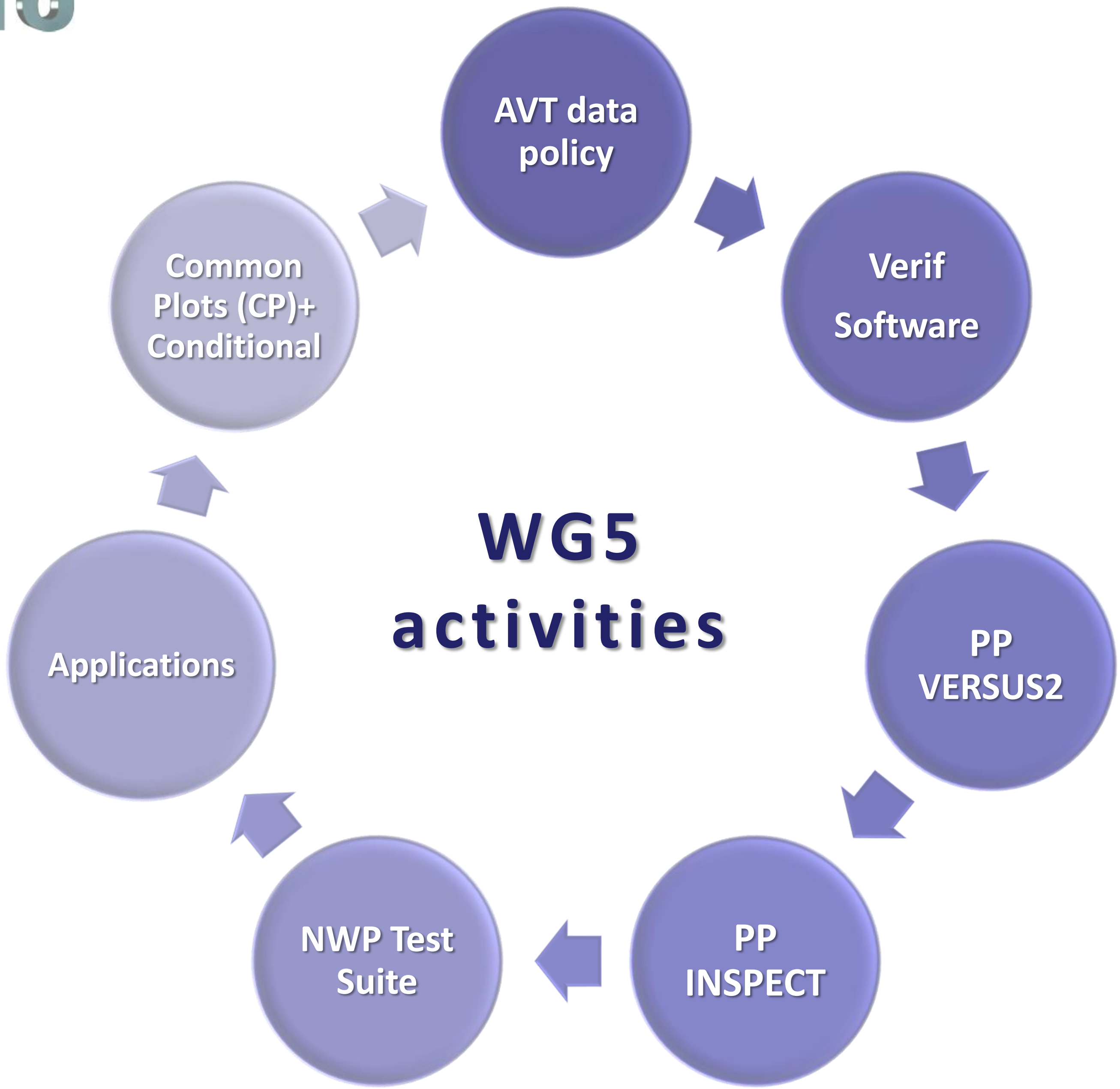
Verification and Case studies

Overview of activities

Flora Gofa

18TH COSMO General Meeting , 5-8 Sept 2016, Offenbach, WG5 Overview


AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities


AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities

Requirements for model output and observation data formats for Verification Tools

Additional Verification Tools (AVT) are targeted to provide code for the implementation of any "useful" verification methods (in addition to the ones included in the CVS).

The adaptation to input datasets, the connection to local Databases, installation specifics and any maintenance issue will be the responsibility of each member that is interested to use them.

AVT in relation with INSPECT, includes spatial methods software (VAST for fuzzy verification, SpatialVX various libraries for displacement or feature based methods), it is anticipated however more verification code to be included in the future (Rfdbk ?)

Data preprocessing of FORECAST data

Forecast model output is used in either GRIB1/2 (e.g. VERSUS), NetCDF (e.g. Rfdbk) or textual format (e.g.CSV)

Processing as AVT input

- Fieldextra used for GRIB1/2 to text (CSV) conversion and for various processing of GRIB1/2 files
- GrADS (Grid Analysis and Display System) to convert and/or interpolate GRIB1/2 to ASCII (also to NetCDF for one variable). PPINSPECT: convert COSMO grib data into ASCII input for R SpatialVx (examples with templates available in WG5 repository).
- R-packages RNetCDF and NCDF to import the NetCDF data in R. For data in netCDF4, an R-package ncdf4 is also available.
- Other R libraries exist for some of the preprocessing tasks. *rNOMADS* used for reading GRIB files in R tuned to NOAA model that allows R users to download global and regional weather model data for processing.
- NetCDF Java library can be used to convert a GRIB file into a NetCDF format on Windows platform. NCL (NCAR command language) can read and write netCDF-3, netCDF-4 classic, netCDF-4, HDF4, binary, and ASCII data. It can read HDF-EOS2, HDF-EOS5, GRIB1, GRIB2, and OGR files

Data preprocessing of OBSERVATION data

BUFR is the WMO standard binary code for the representation and exchange of observational data (point observations)

Radar data (HDF4-5, XML)

OPERA Data Center creates 3 composite products: instantaneous surface rain rate, instantaneous max reflectivity and 1 hour rainfall accumulation. Data are available in both **BUFR** and **HDF5** formats.

Like all observational data, many issues (e.g. problems with parallax, radar beams above the precipitation at long ranges, evaporation of rainfall at lower levels beneath the beam, anomalous propagation, etc.) have to be considered when using such data for verification purposes.

Satellite data sources (HDF4, GRIB, BUFR)

For PP INSPECT and other WG5 applications, satellite data from the following sources/formats are used:

- SEVIRI/Meteosat (EUMETSAT) data: cloud cover, rain intensity, surface temperature and albedo. Formats: HDF4, GRIB, and ASCII.
- Polar-orbiting satellites (NOAA, Terra, Aqua, Suomi NPP): vertical profiles, Total Precipitable Water (TPW), and ozone. Formats: NetCDF, HDF4, and ASCII.
- EUMETCast products (SAF EUMETSAT): rain intensity, precipitation accumulations, snow cover, and surface humidity. Formats: GRIB, BUFR, and HDF4

Suggestions

<table>
<tr><th></th><th>Common formats</th><th>AVT format</th><th>Actions</th><th>Code</th><th>Sources</th></tr>
<tr><td>Forecast</td><td>GRIB1/2
NetCDF</td><td>GRIB1/2
Textual
NetCDF</td><td rowspan="2">QC
Decode
Interpolate
Extrapolate
Change of projection
Aggregation (spatially/timely)</td><td>R libraries
Fieldextra
LIBSIM</td><td>COSMO</td></tr>
<tr><td>Obs</td><td>BUFR
GRIB1/2
HDF4
NetCDF</td><td>Textual</td><td>R libraries
Fieldextra
LIBSIM
GRADS</td><td></td></tr>
</table>

• Due to the wide variety of formats and other issues that must be considered depending on the instrument, measured parameter/index, volume of data, geographic projection, etc., **many approaches can be followed to prepare information as input for AVT**

• **While it would be desirable to have a unique COSMO tool dedicated to preparation of verification data, it is currently not deemed possible given the diversity of database structures, types of observational data used, requirements to interface with models other than COSMO, etc.**

•Nevertheless, **any future COSMO data processing software development should also aim to fulfill the requirements of verification applications**. Treatment of non-gridded binary observations (BUFR) as well as the handling of data formats such as NetCDF or HDF4/5 are considered to be desirable features of any such preprocessing software. It is suggested that NetCDF support should be included in existing COSMO preprocessing software

•Exchange of “community code” for verification input data processing. Expand WG5 repository with code and/or instructions for AVT input data preparation


AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities

Feedback File Based Verification at DWD
-recent developments-

I. Recap on Rfdbk Concept

- Feedback files are produced during DA and contain observations, analysis and past forecast and lots of meta information valuable for observation based verification tasks
- Using feedback files for the verification means a huge reduction in workload as much of the tedious data preparation tasks are done within DA
- Rfdbk (source on https://github.com/rfxf/Rfdbk) is a R interface for COSMO feedback files
- Main purpose of Rfdbk is to load feedback file content with R
- Additional functionality useful for verification is implemented as well
- As guideline a set of verification scripts (used at DWD) can be obtained via https://github.com/rfxf/fdbk_verification

IV. Recent verification progress

Deutscher Wetterdienst
Wetter und Klima aus einer Hand
DWD

Status

- Observation Types: SYNOP, TEMP, GPSRO, SATOB (AMV), PILOT (wind profiler)
- Models: ICON, ICON_P, ICON_P1, ICON-EU, ICON-EU, ICON-EU_P1, ICON-EPS, ICON-EPS_P1, COSMO-EU, COSMO-DE, COSMO-DE_P, COSMO-DE-KENDA, COSMO-DE-EPS, COSMO-DE-EPS_KENDABCEPS, COSMO-DE-EPS_KENDAICON, IFS + experiments
- Verification types: continuous, categorical, ensemble, probabilistic
- Aggregation: by period, by valid-time, by station, time series of monthly means

Missing

- Complete set of SYNOP observations in ICON feedback files
- Observations contained in national bufr file (RR and GUST over Germany)

Visualization examples (COSMO (DE-EPS) routine vs. KENDA)

- Ensemble / Probabilistic
- Continuous / Categorical / Hit rates
- Continuous by Station
- Continuous time series
- DIY/WMO

V. Future Plans

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

- Address issues with SYNOP verification of precipitation and gusts
- Work on confidence intervals for (some) scores (analytically or bootstrap)
- Move the verification scripts into routine/experiment environment to allow for an online verification
- Use aircraft measurements in verification/monitoring

