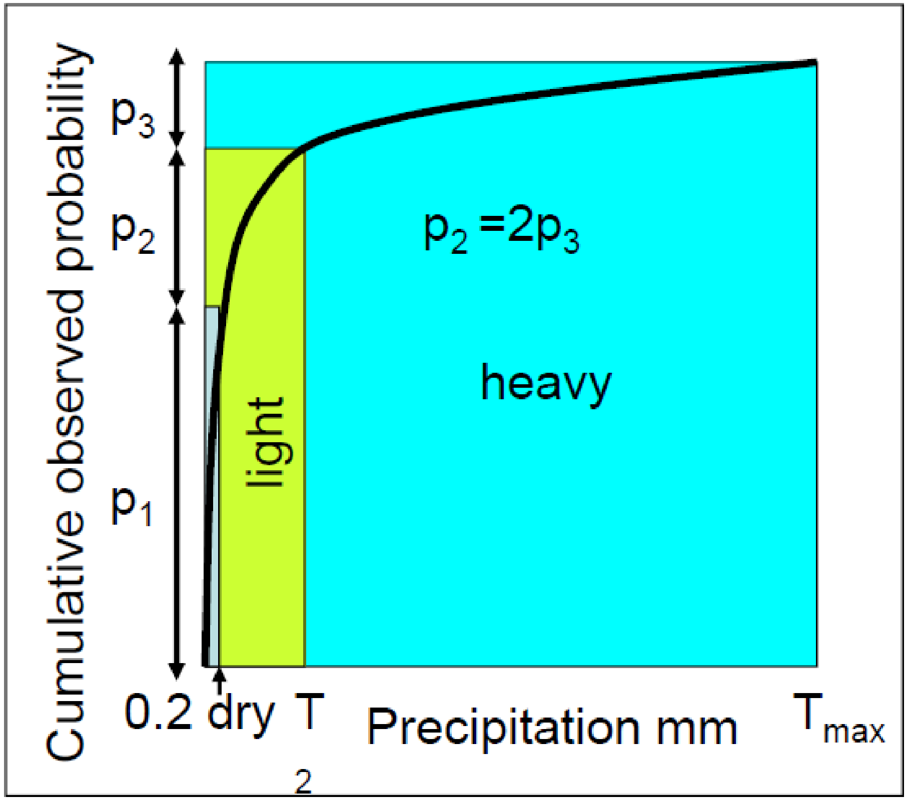


SEEPS precipitation score

SEEPS=Stable Equitable Error in Probability Space

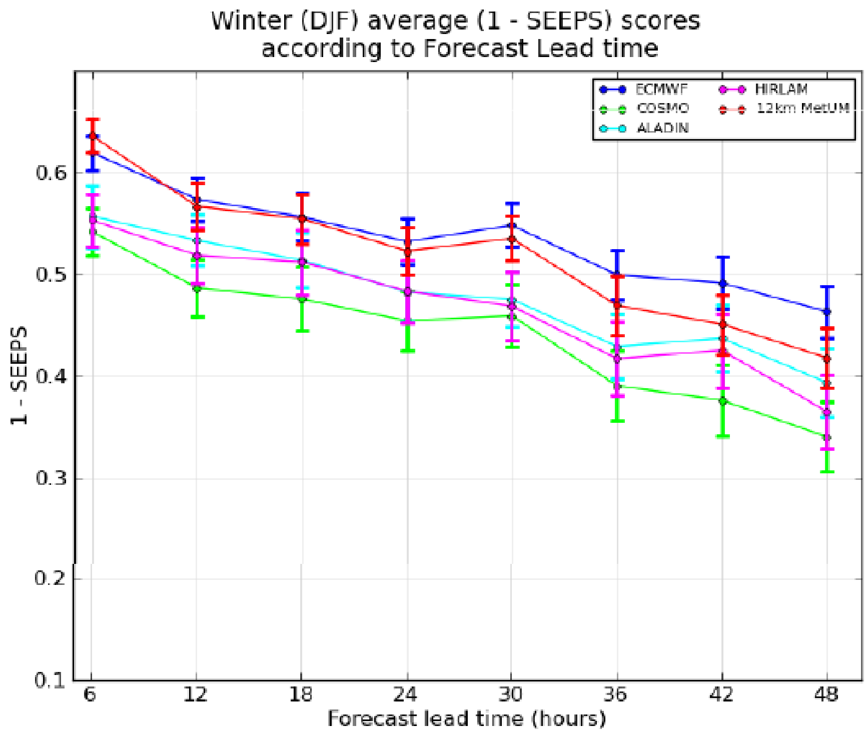
Rodwell et al, 2010, QJRMS 136

- Dry, light , heavy based on observed climatology (24h) at station – p_1 , p_2 , p_3
- Contingency table probabilities based on these categories
- Scoring matrix – stable, equitable
    - SEEPS=0 (perfect) , =1 (no skill - , eg constant)
- Now applying to 6h accumulations in SRNWP-V
    - 6h climatology (courtesy Mark Rodwell)



Winter (DJF) 2009-2012 data
$3^2/_3$ winters

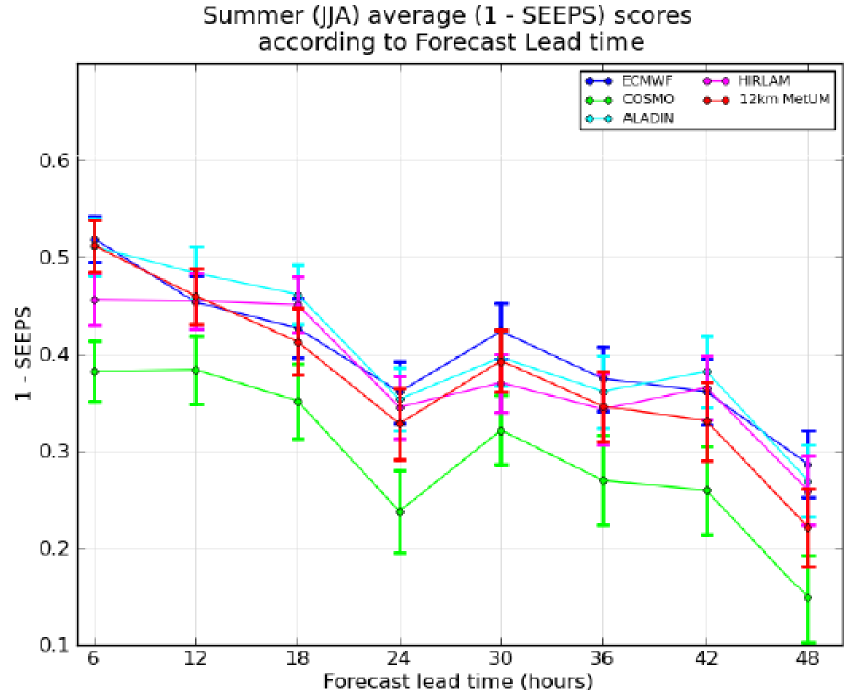
- Higher skill than summer
- EC best at longer range
- 3 groupings
  - EC/UM
  - Aladin/Hirlam
  - COSMO



