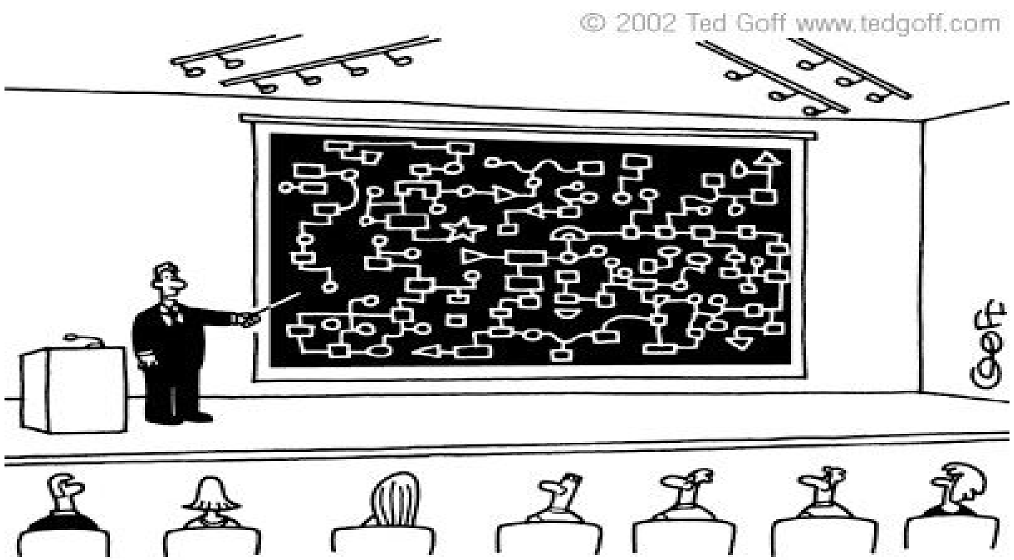
WG5: Verification and Case studies

Overview of activities

Flora Gofa

COSMO General Meeting Sibiu 2013

Science Plan Overview of activities



"This is our plan for the next 1,000 years."

"STC considers VERSUS as the tool for both operational and scientific verification and recommends that the verification strategy focuses on <u>high-resolution and ensemble applications</u>"

Science Plan Overview of activities

Short (..Long) term plans

Tackling model performance improvement issues through the use of conditional verification (CV)

As model errors should be related to specific inaccurately simulated processes, verification under specific conditions (CV) have to be chosen in order to extract selected model uncertainties due to simulation errors, isolating single processes or uncertainties responsible for measured simulation errors. This procedure is based on the selection of forecast products and "mask variables" (model variables, observations or external variables) and application of arbitrary thresholds (conditions) to produce verification.

Statistical methods to identify the skill of convection permitting and near-convection-resolving model configurations

Increasing of models resolution can lead, especially for precipitation but also for continuous surface parameters, to forecasts detail more realistic but inaccurate, the so-called double penalty effect. For this reason neighborhood methods were employed to compare forecasts in appropriate selected size neighborhoods with the gridded radar data for precipitation. For this reason a verification framework needs to be defined (even probabilistic, BS, RPS). Statistical methods proposed should lead to the estimation of the relative skill gained using higher resolution, to the assistance in the decision-making process for model upgrades for similar horizontal resolution and to the comparison between the determinist forecasts with ensemble ones.

Exploitation of available observational dataset for operational and scientific purposes

For model-oriented verification, processing of the observation data needs to be done to match the spatial and temporal scales resolvable by the model. This requires the availability of high spatial resolution observations (satellite or radar post-processed data) to be used to produce vertical profiles or gridded surface analysis. Furthermore particularly important is the exploitation of controlled and possibly homogenous set of surface observations, concerning fluxes, radiation and soil characteristics, such as those available from SRWNP Data Pool Exchange.

Development of tools for probabilistic and ensemble forecast verification

The challenges in verifying "convection-permitting" ensembles are basically the same as in mesoscale "convection-parameterisation" ensembles, with some **added complexities**. Due to their nature, convection-permitting ensembles focus on the shortest range (0-24h) and large error growth in such systems which are correlated strongly to the highly non-linear physical processes of convection, thus verification measures must focus on the relevant gain of the use of such systems toward better representation of convection-based parameters. As for deterministic forecasts, neighborhood methods are proposed to be employed to account for the spatial mismatches between forecasts and observations, especially for precipitation, even though ensemble forecasts can address uncertainties of small-scale processes more adequately.

Severe and High Impact Weather

As there is an increased demand that meteorological services provide accurate forecasts of extreme weather, it is therefore important to be able to objectively evaluate the model performance in these cases. Severe events are rare and this is the reason that standard skill scores are not useful as they depend on base rate. Dependency scores like SEDS and SEDI have been extensively used by the NWP community for some time, but the use of other scores and methods will also be evaluated.

User-oriented Verification products

With increasing model resolution, the number of products the users will ask, as well as their objective performance in terms of their expected quality is only going to rise. Different users might have needs for different verification information (e.g. administrative decisions may depend on model performance), so different verification strategies have to be chosen. It will be necessary to diversify verification methodologies to match the different needs and to this end, the scientific community will have to work more closely with the user community in the design of such verification strategies.

These main activities could be reviewed and updated in the light of future developments in the main fields of model improvements concerning physics and data assimilation, in order to respond to the actual needs of developers and users alike.

COSMO resources for new developments

● Need to optimize the available resources, by monitoring the efforts of the various European Consortia (e.g. through SRNWP collaboration) and of the scientific community in the field of verification.

● The long term continuation of **VERSUS project** in the framework of PT-Support will contribute to the realization of the actions planned (VERSUS project is prolonged already for another year (**phase 6**) and a long-term maintenance plan is set).

● A new priority Task was prepared with the aim to provide the common platform (at ECMWF) for a **standardized meteorological verification framework** of each new COSMO reference version against the existing one. Special Project has been proposed and approved by the European Centre.

PT NWP Test Suite, PL: Amalia Iriza (NMA)

Amended Work Group 5 Task List

1. Common Verification Framework

1.1 Operational Verification *Responsible: ALL*

1.2 Responsibility for Common Plots Reports *Responsible: J.Linkowska,IMGW*

1.3 Verification of vertical profiles using TEMP observations, aircraft data (AMDAR) and wind-profiler data *Responsible: ALL*

1.4 Dissemination of daily Grib model output Files *Responsible: De Morsier, MCH*

2. Exploitation of observational dataset for operational and scientific purposes

2.1 High density verification of precipitation over Italy *Responsible: E.Oberto, ARPA-PT*

2.2 Exchange of a common data set of non-GTS data DWD *Responsible:U.Damrath*

2.3 Evaluation of COSMO models in the lower PBL Responsible: Raspanti, Gofa, Kaufmann

3. Evaluation of convection permitting models performance

3.1 Long Term Trend Verification *Responsible: ALL*

3.2 Conditional Verification *Responsible: ALL*

3.3 Weather Dependant Verification (WDV) *Responsible: ALL*

3.4 Severe and High Impact Weather *Responsible:*

Amended Work Group 5 Task List

4. Neighborhood method techniques

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

Responsible: D. Leuenberger, MCH

4.2 Precipitation verification using radar composite network with neighborhood methods

Responsible: N. Vela, ARPA-PT

5. Verification of EPS products (Cooperation with WG7)

6. Other

6.1 Annual Workshop/Tutorial on VERSUS2 & WG5

Overview of verification activities

Authors: ALL

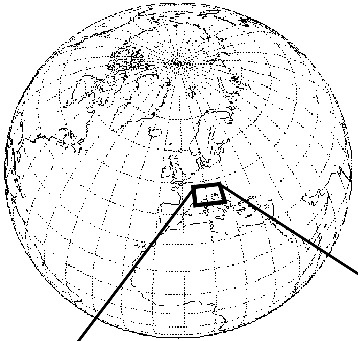
WG5

1. Common Verification Framework

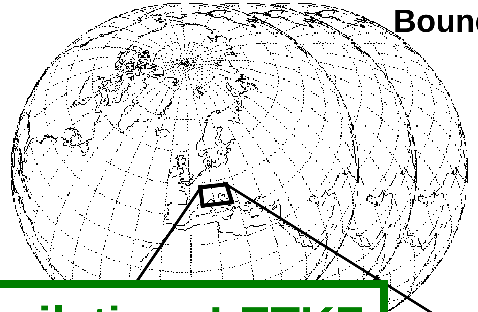
Operational verification (Task 1.1)

Verification of surface weather elements in various time steps (for each forecast hour) at SYNOP stations for: MSLP, T2m, Td2m, 10m-windspeed, TCC, Precipitation (mandatory), Wind gust and radiation verification (desirable). New techniques for wind direction representation should be investigated and applied. Representative scores for precipitation properties. Intercomparison of various resolutions of COSMO model with driving model. Customer-focused verification products.

Project COSMO-NExT



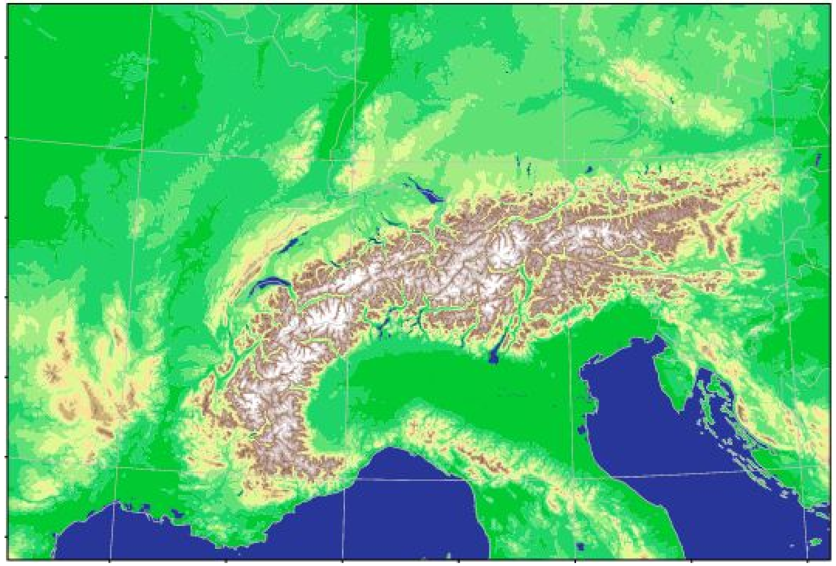
Boundary conditions: IFS
10km
4x daily



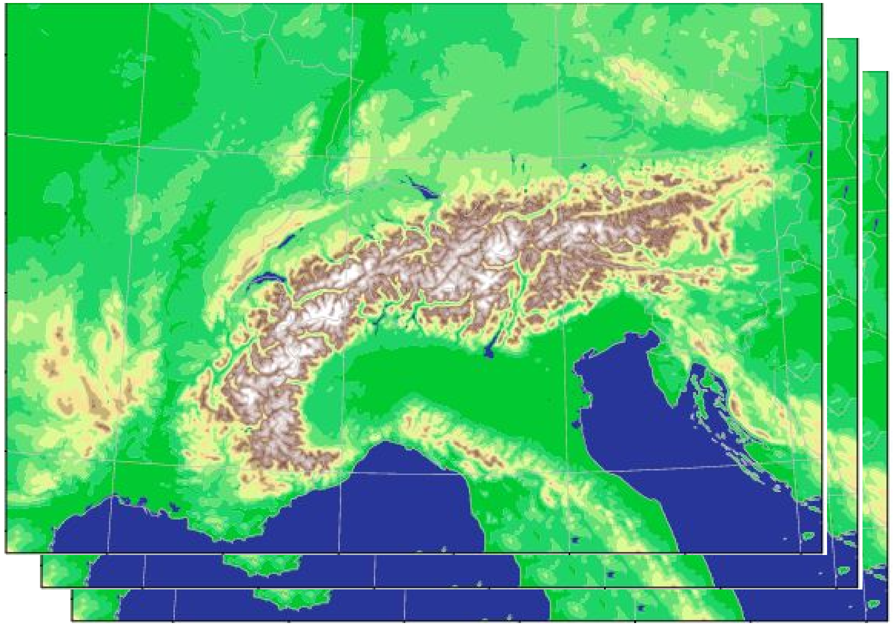
Boundary conditions: VarEPS
20km
2x daily

ensemble data assimilation: LETKF

COSMO-1: 8x daily O(24 hour) forecasts
1.1km grid size (convection permitting)



COSMO-E: 2x daily 5 day forecasts
2.2km grid size (convection permitting)
O(21) ensemble members



Verification 00 & 12 UTC forecasts

1) SYNOPs

(a) COSMO-2 domain («Alps»)
(b) Switzerland («CH»)

2) TEMPs

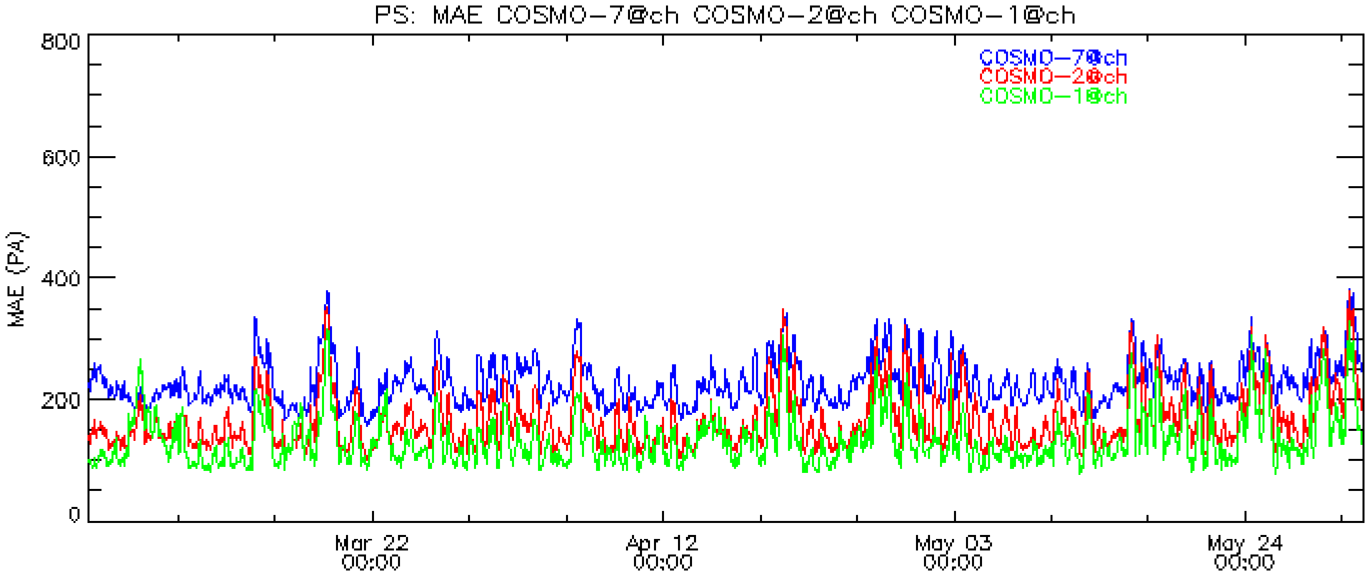
COSMO-7

COSMO-2

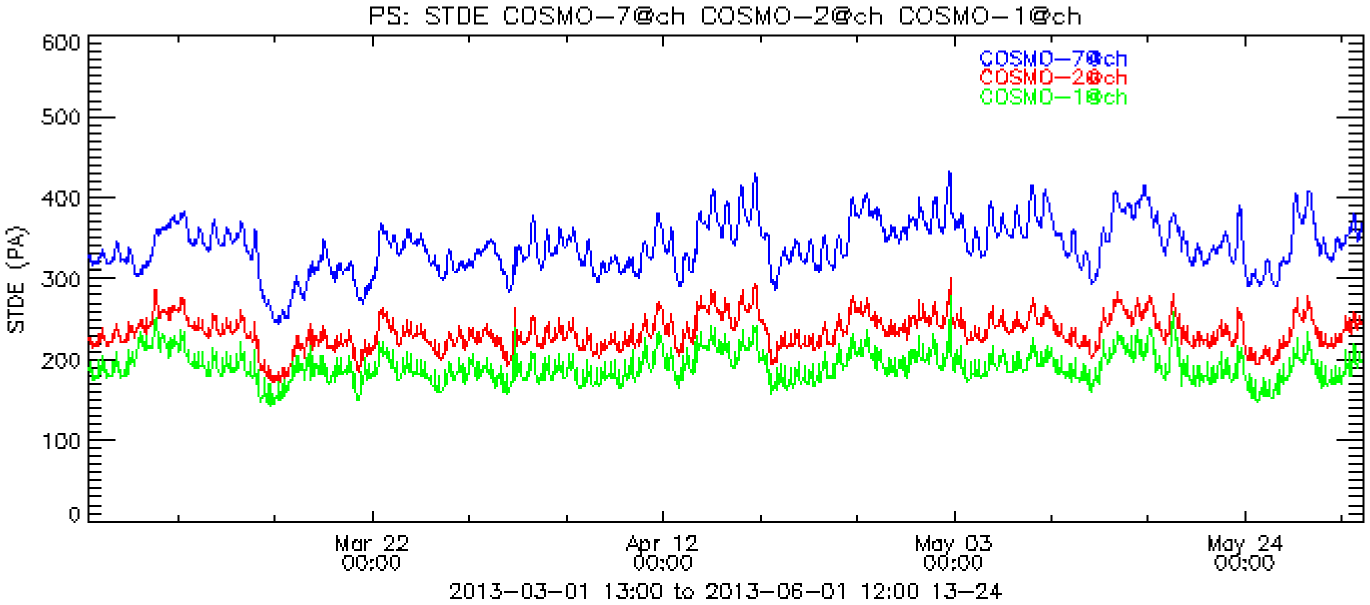
COSMO-1

forecasts missing for the verification:
17 in Autumn 2012, **1** in Winter 2012/13, **0** in Spring 2013

Surface pressure Spring 2013 CH



mean absolute error

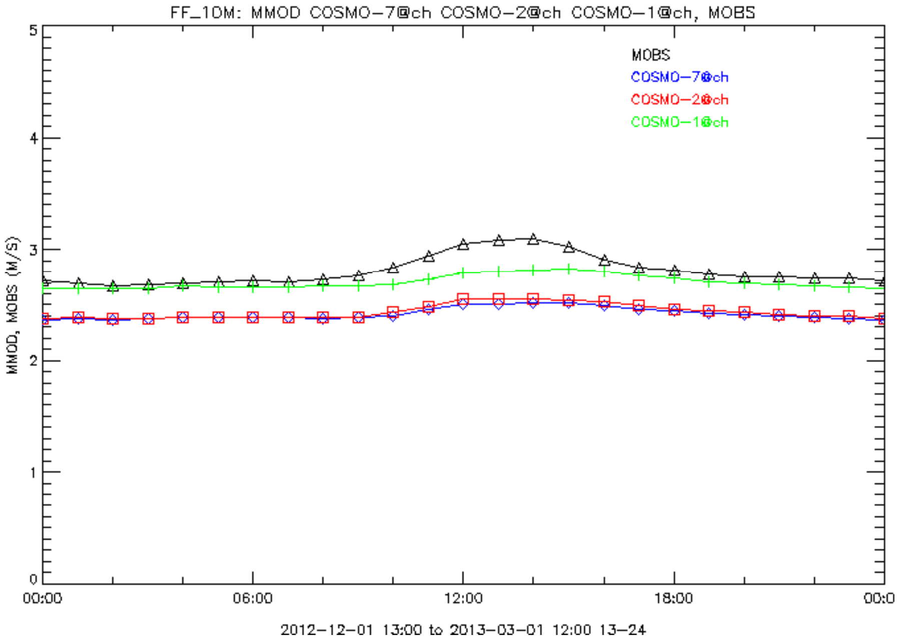


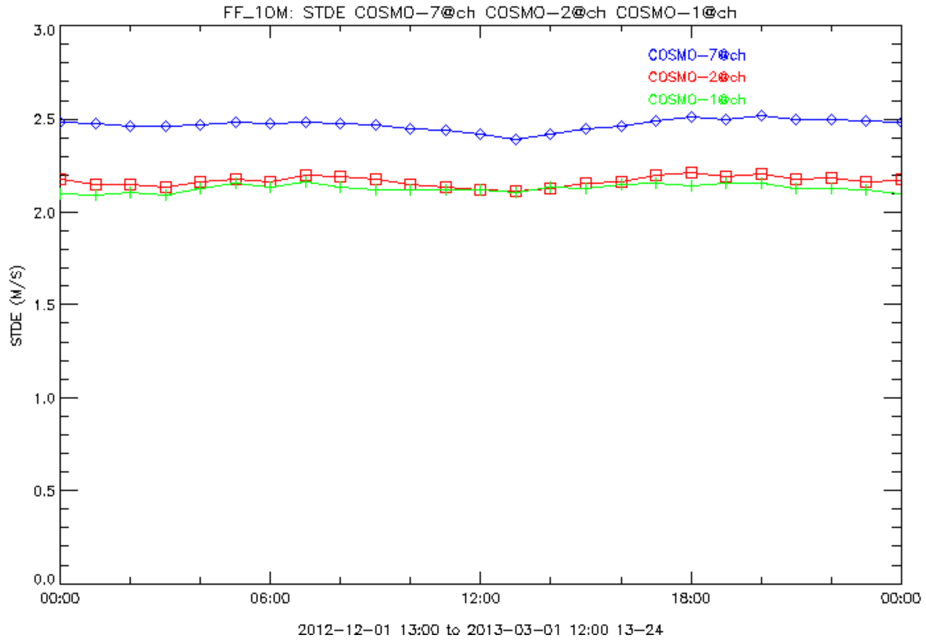
std dev

10m-wind speed Winter 12/13

values

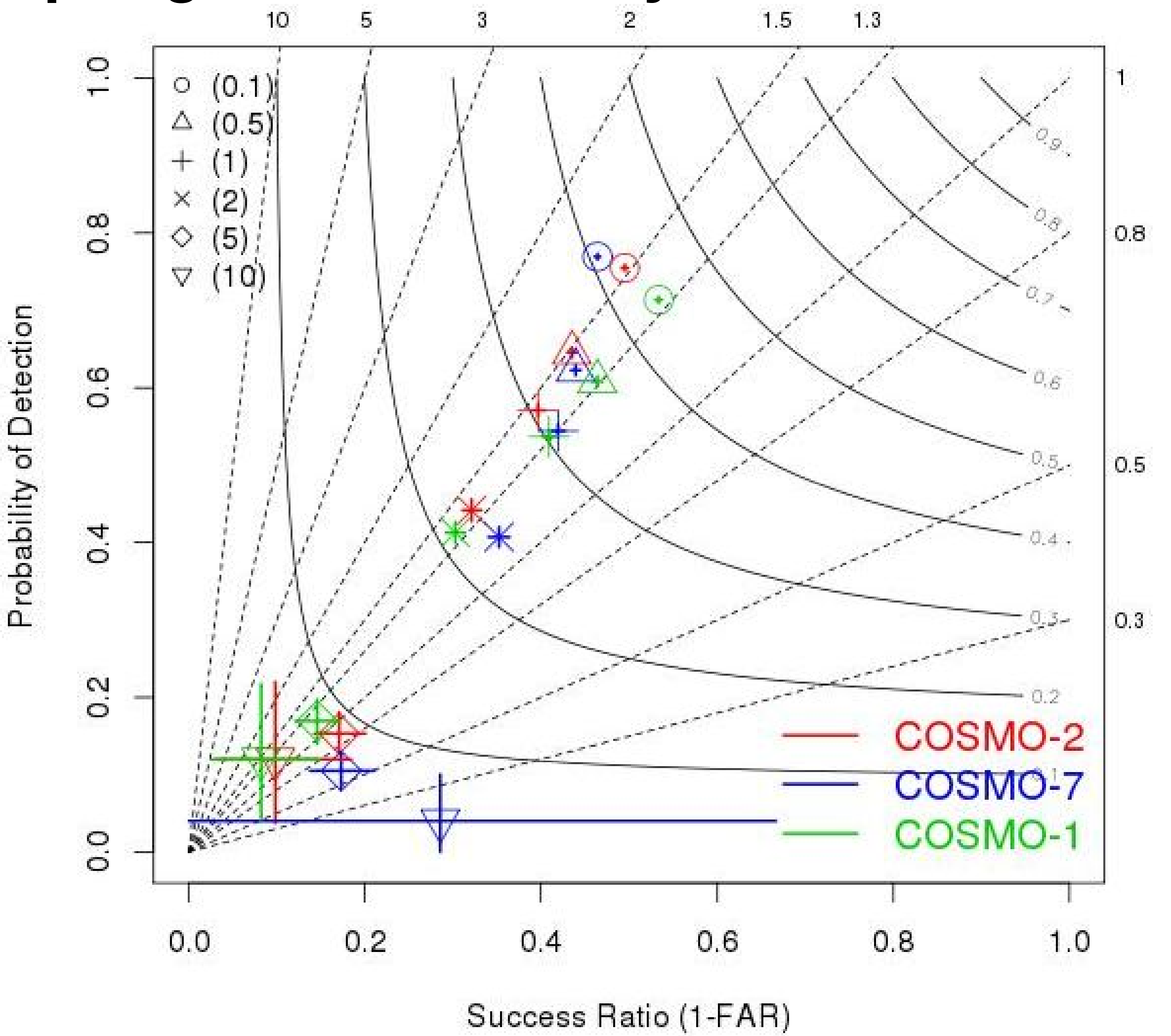
std dev





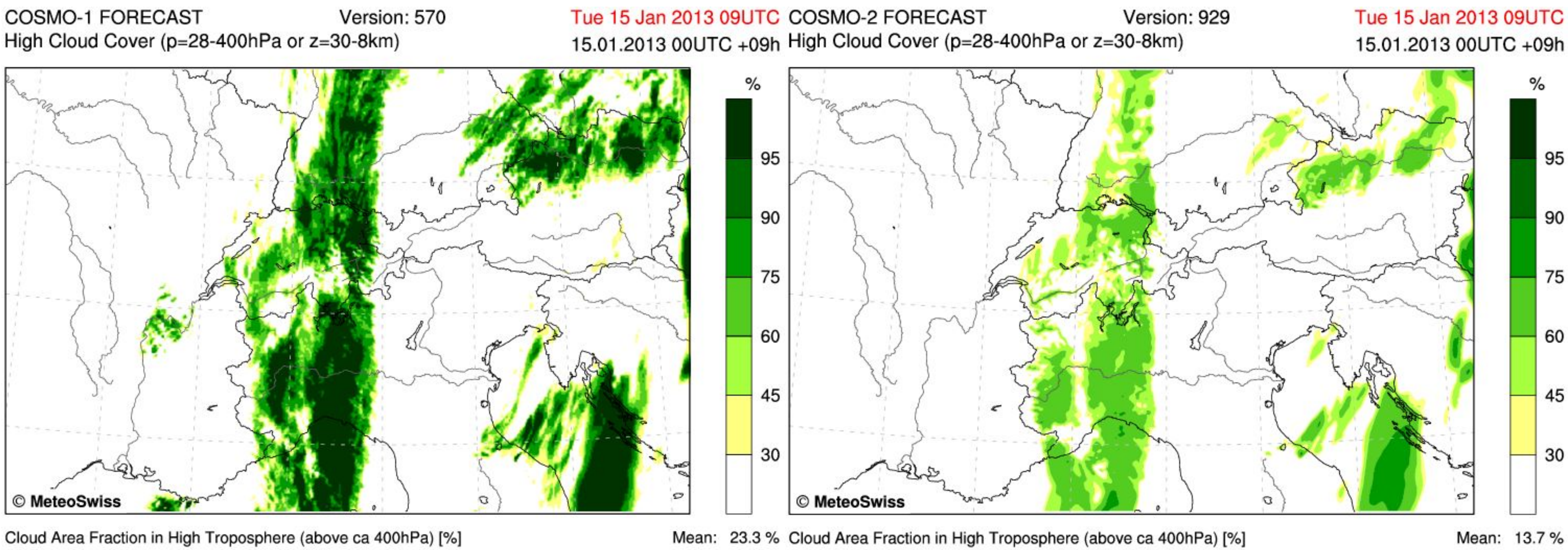
Higher windspeed due to missing of subgrid scale orography

Performance diagram: 1h precipitation sums
Spring 2013: all hourly sums from +12 to +24h



High cloud cover: 15.01.13 00 UTC +9h

COSMO-1 COSMO-2

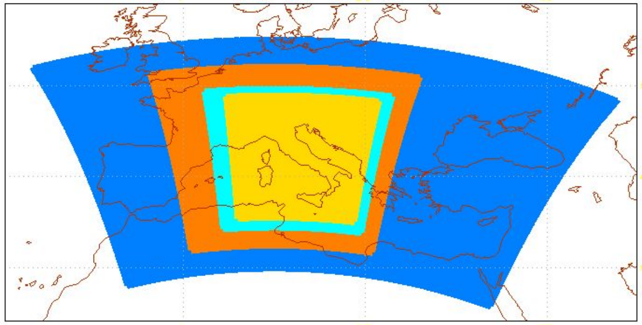


new version 4.23 (in COSMO-1): now the model can really achieve 100% cloud cover for cirrus clouds. Before the maximal cloud cover has been at about 80%, which is not realistic

1. Common Verification Framework

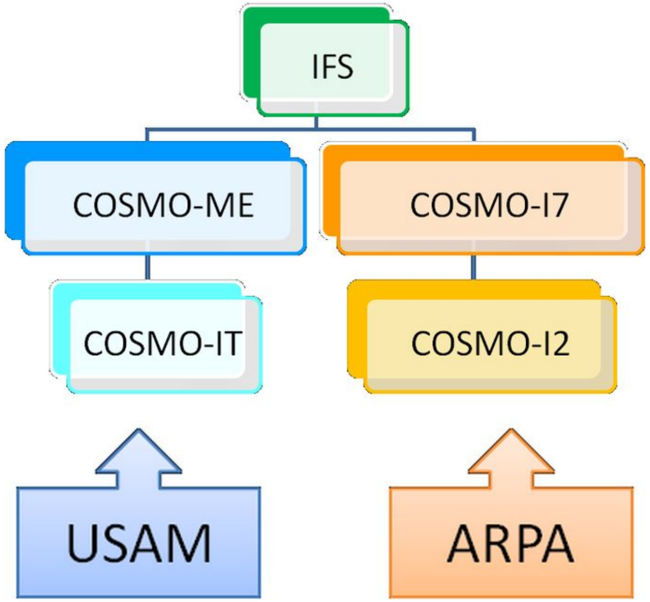
Verification of vertical profiles using TEMP observations, aircraft data (AMDAR) and wind-profiler data (Task 1.3)