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

Status of UA verification with MEC+Rfdbk at MCH

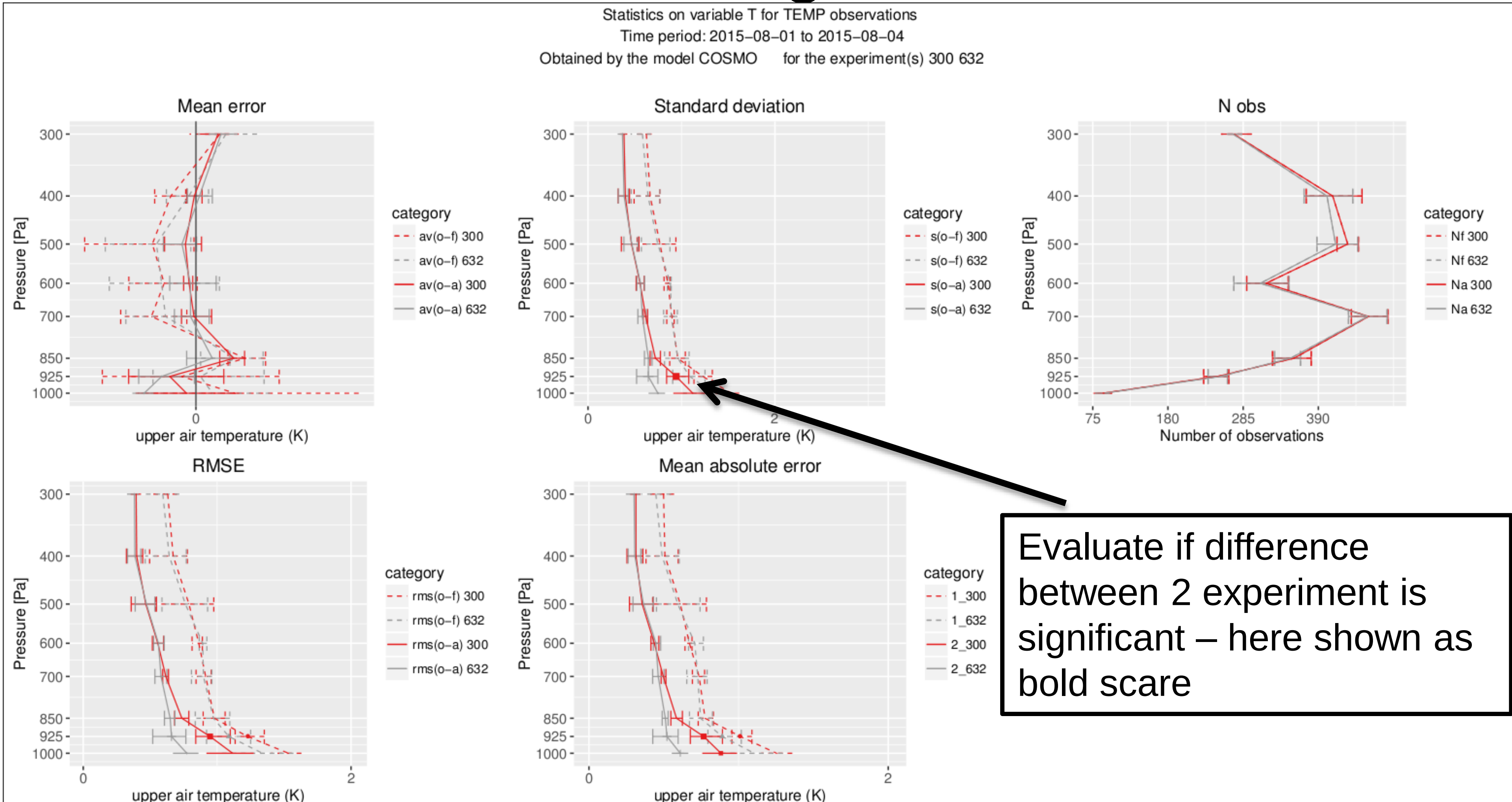
Xavier Lapillonne, Daniel Leuenberger, Josué Gehring, Yann Lepoittevin

Current usage and plan

- MEC is run daily to generate feedback files from the latest operational run
  - currently COSMO-1 and 7
  - Observation types: TEMP, PILOT, AIREP, SYNOP

- Rfdbk verification package is used for upper-air seasonal verification of operational forecast:
  - COSMO-1, COSMO-7 (TEMP only – will start for JJA 2016)
  - COSMO-E (currently only control tested – investigation needed for probabilistic scores)

Other features: significant difference



Evaluate if difference between 2 experiment is significant – here shown as bold scare

User experience

- MEC : learning curve for using it. The software is still not used in many configurations. We had some issues with MCH specific configurations (e.g. sleeve – now fixed thanks to Andreas Rhodin)

- Rfdbk : powerful and flexible package – however it requires to invest time in learning “R” in order to use it efficiently

- We have developed our own functions/driver scripts/unittests that we could share (currently on private github at MCH)

VAST, what's new

Naima Vela

COSMO General Meeting 2016, Offenbach

September the 5th 2016

Software improvements version 1.4 (after test phase)

- Manual renovation
- BUFR reading problems
- Automatic delete of configuration files
- Possibility to re-run verification without re-load input files
- Check of the verified area
- Improvement in the naming of the files
- Absolute and common paths
- PDF/PNG output choice
- KM/GRID POINTS selection for the space scale

Future developments

- Possibility to verify other variables:
  - Total cloud cover
  - 2 m temperature
  - Wind speed
- Extention of the verification to the time dimension
  - Possibility to add a dimension to the verification
- Code refinements

![COSMO logo]


AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities

PP VERSUS2 extension – End of the project

An extension project plan for VERSUS2 was agreed in order to complete pending implementations (GRIB2 implementation, EPS scores consolidation, documentation, final testing) and to complete all test phases

Presentation to follow by VERSUS SCA : Antonio Vocino

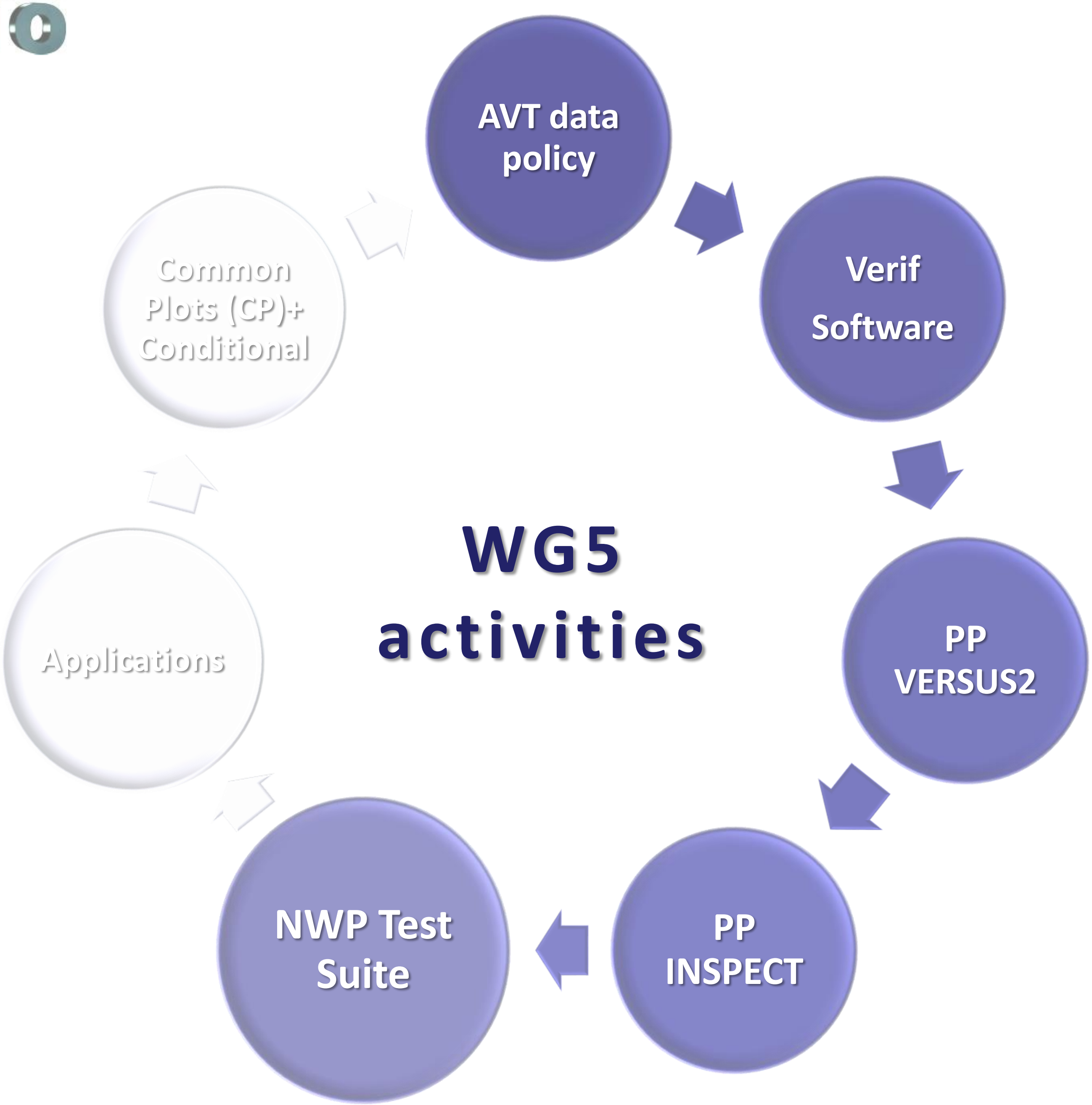

AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities

INSPECT: INtercomparison of SPatial vErification methods for COSMO Terrain

Priority Project: April 2015 – Sept 2017

A. Bundel, F.Gofa

Presentation to follow by Anastasia


AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities

Testing each new model version against last official version

• Responsible for verification and the evaluation report with the verification summary and the recommendation based on relative model performance