Error bars 70% confidence intervals

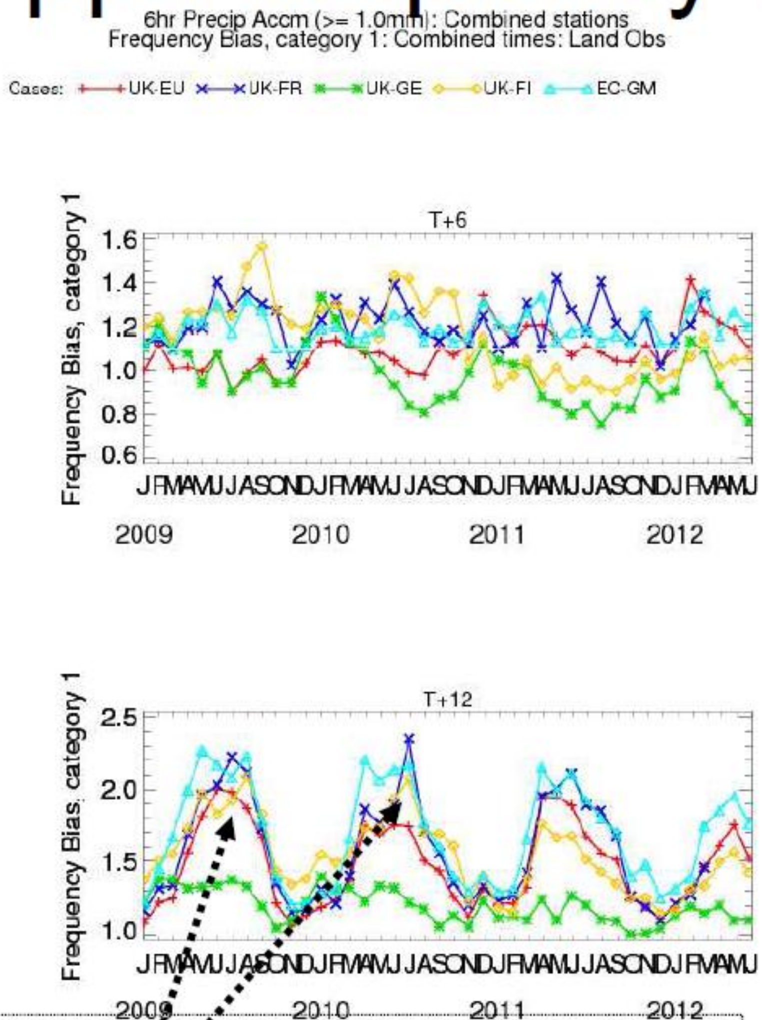
Summer (JJA) 2009-2011 data
3 summers

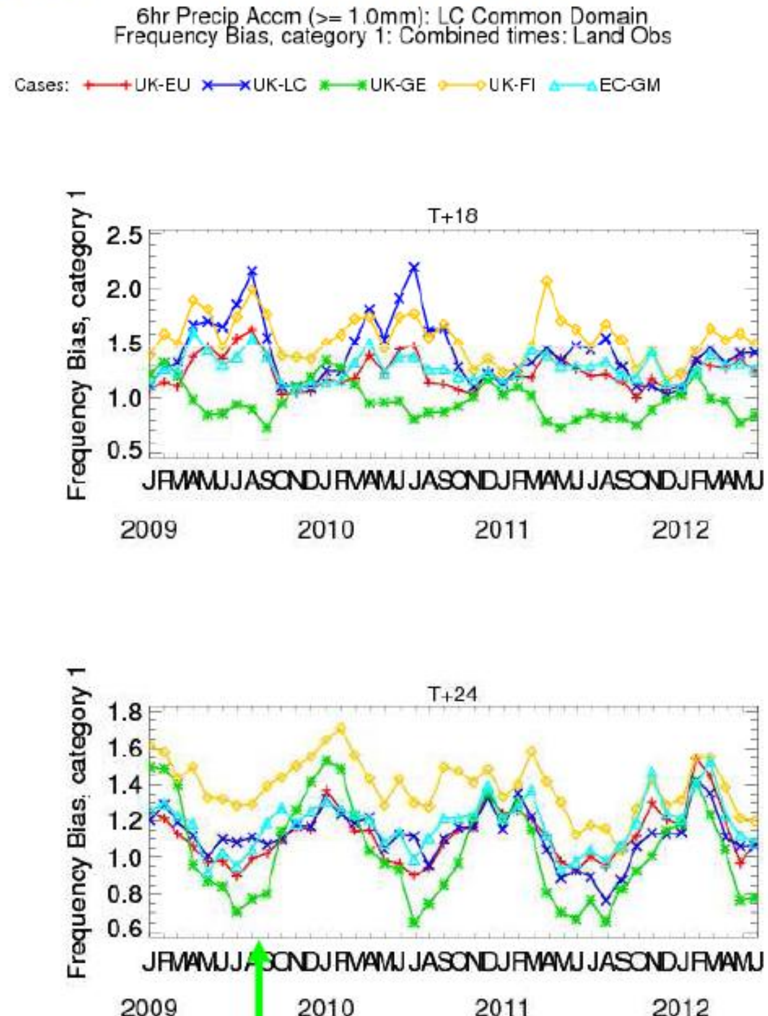
- Lower skill (≈ -0.1) than winter
- 2 groupings
  - EC/UM/Aladin/Hirlam
  - COSMO
  - Aladin >EC
- Dip in skill at T+24 (evening-night)
- Lack of showers persisting into late evening



Error bars 70% confidence intervals

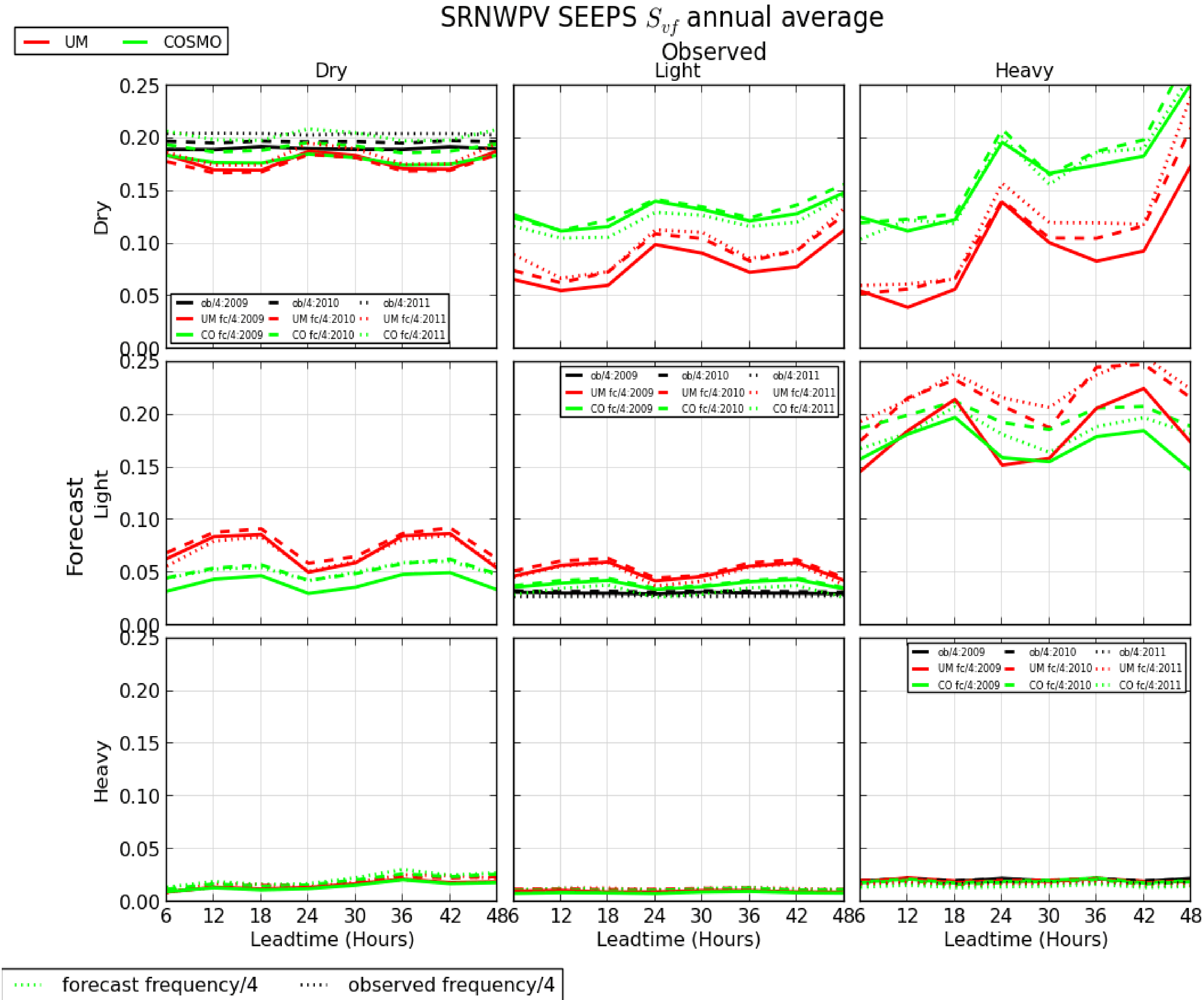
ppn frequency bias >1mm/6h



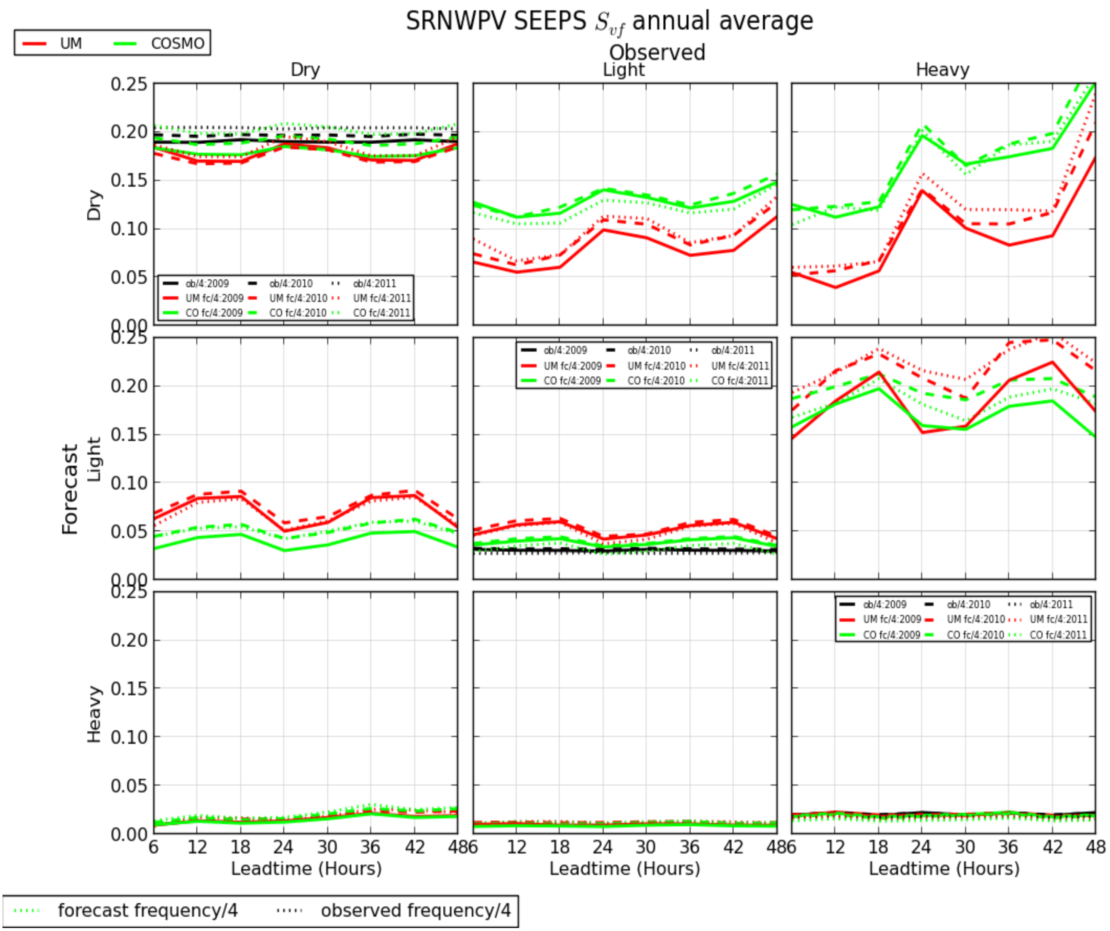


Summer day bias, convection

Night negative bias



SRNWPV SEEPS S_{vf} annual average
UM
COSMO
Observed
Dry
Light
Heavy
Forecast
Dry
Light
Heavy
0.25
0.20
0.15
0.10
0.05
0.00
ob/4:2009
ob/4:2010
ob/4:2011
UM fc/4:2009
UM fc/4:2010
UM fc/4:2011
CO fc/4:2009
CO fc/4:2010
CO fc/4:2011
6
12
18
24
30
36
42
48
Leadtime (Hours)
forecast frequency/4
observed frequency/4



Higher positive bias of the UM (and other models) leads to a better SEEPS **as fewer heavy events are missed**. Although the UM is relatively worse at overpredicting light precipitation when dry observed (row 2, column 1), this is less of a penalty than COSMO gets for predicting dry when either light or heavy observed.

In the original paper of Rodwell on SEEPS is stated that:
"Case studies demonstrate that <u>SEEPS is sensitive to overprediction of drizzle (our case) and failure to predict heavy large scale precipitation and incorrectly locating convective cells</u> (again our case, where COSMO underpredicts heavy rain).