Datasets

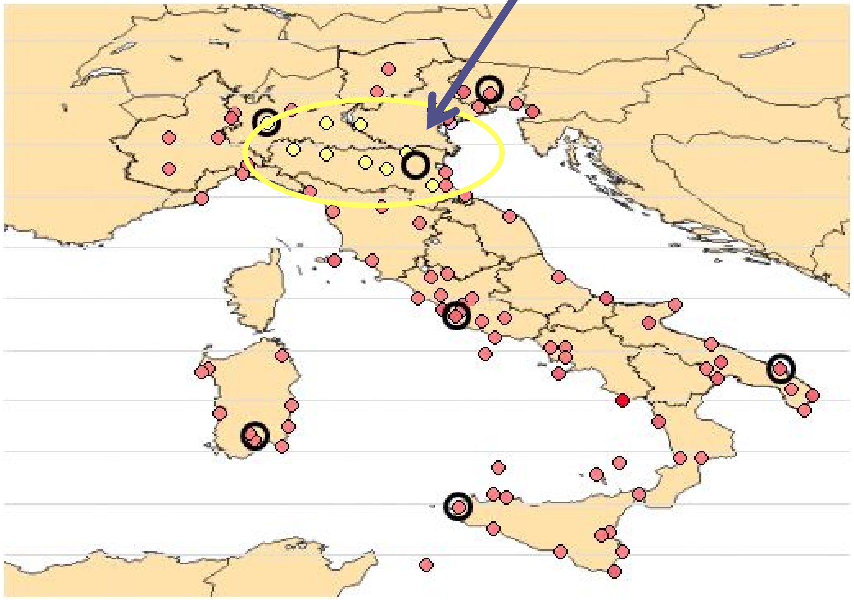


7 km

2.8 km



San Pietro Capofiume



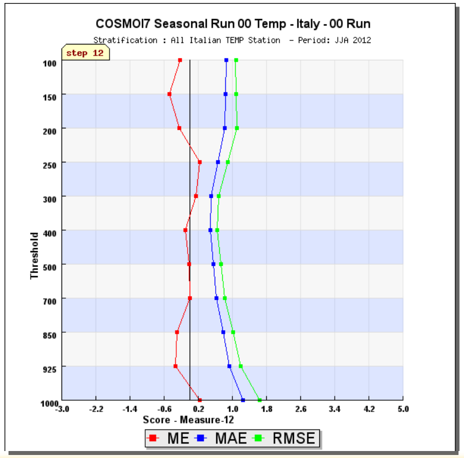
Surface stations:

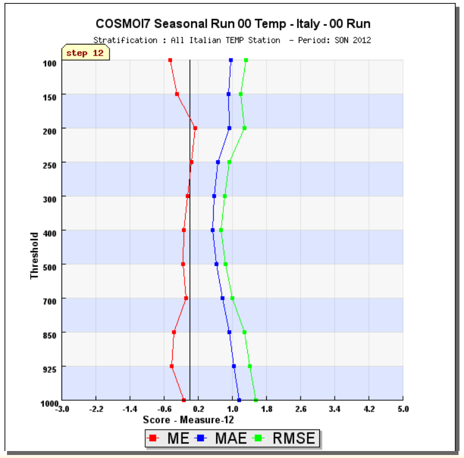
- All Italian Stations (~130)
- Po valley Stations (9)

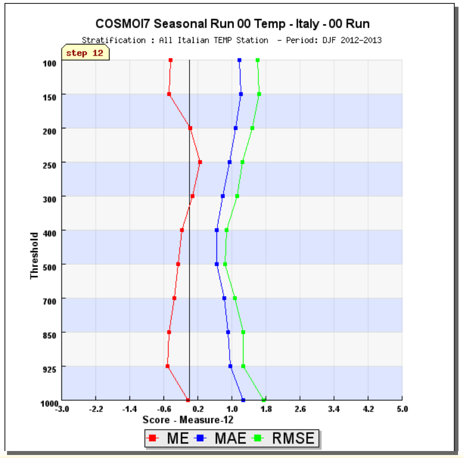
O Sounding Stations (7)

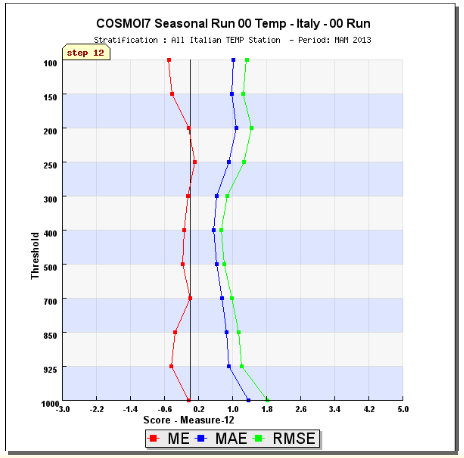
00 UTC and 12 UTC

Upper air: Temperature COSMO-I7



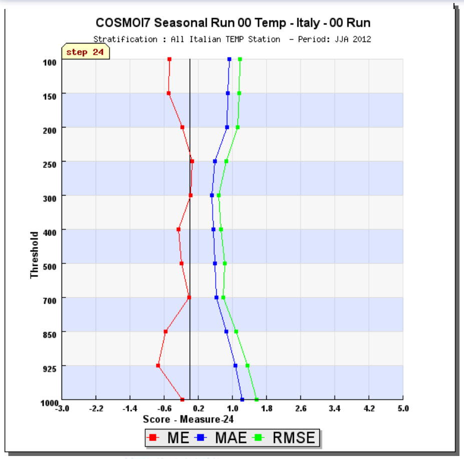


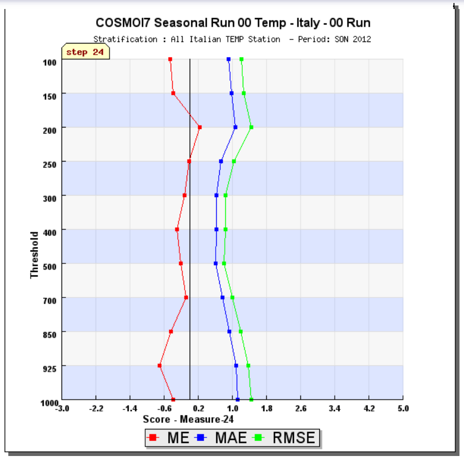


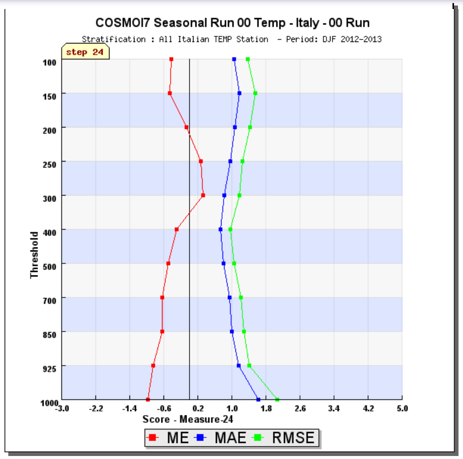


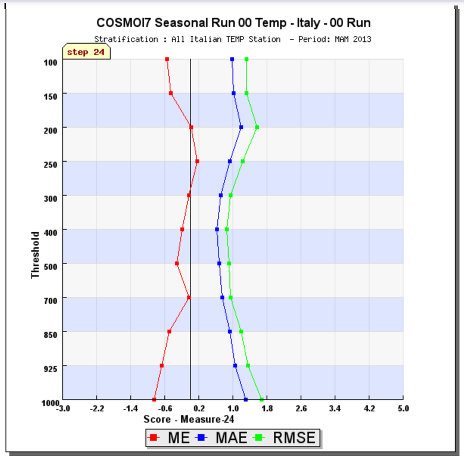
00 UTC

12 UTC



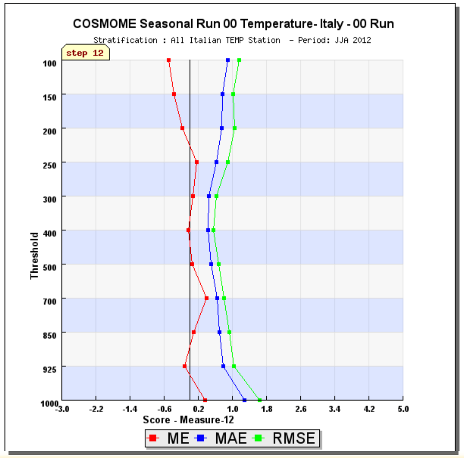


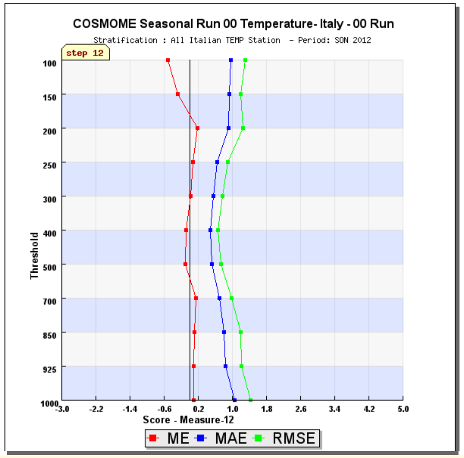


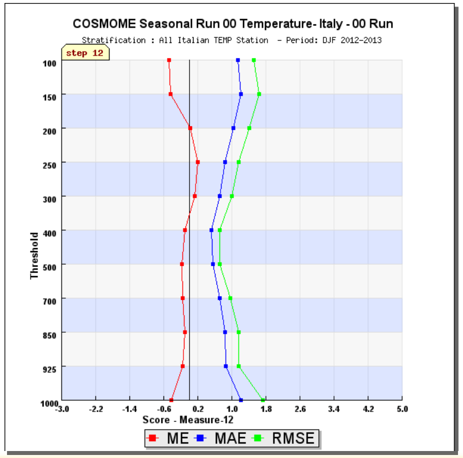


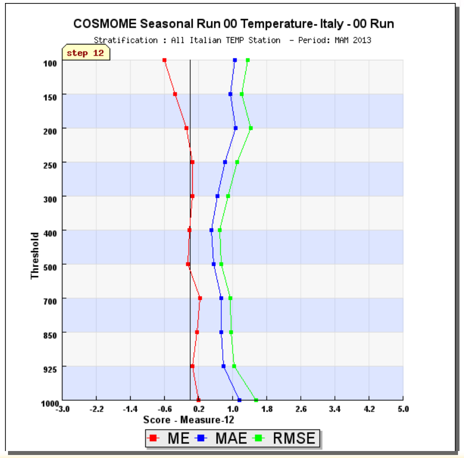
RUN 00 UTC -FORECAST DAY 1

Upper air: Temperature COSMO-ME



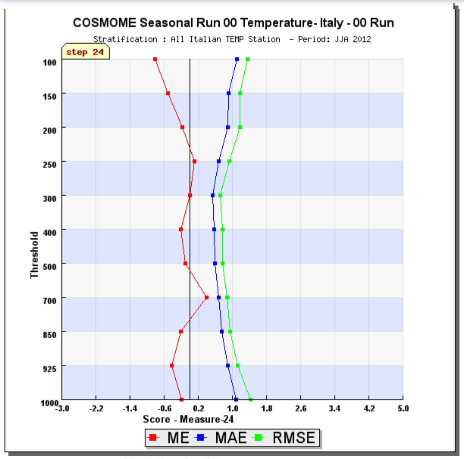


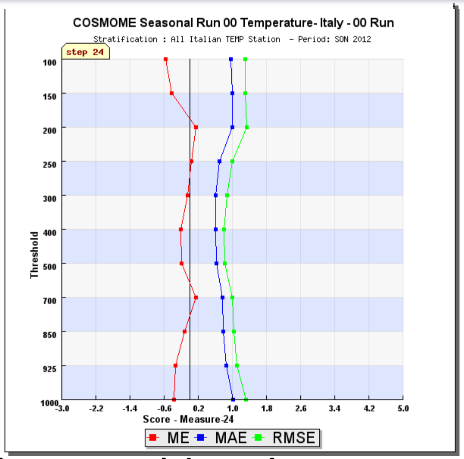


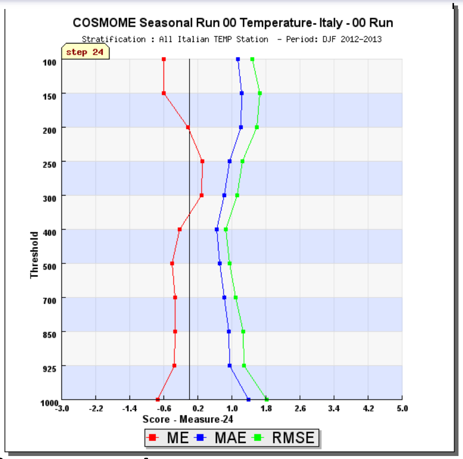


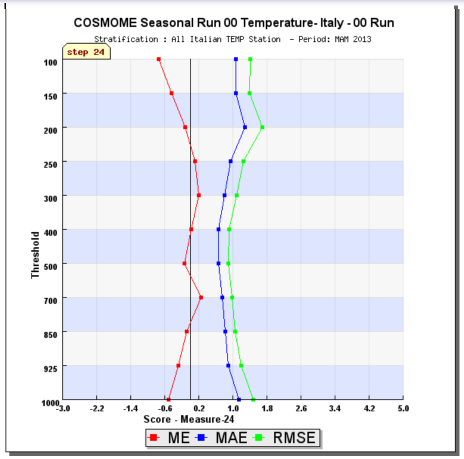
00 UTC

12 UTC







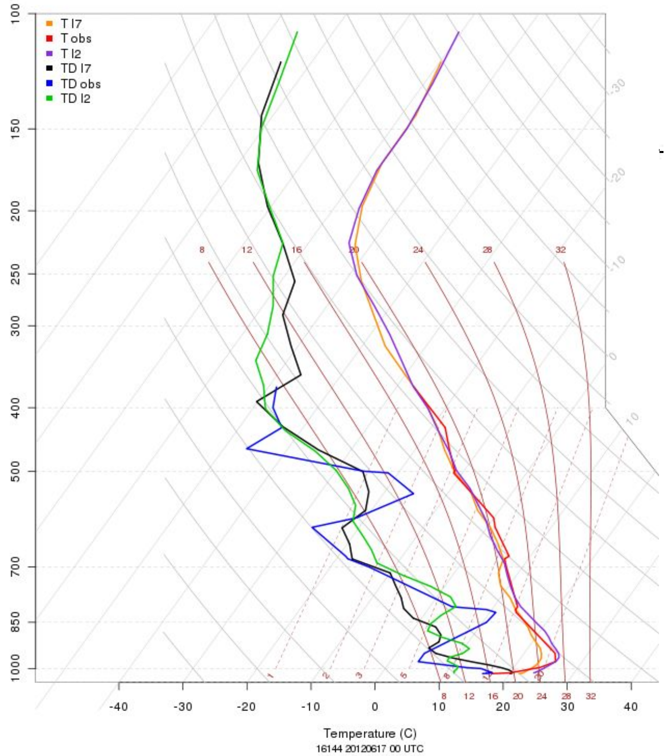


COSMOI7 always colder close to surface with respect to COSMOME

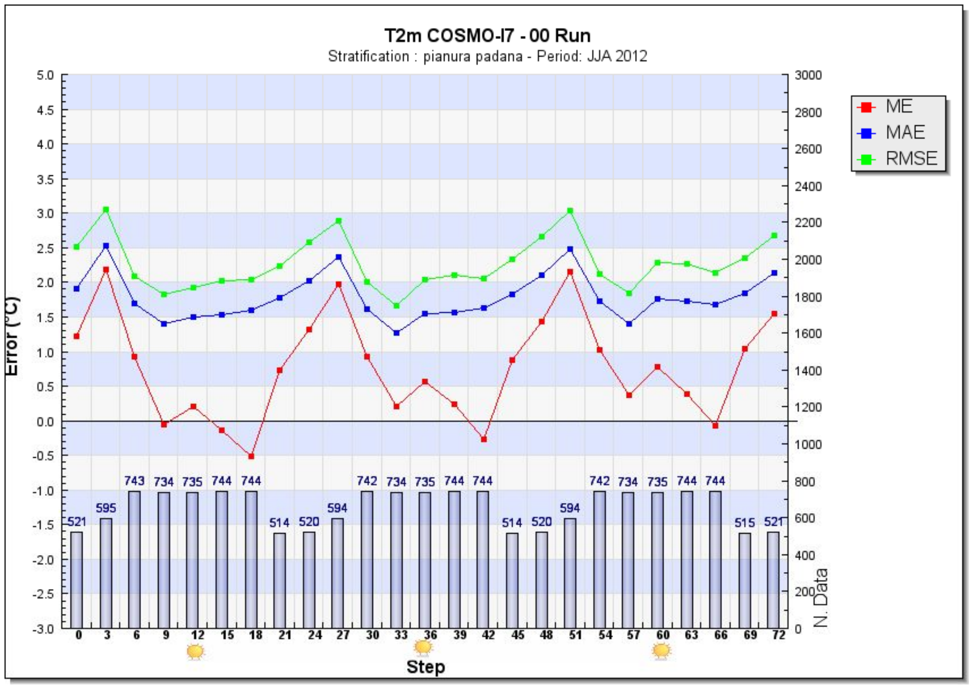
RUN 00 UTC -FORECAST DAY 1

Typical situation in Po Valley

Example of sounding at 00 UTC in San P.Capofiume (near Bologna)

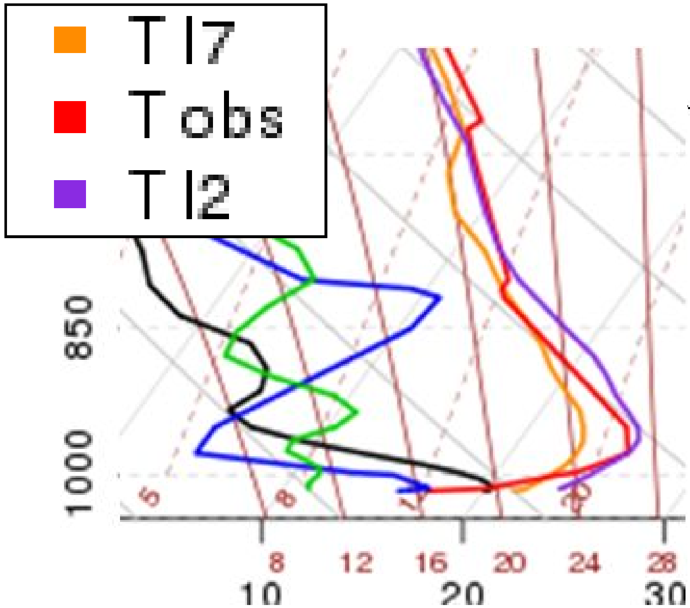


COSMO-I7 T2m verification in the Po Valley

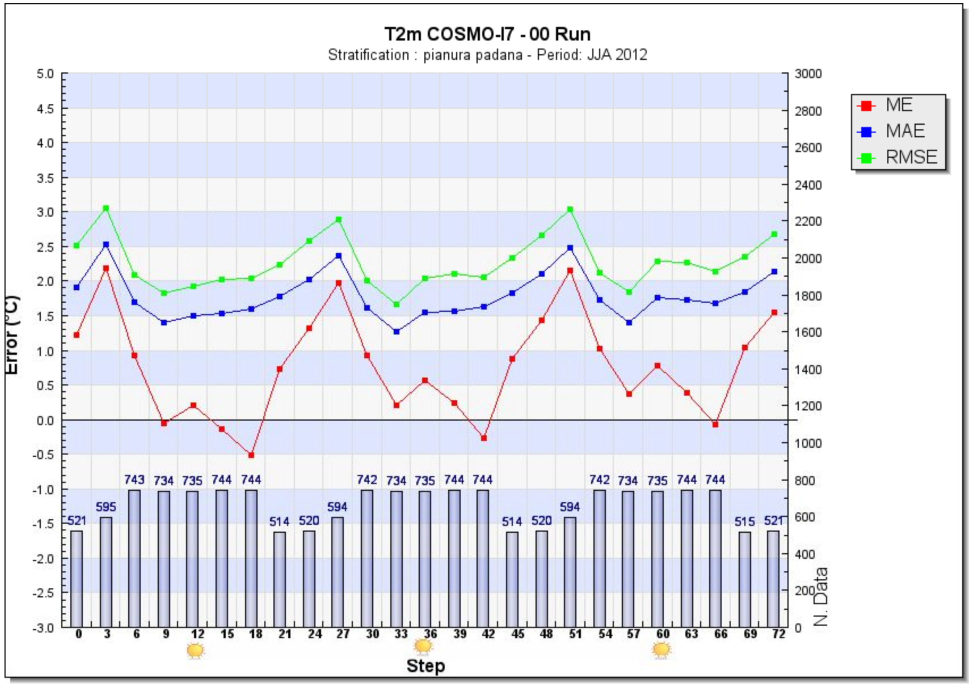


Typical situation in Po Valley

Example of sounding at 00 UTC in San P.Capofiume (near Bologna)