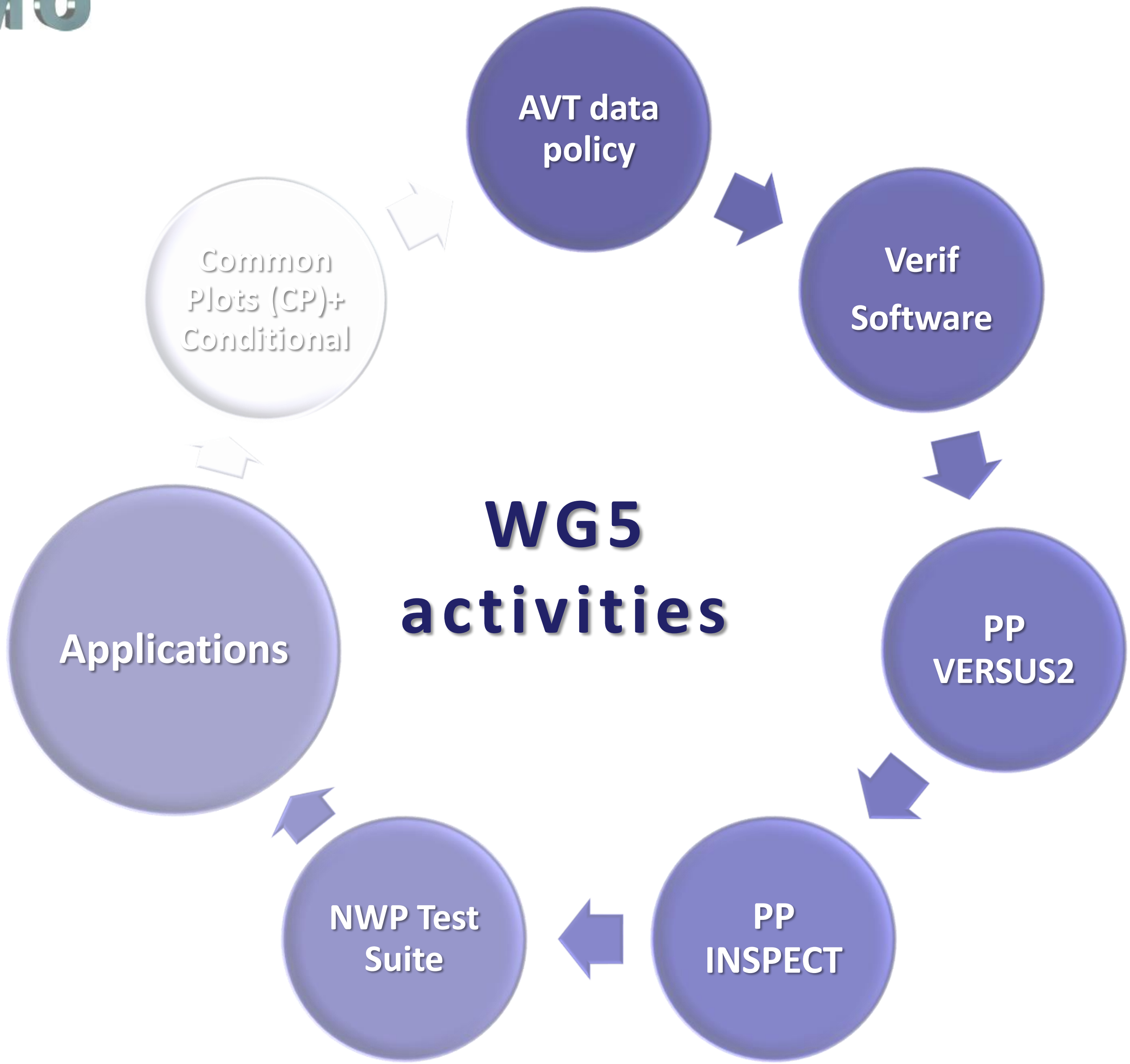
• For 2015-16:

✓Test 5.03 vs. 5.02

✓Test 5.04a vs 5.03

•Joint meeting with WG6: Proposals for report improvement as well as of the verification metrics/methods

Plans for next NWP test suite tests to be presented by Massimo Milleli (WG6)

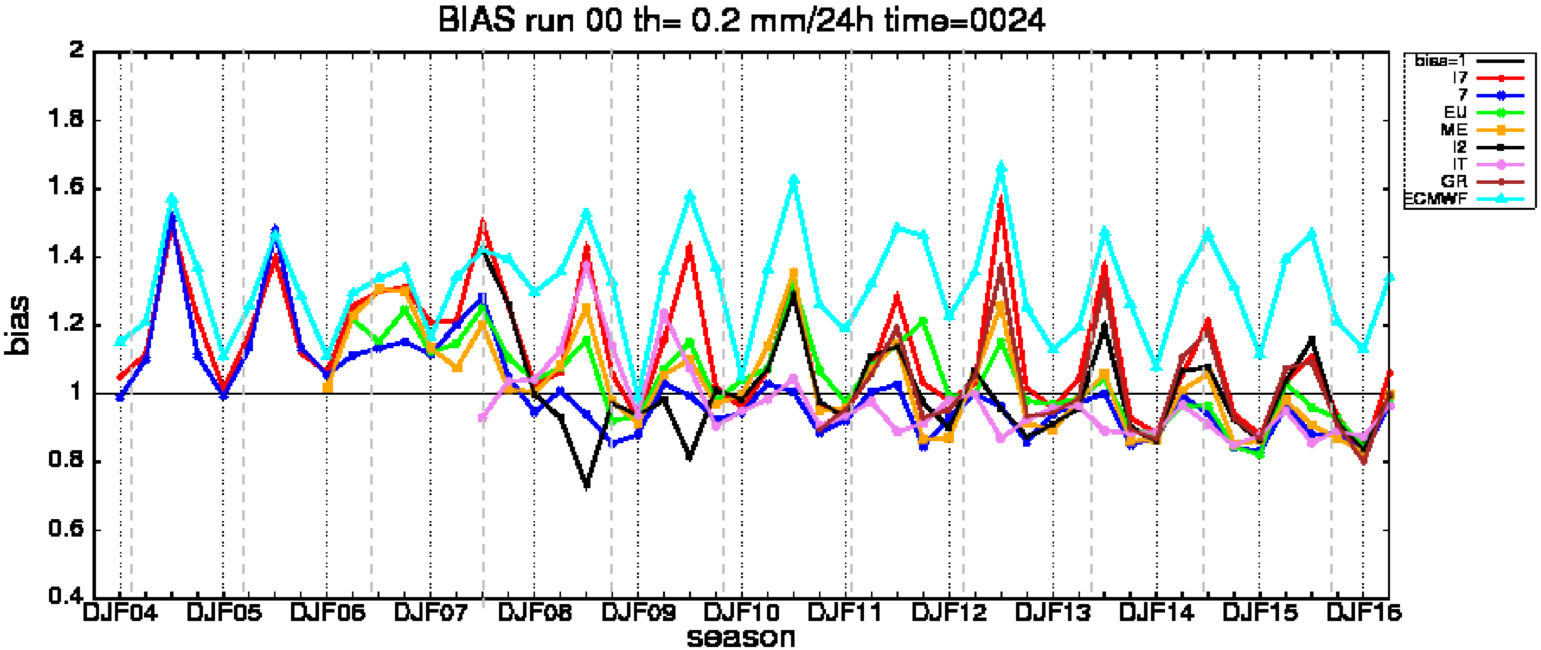

AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities

The methodology

Precipitation- high resolution network

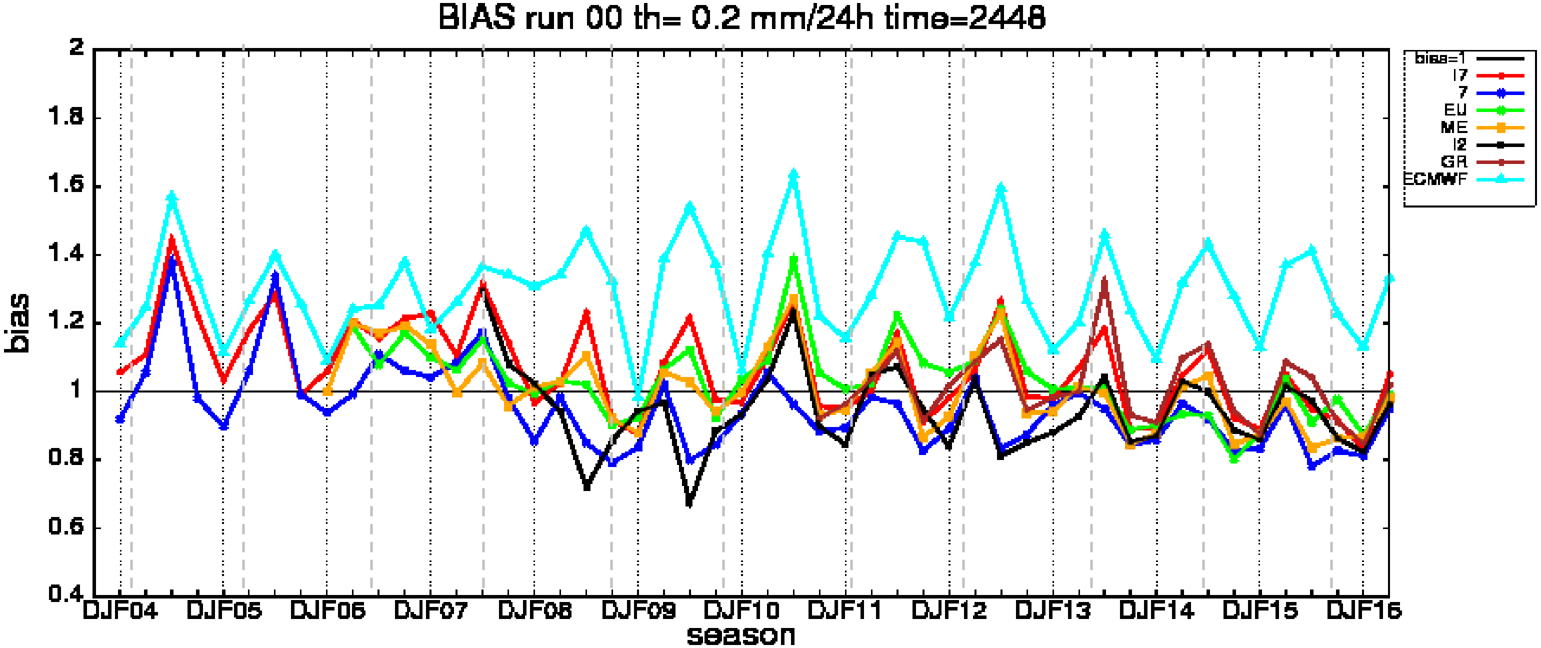
•Common area → Italy
•Dataset → high res raingauges
•Method → 24h/6h averaged cumulated precipitation or maximum values (both observed and forecasted) over meteo-hydrological basins