Considerations on SEEPS results

Note: the higher SEEPS, the worse the verification

The diagonal plots show frequency of observed and forecast in each category so no contribution in SEEPS. The off diagonal plots show how the SEEPS contributions arise.

<u>COSMO is being penalised by</u>:

- Missing heavy events as u can see from both the forecast-dry (row 1 column 3) and forecast-light-precipitation (row 2, column 3) categories

- Missing light-precipitation events (row 1, column 2) from forecast-dry category

THESE EVENTS SHOW UNDERESTIMATION

Also COSMO is penalised, even if to less extent, by:
- predicting light precipitation when it is observed dry (row 2, column 1)

THIS EVENT SHOWS OVERESTIMATION IN SMALL THRESHOLDS (but clearly has less weight than the others above numerically)

Priority Task - NWP Meteorological Test Suite Plan

Goal

• Build up a software environment to perform carefully-controlled and rigorous testing
• Calculation of verification statistics for any COSMO model test – version
• Offer necessary information on the model forecasting Performance
• Provide the COSMO community with standards against which the impacts of new developments in the model should be Evaluated
• Benchmark to monitor the progress of mesoscale forecast improvement (periodic testing as COSMO evolves)

Common Plot Reports

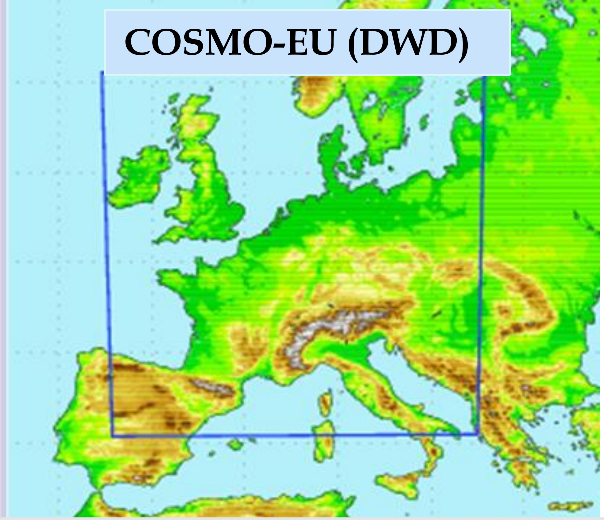
CONSORTIUM FOR SMALL SCALE MODELING

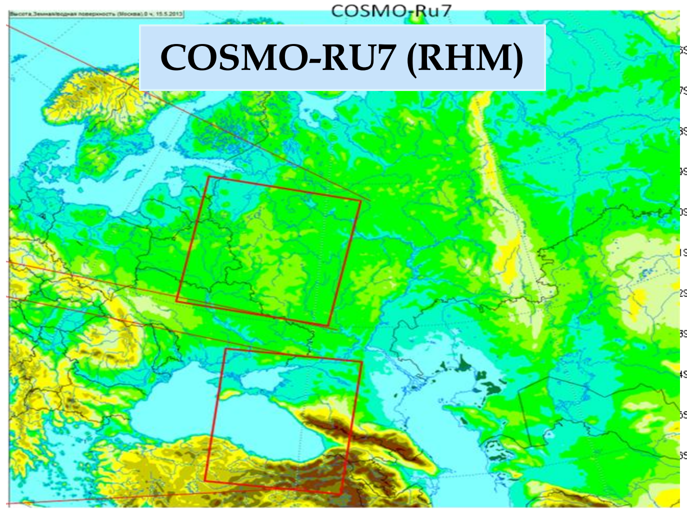
Standard Verification

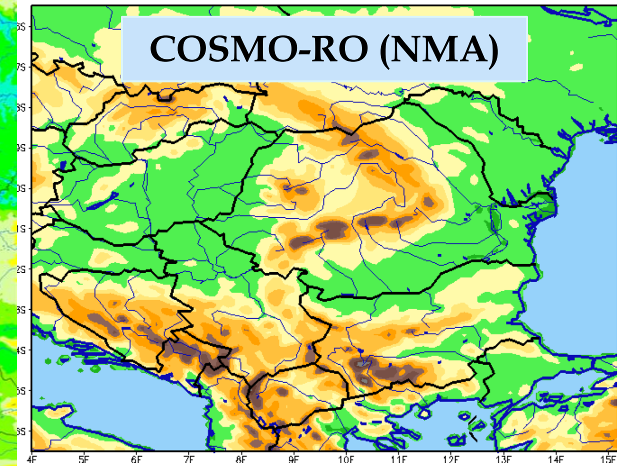
- **Period: JJA 2012, SON 2012, DJF2012/2013, MAM 2013**
- **Run: 00 UTC run**
- **Continuous parameters - T2m, Td2m, Mslp, Wspeed, TCC**
  - Scores : ME, RMSE
  - Forecasts Step: every 3 hours
- **Dichotomic parameters - Precipitation:**
  - Scores: FBI-POD-FAR-TS with Performance Diagram
  - Cumulating: 6h and 24h
  - Thresholds: 0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20 mm/6h and mm/24h

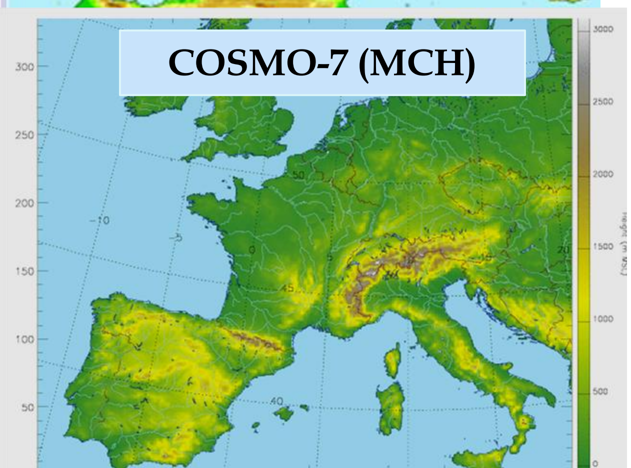
Conditional Verification (focus on the next slides)

- **<u>2mT verification with the following criteria (1 condition)</u>:**
  - Total cloud cover >= 75% (overcast condition) (condition based on observations)
  - Total cloud cover <= 25% (clear sky condition) (condition based on observations)
-
- **<u>2mT verification with the following criteria (2 conditions)</u>:**
  - Total cloud cover >= 75% (overcast condition) **AND** Wind Speed<2.5 m/s (condition based on observations)
  - Total cloud cover <= 25% (clear sky condition) **AND** Wind Speed<2.5 m/s (condition based on observations)

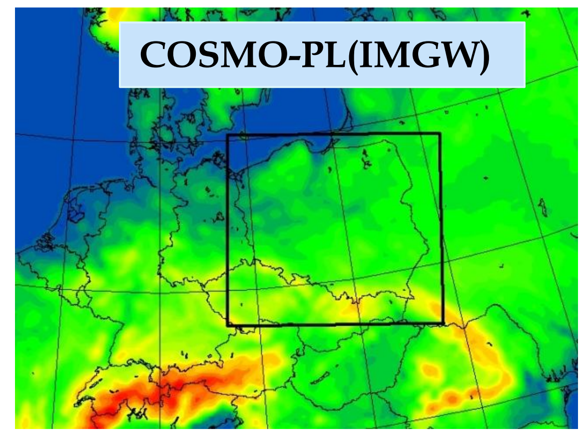

COSMO-EU (DWD)

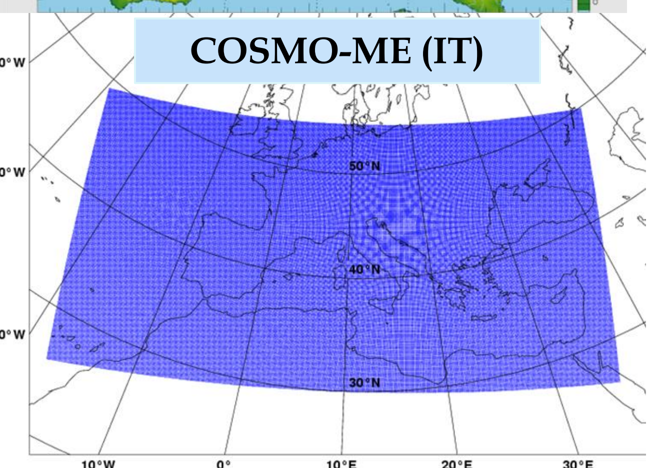

COSMO-Ru7
COSMO-RU7 (RHM)

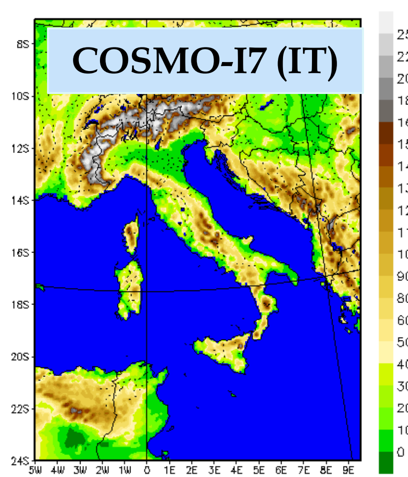

COSMO-RO (NMA)

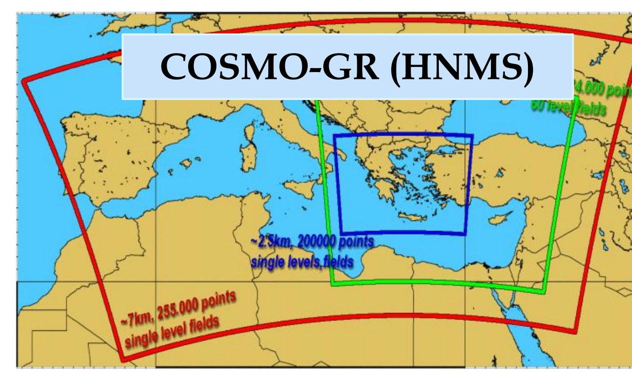

COSMO-7 (MCH)