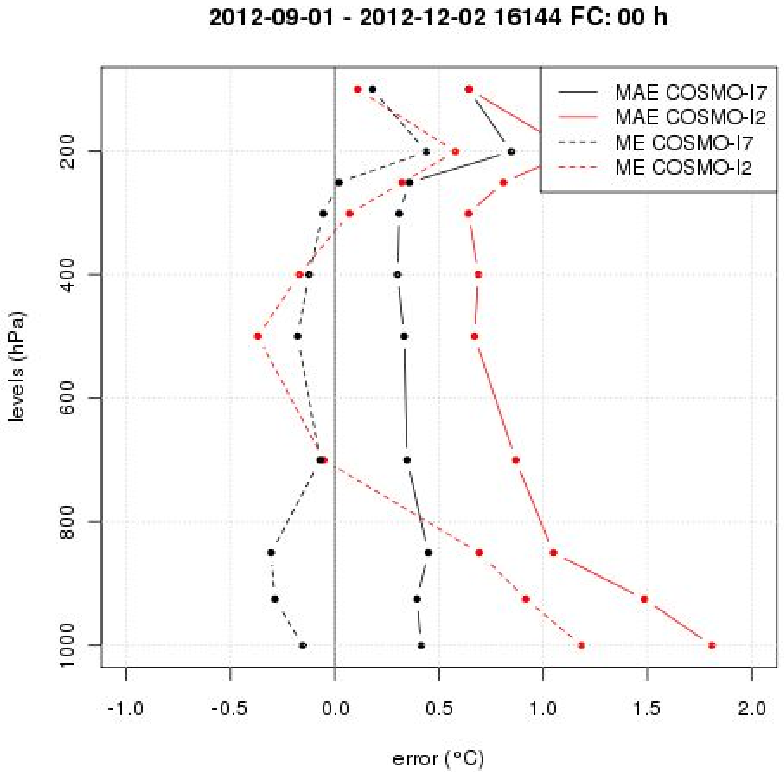


COSMO-I7 T2m verification in the Po Valley

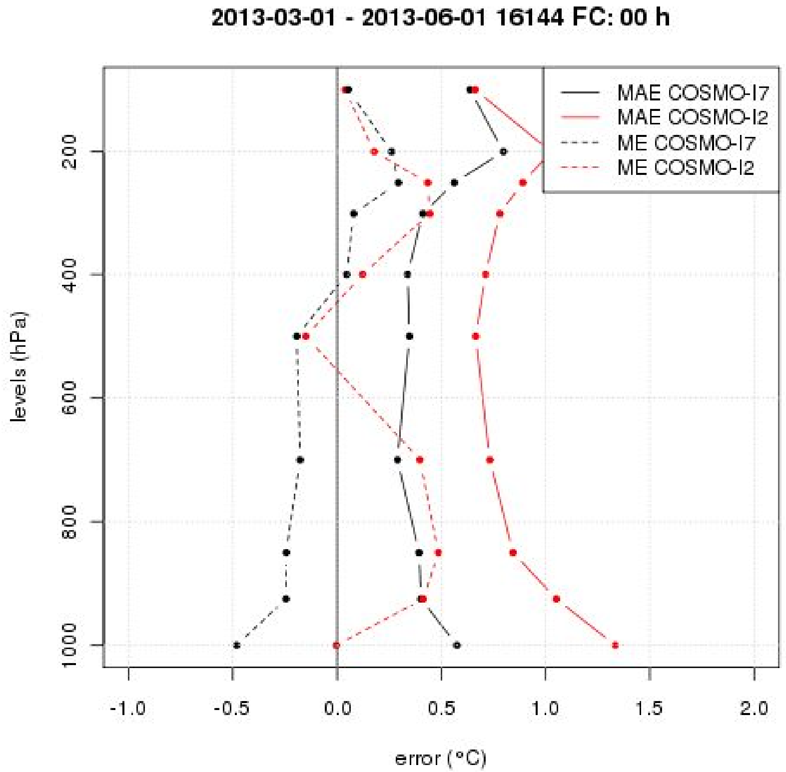


Upper Air Temperature verification: San Pietro Capofiume

SON 2012 – FC +00h 00UTC

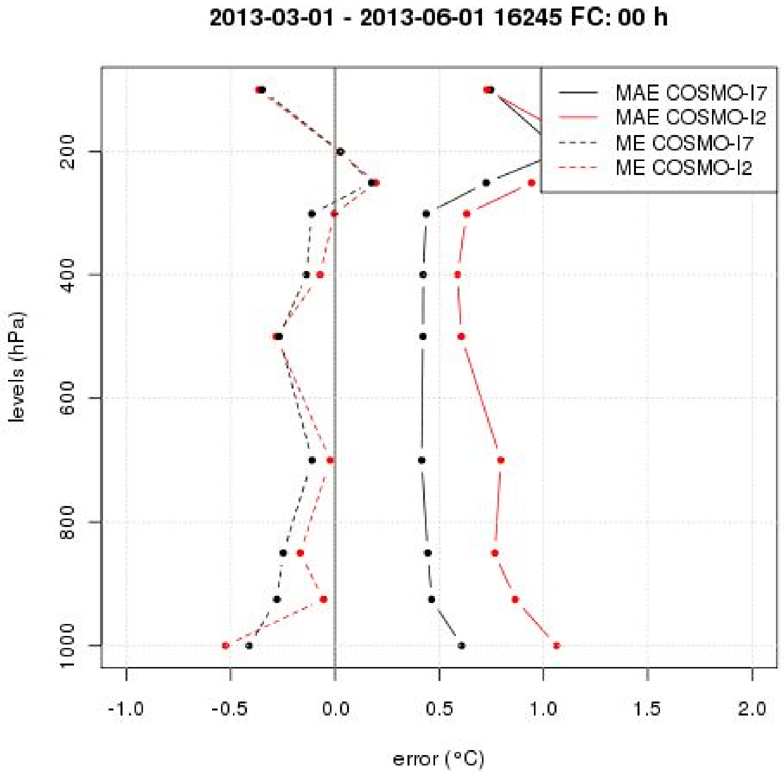


MAM 2013 – FC +00h 00 UTC

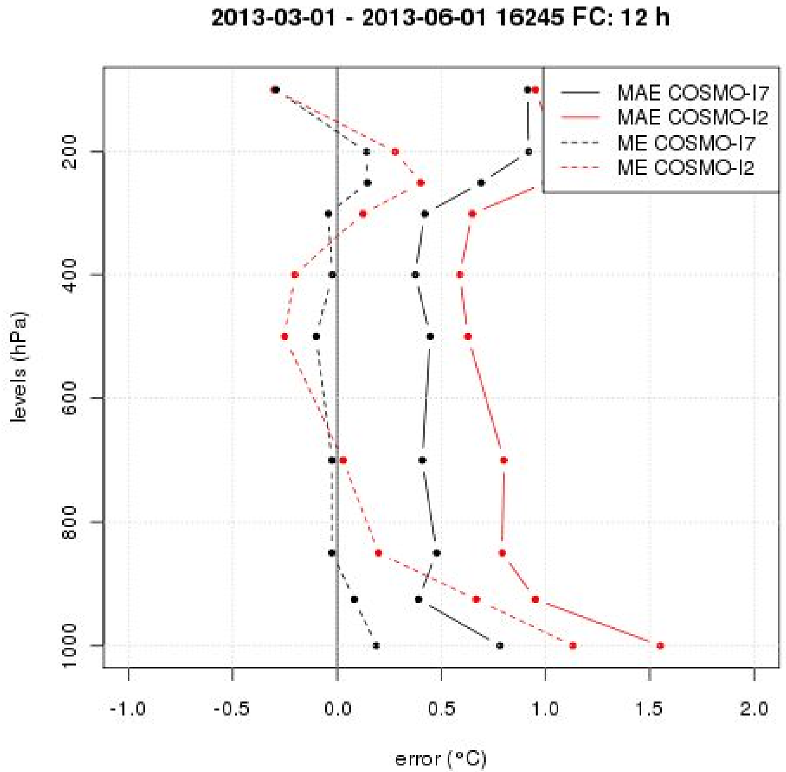


Upper Air Temperature verification: Pratica di Mare

MAM 2013 – FC +00 h 00 UTC

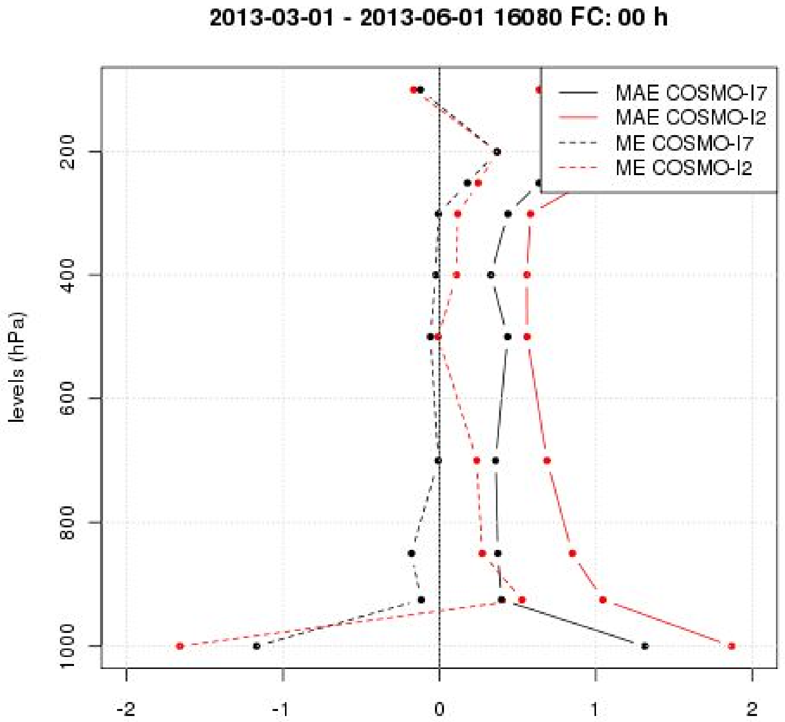


MAM 2013 – FC +00 h 12 UTC

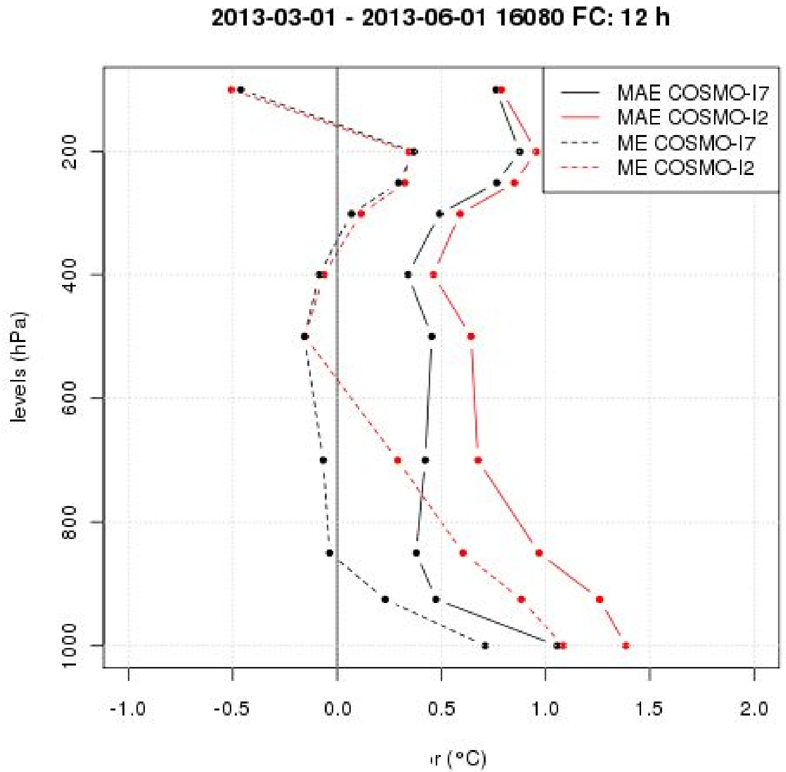


Upper Air Temperature verification: Milano

MAM 2013 – FC +00 h 00 UTC

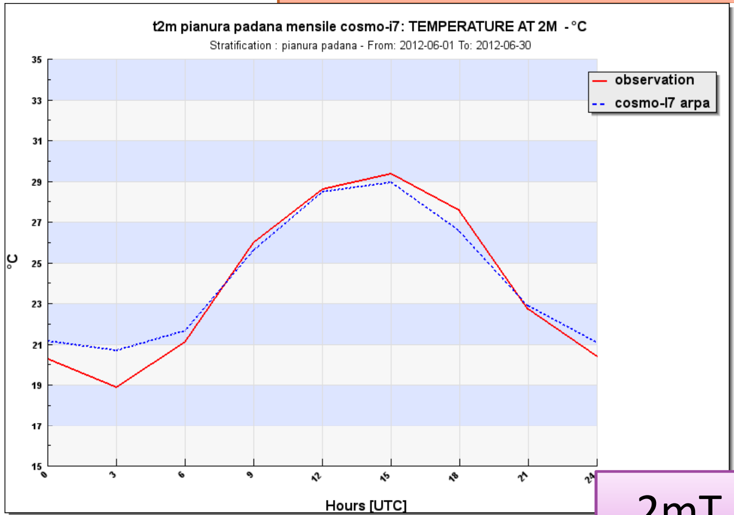


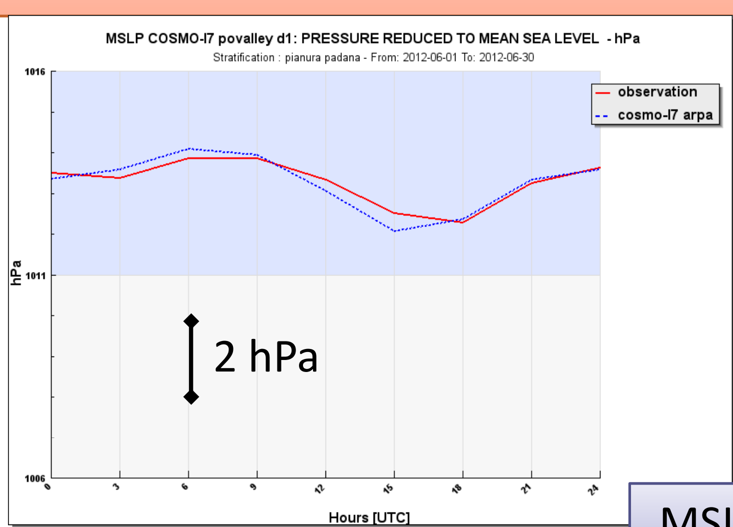
MAM 2013 – FC +00 h 12 UTC



Daily Cycle: COSMO-I7/I2
Po Valley June 2012

7 Km

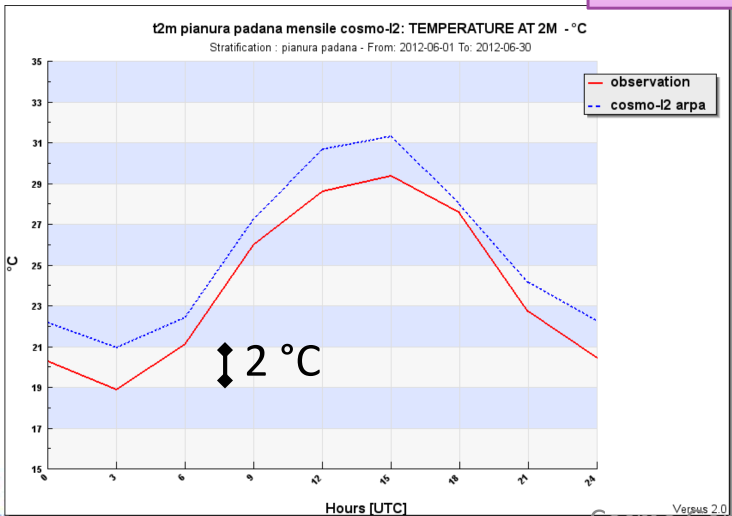


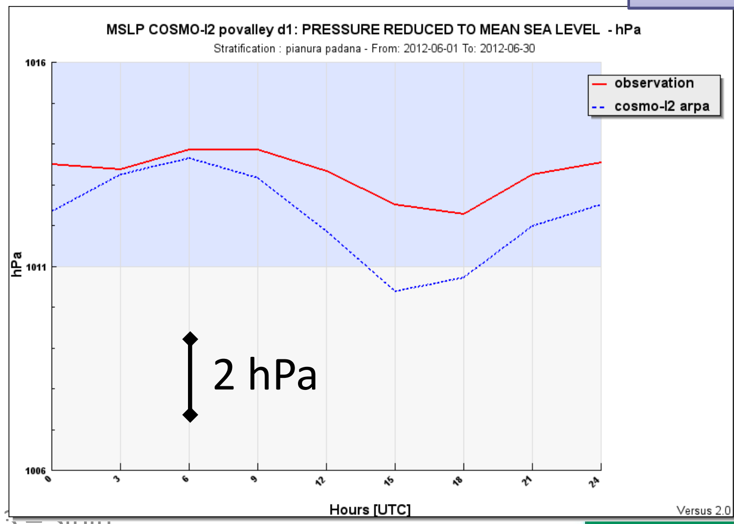


2mT

MSLP

2.8 Km





Overestimation of 2mT in both CI7, CI2 for early morning and night

Daily Cycle: COSMO-2 Po Valley June 2012

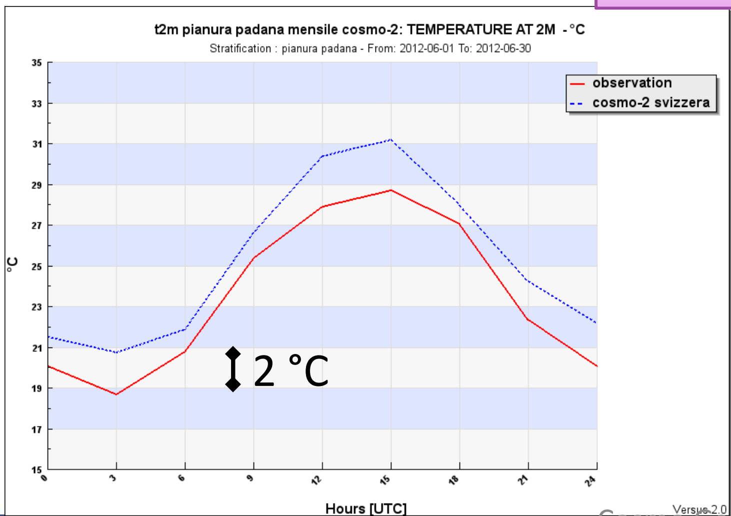
sorry
not available
at this time

7 Km

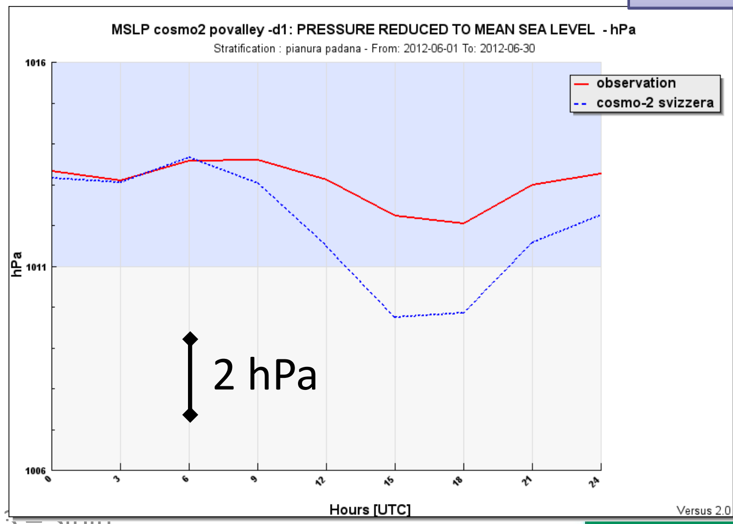
sorry
not available
at this time

2mT

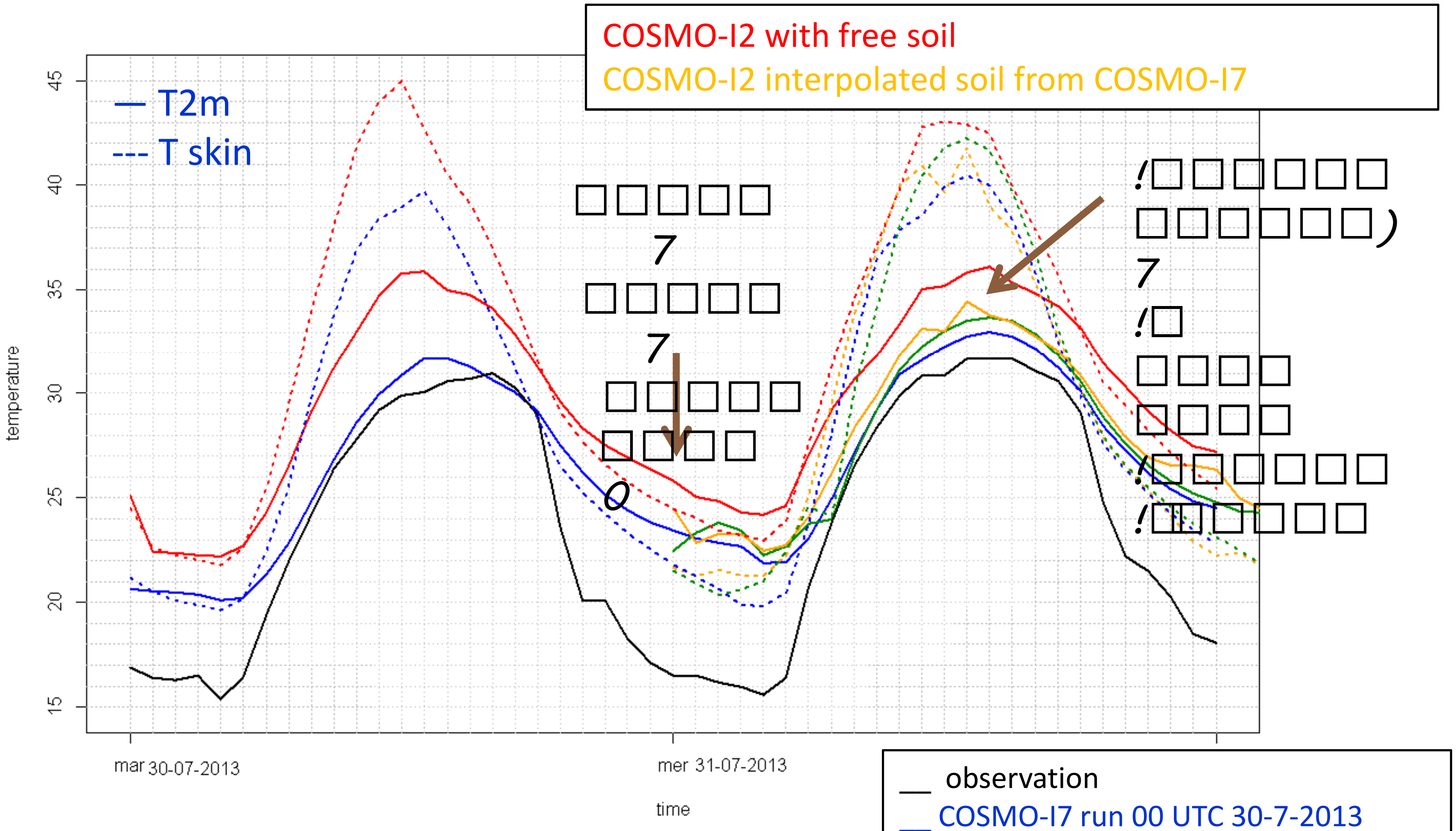
MSLP



2.8 Km

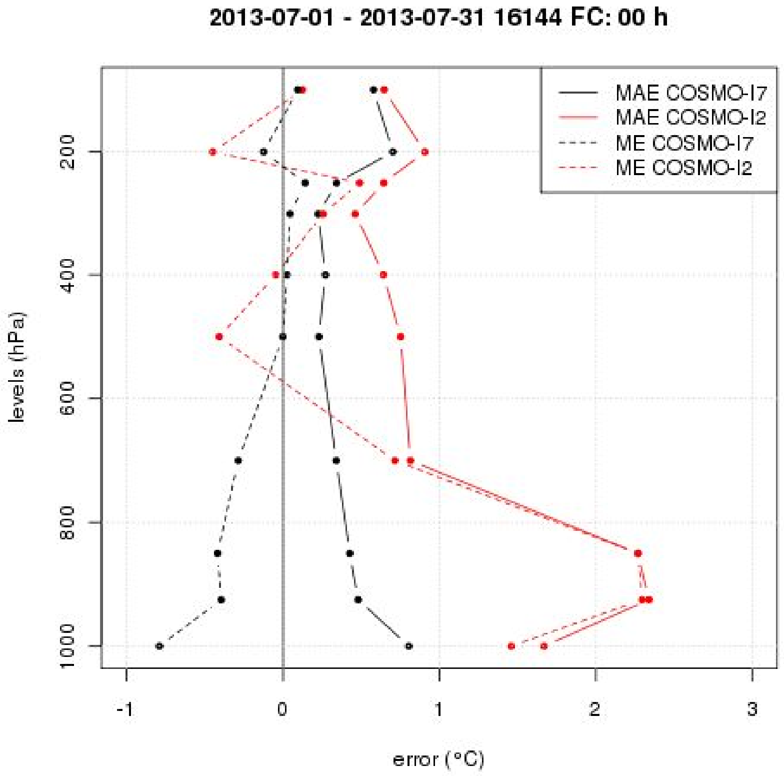


Impact of different soil on T2m: a quick test

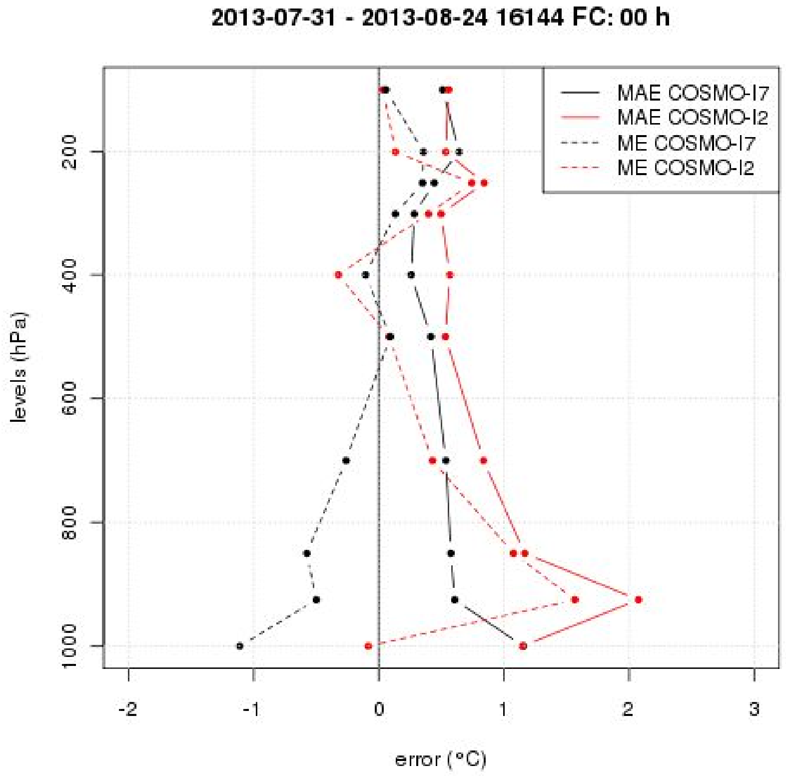


Impact of different soil on vertical temperature profile: quick verification

San Pietro Capofiume (16144)
COSMO-I2 free soil

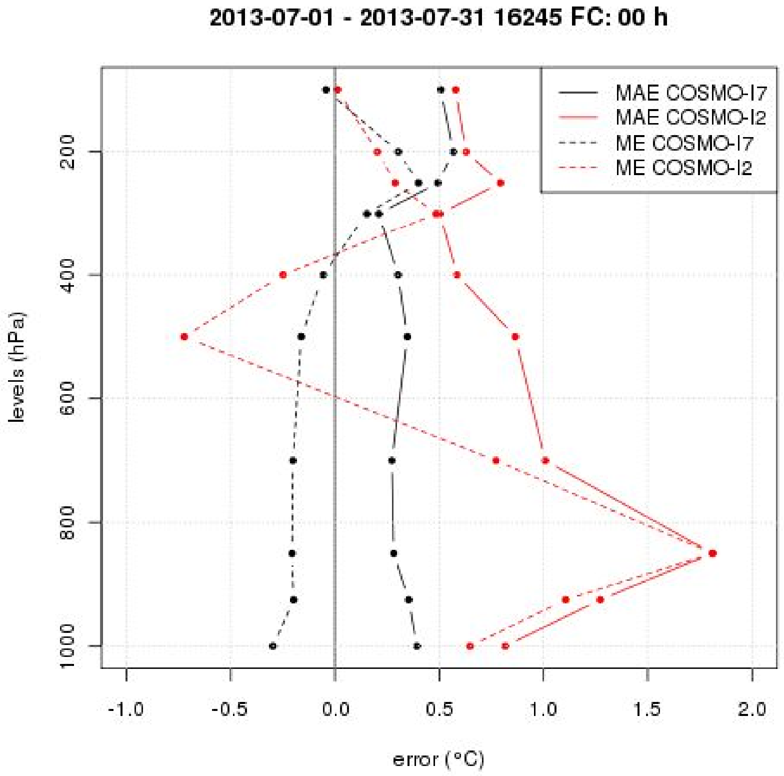


San Pietro Capofiume (16144)
COSMO-I2 interpolated soil

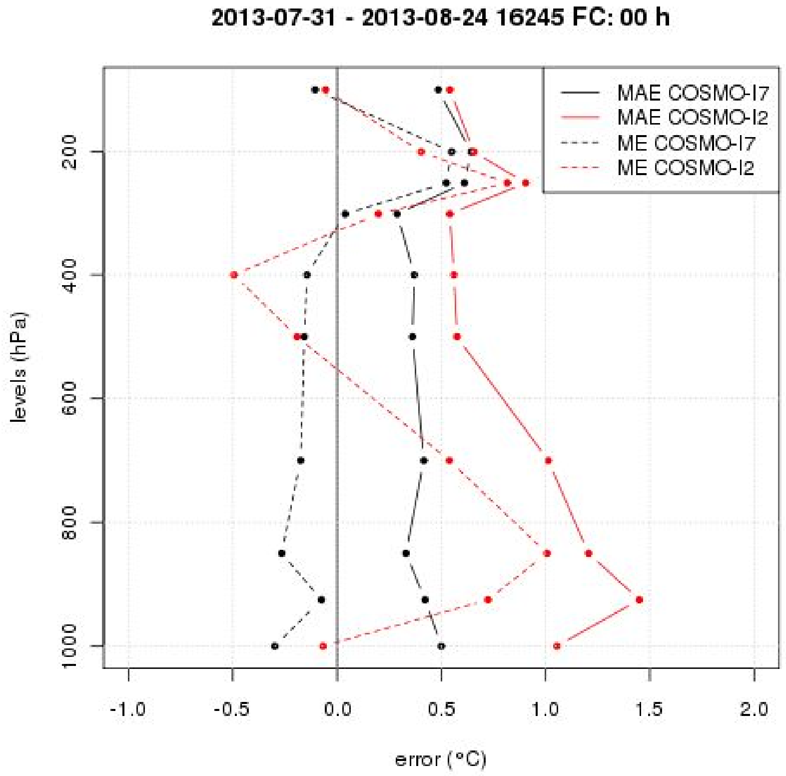


Impact of different soil on vertical temperature profile: quick verification

Pratica di Mare (16245)
COSMO-I2 free soil



Pratica di Mare (16245)
COSMO-I2 interpolated soil

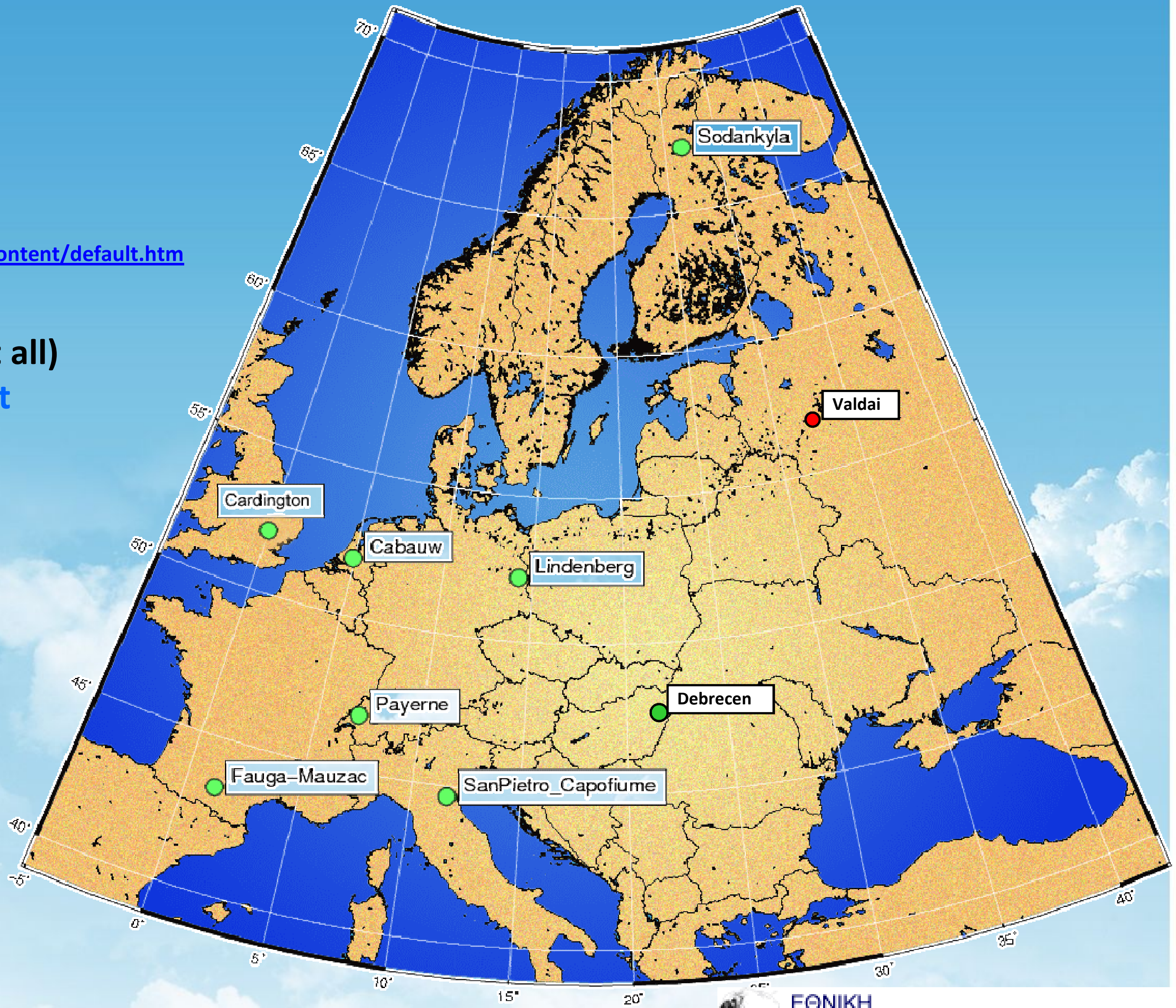


Evaluation of COSMO models in the lower PBL (Task 2.3)

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.

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

VERSUS Implementation

- Front-end to upload Data pool ASCII files for obs
- Calculation of Radiation balances from obs
- Calculation and storage of hourly averaged quantities for obs and fcs
- Daily Cycles
- Time series

Next implementations:

- Scatterplots
- Average on different period (3, 6, 12 hours) – if necessary
- Use of Obs and Fcs for standard and Conditional verification

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

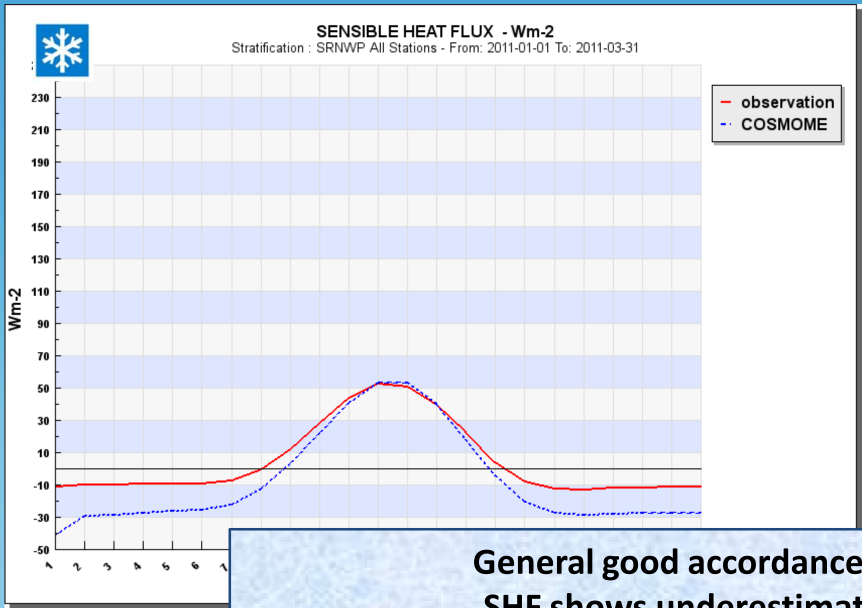
Obs and Fcs data availability

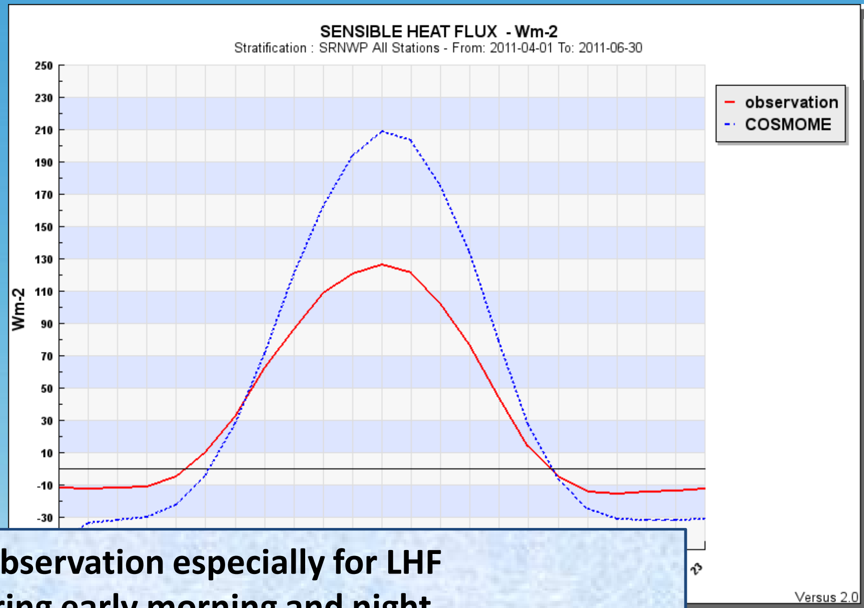
OBS data	**FCS data**
RSWD: incoming solar radiation	ASWDIR_S aver. direct downward Sw rad. surface
RSWU: reflected solar radiation	ASWDIFD_S aver. diffuse downward Sw rad. Surface
	Avg. Balance of SW
RLWD: incoming thermal radiation	ALWD_S aver. downward Lw radiation at the surface
RLWU: outgoing thermal 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

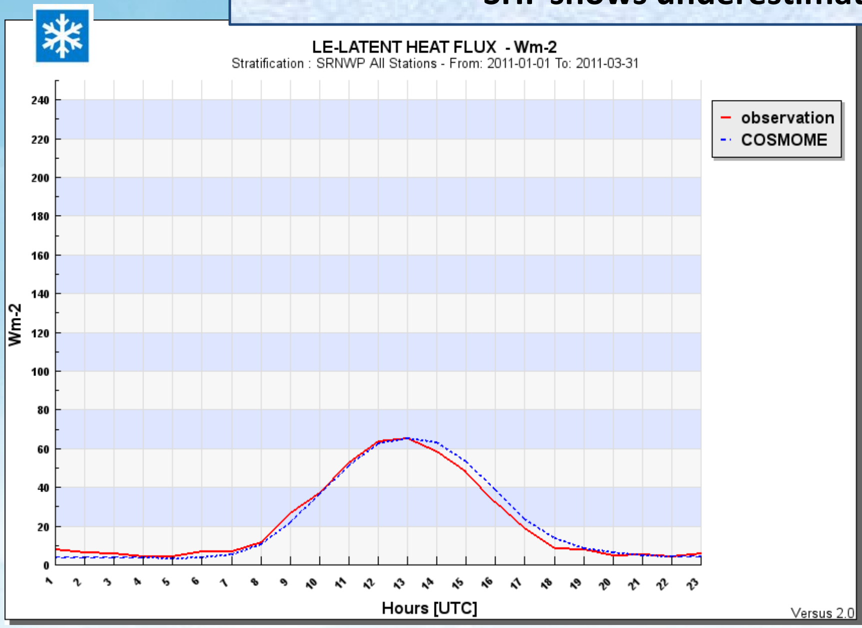
ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