LONG TREND PRECIPITATION with high resolution stations



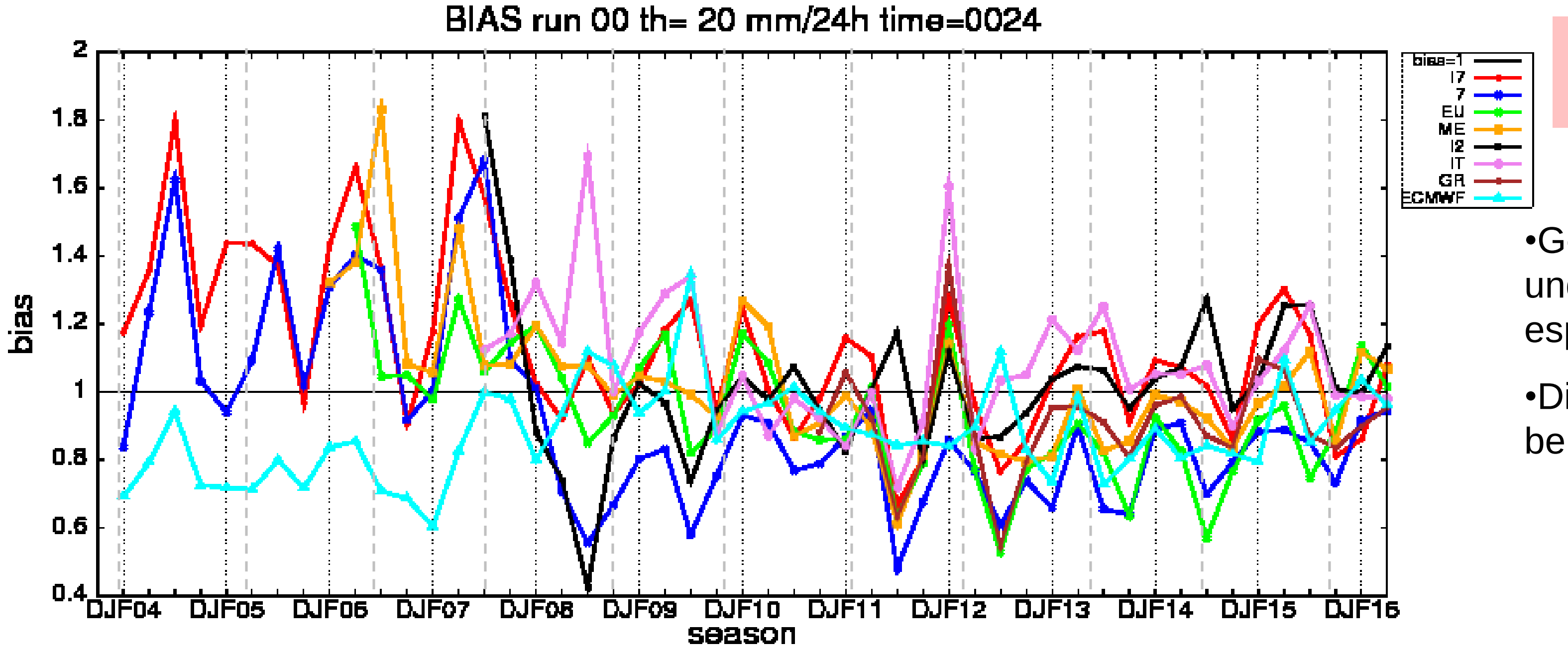
LOW THRESHOLDS

- •Ecmwf overestimation
- •Summer overestimation
- •Reduction of the overestimation for LAM



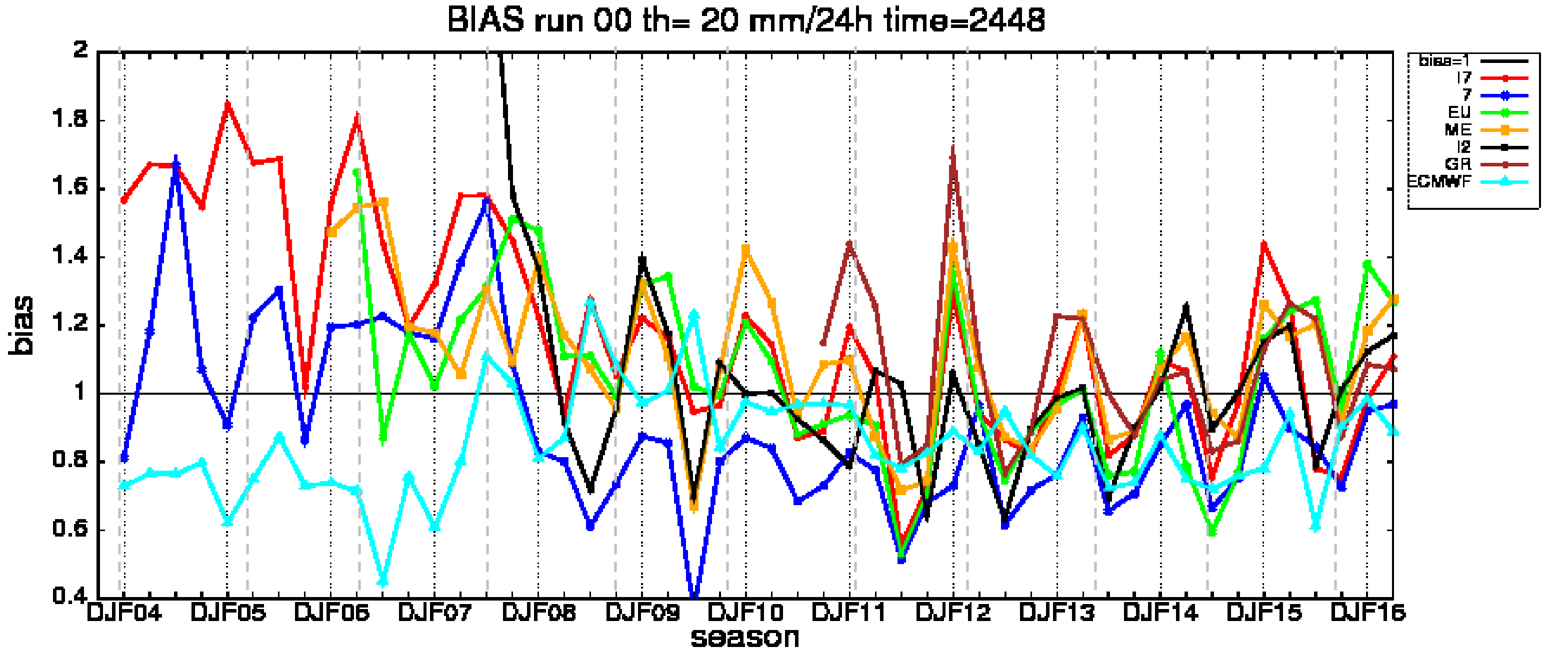
- •Ecmwf overestimation
- •Reduction of bias
- •Increasing winter underestimation