THE MODELS

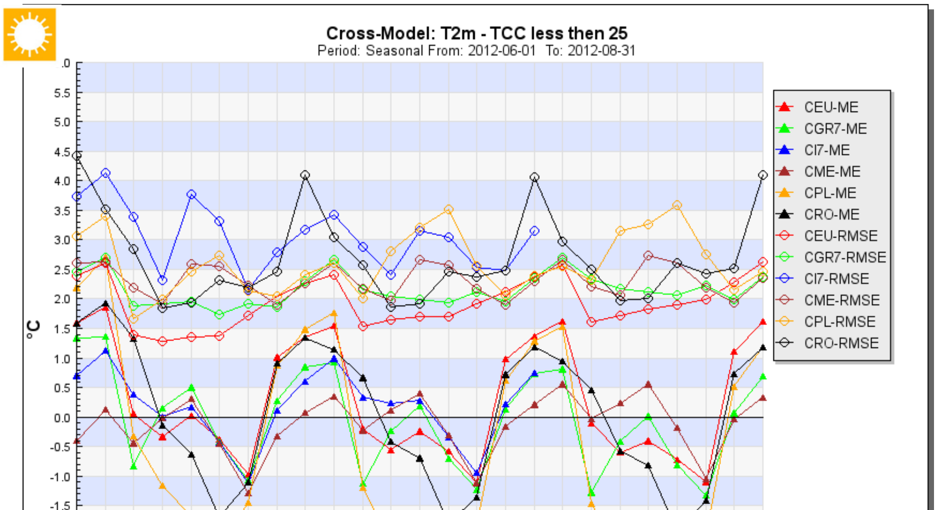

COSMO-PL(IMGW)

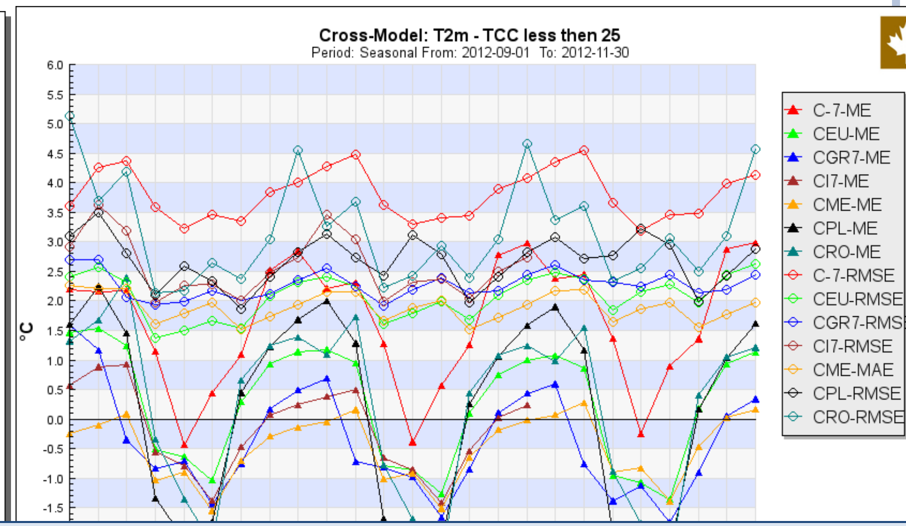

COSMO-ME (IT)
50°N
40°N
30°N
10°W
0°
10°E
20°E
30°E


COSMO-I7 (IT)

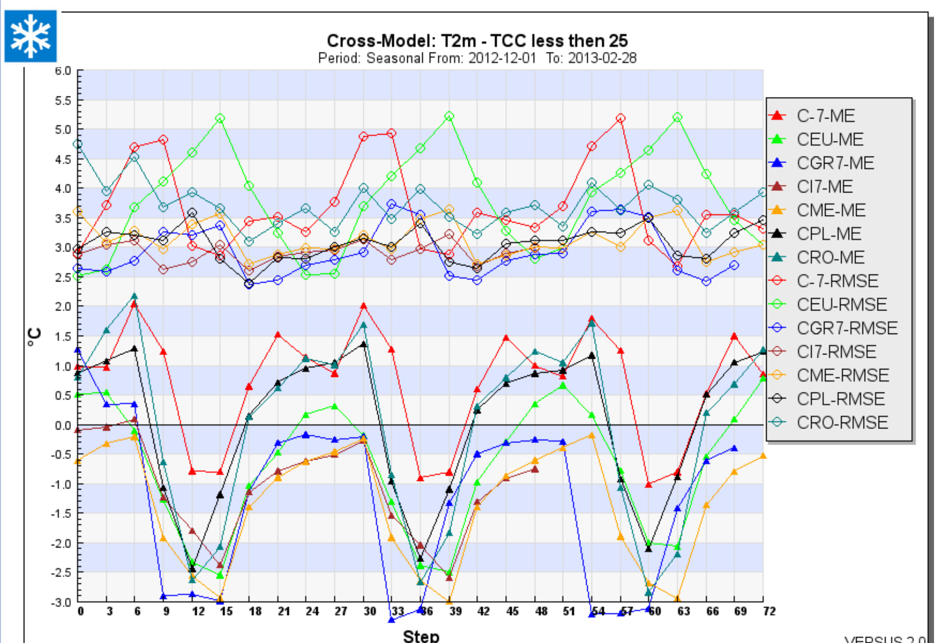

COSMO-GR (HNMS)
~7km, 255.000 points
single level fields

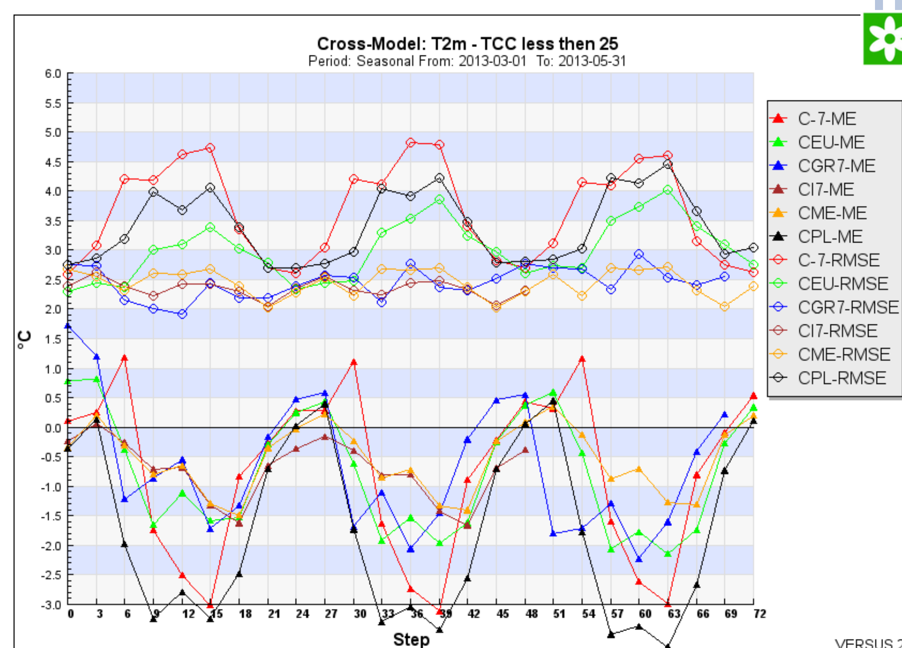
2mT in Sky Clear conditions - JJA 2012 – MAM 2013



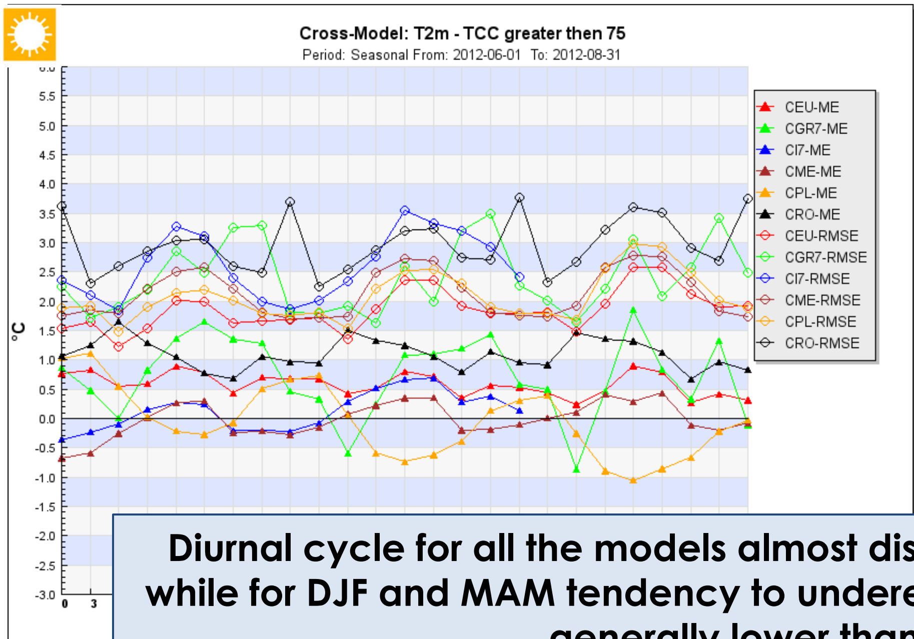


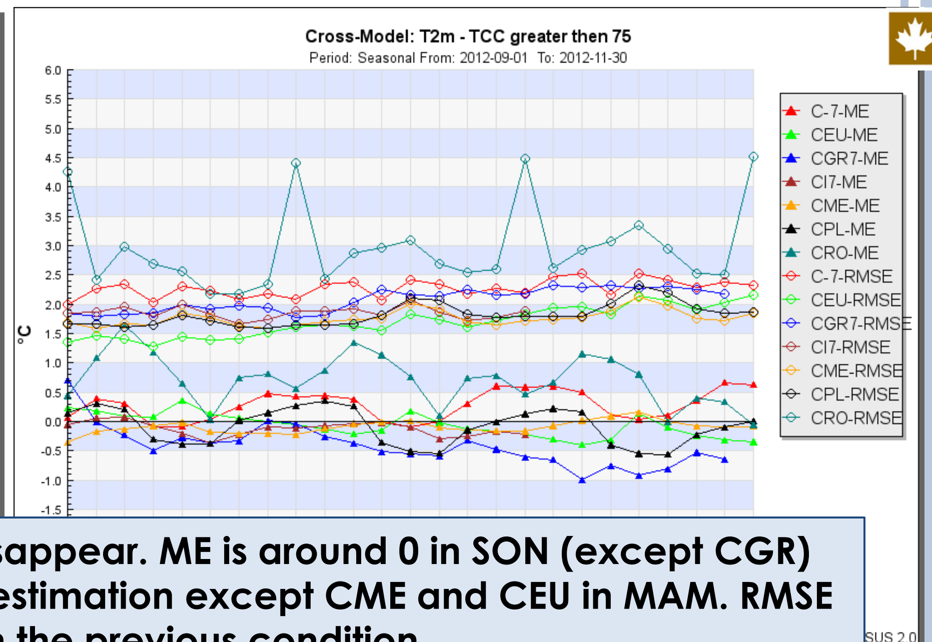
Clear diurnal cycle for all the models with a general tendency to underestimation in DJF and MAM (maybe poor sample) and amplitude of the error pronounced. RMSE between 2° and 4-5°.