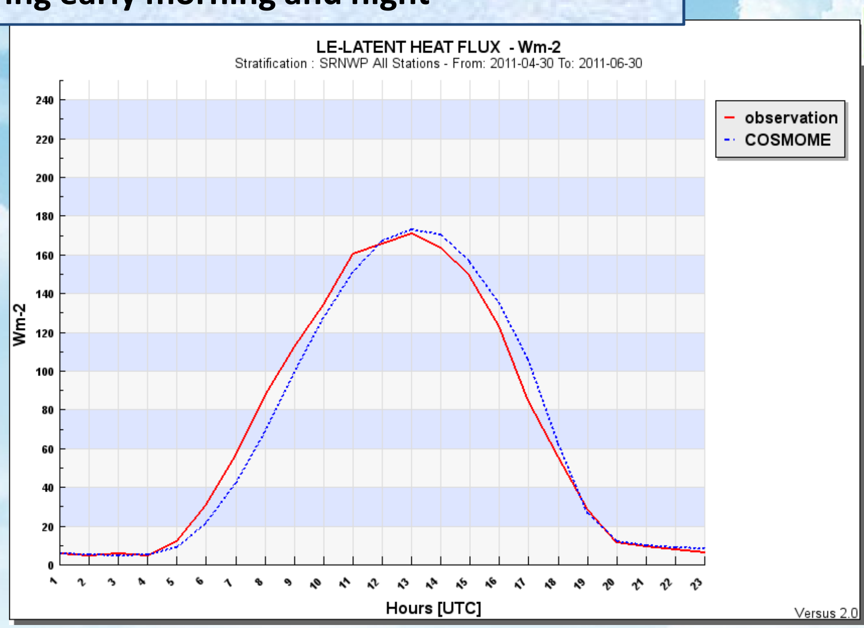
Some Results



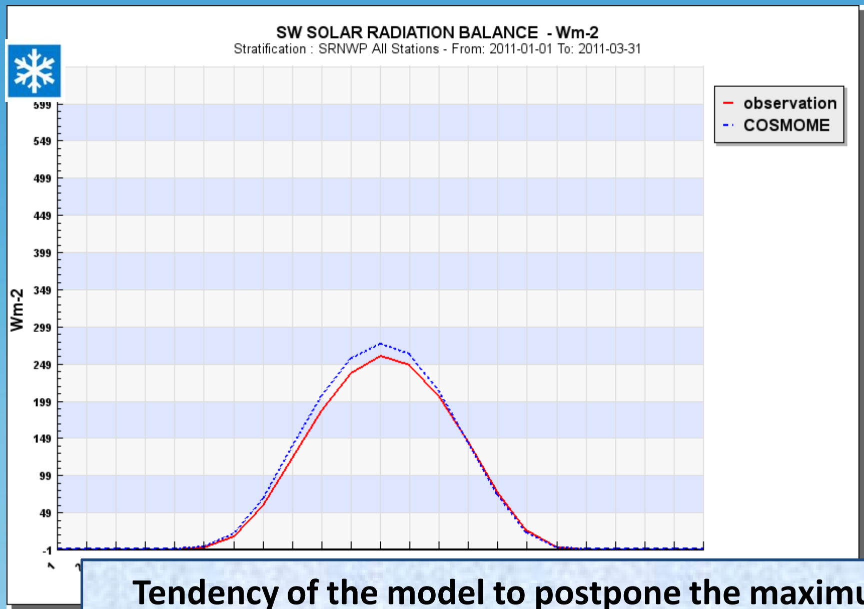


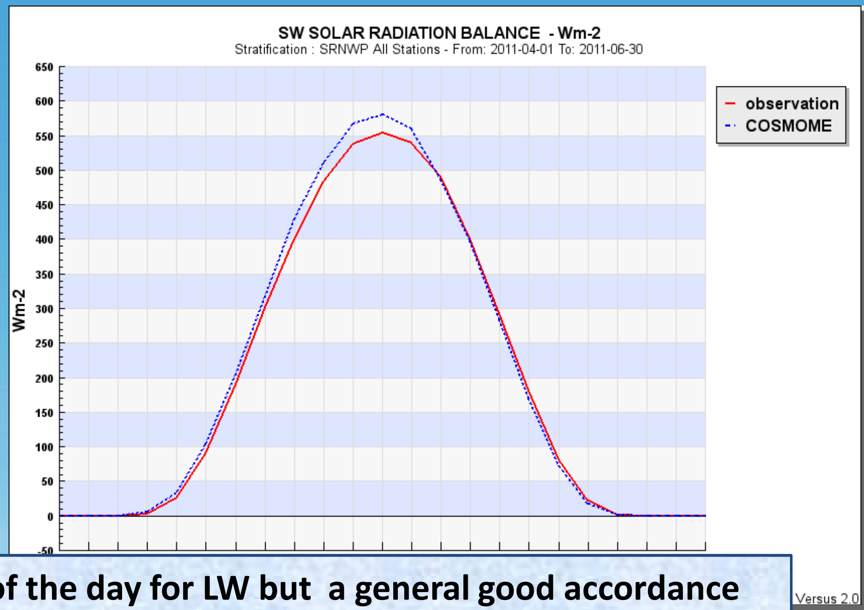
General good accordance with observation especially for LHF
SHF shows underestimation during early morning and night



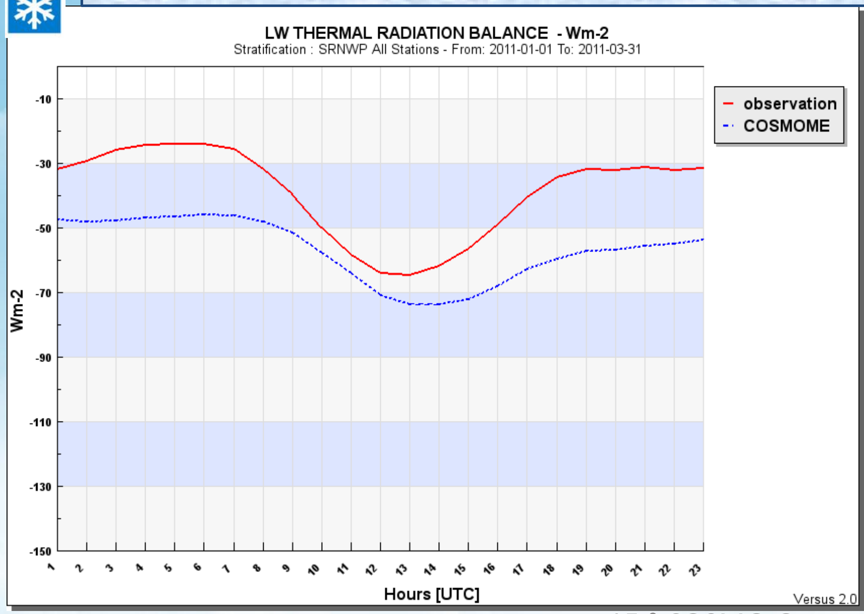


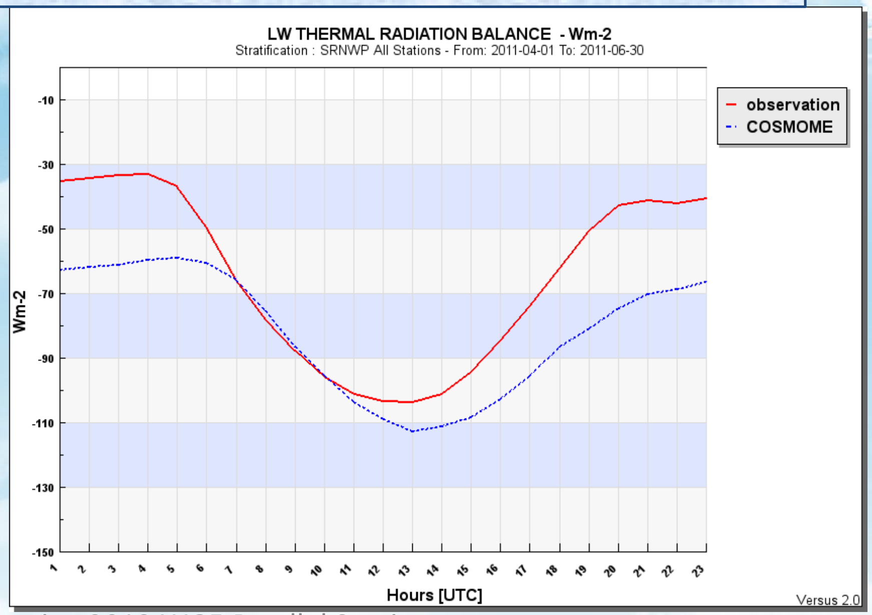
Some Results



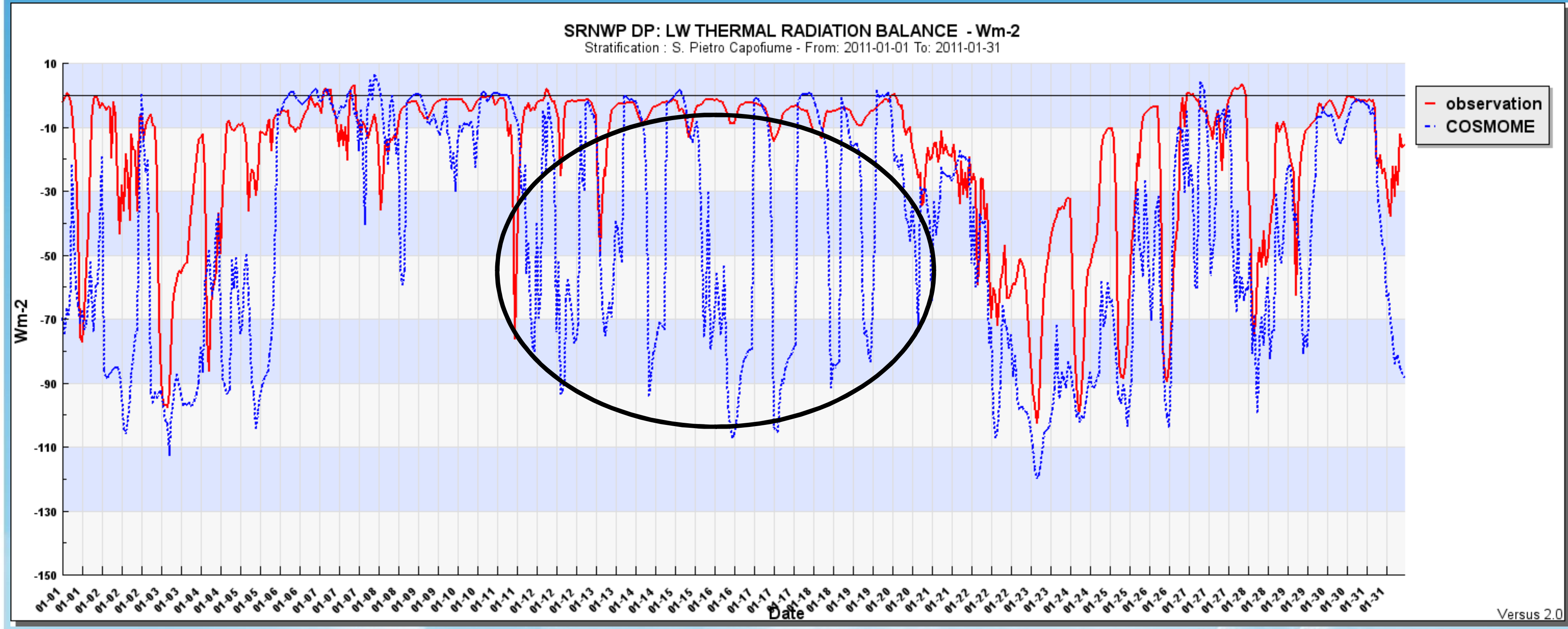


Tendency of the model to postpone the maximum of the day for LW but a general good accordance with observation. Minimum in LW more pronounced in Spring





Something interesting
Time series for July 2011 LWR balance

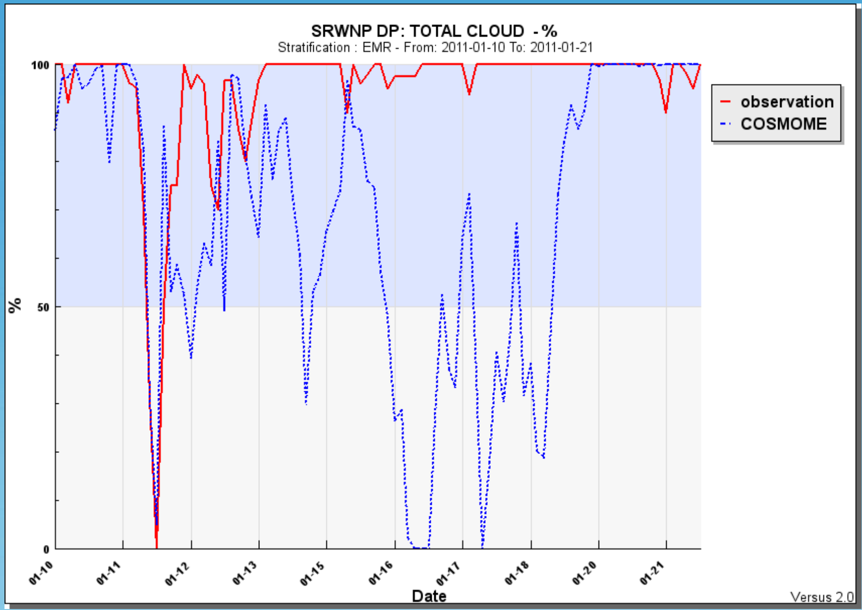


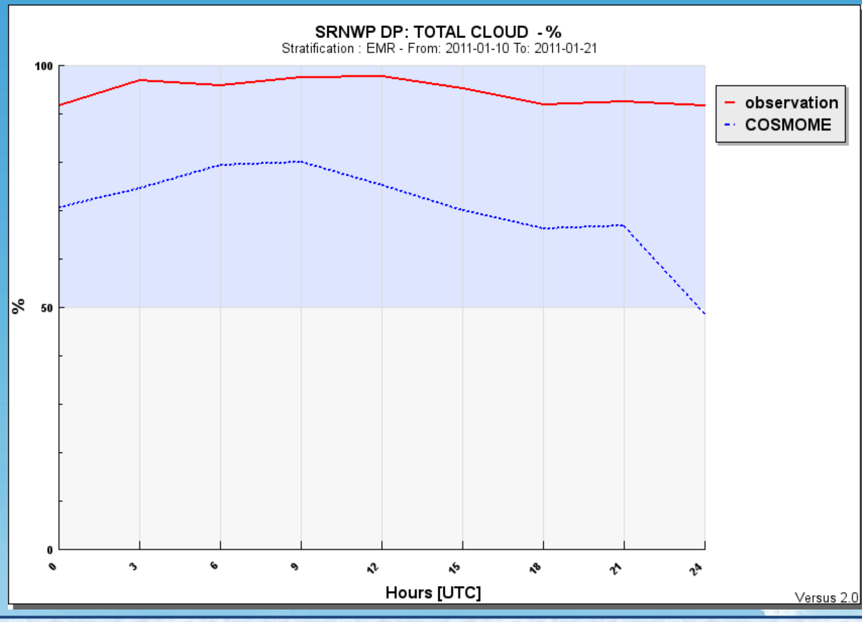
From 10/01/2011 to 20/01/2011 the model gave a completely wrong LW

SERVIZIO METEOROLOGICO DELL'AERONAUTICA

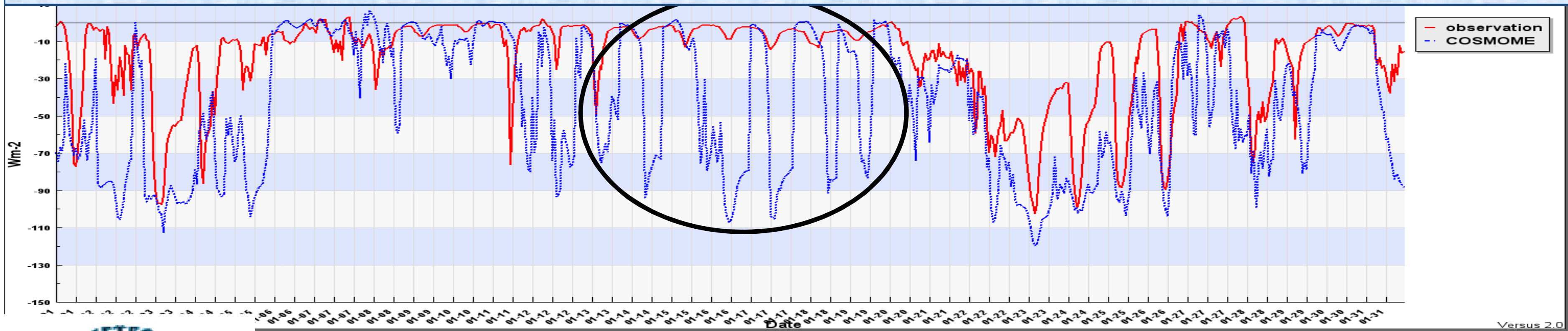
ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

Something interesting





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



ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

Verification of Global Radiation With Hourly Measurements Over Switzerland

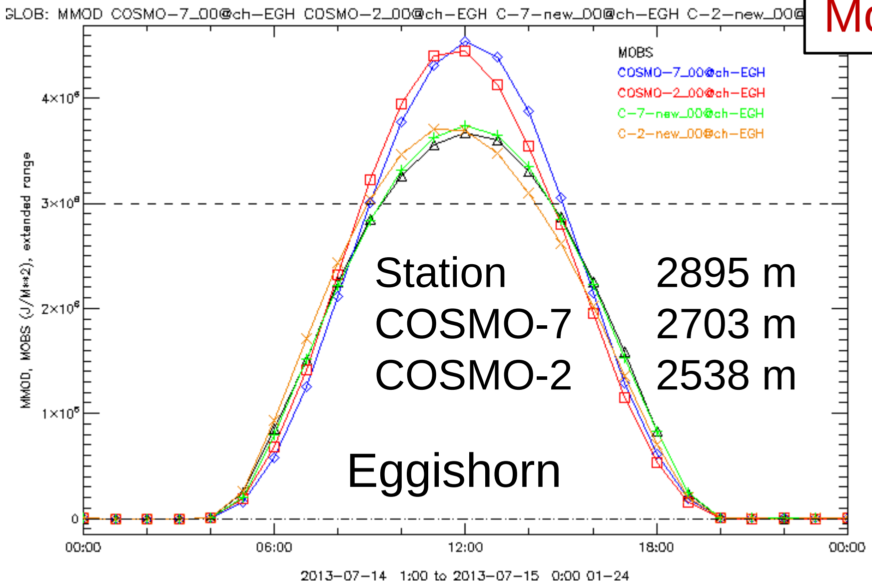
COSMO GM - WG5 Parallel Session

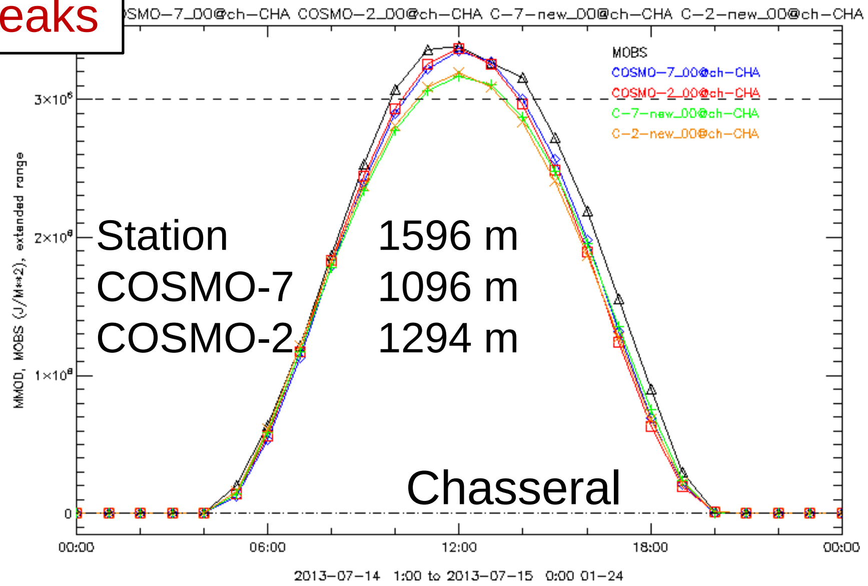
Global Radiation from the Model

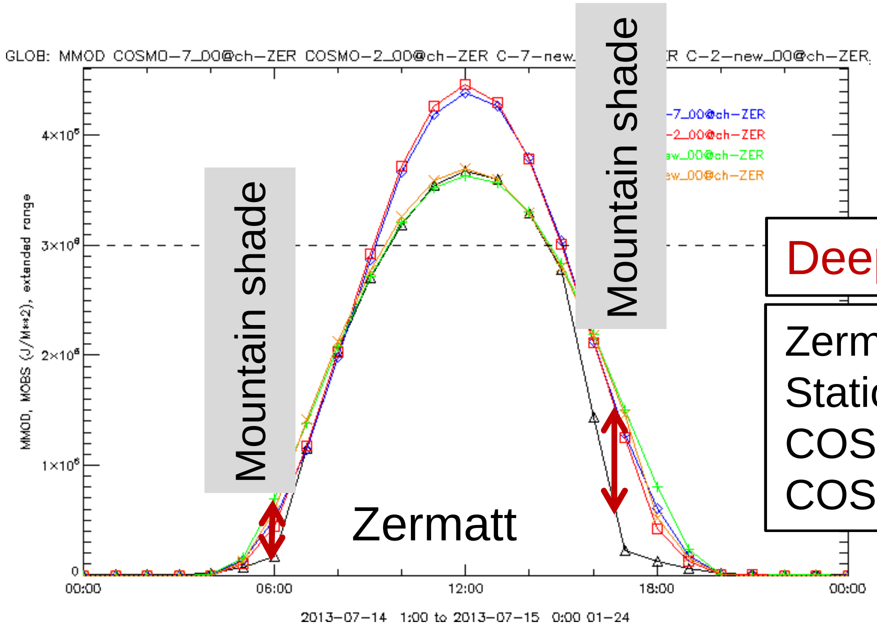
- **Old** approximation (e.g. "Beschreibung des COSMO-DE-EPS und seiner Ausgabe in die Datenbanken des DWD", 2012)
  - GLOB = ASOB_S / (1 - ALB_RAD)
  - Caveats:
    - ALB_RAD is the albedo for the diffuse radiation only
    - ALB_RAD is an instantaneous value, ASOB_S an accumulated value → inconsistency
- **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





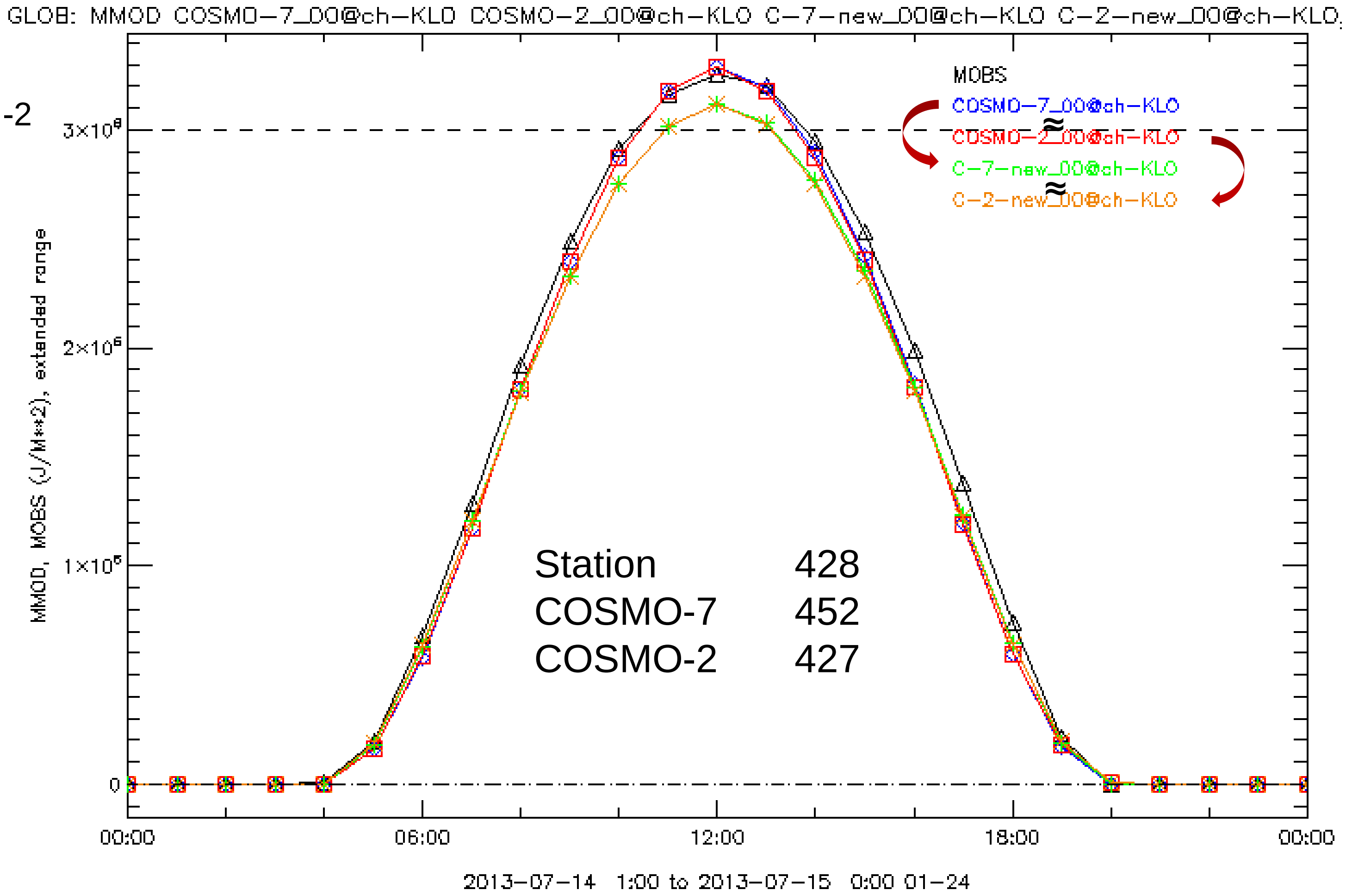


Deep narrow valley

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}

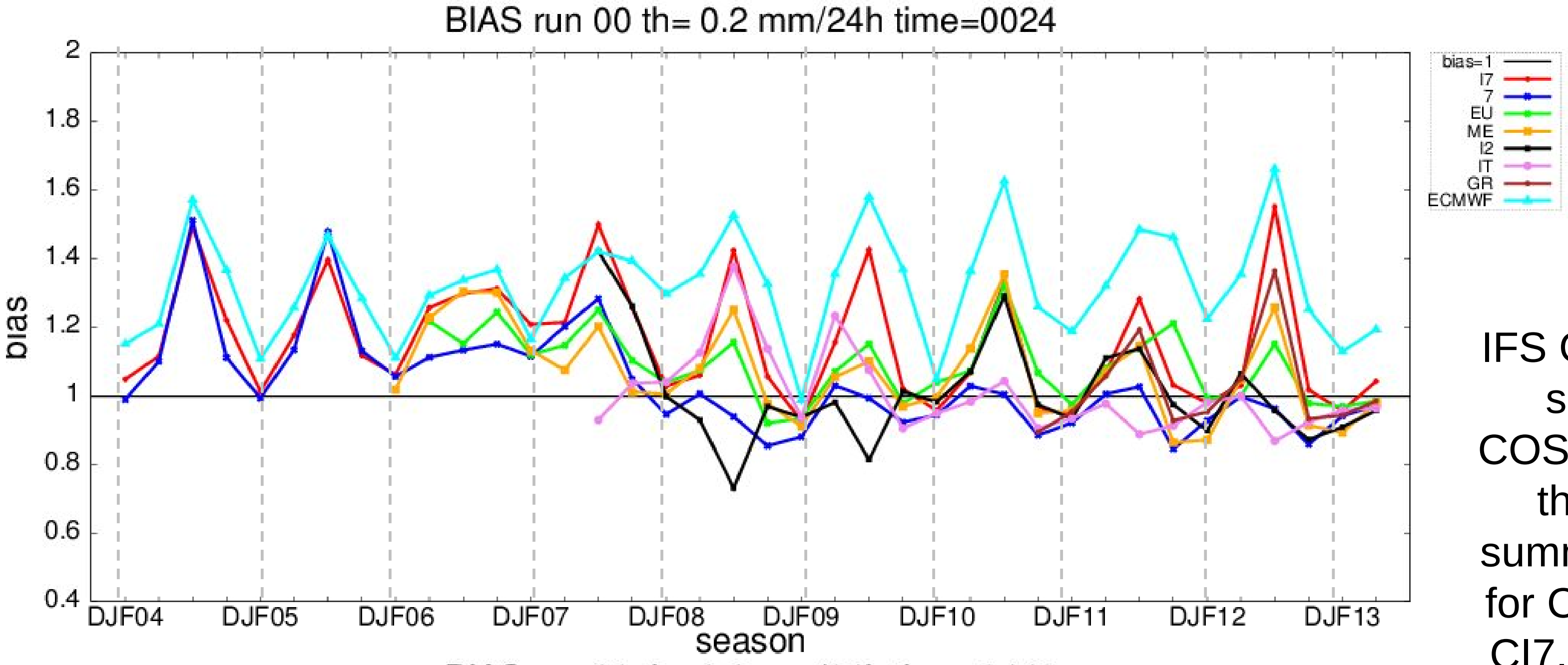


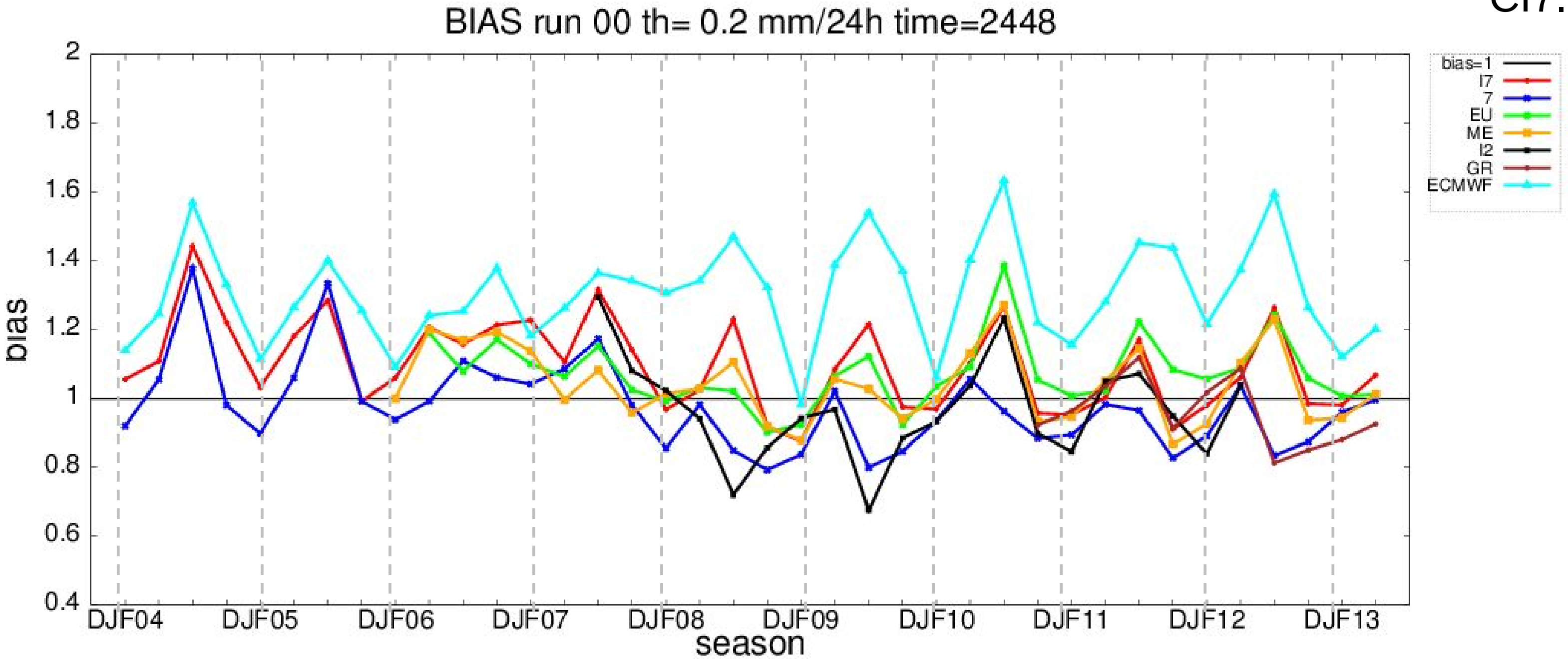
Summary

- Global radiation in COSMO-7 and COSMO-2 is almost equal (if cloudiness is equal) -> topographical effect is small
- Old approximation using diffuse albedo (ALB_RAD) leads to systematically higher values during clear days of at least 5-10% or even greater at some locations
- New direct summing of radiation components should be used for consistency with what the model uses internally
- The global radiation is generally slightly underestimated compared to measurements during fair weather conditions especially at lower elevations

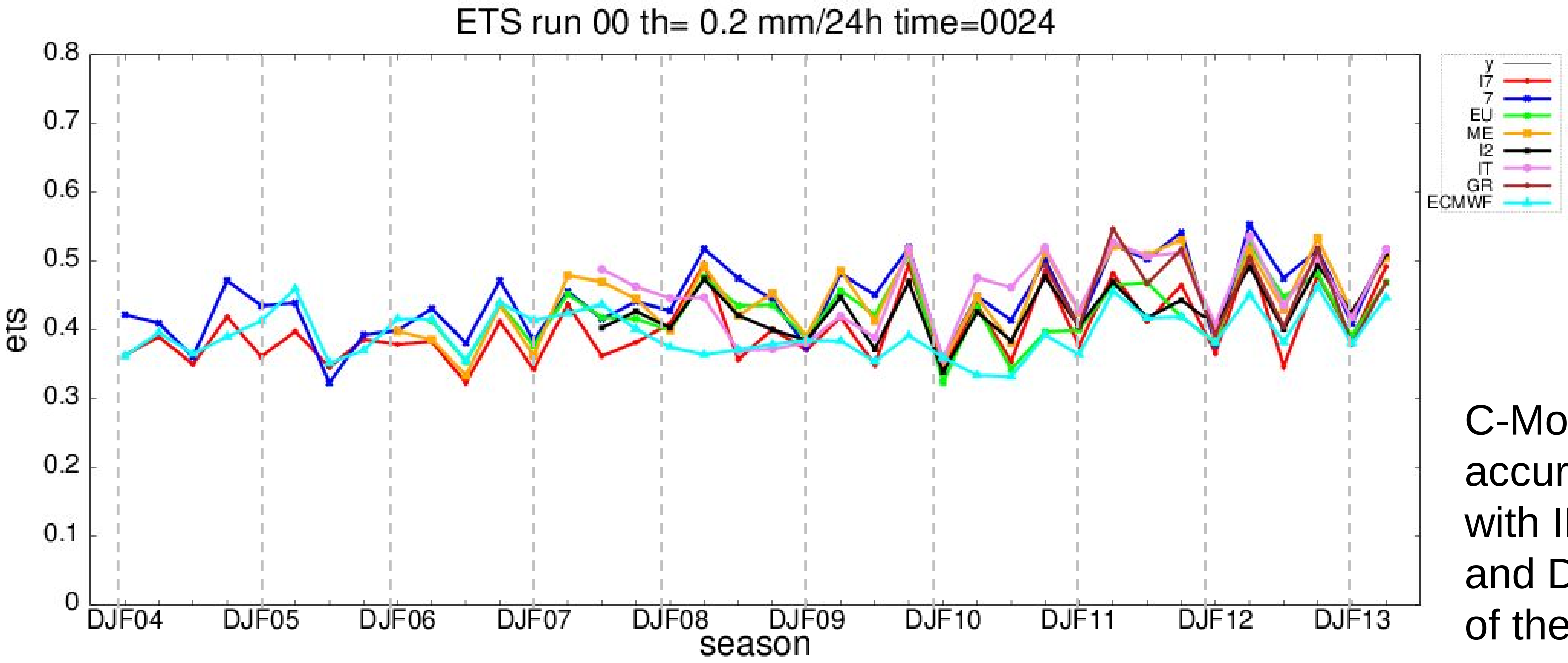
Long Term Trends (Task 3.1)

Long term trend verification of various statistical indexes (COSI index, ETS, FBI, RMSE...) of COSMO models implementation when large period set of data are available to estimate the general improvement of each COSMO implementation.