LONG TREND PRECIPITATION with high resolution stations

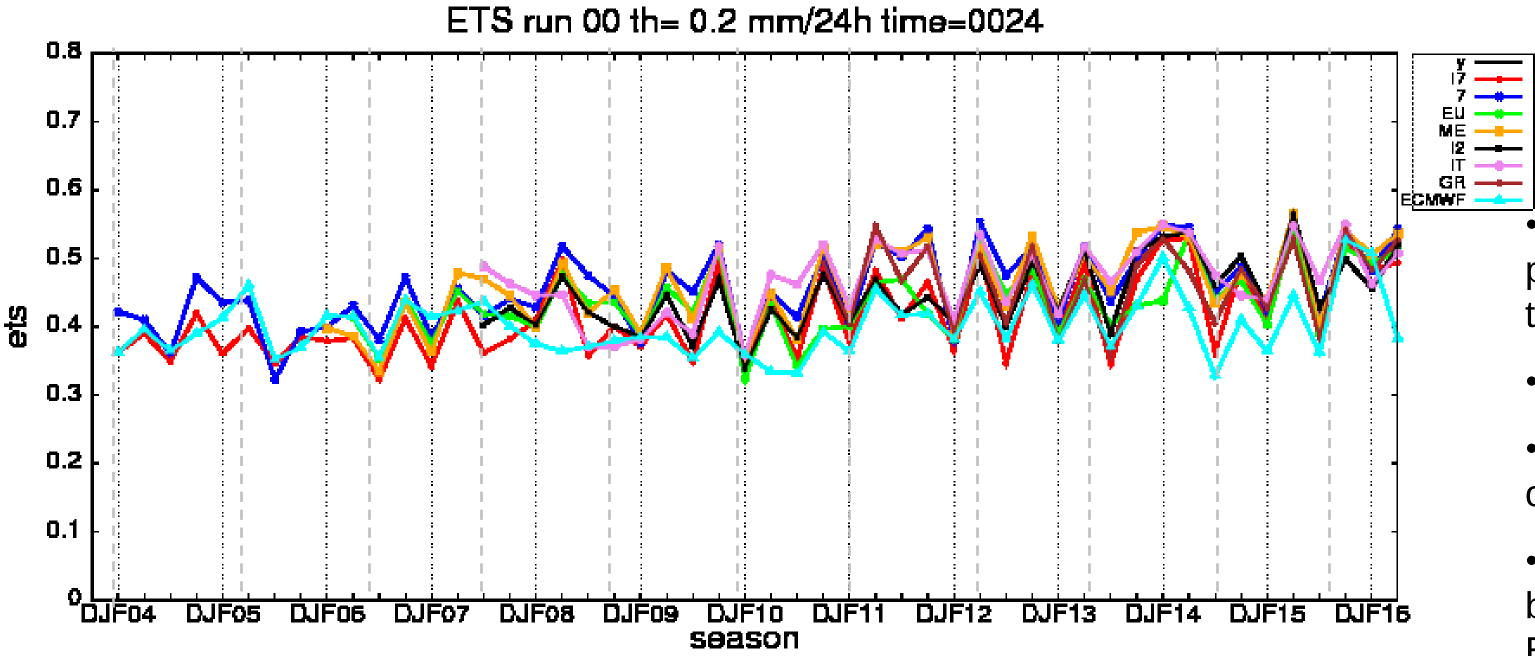


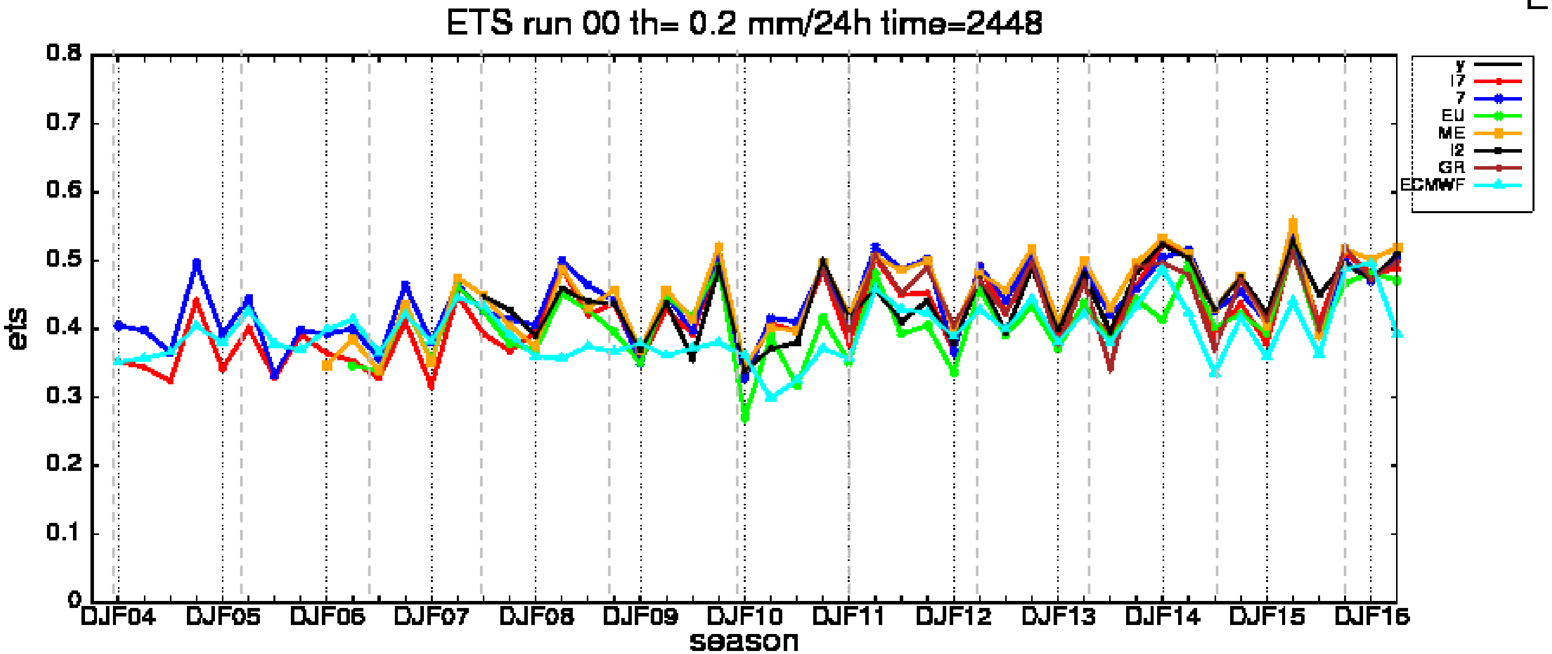
HIGH THRESHOLDS

- General underestimation, especially 7, EU
- Different behavior



LONG TREND PRECIPITATION with high resolution stations

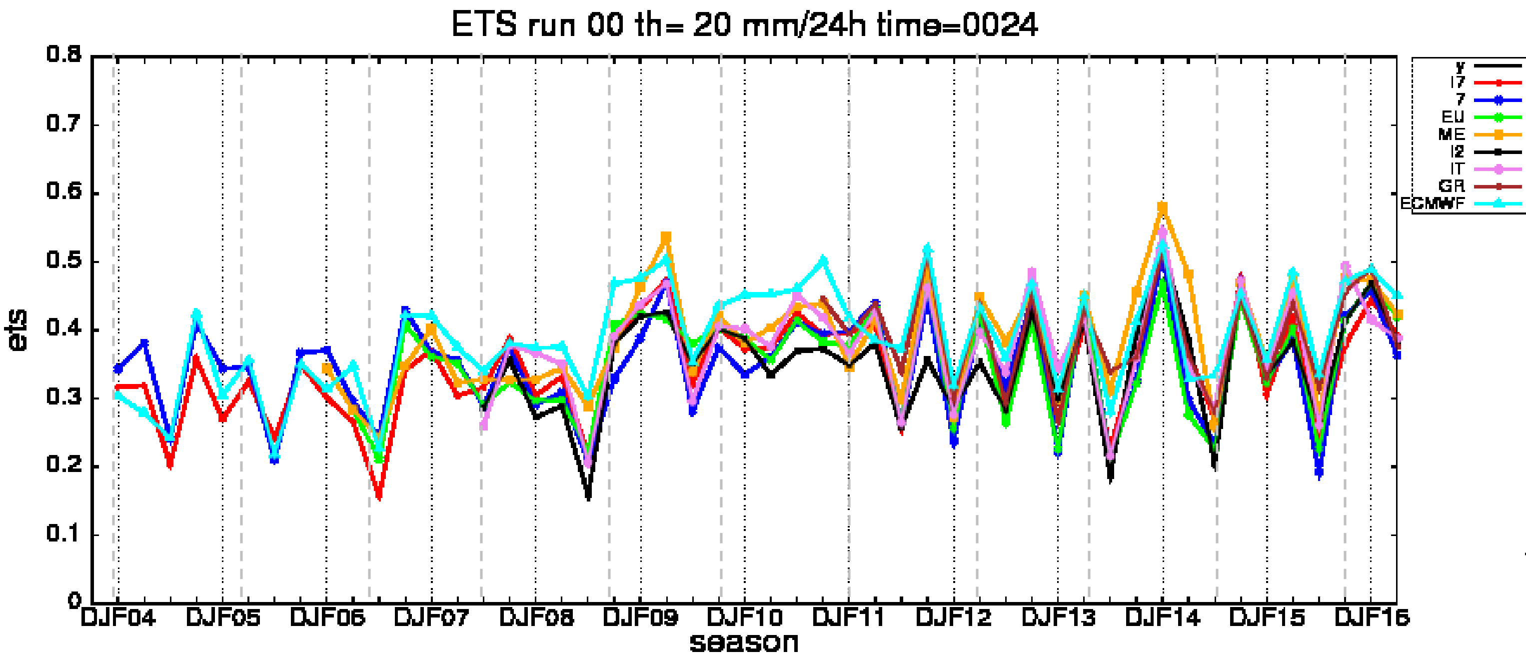




LOW THRESHOLDS

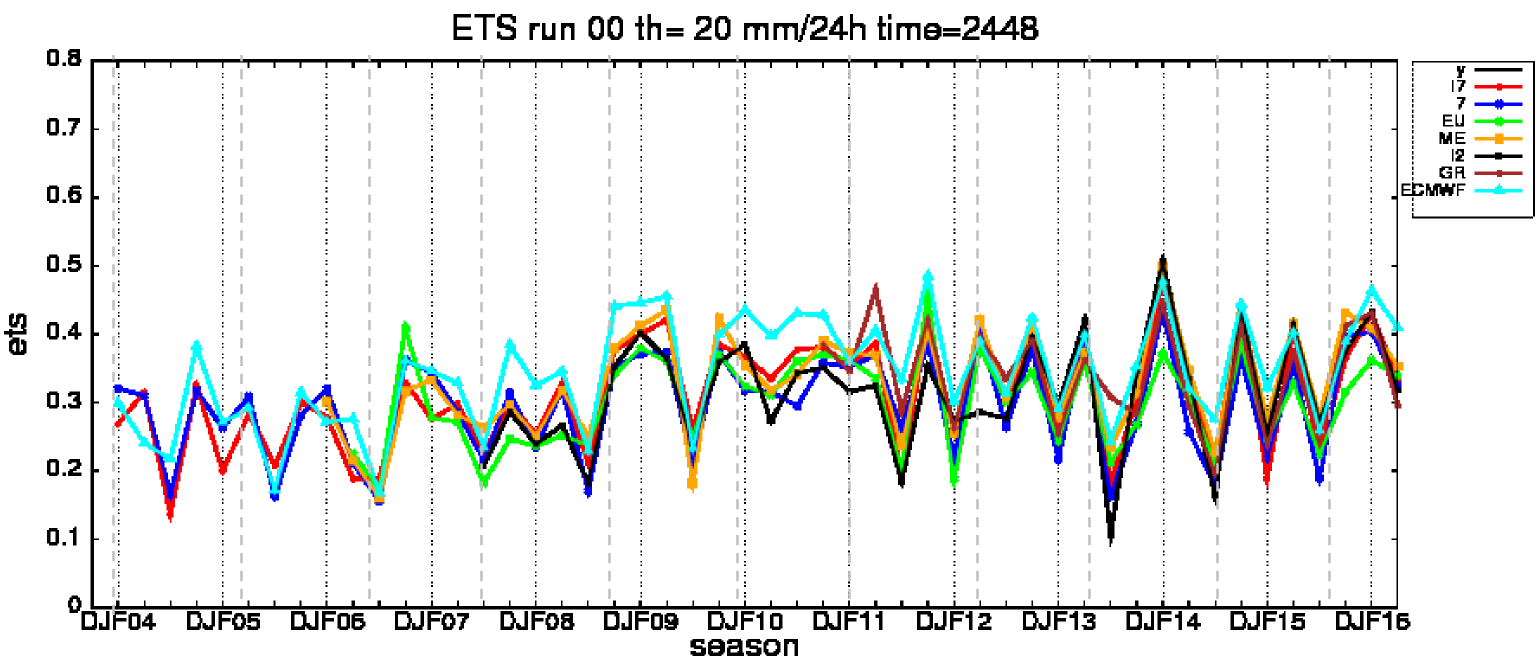
- Very slightly positive/steady trend
- Good ME,7
- Big seasonal oscillation
- LAM perform better than ECMWF

LONG TREND PRECIPITATION with high resolution stations



HIGH THRESHOLDS

- Very slightly positive trend
- Big seasonal oscillation
- ECMWF often performs better than LAM



Complimentary assessment of forecast performance with climatological approaches

F.Gofa, V. Fragkouli, D.Boucouvala

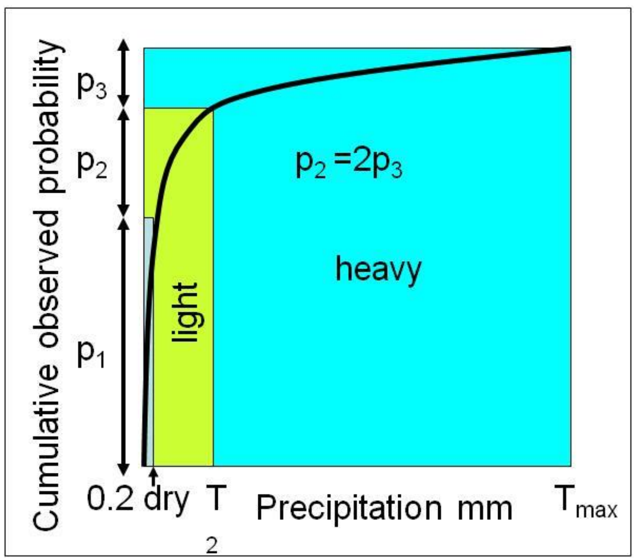
The use of SEEPS with metrics that focus on extreme events, such as the Symmetric Extremal Dependence Index (SEDI) that is adjusted to the climatological distribution of precipitation at each location, enables assessment of locally important aspects of the forecast while providing a reliable performance measure.