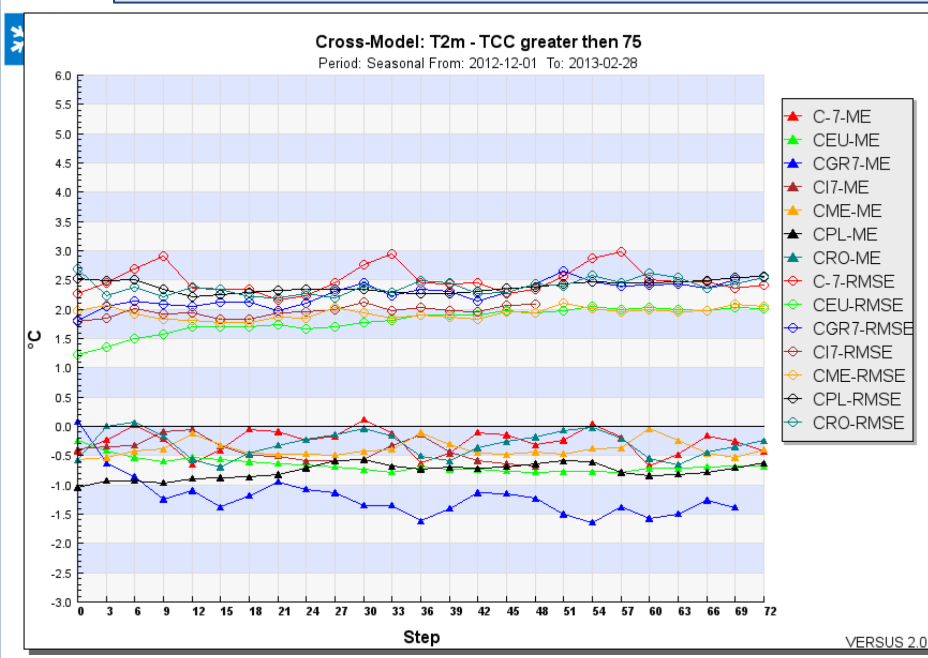


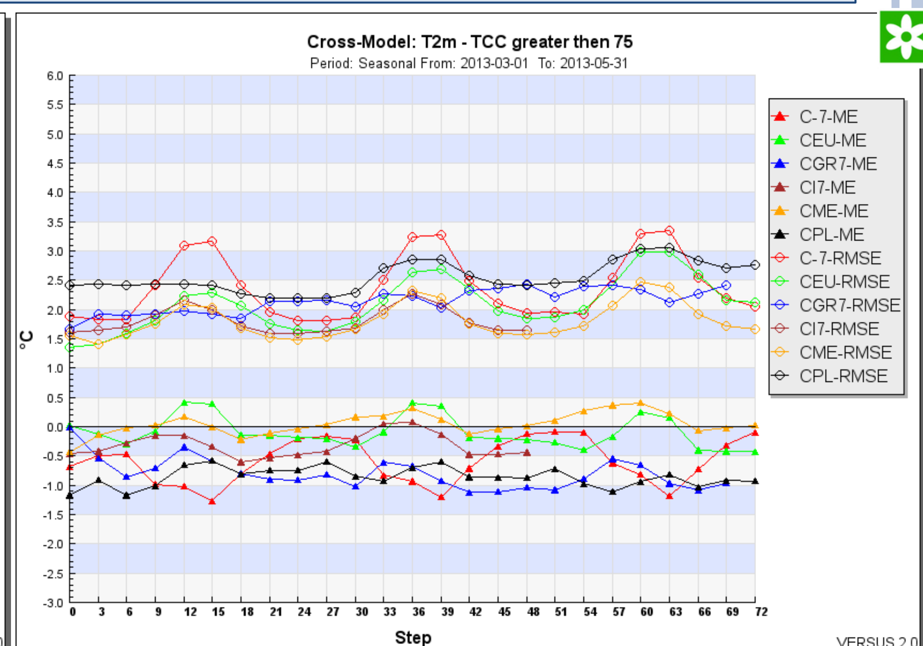
2MT IN OVERCAST CONDITIONS - JJA 2012 – MAM 2013



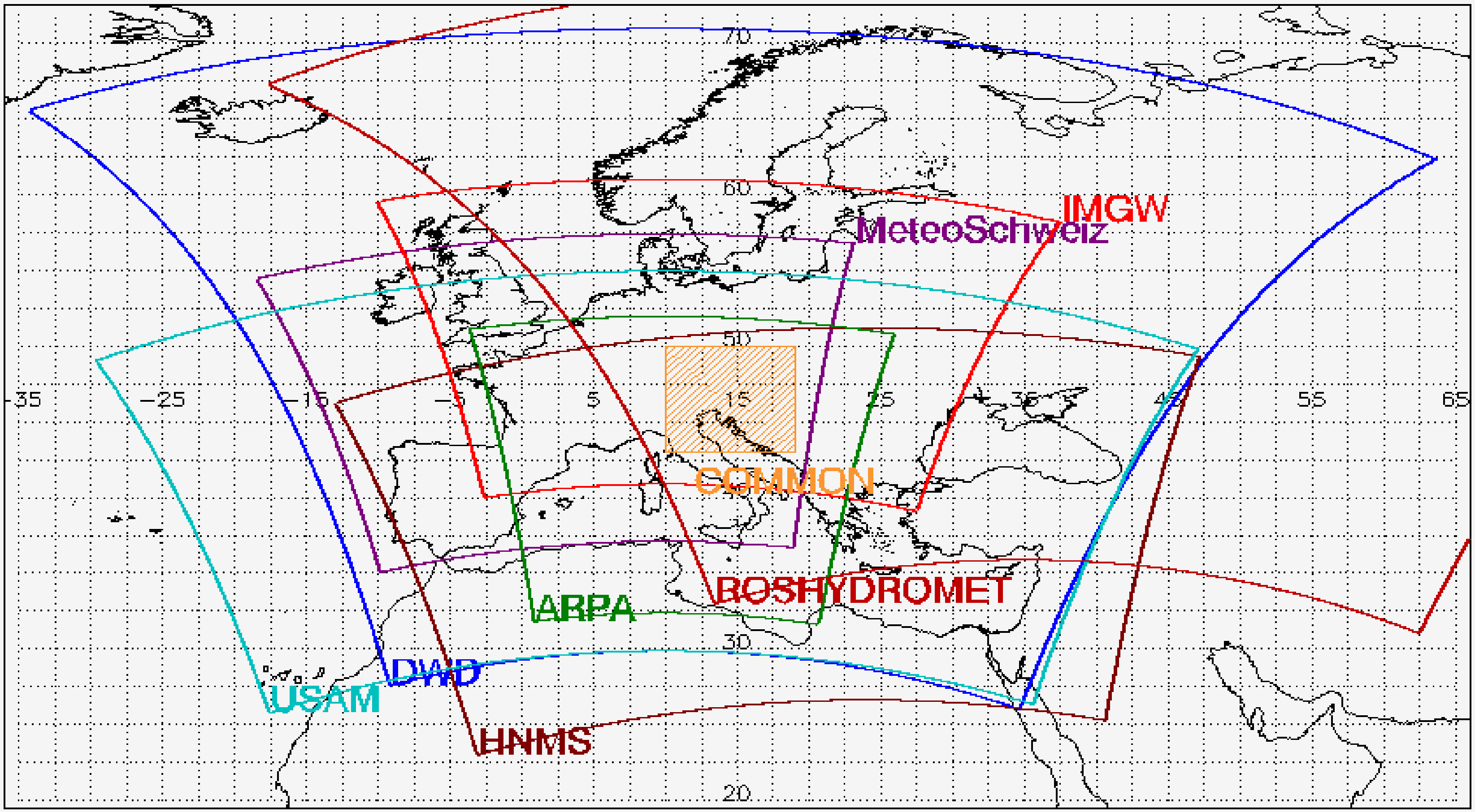


Diurnal cycle for all the models almost disappear. ME is around 0 in SON (except CGR) while for DJF and MAM tendency to underestimation except CME and CEU in MAM. RMSE generally lower than the previous condition.