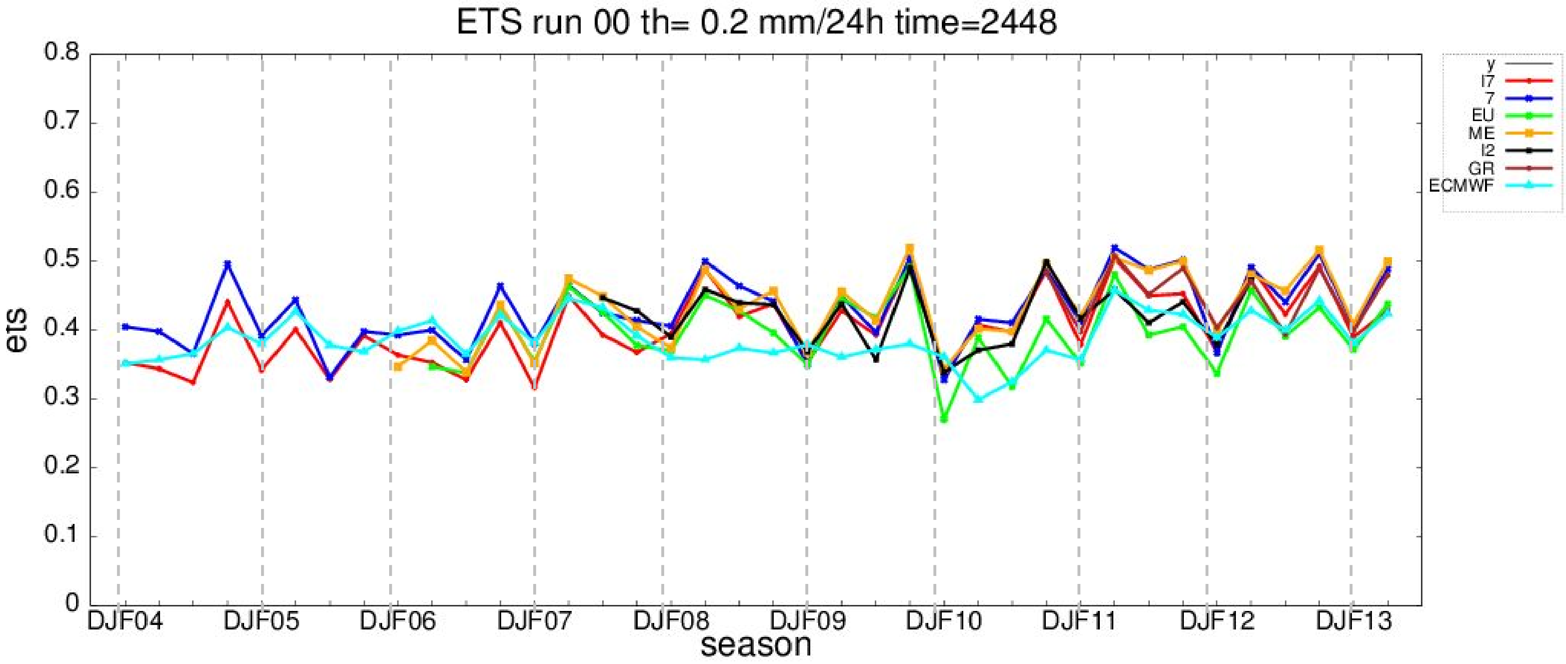


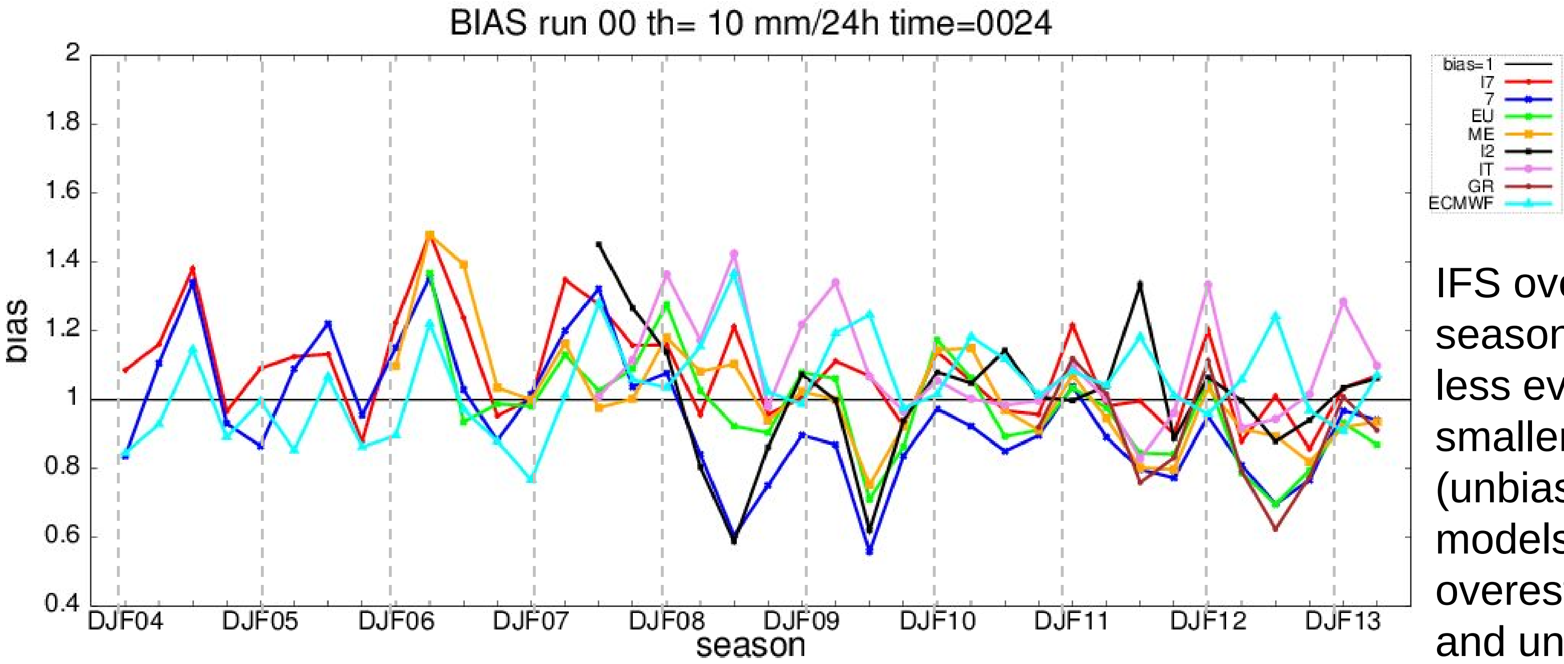


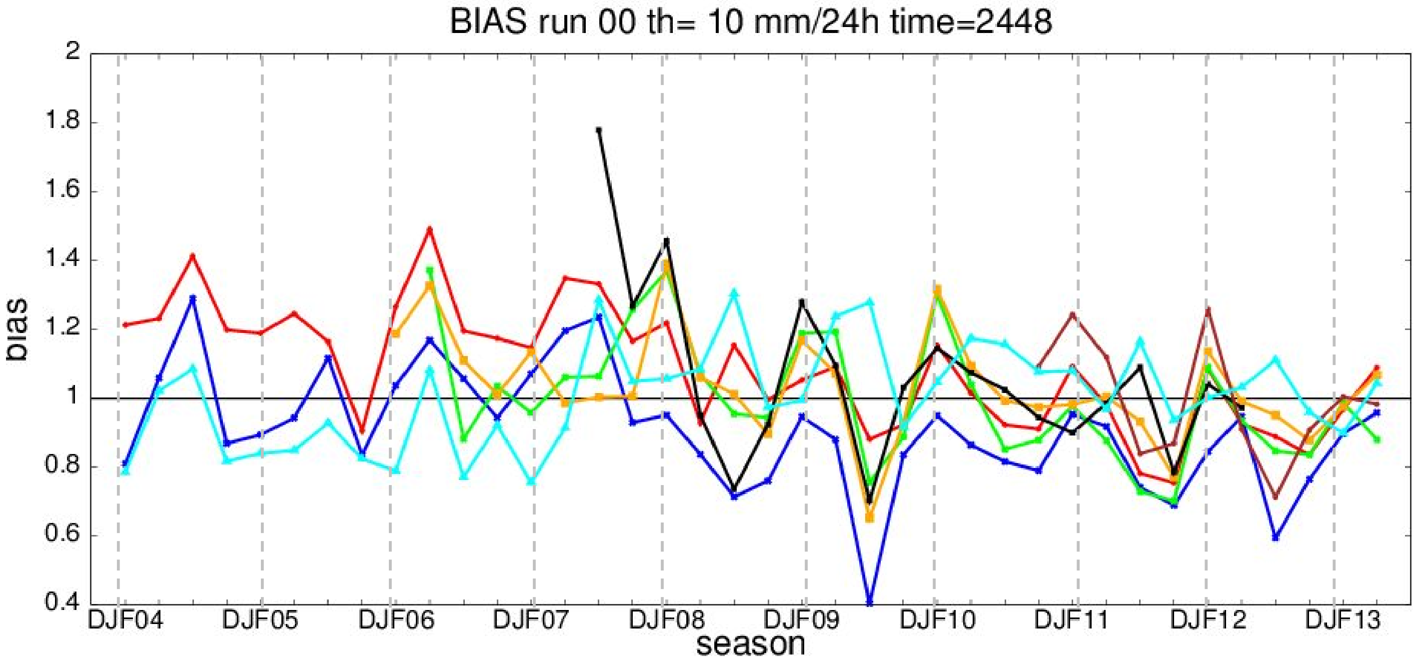
IFS Overestimates all seasons, while COSMO models keep this tendency in summer, eg JJA 2012 for CEU, CME, CGR, CI7. Less evident for D+2



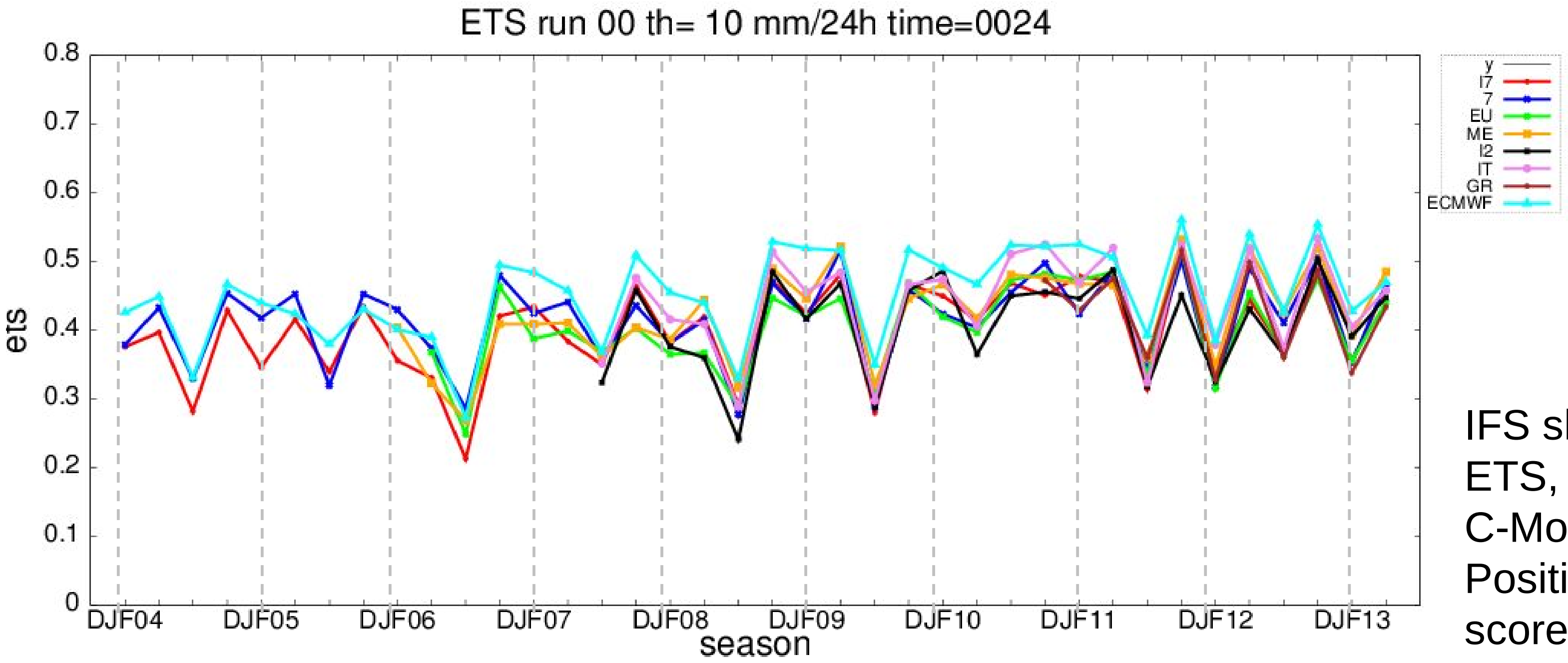
C-Models show higher accuracy compared with IFS for both D1 and D2. Positive trend of the score.



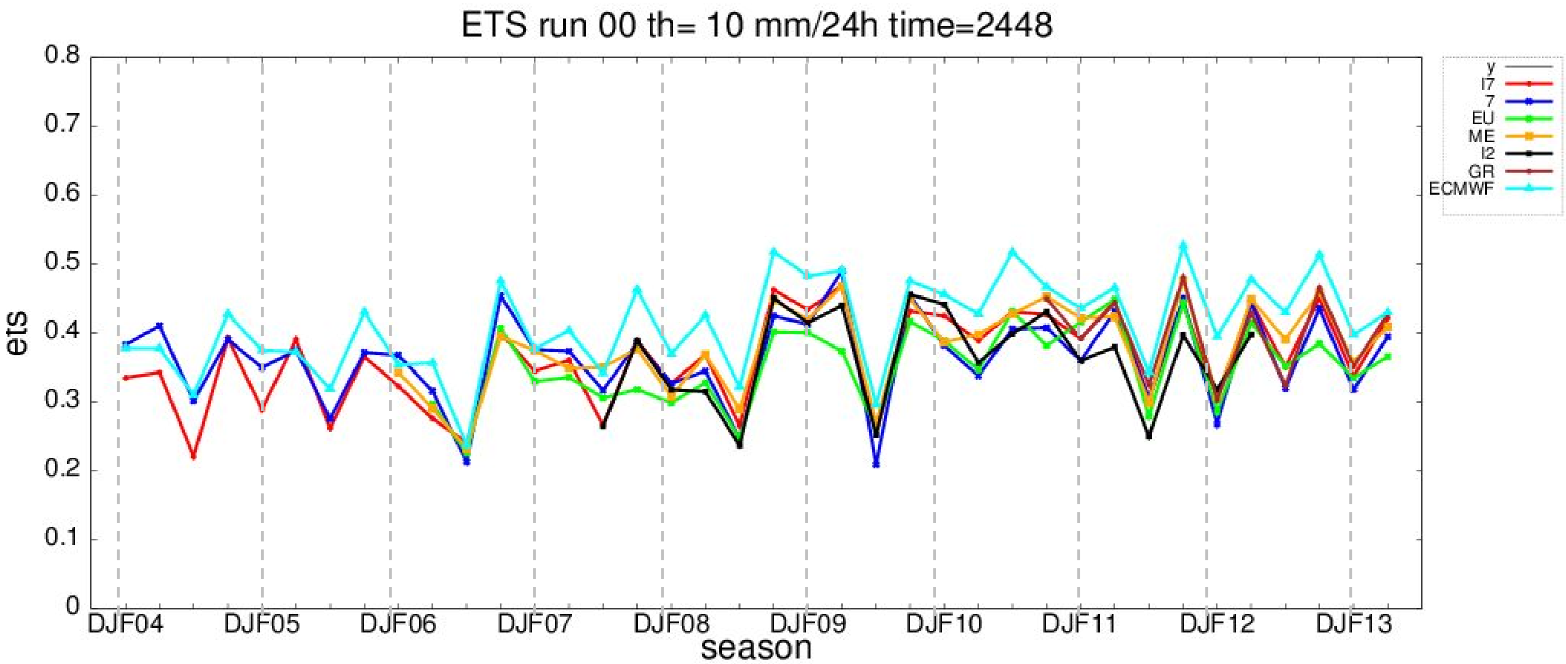




IFS overestimates all seasons in D1, but it is less evident than smaller threshold (unbiased in D2). C-models tend to overestimate in DJF and underestimate in JJA (more), but not all of them. General tendency for C7 to underestimate the event for all the thresholds



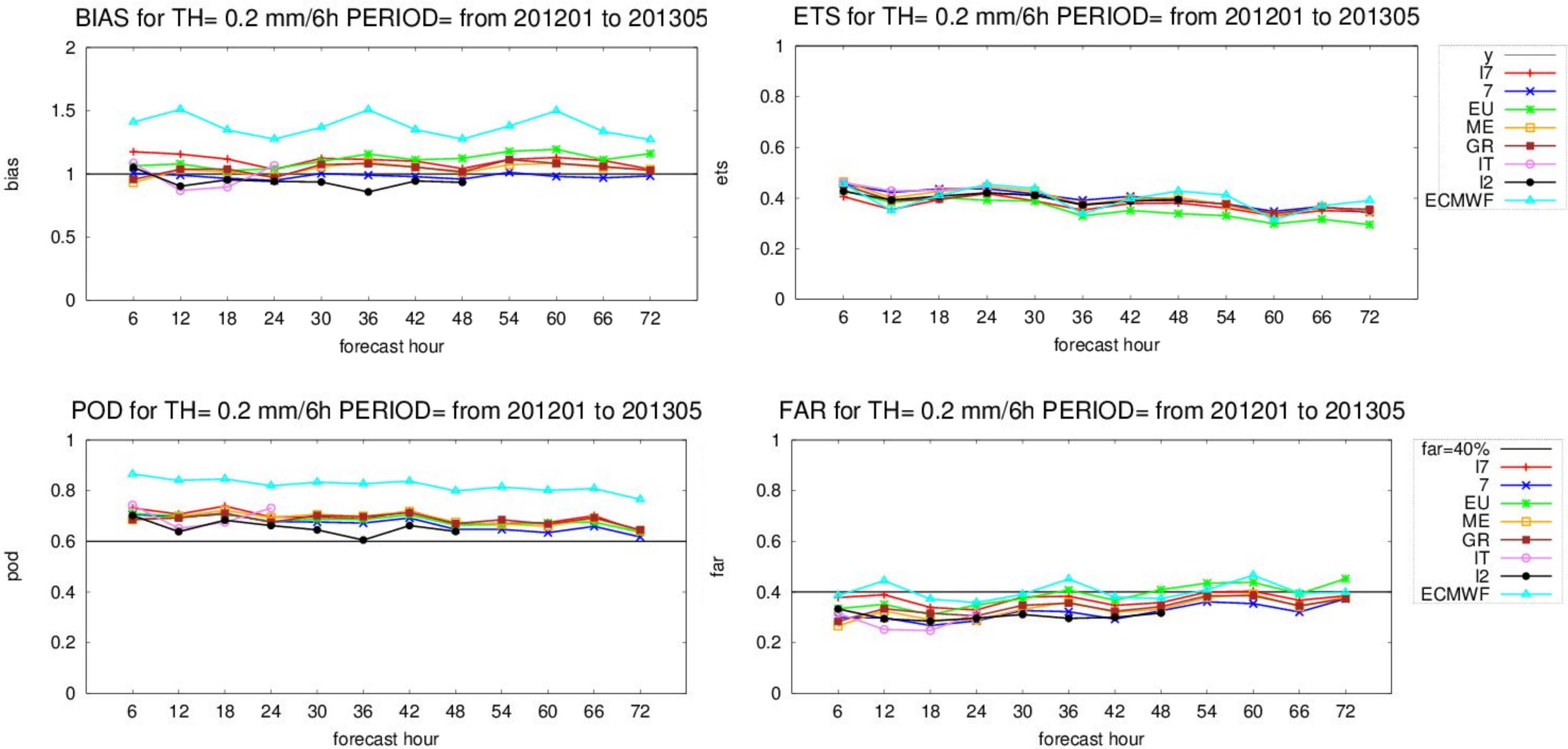
IFS shows higher ETS, but very close to C-Models.
Positive trend of the score.



6h cumulated precipitation average over areas: 201201-201305

Rain/NoRain case 201201 - 201305

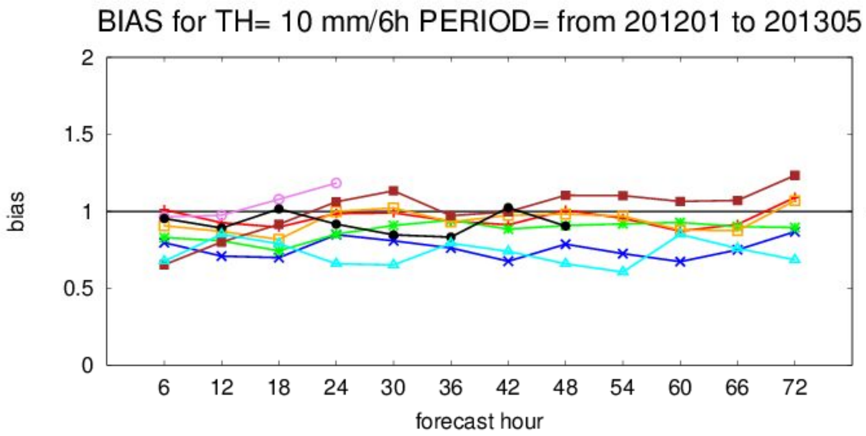
Overestimation for IFS -> higher POD. Low Bias for C-Models

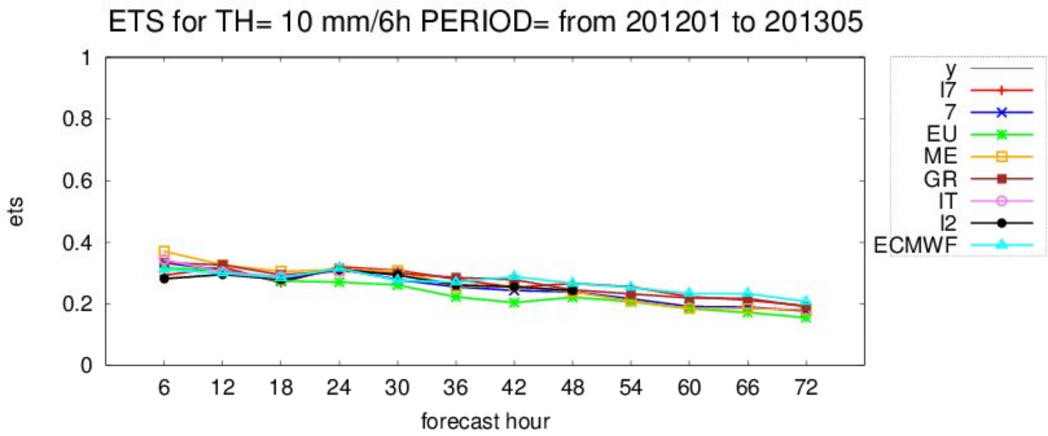


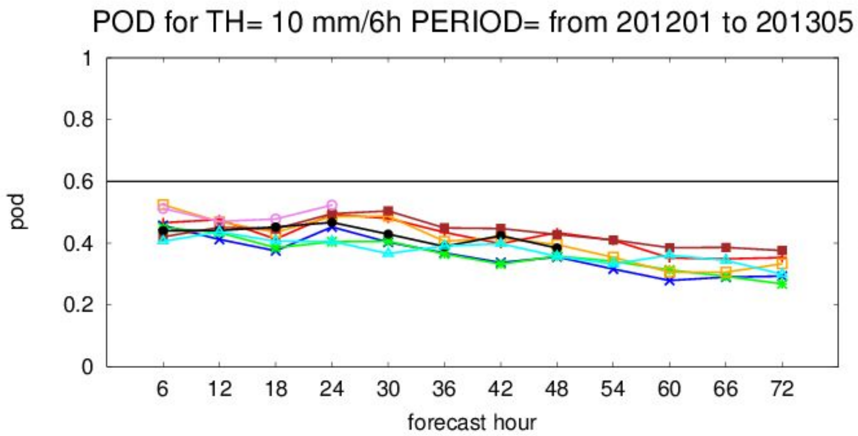
6h cumulated precipitation average over areas: 201201-201305

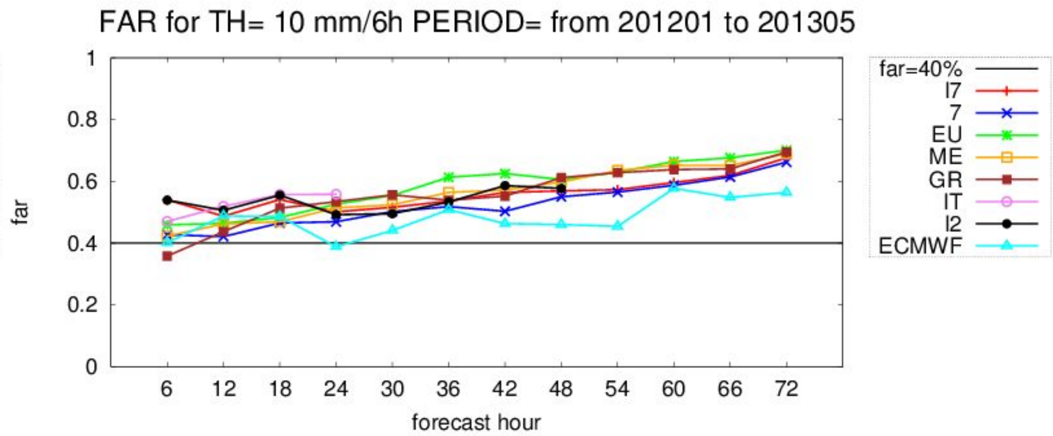
10mm/6h case 201201 - 201305

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





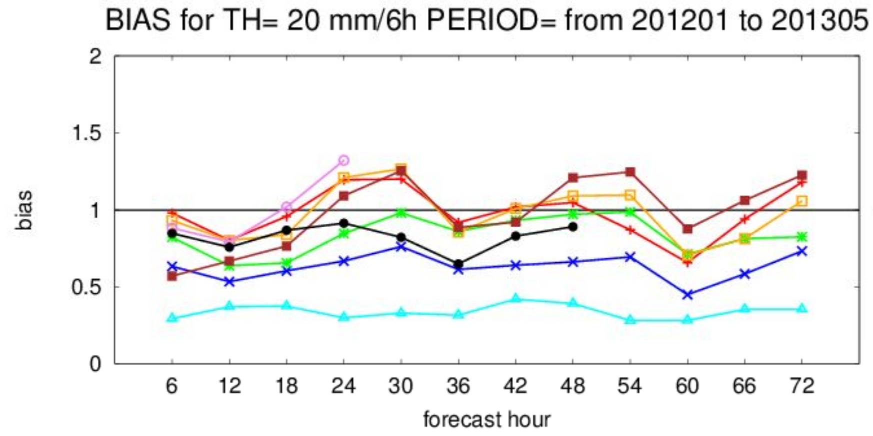


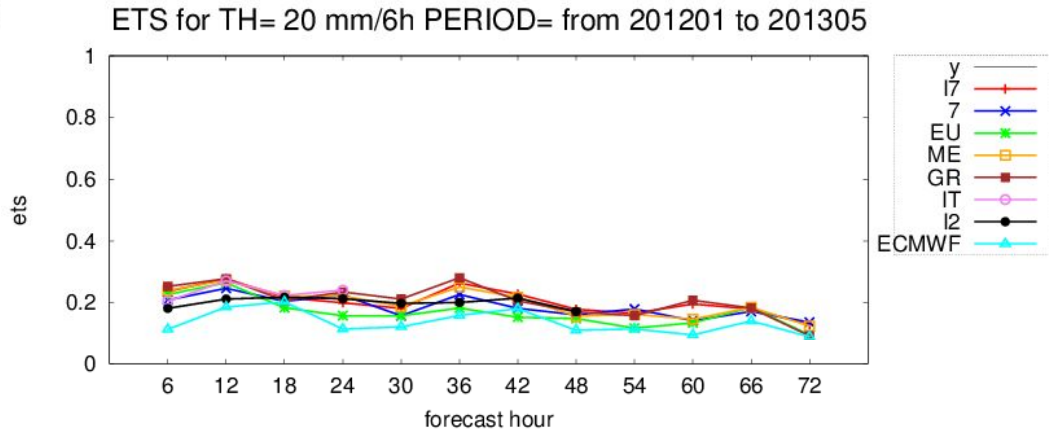


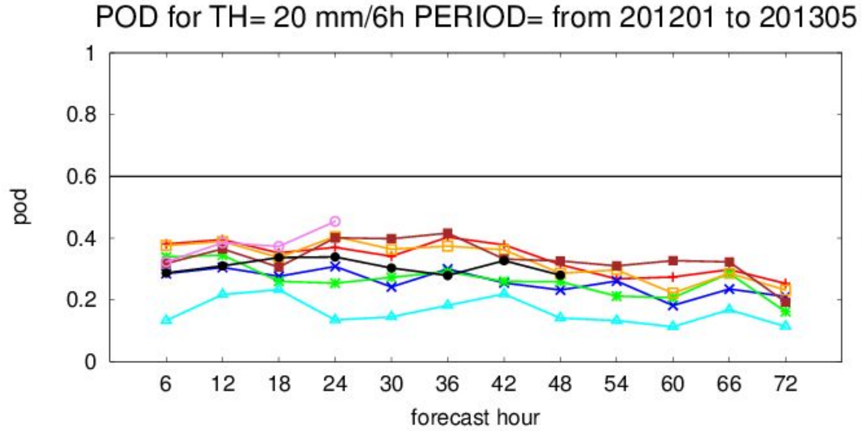
6h cumulated precipitation average over areas: 201201-201305

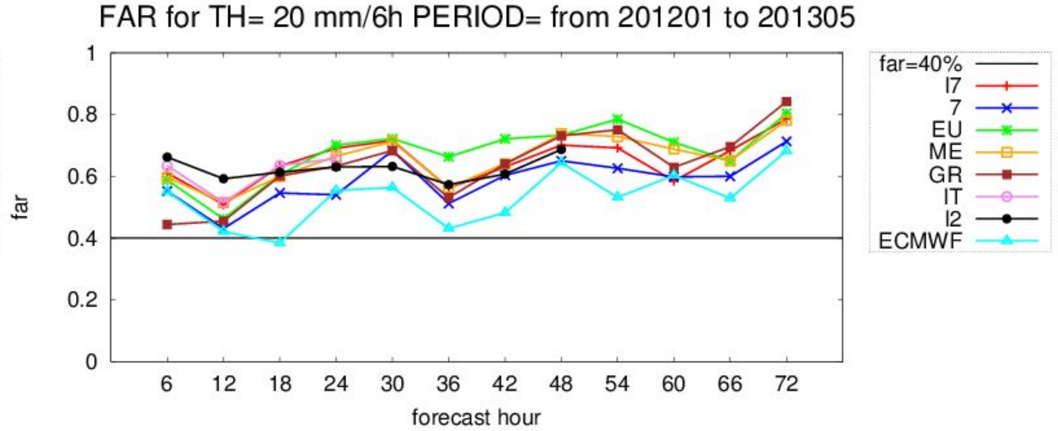
20mm/6h case 201201 - 201305

Very high threshold penalised IFS. C-Models are better in general, but CI2, C-7 and CEU (less) underestimate the event.