ΕΘΝΙΚΗ
ΜΕΤΕΩΡΟΛΟΓΙΚΗ
ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

Stable Equitable Error in Probability Space (SEEPS)

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

- 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 - , e.g. constant)



•*The SEEPS index matrix was calculated as the scalar product of the SEEPS weights matrix and the contingency table of total available model/observation pairs for each station averaged over the number of the days of the month.*

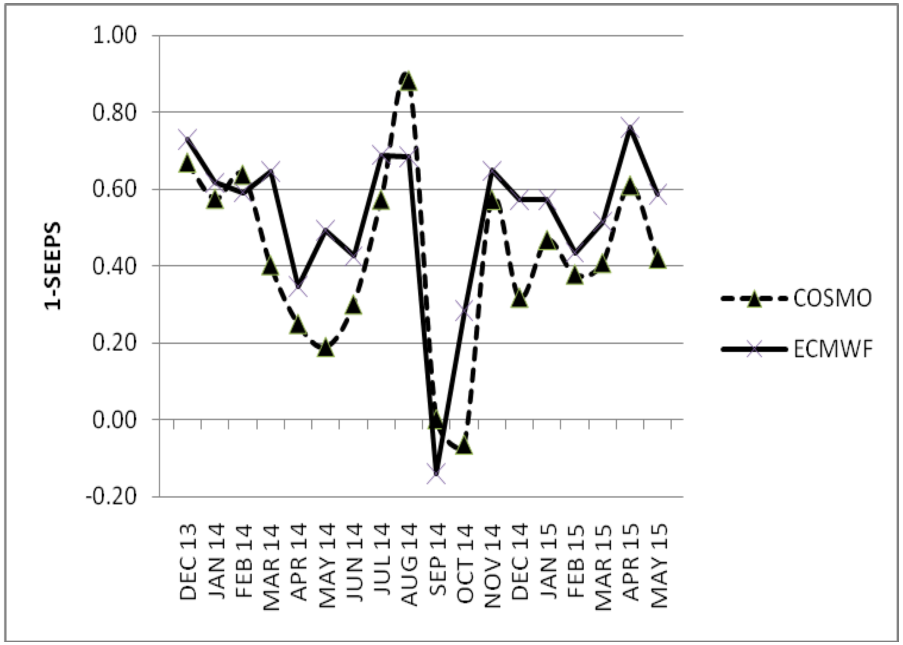
Error scoring matrix

$$\{s^S_{vf}\} = \frac{1}{2}\begin{Bmatrix} 0 & \frac{1}{1-p_1} & \frac{1}{p_3}+\frac{1}{1-p_1} \\ \frac{1}{p_1} & 0 & \frac{1}{p_3} \\ \frac{1}{p_1}+\frac{1}{1-p_3} & \frac{1}{1-p_3} & 0 \end{Bmatrix}$$

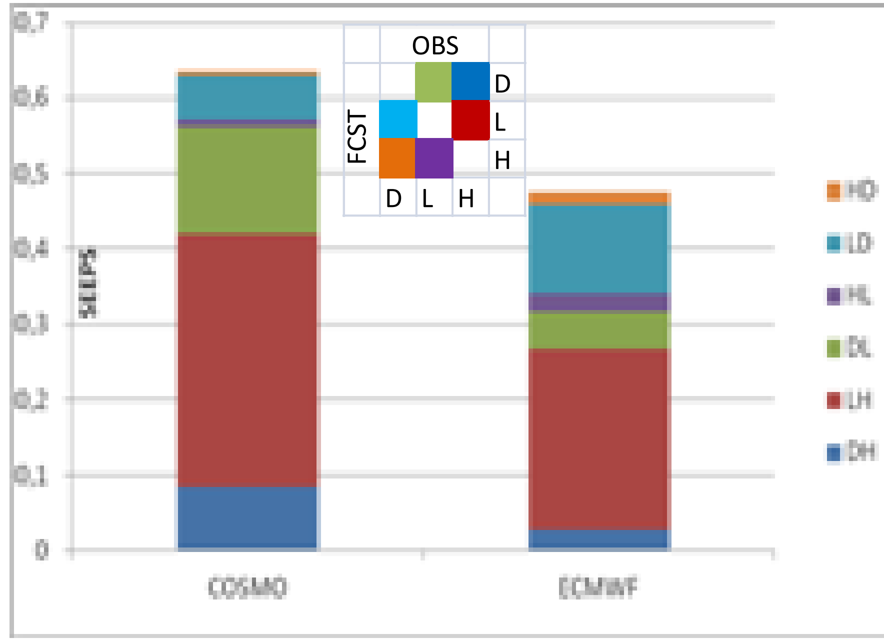
p1+p2+p3=1, p2=2p3

- Dry weather is defined as less or equal 0.2mm/24h
- The SEEPS index matrix elements are **HD** (modeled Heavy-observed Dry), **LD** (modeled Light, Observed Dry), **LH** (modeled light, observed Heavy), **DH** (modeled Dry, observed Heavy).

Monthly variation of 1-SEEPS during the observational period

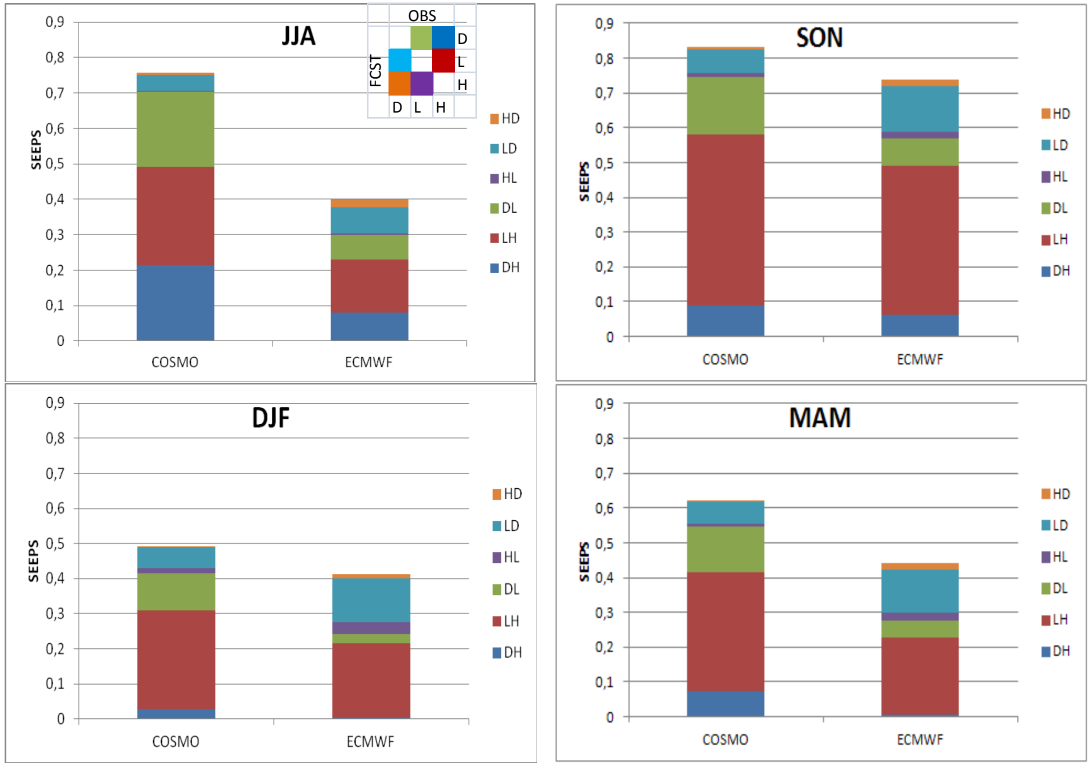


Decomposition of SEEPS for the whole period analyzed



Time series for SEEPS (24h rain) exhibits poorer performance during the summer months while the ECMWF model consistently delivers better performance than the COSMO model. Both models have largest SEEPS error contribution for the ‘light’ category when ‘heavy’ was observed.

Seasonal Decomposition of SEEPS for COSMO and ECMWF models



For stations with moderate-to-dry climatologies (p1>0.5), such as Greece, predicting ‘light’ rainfall when ‘heavy’ is observed is penalized considerably more than predicting ‘light’ when ‘dry’ is observed (blue). For IFS/ECMWF, SEEPS is mainly connected with LD and LH categories, indicating that has the tendency to smooth out preci forecasts. **COSMO model is penalized for LH and DL categories, leading to the conclusion that its forecast is usually ‘drier’ than that of the ECMWF model, and that the SEEPS score is strongly influenced by this attitude.**