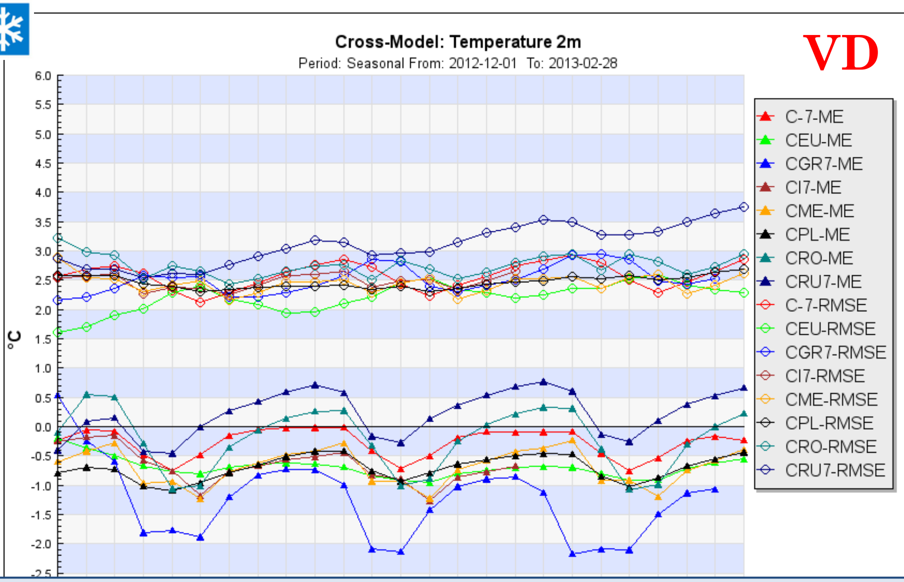


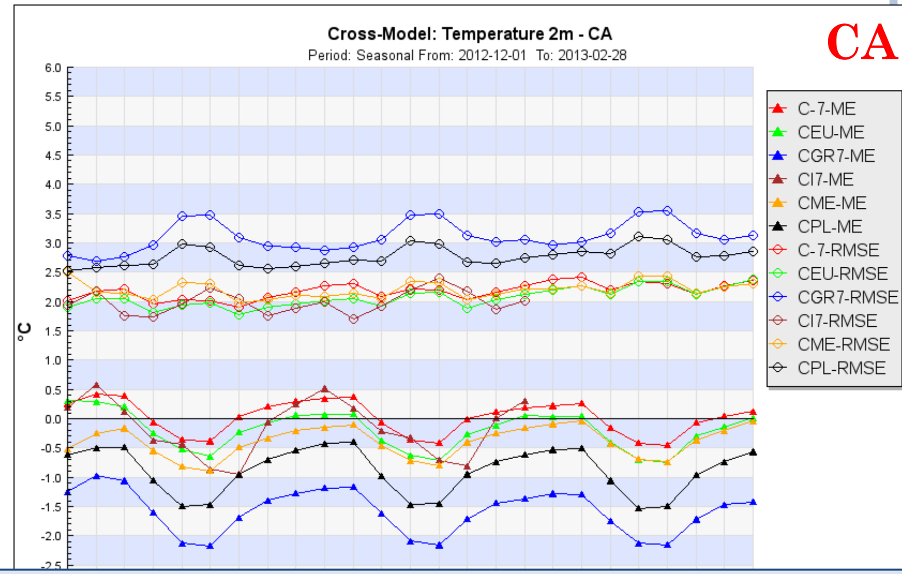


Standard Verification on Common Area

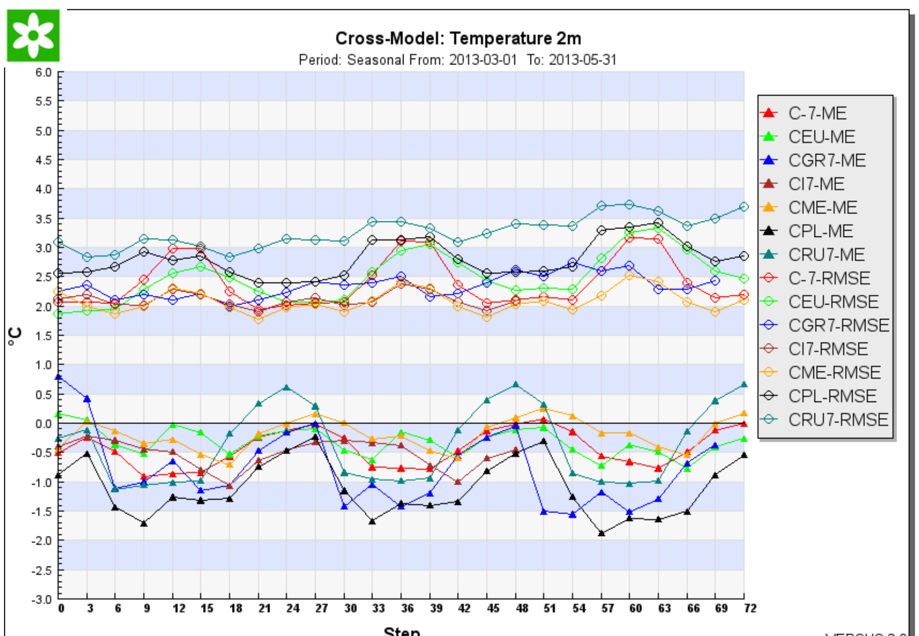


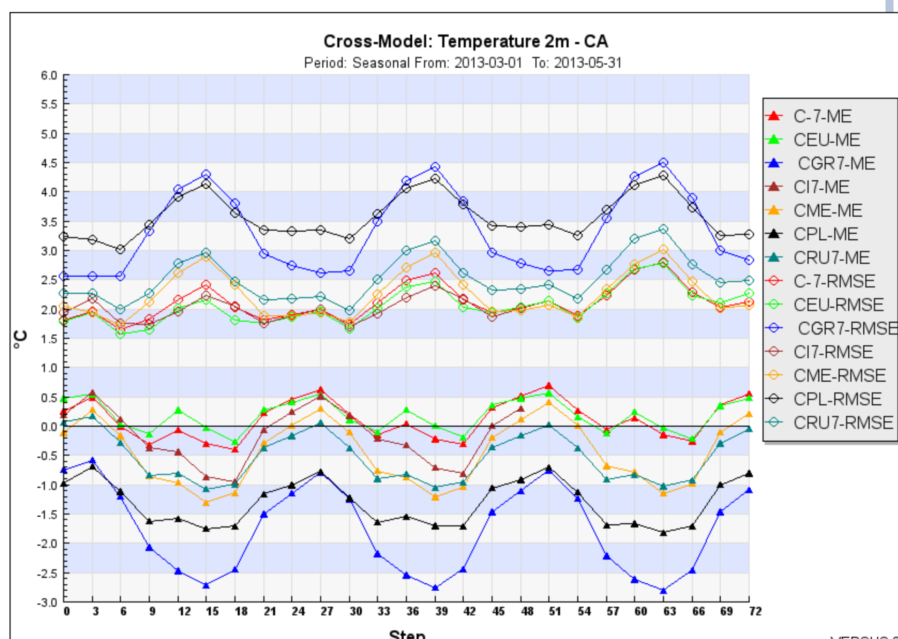
2M TEMPERATURE - DJF 2013 – MAM 2013





DJF and MAM: CPL, CGR increase underestimation in the CA while CEU, CME and CI7 decrease this tendency. RMSE in CA worse for CPL and CGR, while CI7, CEU, CME slightly improve.





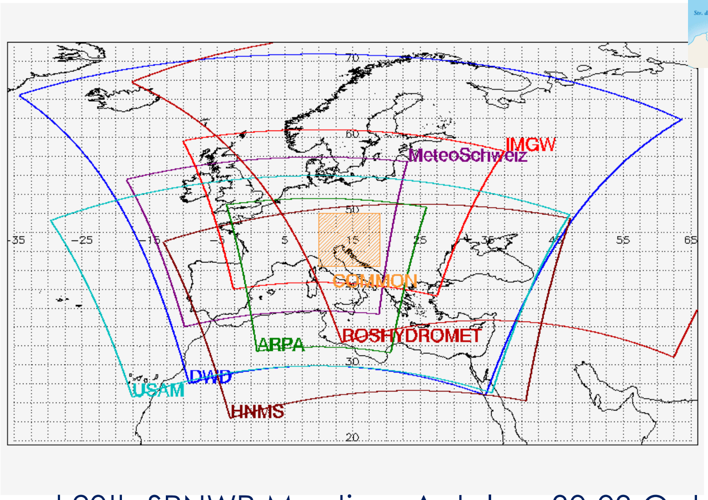
3 methods

1

•Common area → Italy
•Dataset → high res raingauges
•Method → 24h/6h averaged cumulated precipitation or maximum values (both observed and forecasted) over 90 meteo-hydrological basins

2

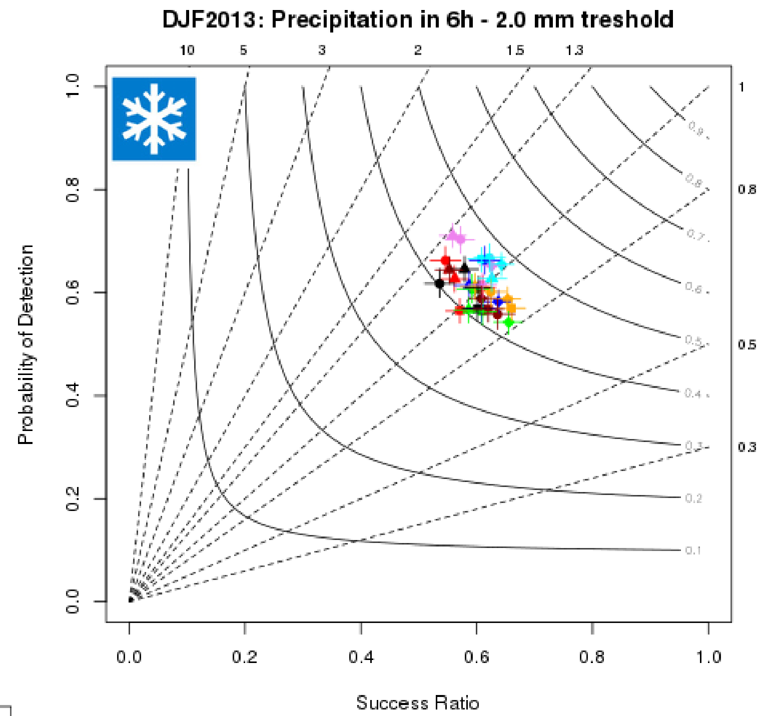
•Common area → decided in Lugano
•Dataset → synop stations
•Method → 24h/6h averaged cumulated forecasted precipitation values over 15 km radius, 24h/6h cumulated observed precipitation values over station point



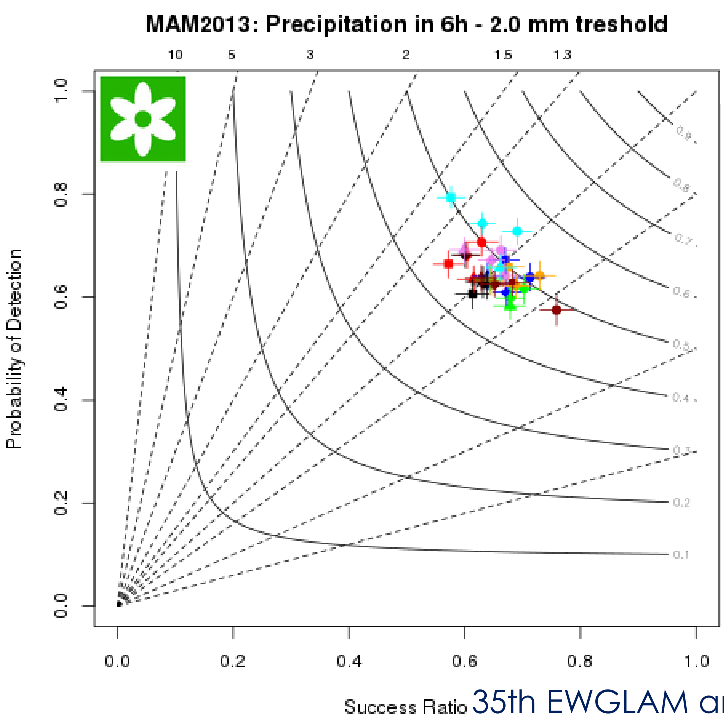
3

Various domains → each countries dataset → synop stations
Method → 24h/6h averaged cumulated forecasted precipitation values over 15 km radius, 24h/6h cumulated observed precipitation values over station point