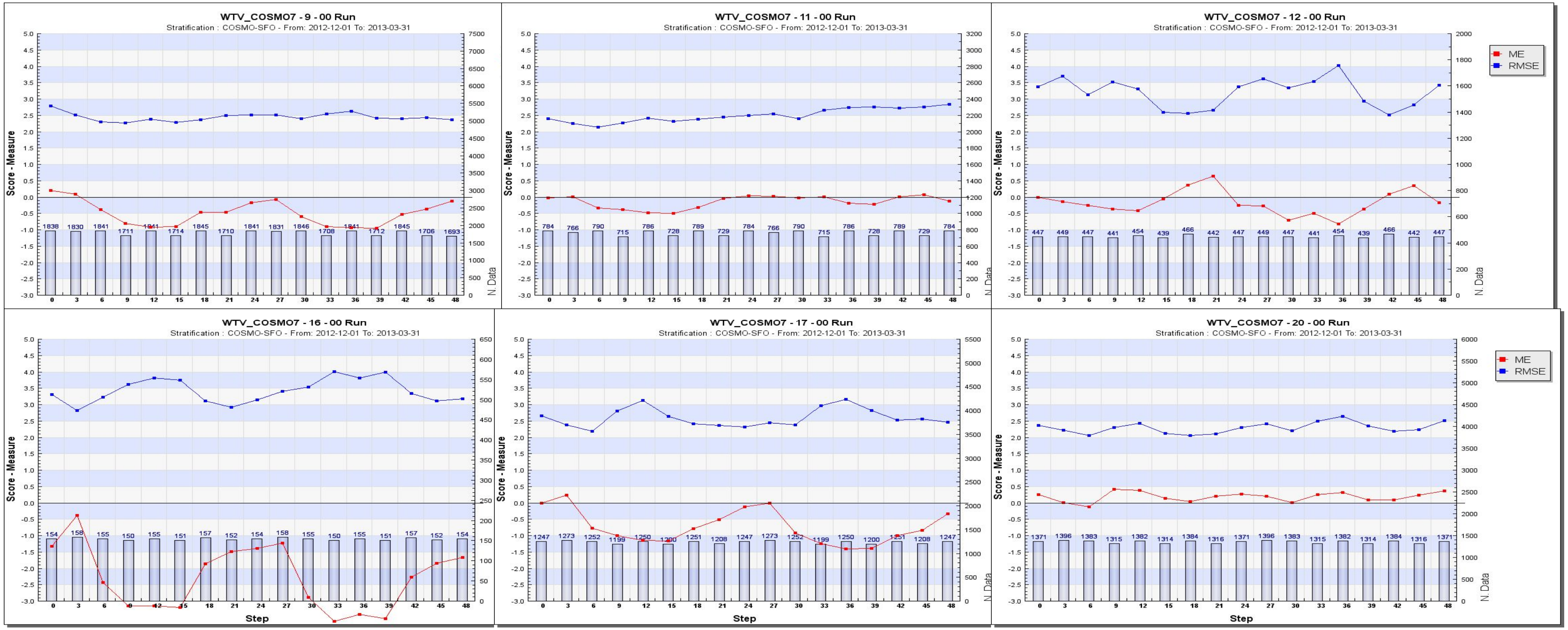


Weather Defined Verification (Task 3.3)

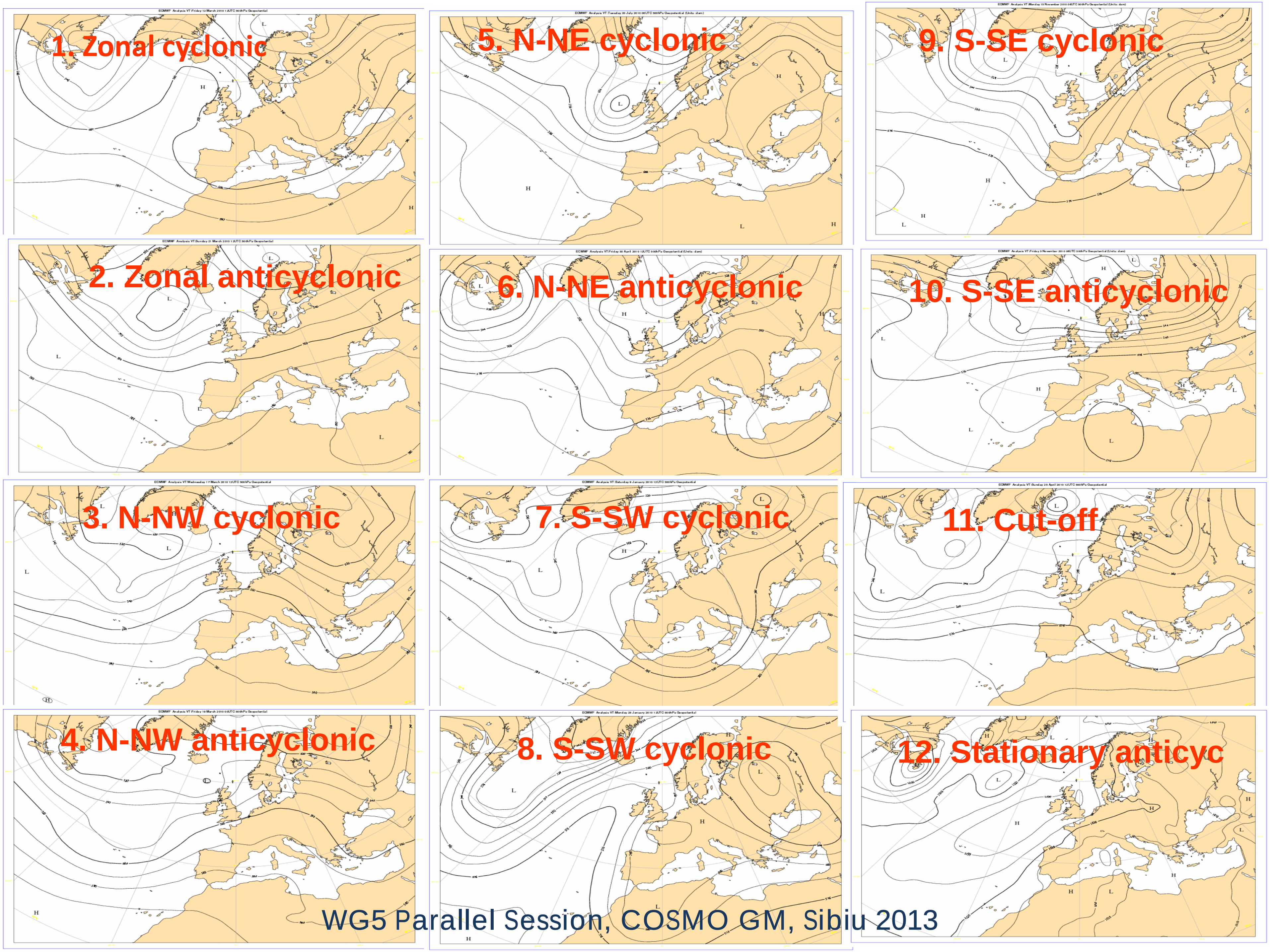
Weather type verification

- Large number of experiments –
  - 10, 20, 30 types
  - CKM, DKM, PCT, PTT
  - three domains of different scales
  - pmsl and pmsl anomalies as classification variables (ECMWF ERA40 and interim reanalysis, DJFM 01.09.1957 – 31.01.2013)
- To evaluate "discriminative power" of classifications, Kolmogorov-Smirnov criterion was used for temperature and precip distributions
- Finally, a classification with 20 weather types was chosen: the distance k-mean (DKM) method, domain of 0°-75° E, 30°-72° N, pmsl variable.

Weather type verification. COSMO-RU7, 2nd test period, whole Sochi region.

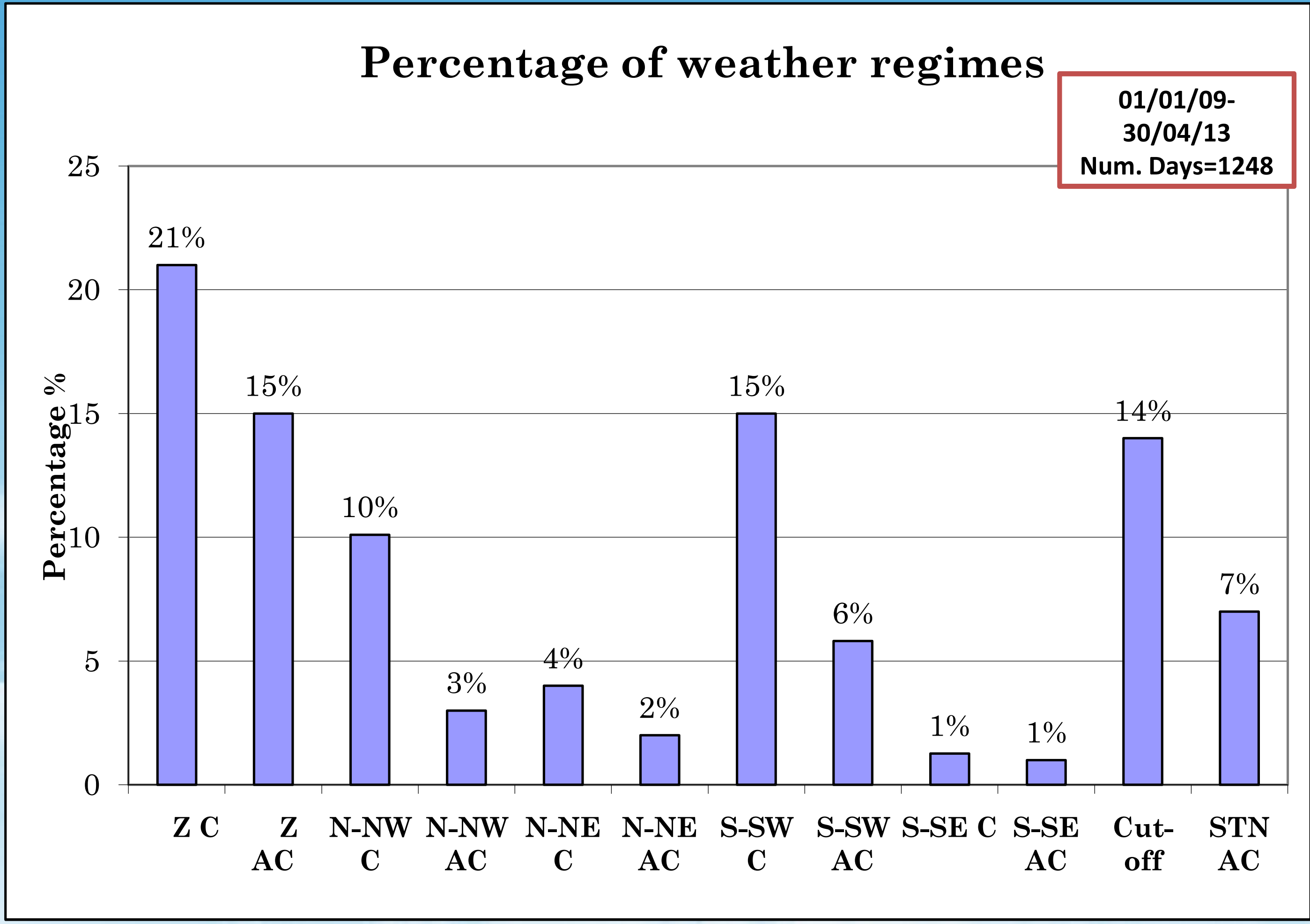


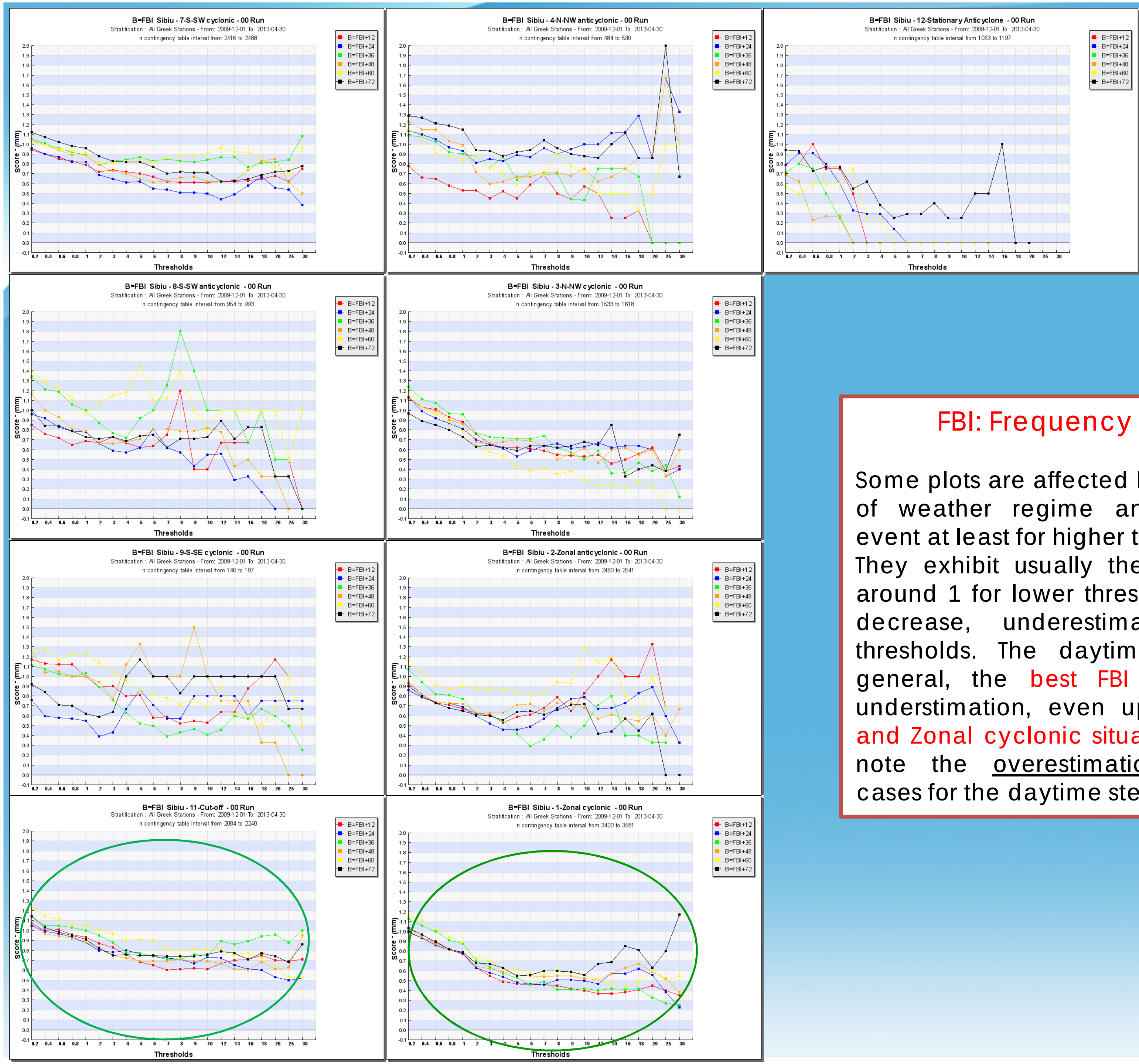
- There are differences in error cycles.
- Diurnal error cycle is most pronounced for some types.
- Type 20 – Sochi is in the rear of a cyclone with NNW flow – is the only type with mostly positive ME.
- **Such scores will be part of forecaster reference guide.**


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 cyclonic
9. S-SE cyclonic
10. S-SE anticyclonic
11. Cut-off
12. Stationary anticyc

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

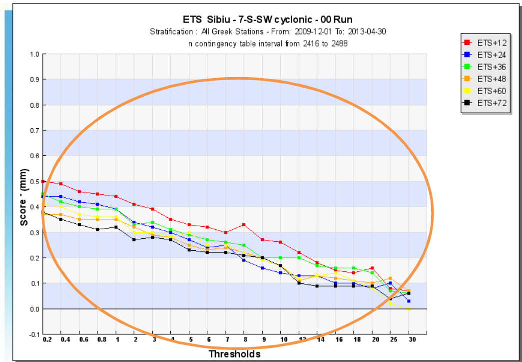


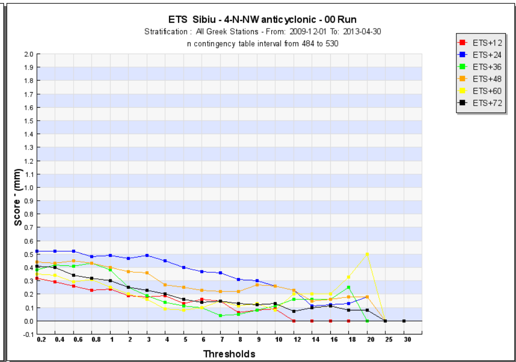


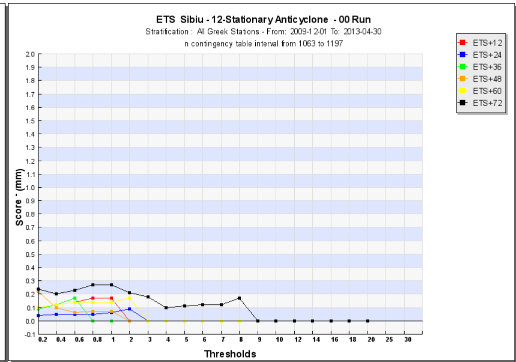
FBI: Frequency Bias Index

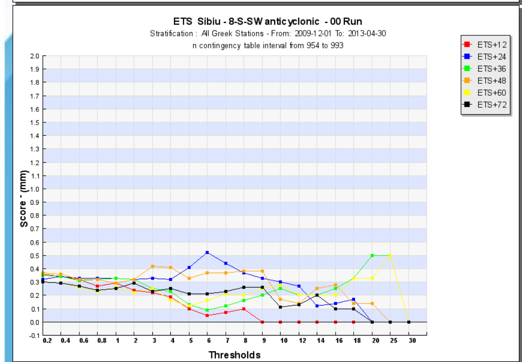
Some plots are affected by the poor sample of weather regime and/or precipitation event at least for higher thresholds.

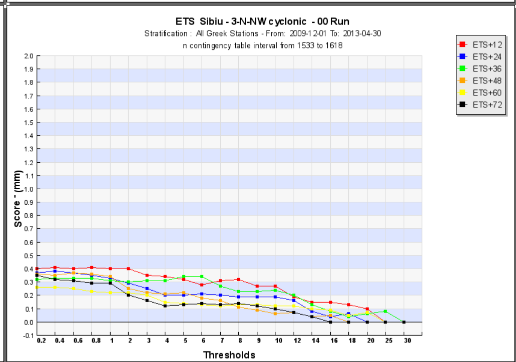
They exhibit usually the tendency for FBI around 1 for lower thresholds that tends to decrease, underestimating the higher thresholds. The daytime steps show, in general, the best FBI in terms of less understimation, even up to +72h Cut-Off and Zonal cyclonic situations. It is worth to note the <u>overestimation</u> of rain/norain cases for the daytime steps .

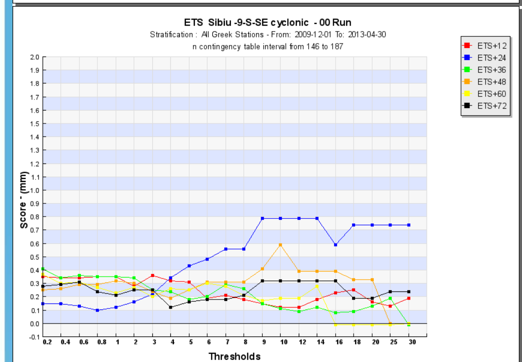


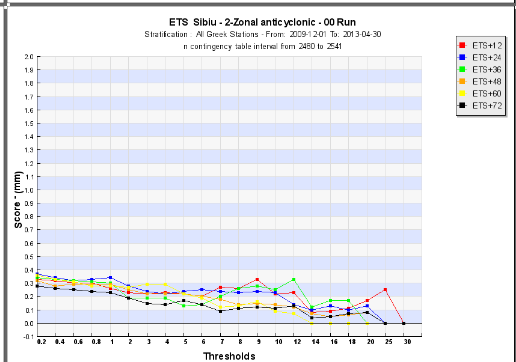


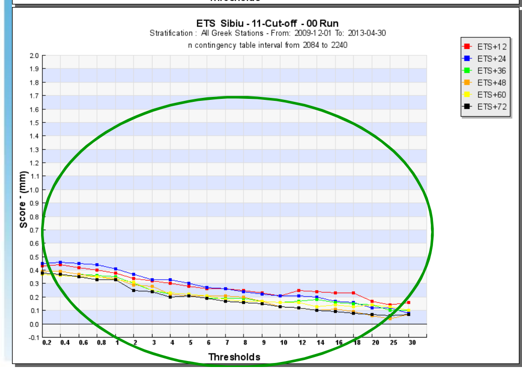


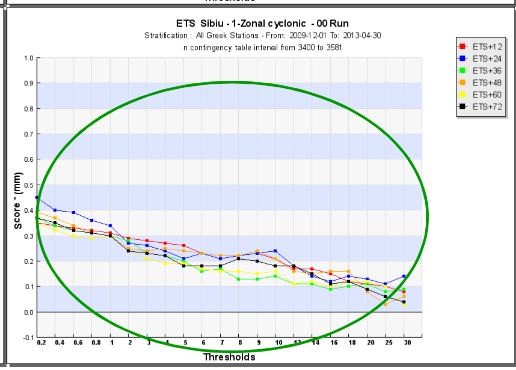












ETS (Equitable Threat Score)

ETS reflects the FBI curves behaviour, giving the highest values for cyclonic and cut-off siutations, where also the FBI shows less spread among the steps.

The first day of forecasts has in general the best performance and ETS goes to 0 only for very high thresholds (here 12h mm sums)

Precipitation verification using radar composite network with neighborhood methods (Task 4.2)

Validation of precipitation from COSMO model using radar composite data corrected with rain gauges values through neighborhood methods (e.g. upscaling, fractions skill score, intensity-scale).

About the work

The aim of the work is to realise a new type of verification for COSMO I7

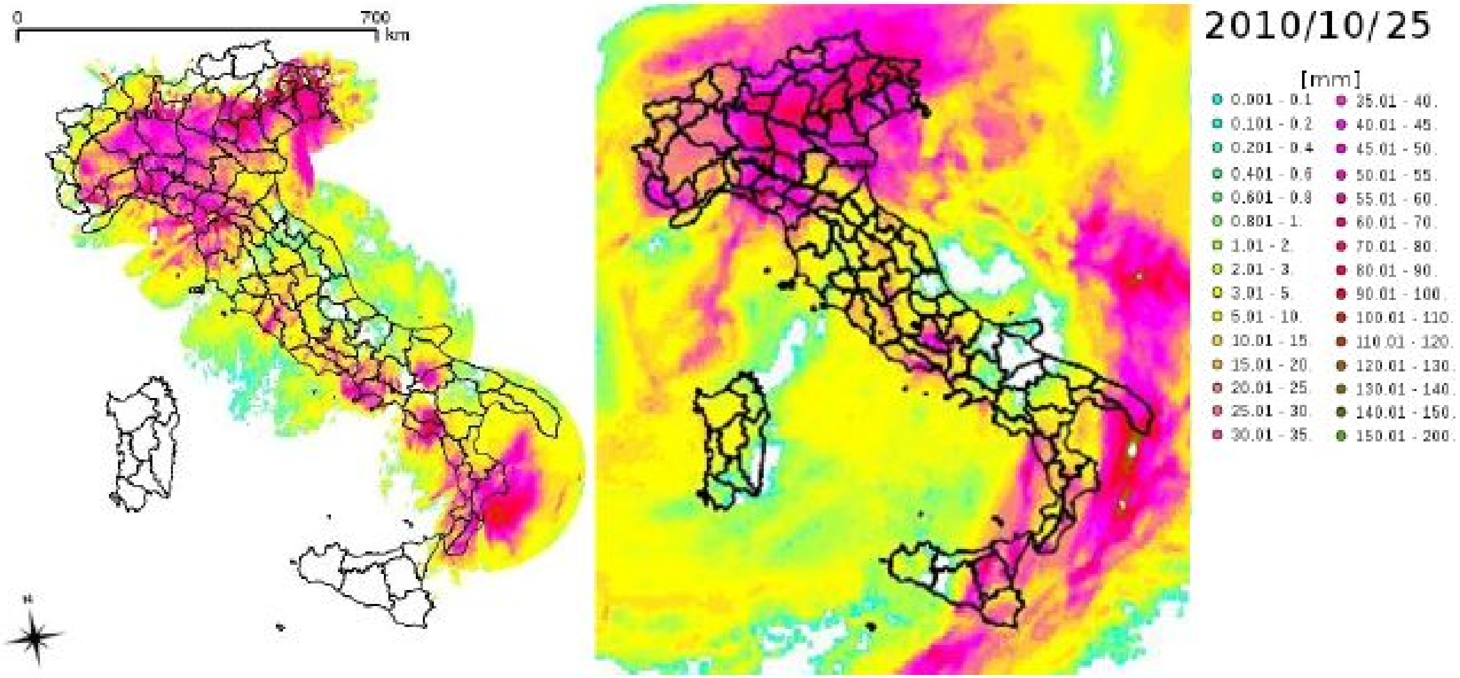
- The verification is made against a precipitation field estimated by the Italian radar mosaic corrected with the data coming from the Italian rain gauges network
- Two phases:
  - Relative error
  - Fuzzy verification
- Case study: October 24-25-26-27, 2010
- Finding a useful methodology for different kind of studies

Future developement

- Complete the study by applyng the same methodology to COSMO I2 (or other models)

Preliminary analysis (eyeball)

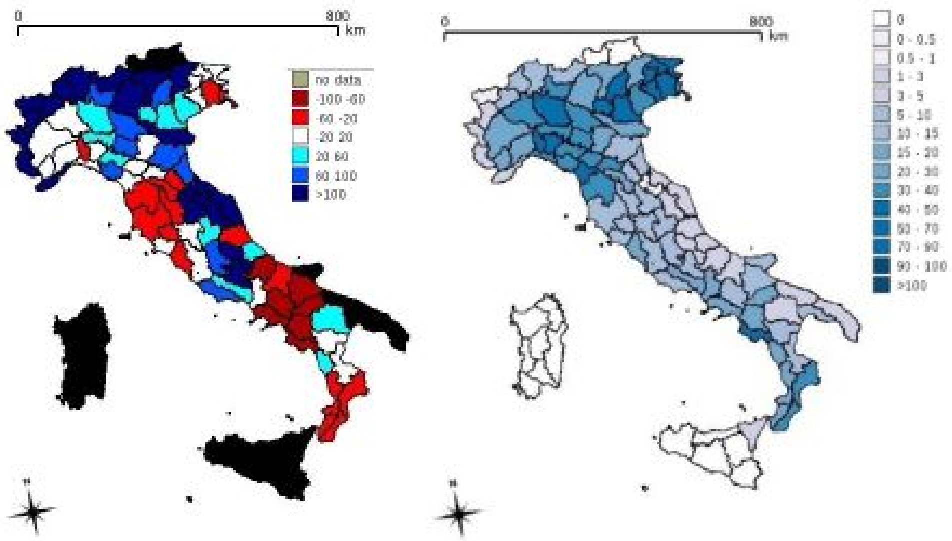
The first step in the verification is an eyeball comparison between the modified radar field and the COSMO-I7 forecast (October the 25^{th})