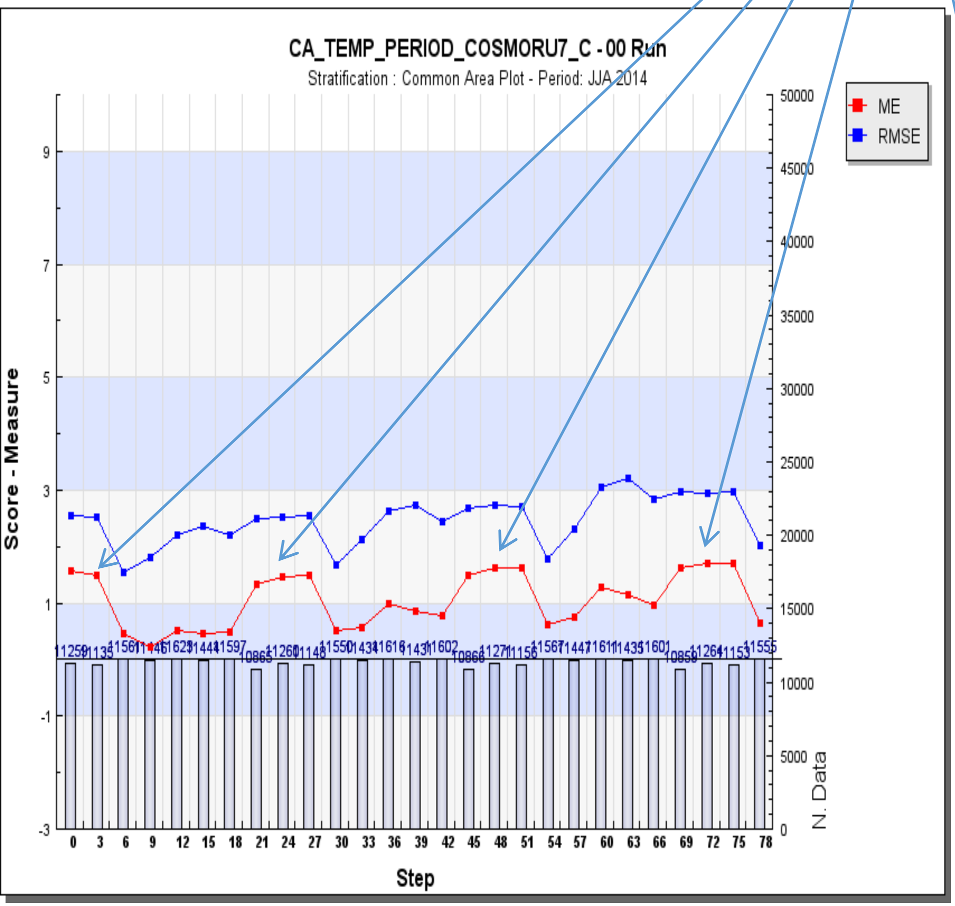
1. Tuning

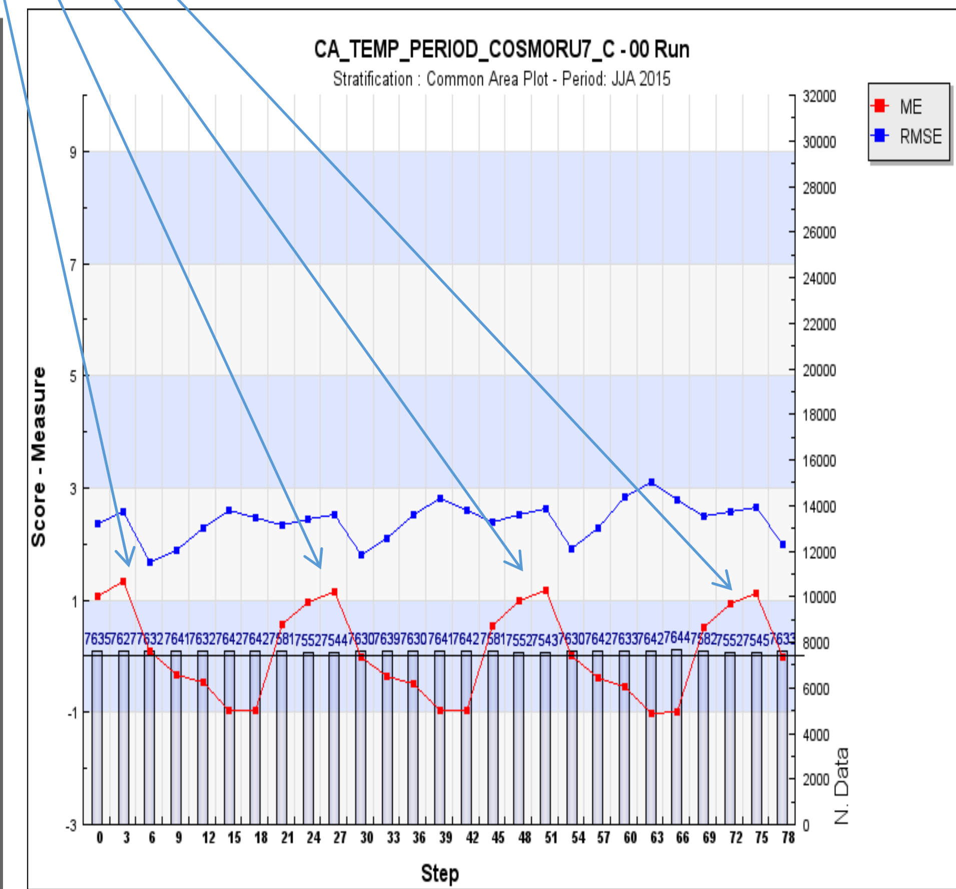
`tkhmin and pat_len`

Operational for all COSMO-Ru versions

since 18 May 2016

Motivation: Too high night near surface temperatures in summers





Experiments

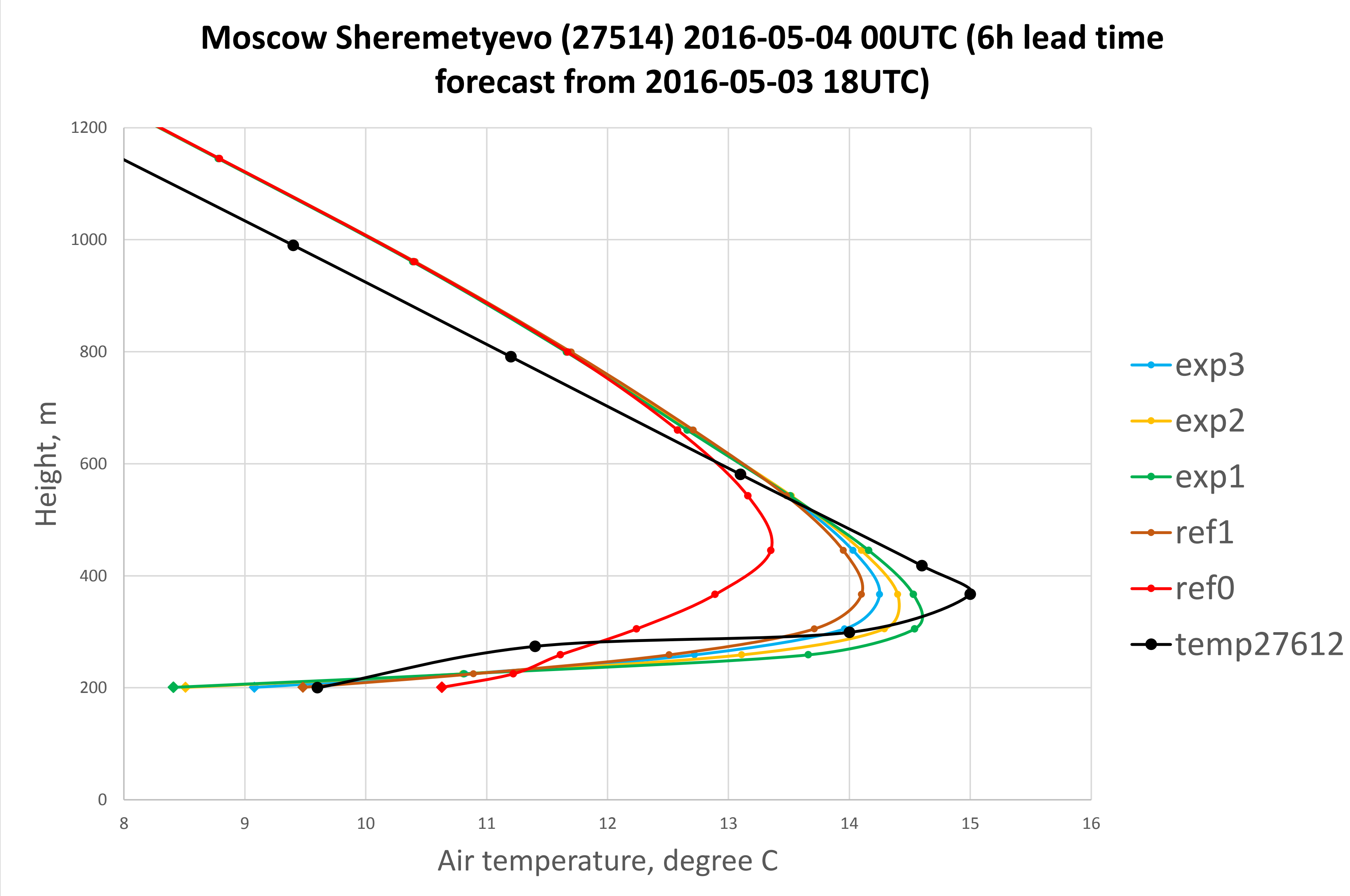
- ref0 – No tuning:

  `tkhmin = 0.4, pat_len = 500`
- ref1 – test on pat_len influence:

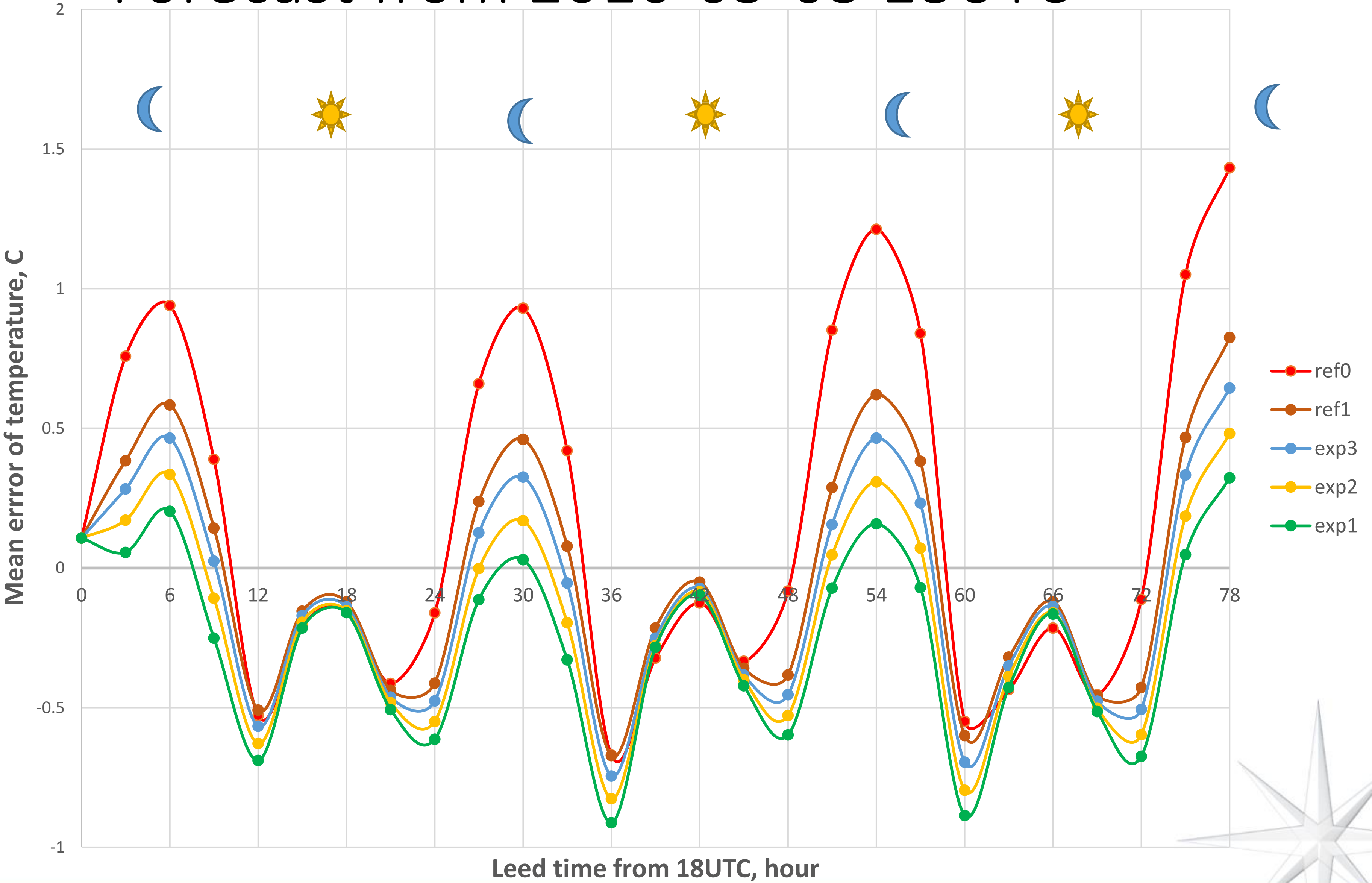
  `tkhmin = 0.4, pat_len =` **`50`**
- exp1 : `tkhmin =` **`0.1`**`, pat_len = 50`
- exp2 : `tkhmin =` **`0.2`**`, pat_len = 50`
- exp3 : `tkhmin =` **`0.3`**`, pat_len = 50`