1

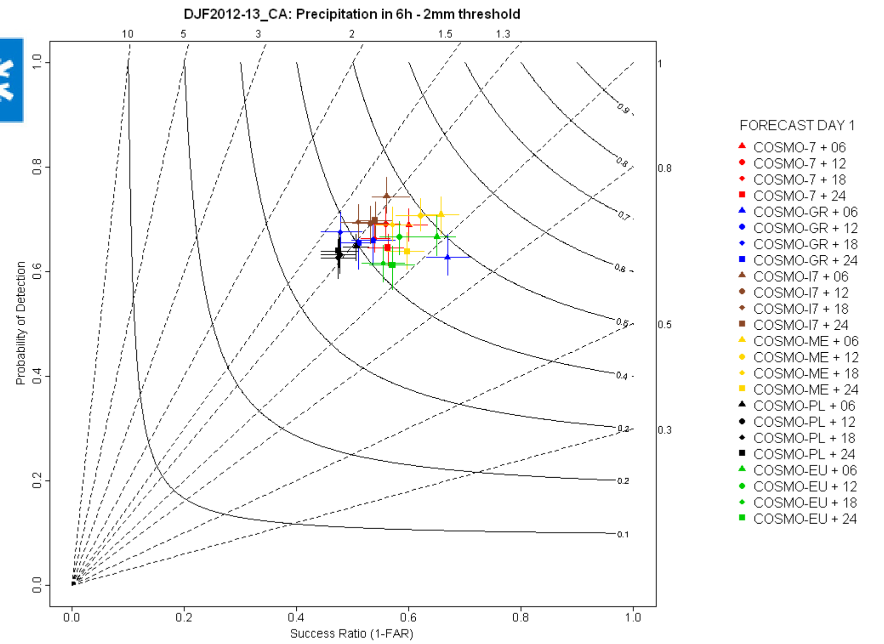


Average over area > 2 mm/6h



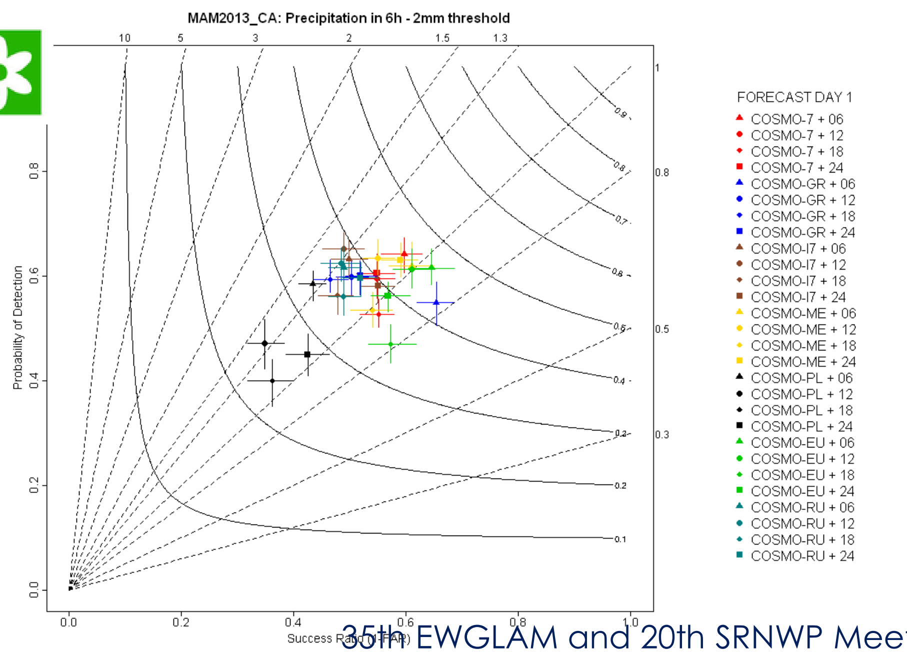
IFS exhibits less overestimation

I7 0006
I7 0612
I7 1218
I7 1824
7 0006
7 0612
7 1218
7 1824
EU 0006
EU 0612
EU 1218
EU 1824
ME 0006
ME 0612
ME 1218
ME 1824
I2 0006
I2 0612
I2 1218
I2 1824
IT 0006
IT 0612
IT 1218
IT 1824
GR 0006
GR 0612
GR 1218
GR 1824
ECMWF 0006
ECMWF 0612
ECMWF 1218
ECMWF 1824