- It is possible to notice a good agreement between the two fields
- The agreement is good for what concerns the dislocation of the precipitation patterns, a little worst if we look at the intensity of the precipitation

Relative error - October 25, 2010

- October the 25^{th}: the precipitation almost covers the entire Italy

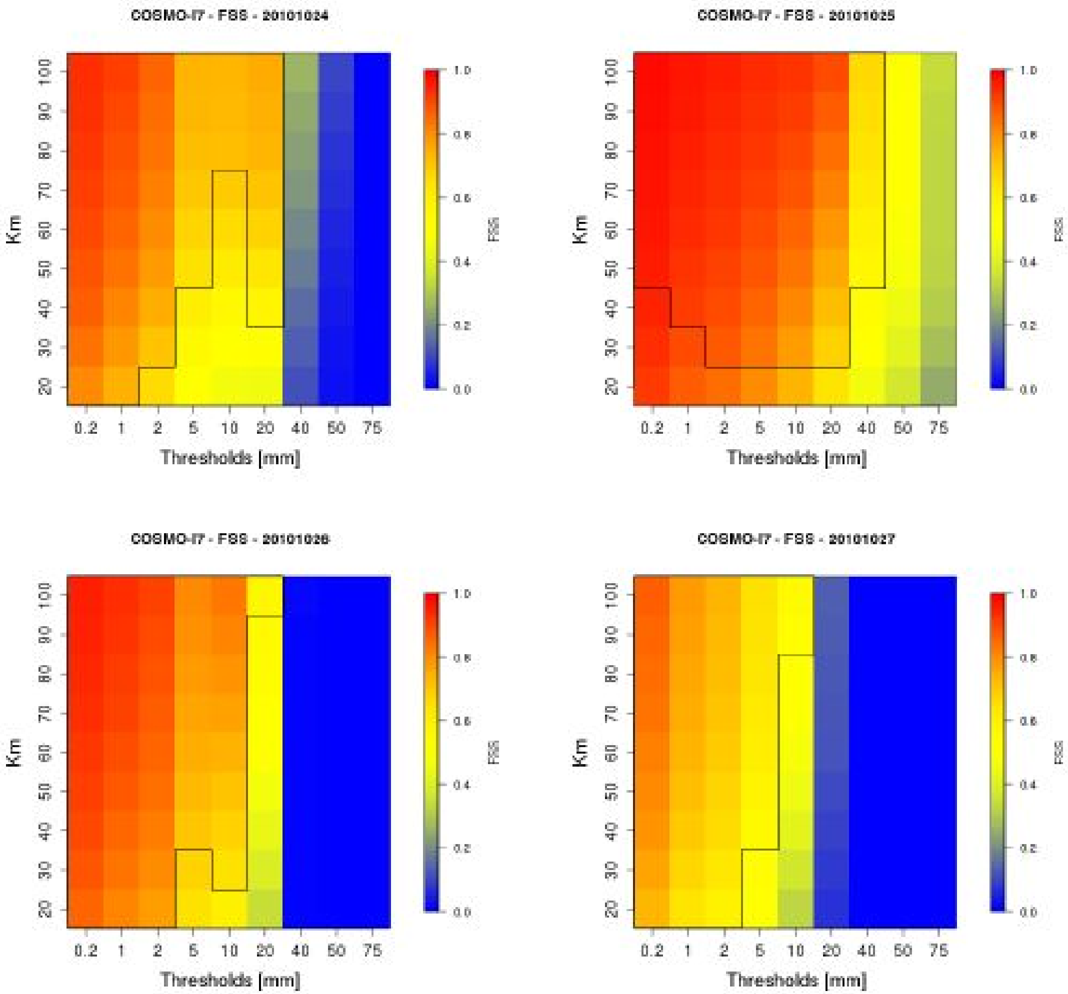


- Blue: overestimation
- Red: underestimation

- General overestimation of the model over northern Italy, more marked in the alpine region
- For what concerns the peninsula, the model underestimates almost everywhere, with the exception of the Marche and part of the Lazio regions

FSS calculation

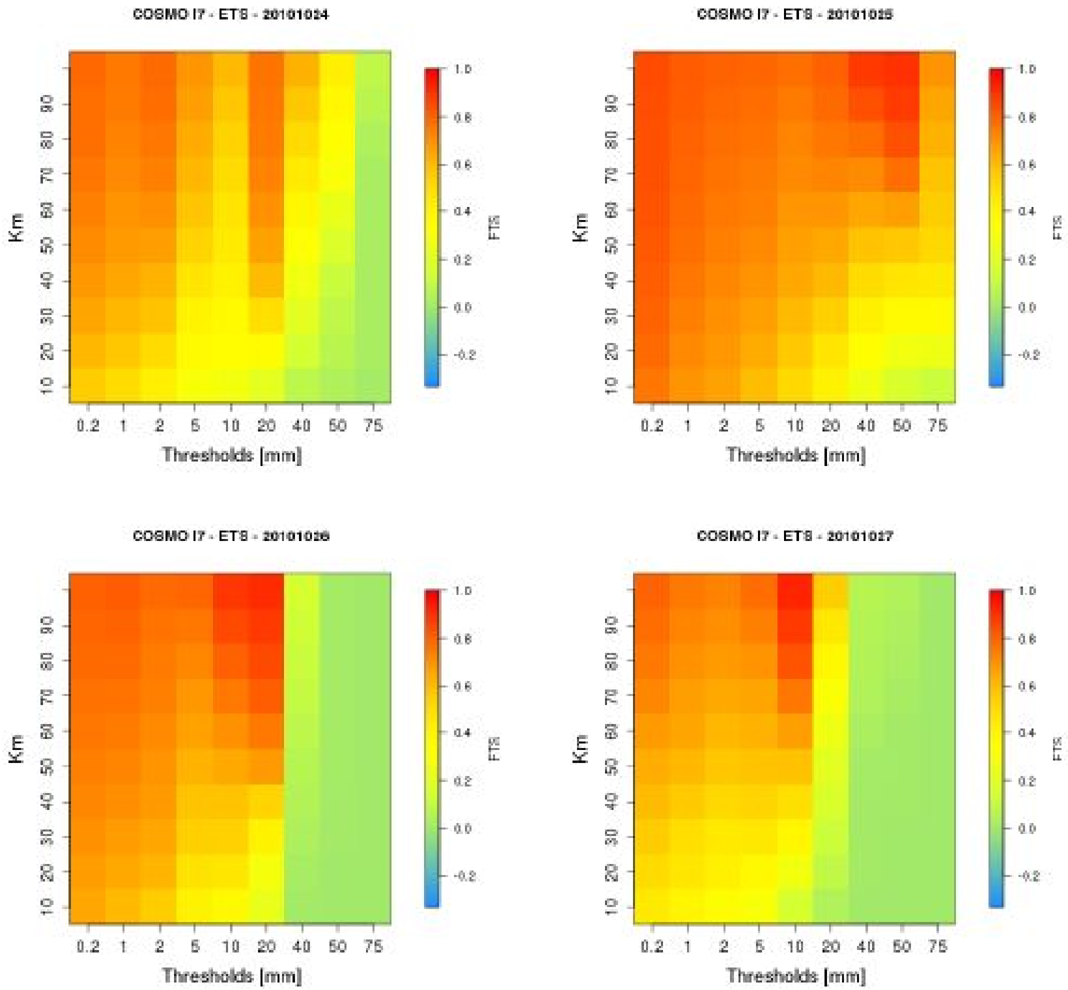
FUZZY VERIFICATION



- X: thresholds [mm]
- Y: box side [km]
- Z: FSS [0 - 1]

Black line surrounds values that are higher than useful FSS limit

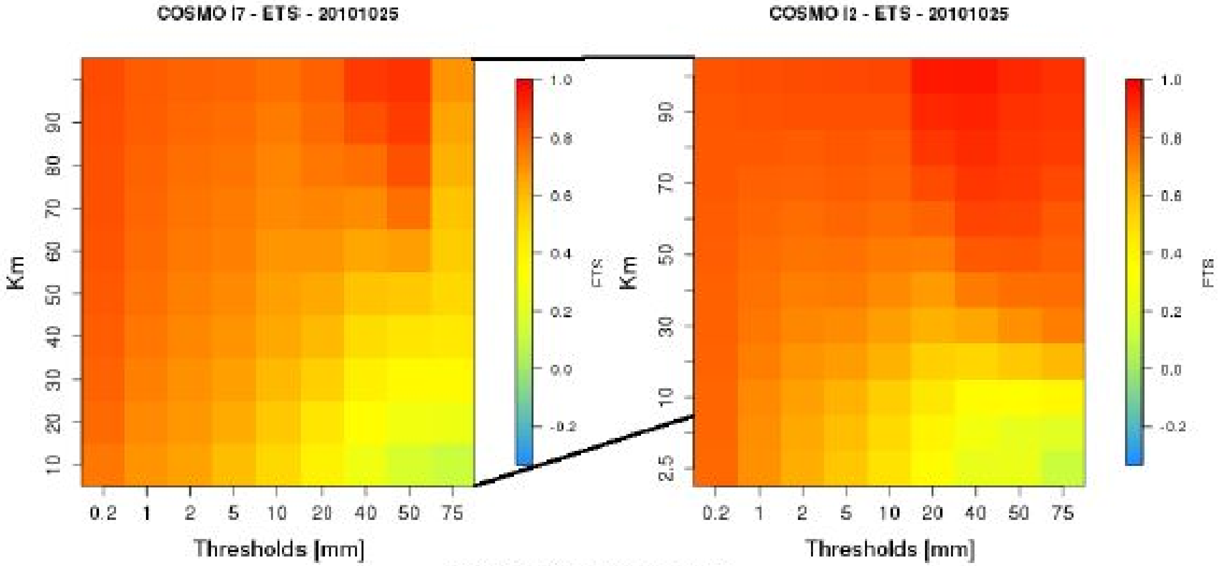
ETS calculation

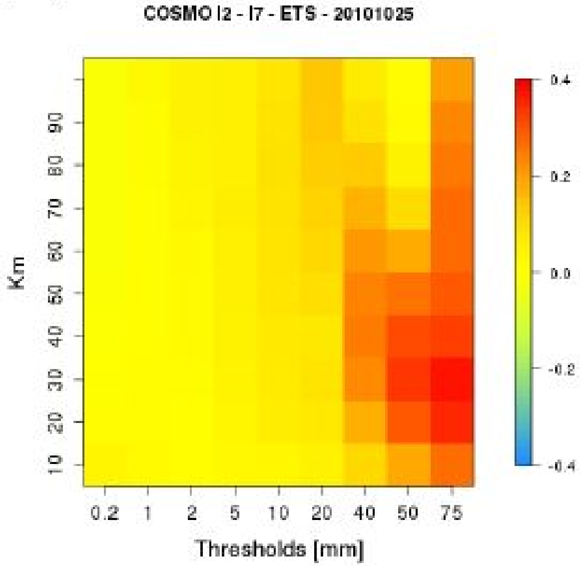


- X: thresholds [mm]
- Y: box side [km]
- Z: ETS [(-1/3) - 1]

ETS

FUZZY VERIFICATION





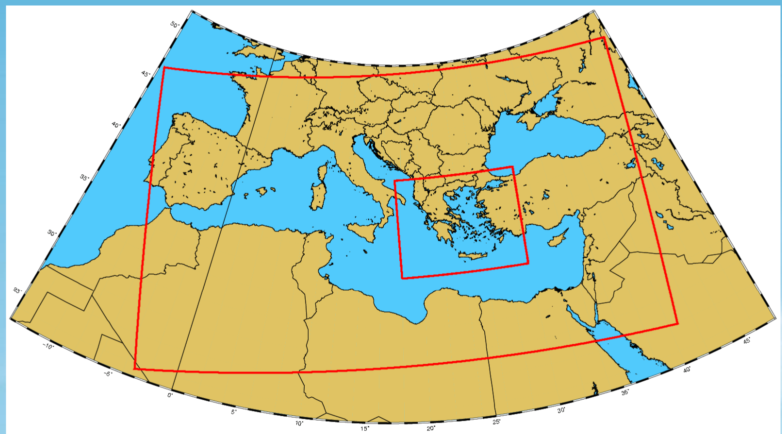
ETS CI2 – ETS CI7

For higher thresholds and for almost all the scales CI2 better than CI7

Verification of EPS systems (Task 5.1)

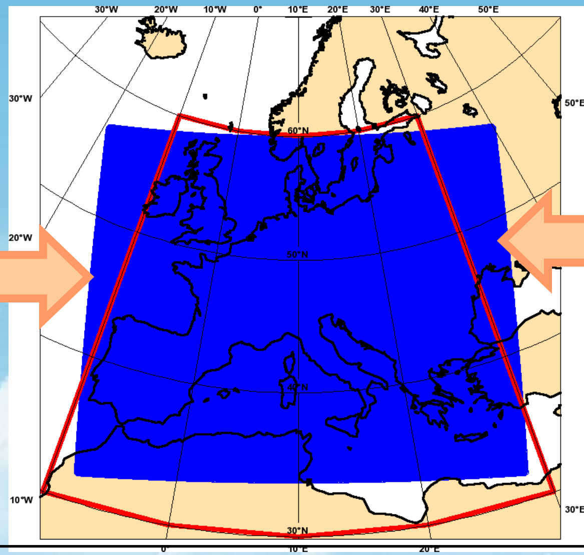
Verification of surface weather parameters from Operational EPS systems based on COSMO models. Identification of measures that emerge the possible gain of the use of such systems.
Main functionalities have been implemented in VERSUS.

Stratifying according to various model implementations…..



COSMOGR7 Grid Area

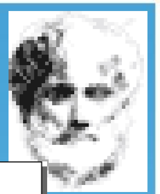
COSMO-LEPS clustering area

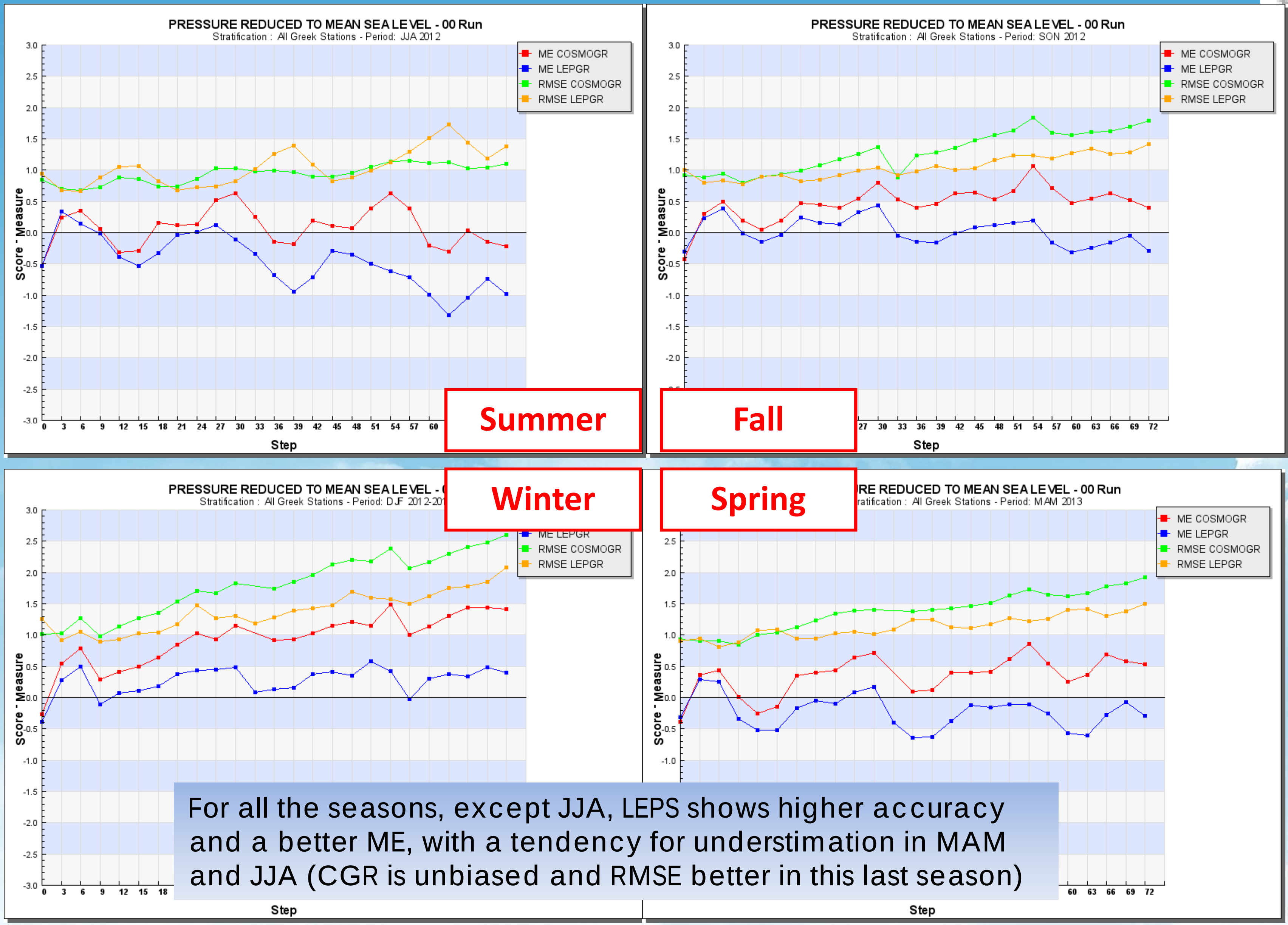


COSMO-LEPS Integration Domain

- suite runs twice a day (00 and 12UTC) as a "time-critical application" managed by ARPA-SIMC;
- Δx ~ 7 km; 40 ML; fc+132h;
- COSM0 v4.21 since July 2012;

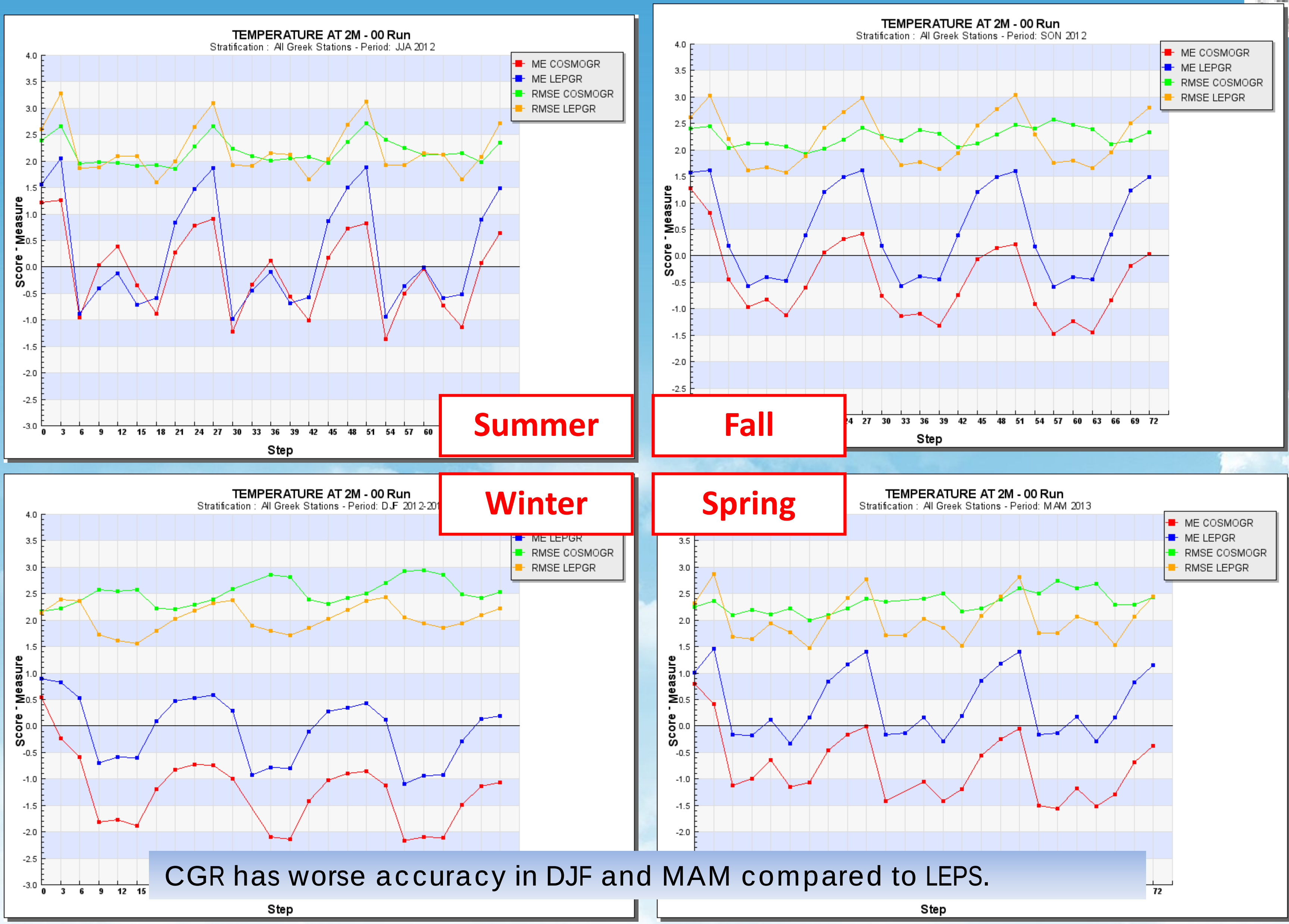
MSLP: COSMOGR7-LEPS (ens mean)





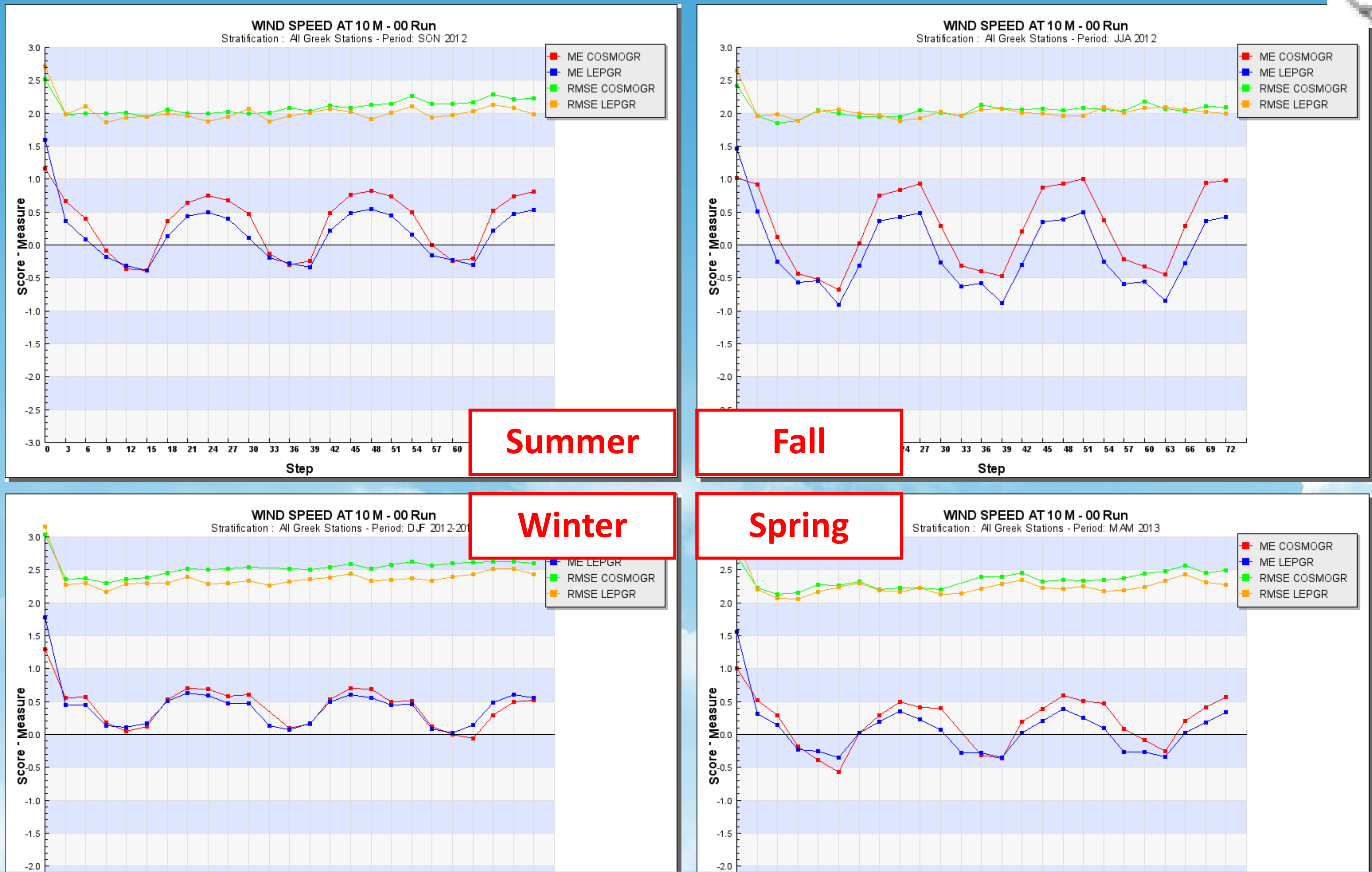
For all the seasons, except JJA, LEPS shows higher accuracy and a better ME, with a tendency for understimation in MAM and JJA (CGR is unbiased and RMSE better in this last season)

TEMP2m: COSMOGR7-LEPS (ens mean)



CGR has worse accuracy in DJF and MAM compared to LEPS.

WindSp: COSMOGR7-LEPS (ens mean)

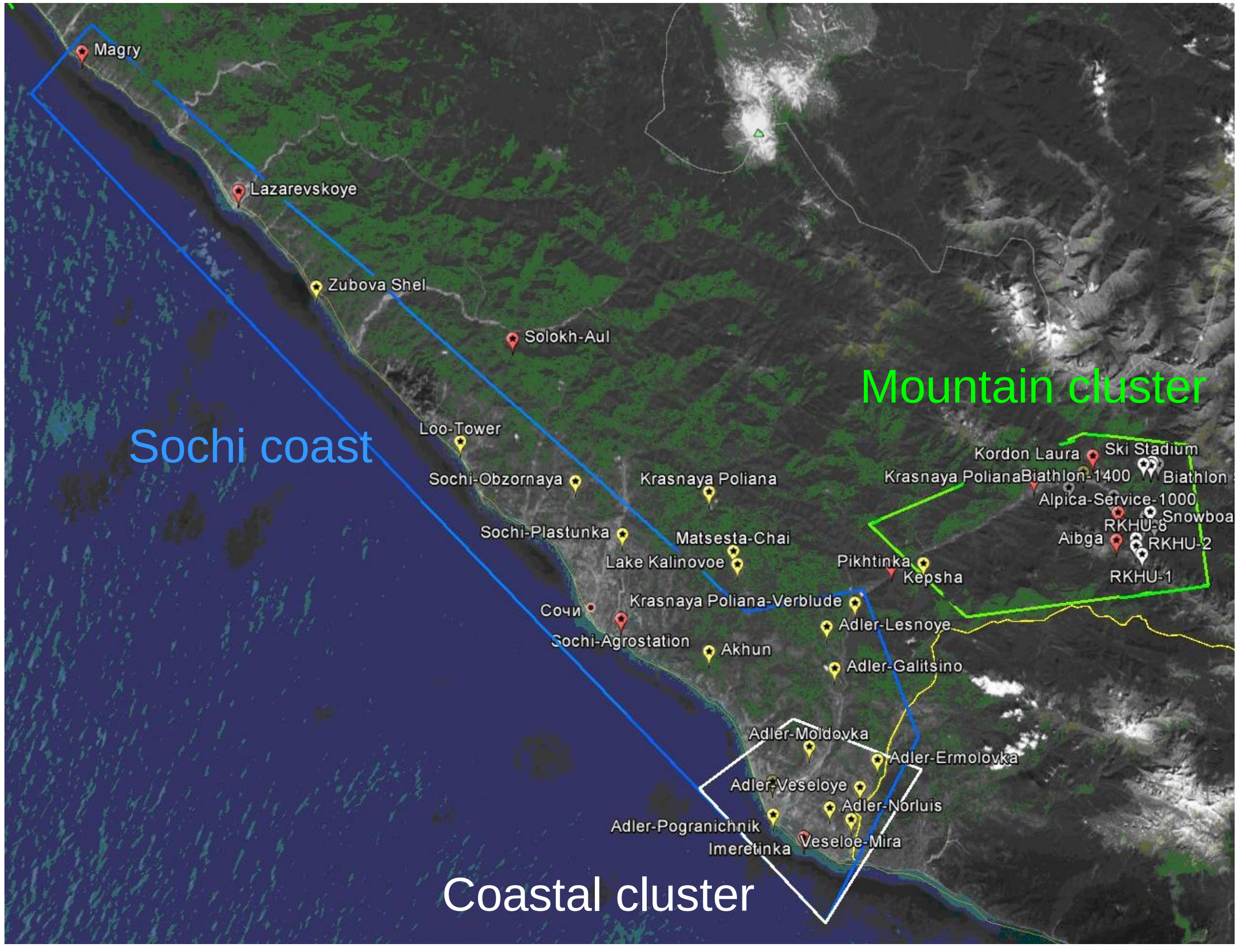


Very similar behaviour for all the seasons, with more overestimation of CGR compared to LEPS.
In terms of RMSE CLEPS exhibits lower values.

Socchi Olympics Verification Experiences (Task 6)

CONSORTIUM FOR SMALL SCALE MODELING
COSMO

Polygons of verification



Forecasts for the Mountain cluster are the most important!