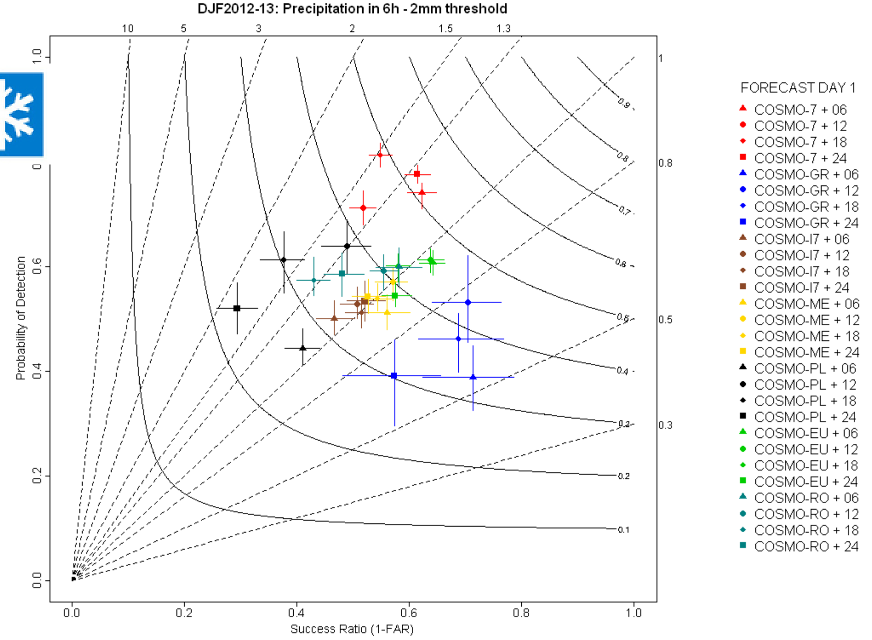


Average over area > 2 mm/6h

2

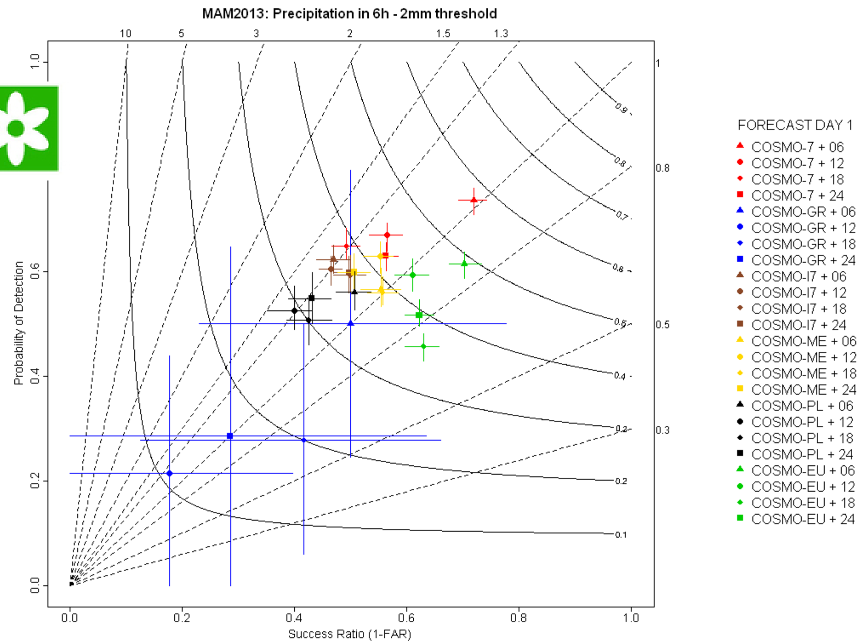


Similar situation of CA

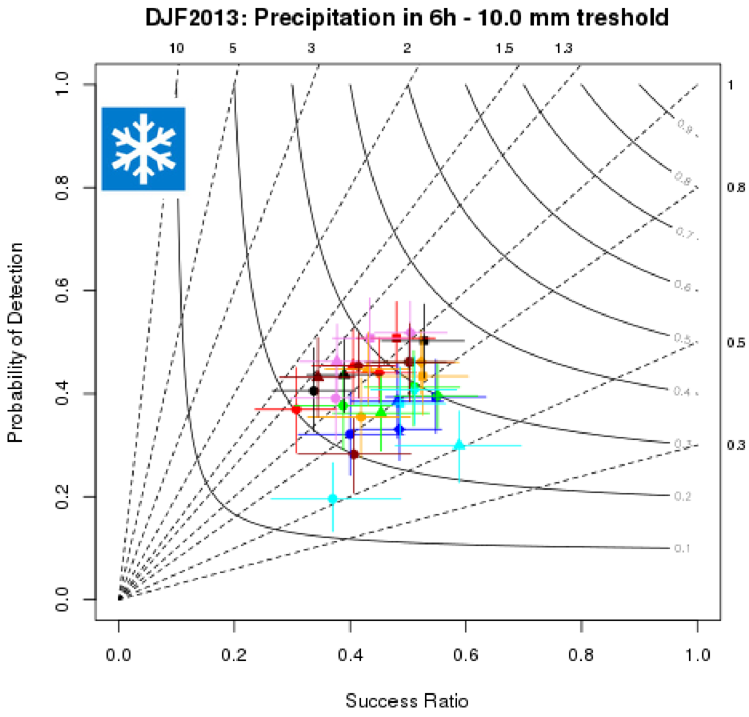


Average over area > 2 mm/6h

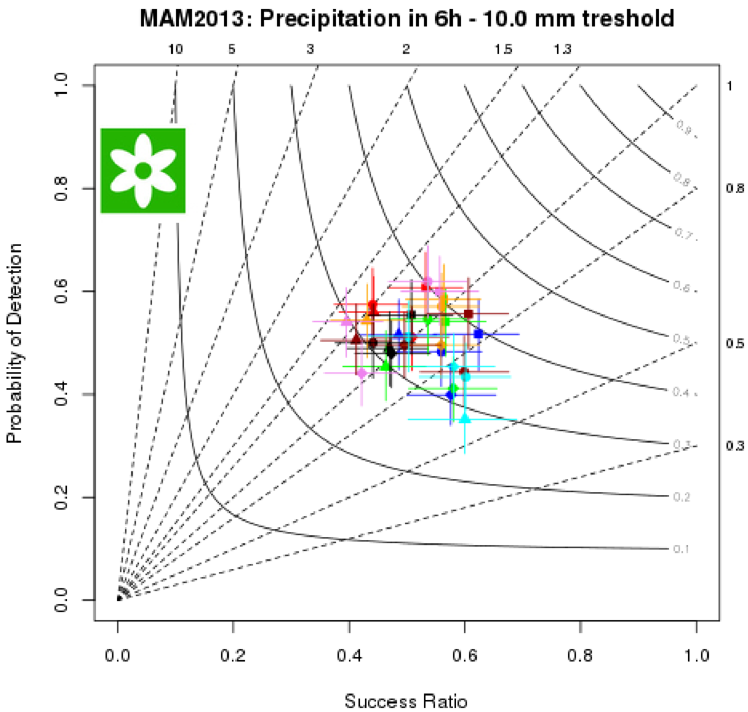
3



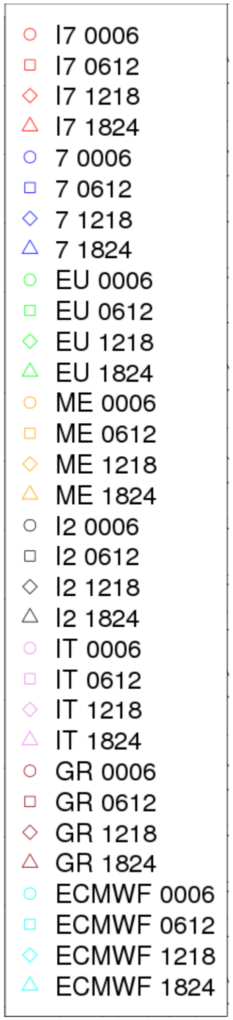
1

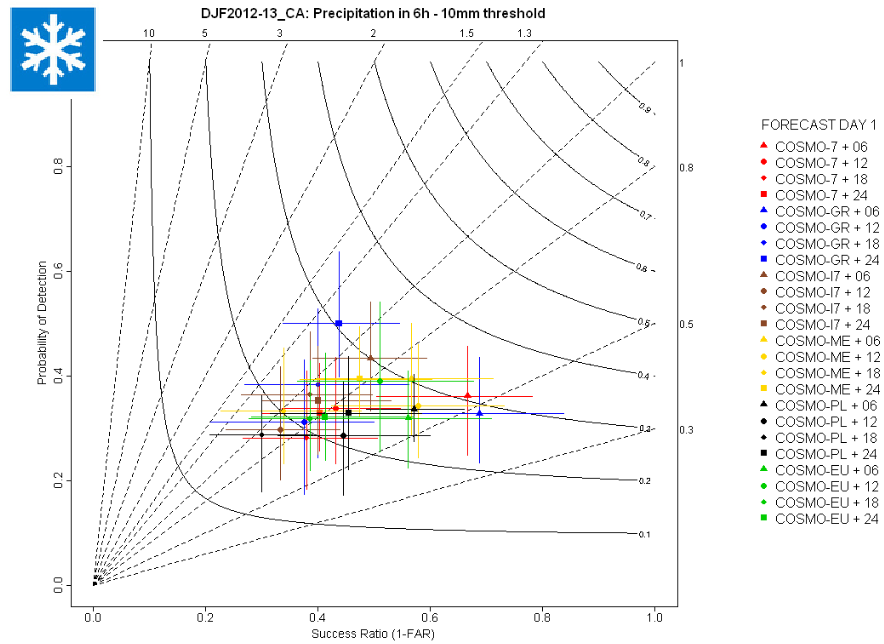


Average over area > 10 mm/6h



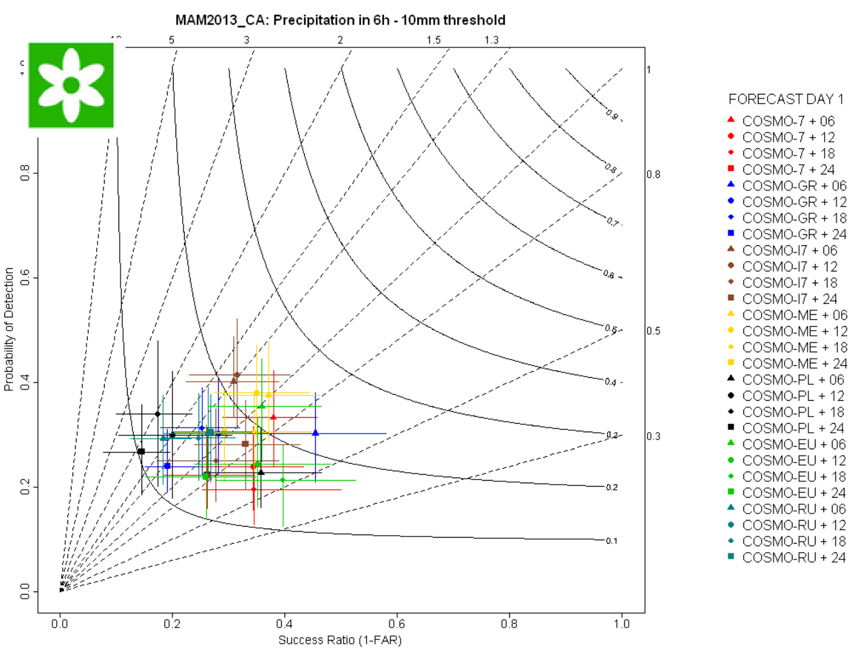
Uncertainty grows. Low scores for IFS



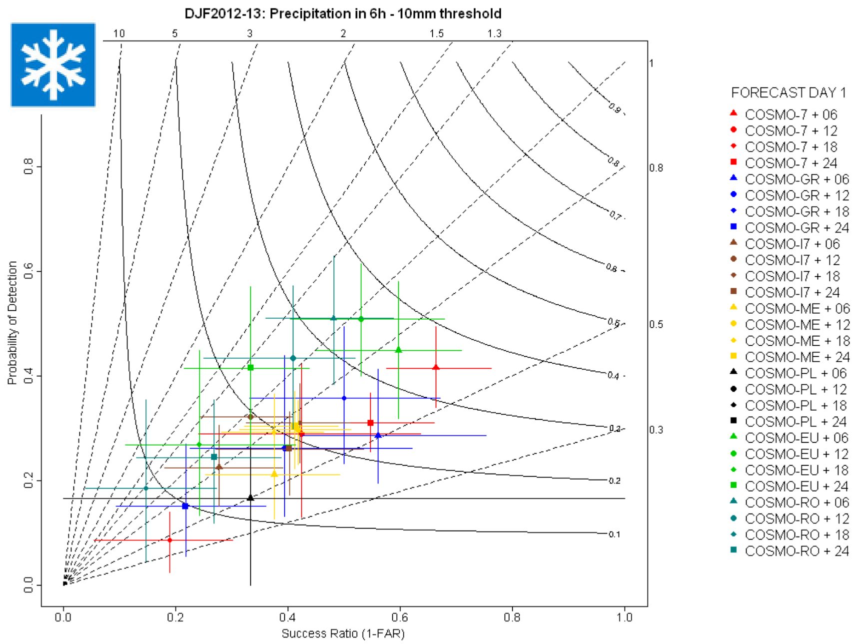


Average over area > 10 mm/6h

2

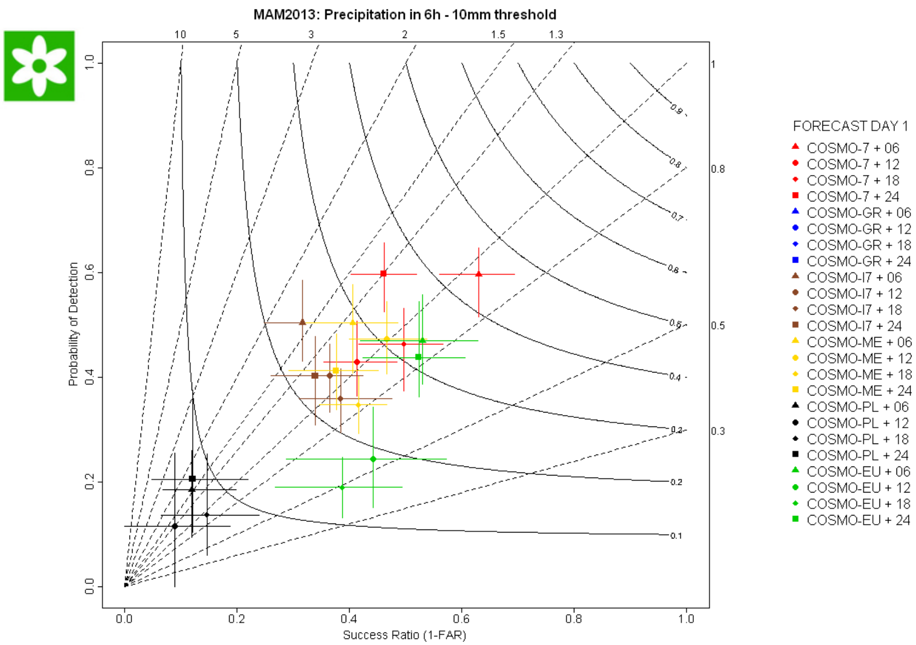


Similar situation of CA-1. Lower scores due probably to the use of SYNOP in CA-2 and high resolution network in CA-1



Average over area > 10 mm/6h

3



Some final considerations

1. Importance of exchange systematically (SYNOP) high quality and homogenous information like Global Radiation.

2. There is not “right” Score for precipitation events. Scores used in Long term trend evaluation and even the experience with SEEPS show that only partial result can be obtain from a single score and the combination of various measures is still necessary.

3. Impact on precipitation verification of different approaches and observational dataset: Common area using synop, common area using raingauges, various domain using synop: how to compare o**bjectively** the results and how to evaluate the impact of the used method?