Models

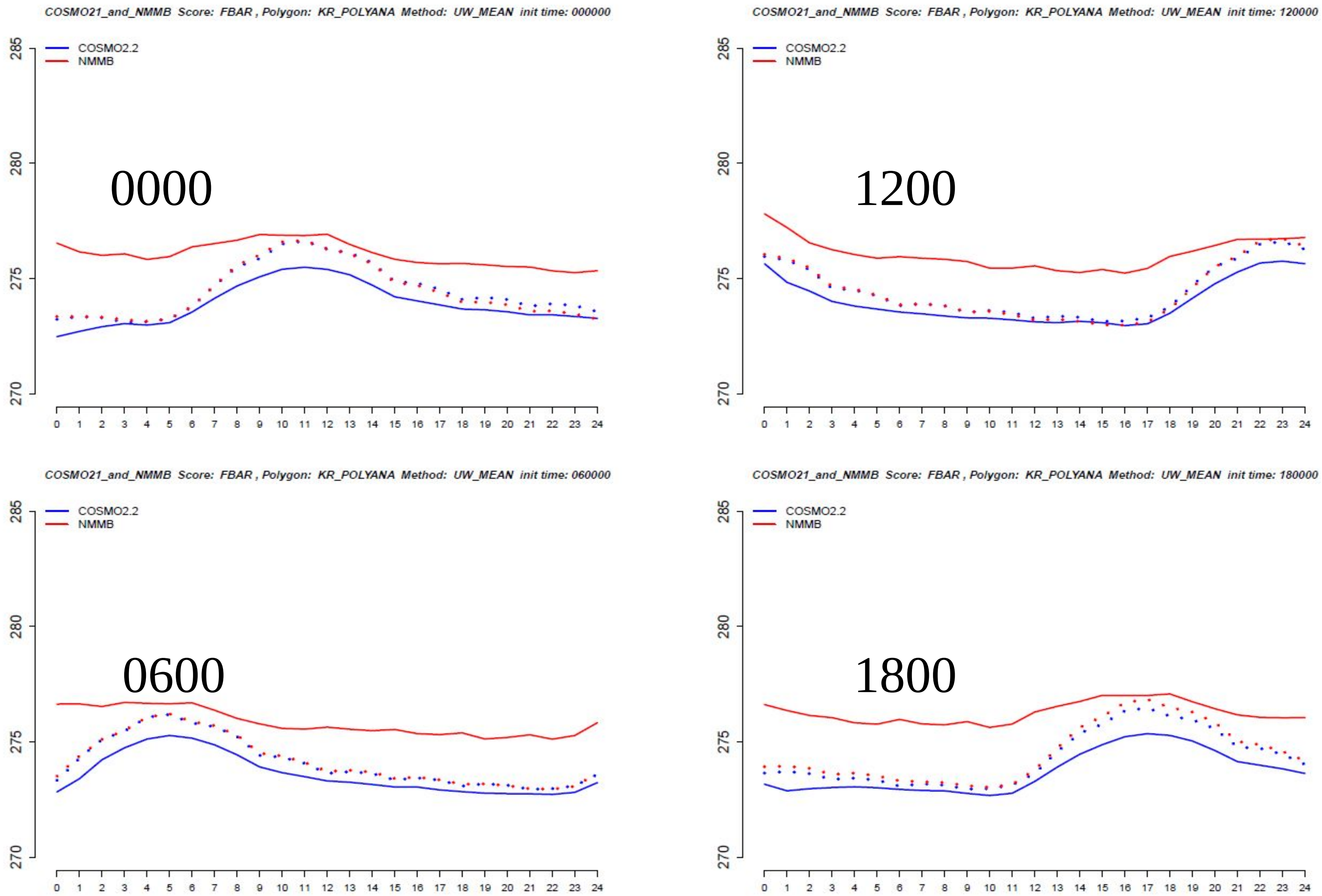
- **2.2-km South region COSMO version** with 40 levels and explicit deep convection calculation (initial and boundary fields from 7-km COSMO-RU) ***interpolated to 1*1-km regular grid using FieldExtra***
- **American 1-km NMMB model**
- Forecast period 24 h
- 4 initial times (00, 06, 12, 18)

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

COSMO blue, NMMB red

Mountain cluster

2nd test period

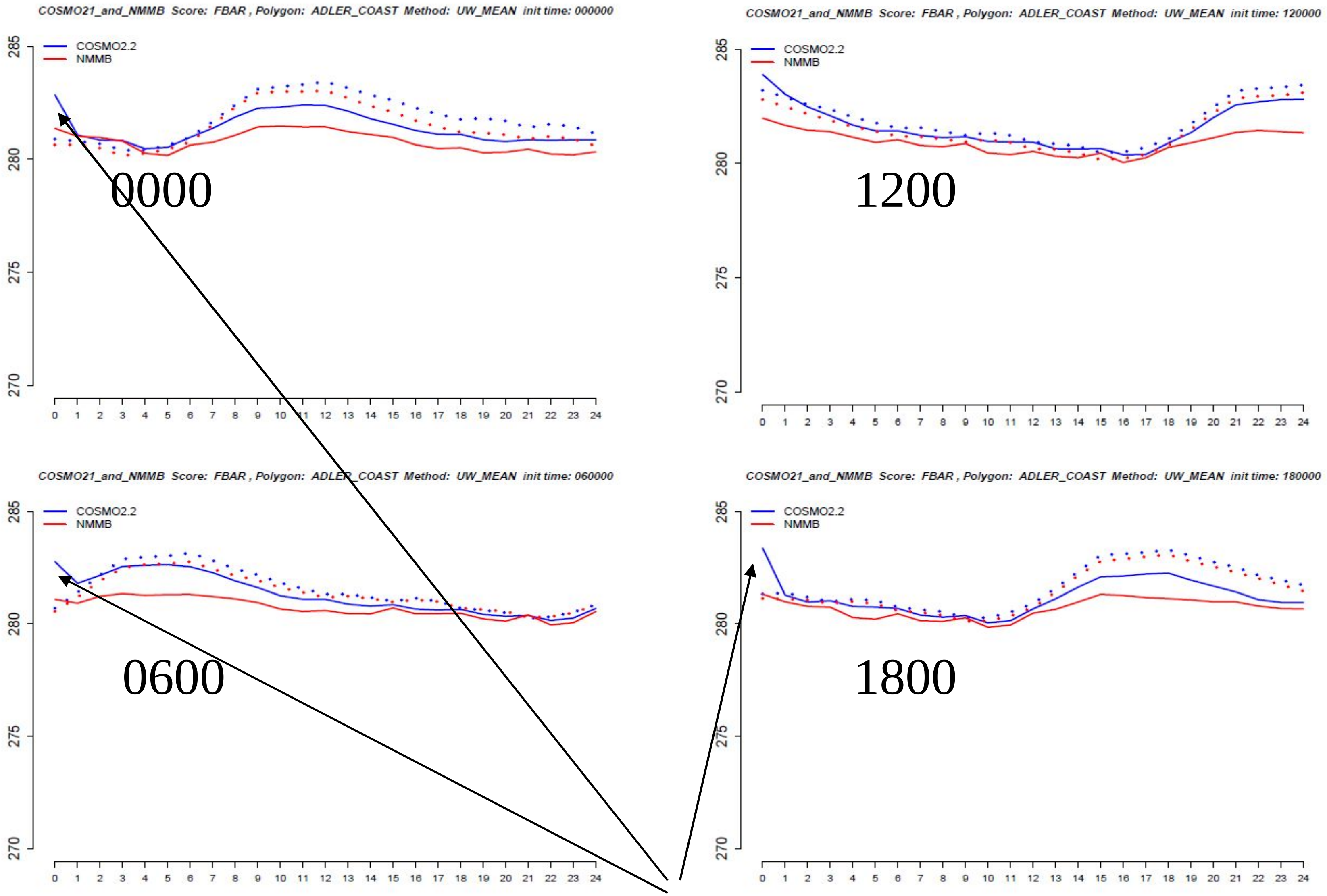


COSMO yields better T2m means and diurnal cycle, **especially in the mountain cluster**

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

Sochi coast

2nd test period

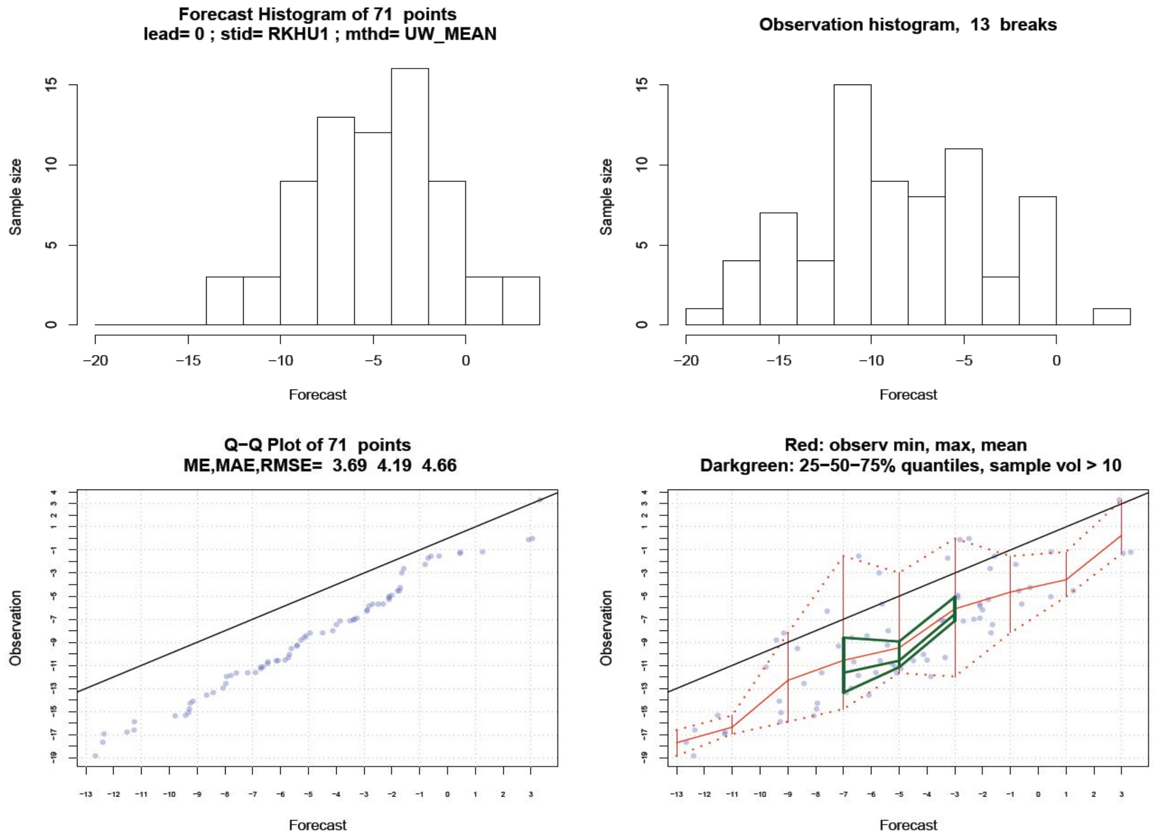


In the coastal polygons, there is a **systematic COSMO error at the initial time** that is likely due to the initial field. It is not detected in the mountain cluster.

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 calibrate the forecasts in the whole variable range including the distribution tails, 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.

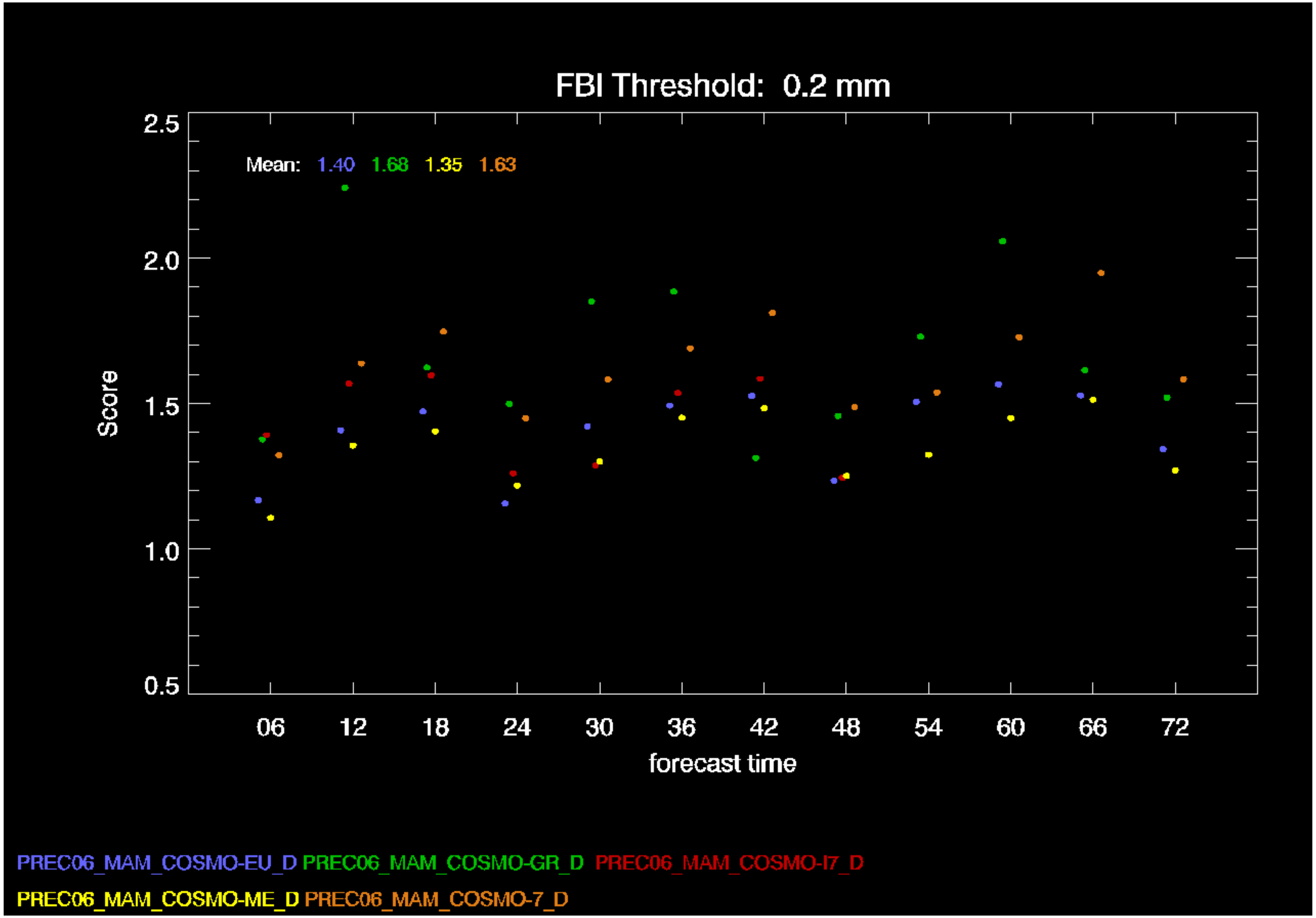
Station “portraits”. Here for T2m RKHU1 station (on the Aibga ridge), nearest point, lead 00 h.



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).

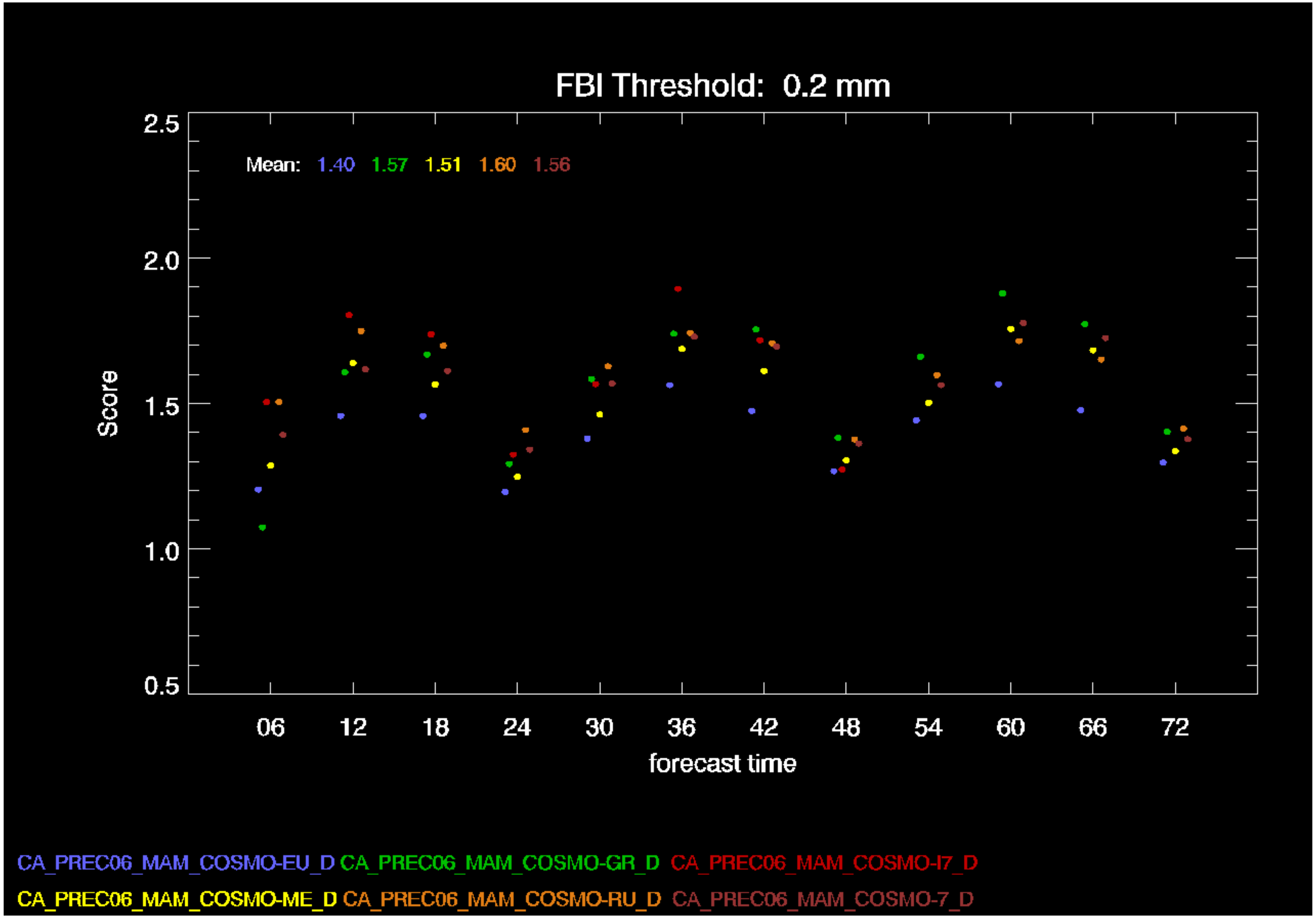
Calibration implies a shift of the frc mean-median to the diagonal.
The T2m area outside the green strip indicates sample instability (calibration uncertainty) due to the small data volume.
Importance of the above diagnostic verification for “critical thresholds” that are crucial for decision-making (distribution tails and small samples)

Steps to show different model behaviour concerning QPF
1.) The basic situation: FBI for threshold 2mm $(6h)^{-1}$
Common plots, MAM 2013, national chosen stations

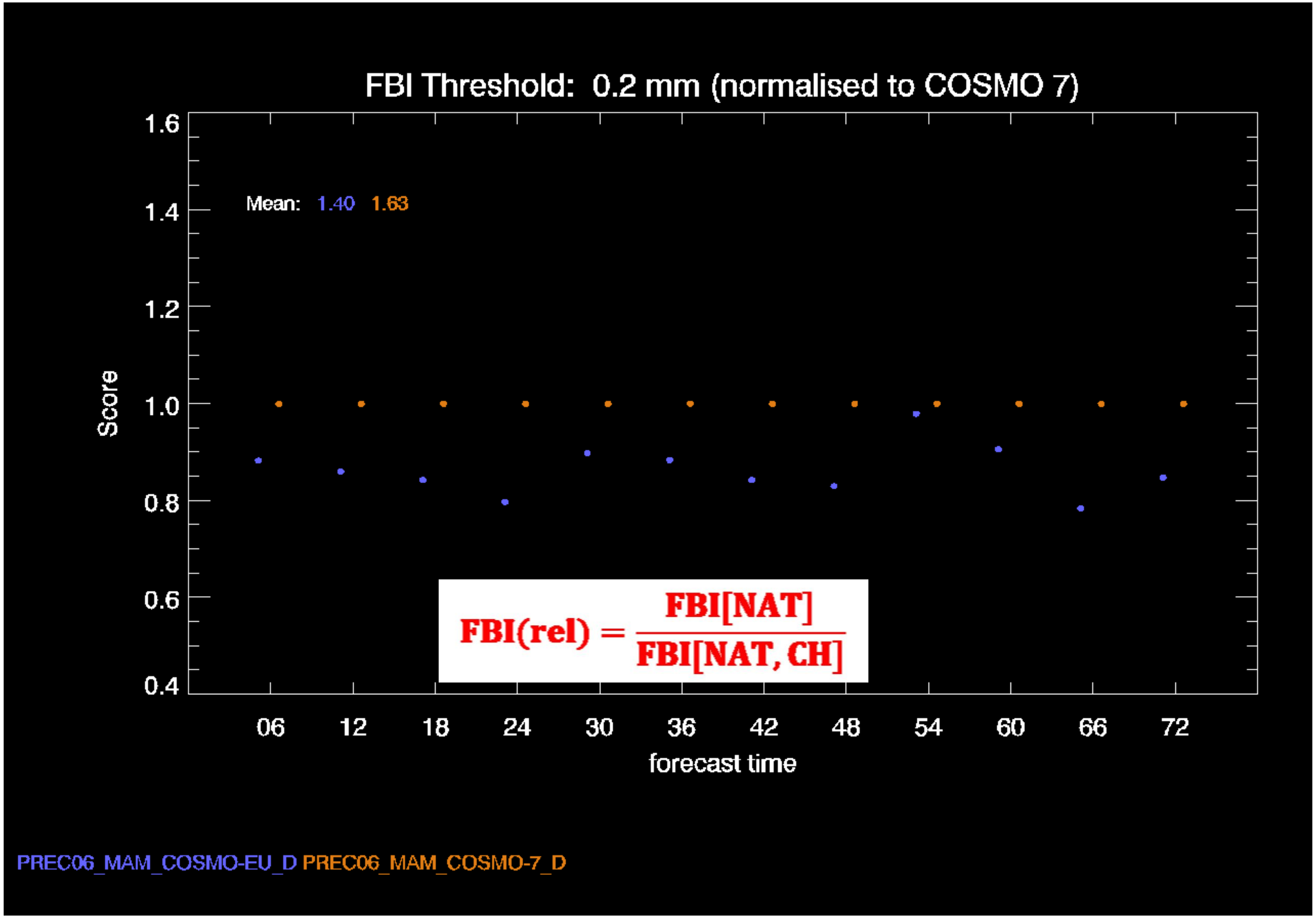


2.)The situation over the common area demonstrates the QPF quality of different models over the same region

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

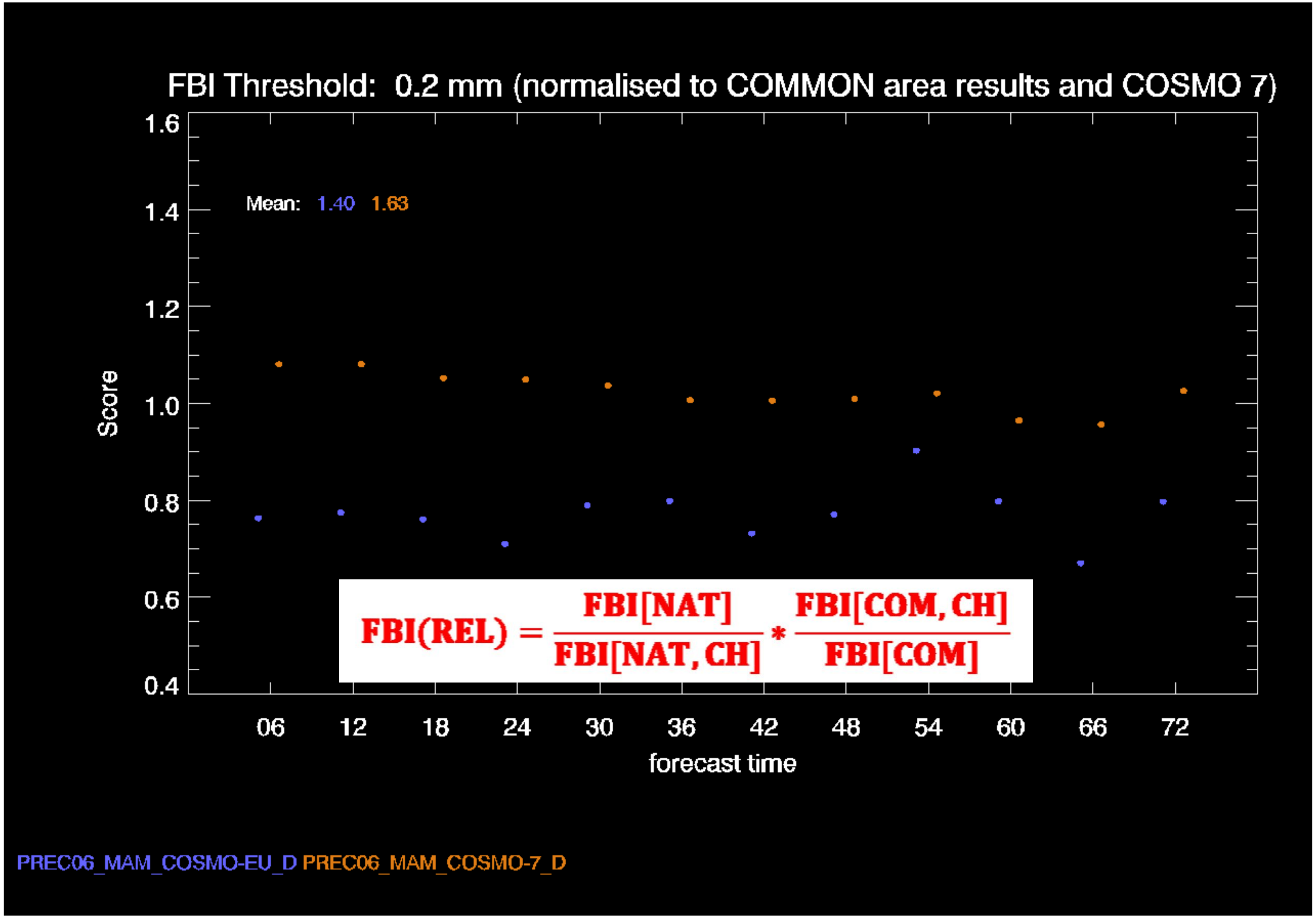


3.) Elimination of region specific properties
national chosen stations: relation COSMO-EU – COSMO-7, demonstrates the QPF quality over different regions

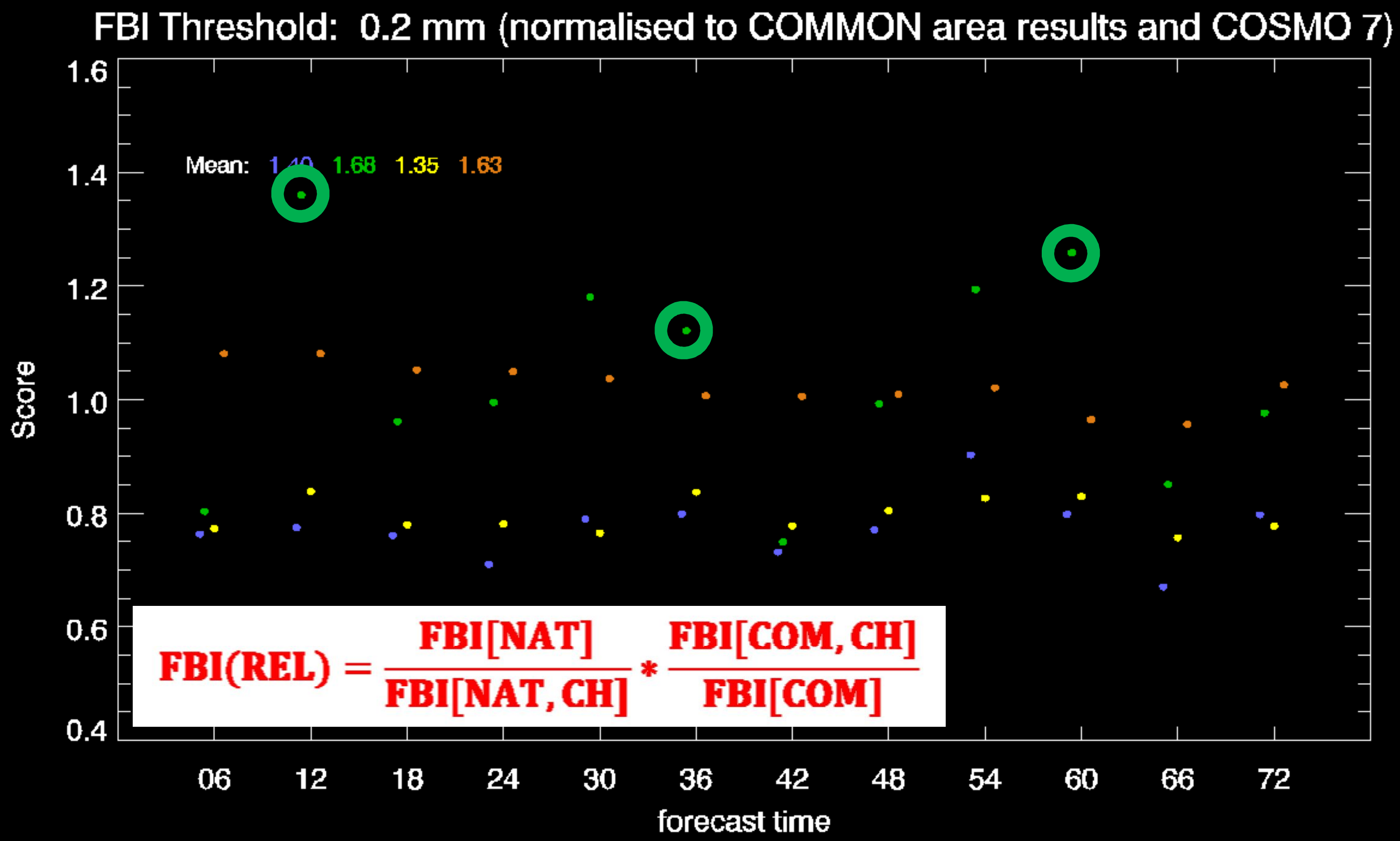


$$\mathbf{FBI(rel)} = \frac{\mathbf{FBI[NAT]}}{\mathbf{FBI[NAT,CH]}}$$

4.) National chosen stations: relation CEU – C7, normalised to common stations, demonstrates the QPF quality over different regions with „elimination“ of model specific errors



5.) Common plots, MAM 2013, national chosen stations
FBI for threshold 0.2mm $(6h)^{-1}$ relation CEU – C7,
normalised to common stations, all model versions



$$FBI(REL) = \frac{FBI[NAT]}{FBI[NAT, CH]} * \frac{FBI[COM, CH]}{FBI[COM]}$$

```
Sun   Rise    Set
Ath  3:52 – 16:58
Rom  4:32 – 17:49
Zür  5:40 – 18:12
Valid 15.04.
```

PREC06_MAM_COSMO-EU_D PREC06_MAM_COSMO-GR_D
PREC06_MAM_COSMO-ME_D PREC06_MAM_COSMO-7_D

➔ **Demonstrates:**

- **Sun rises earlier over Greece than over all other verification regions!**

WG5 Contributing Scientists

Ulrich Damrath, DWD
Francis Schubiger , MCH
Pirmin Kaufmann, MCH
Angela Celozzi, USAM
Adriano Raspanti, USAM
Flora Gofa, HNMS
Dimitra Boucouvala, HNMS
Joanna Linkowska, IMGW
Rodica Dumitrache, NMA
Amalia Iriza, NMA
Anastasia Bundel, RHM
Anatoly Muraviev, RHM
Alexander Kirsanov, RHM
Maria Stefania Tesini, ARPA-SIM
Elena Oberto, ARPA-PT
Naima Vela, ARPA-PT