Case of forecasts from 2016-05-03 18UTC was chosen for experiments, when a strong night overheating was observed in the Central region of Russia.

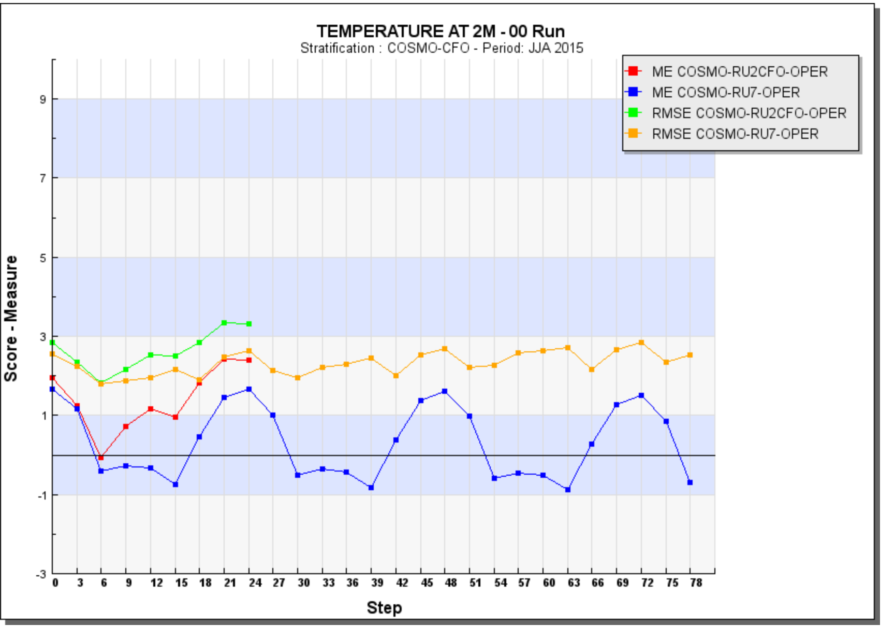
Effect of tuning on night air temperature stratification

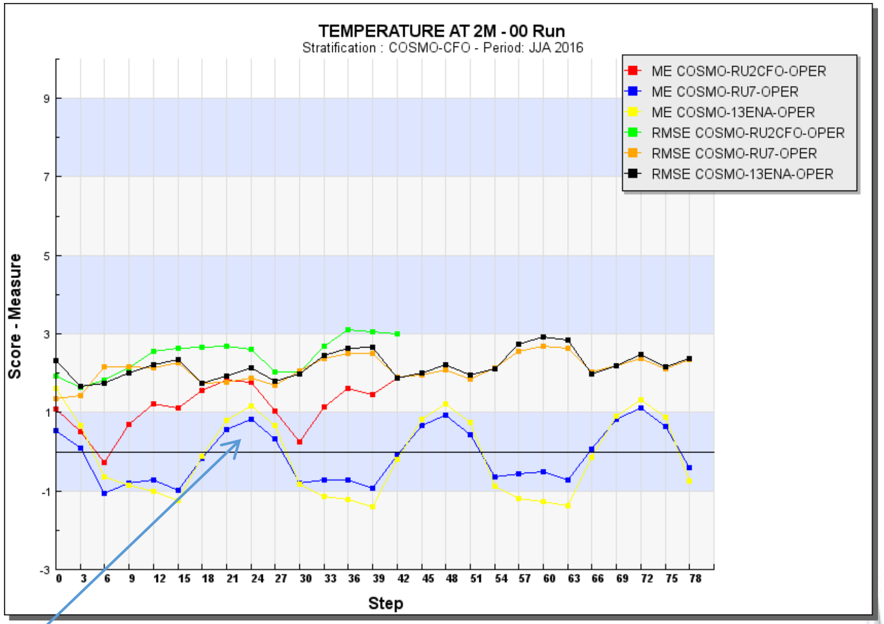


ME for 3155 stations of European Russia, Forecast from 2016-05-03 18UTC

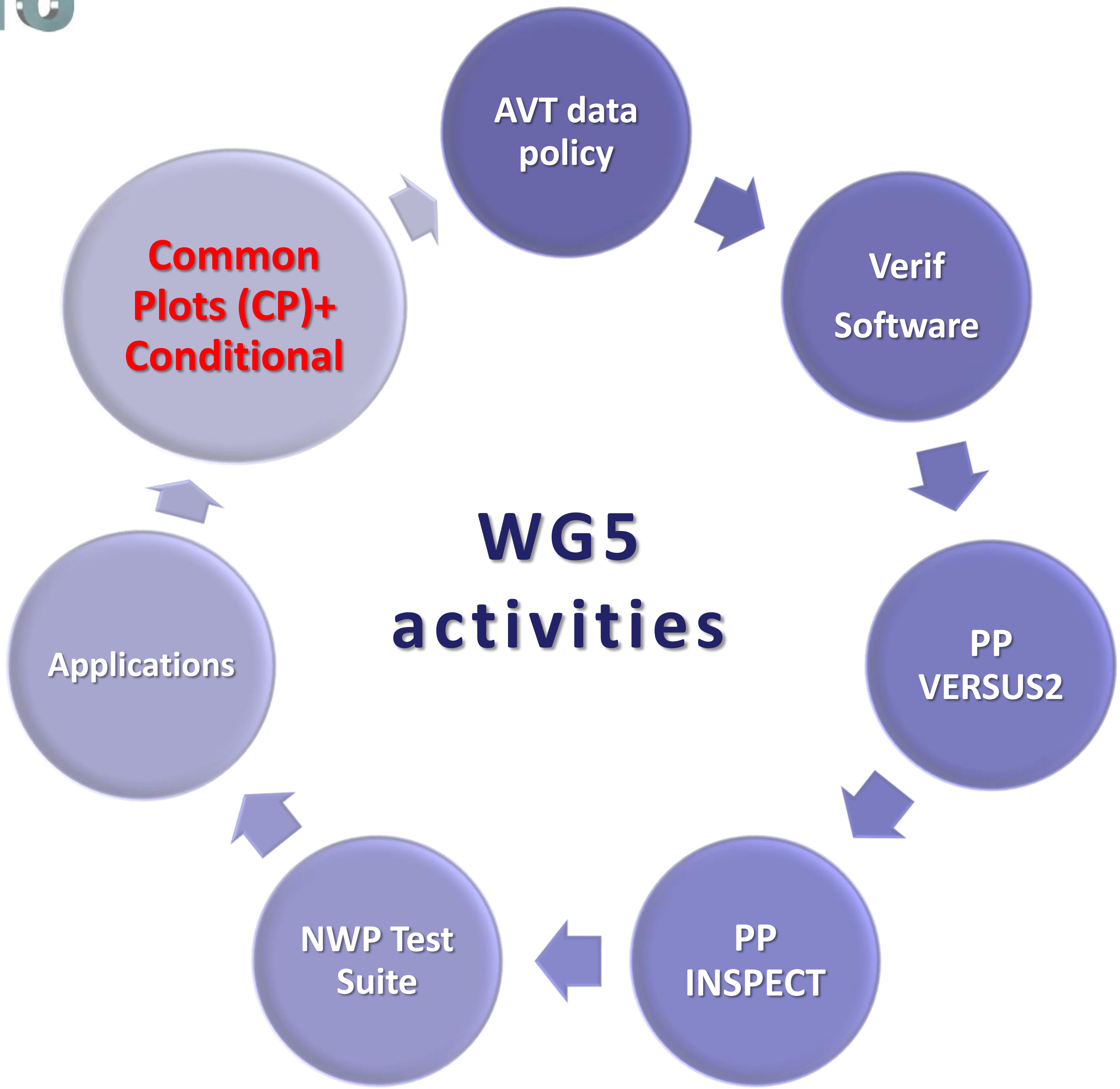


Errors for JJA 2015 (no tuning) and JJA 2016 (new parameters: tkhmin=0.1, pat_len=50, operational from 18 May 2016)





Better scores for night temperature minimums


AVT data policy
Verif Software
PP VERSUS2
PP INSPECT
NWP Test Suite
Applications
Common Plots (CP)+ Conditional
WG5 activities

Common Plot Reports
2015-2016

Presentation of Verification Overview (D. Boucouvala)

Common Plot Activity					
0.2	Reporting			description	status
FTEs	Name	qrt1	qrt2	qrt3	qrt4
0.2	Boukouvala	0.03	0.07	0.03	0.07
0.35	Score Production			description	status
FTEs	Name	qrt1	qrt2	qrt3	qrt4
0.05	Pflüger	0.013	0.013	0.013	0.011
0.05	Lapillonne	0.013	0.013	0.013	0.011
0.05	Vocino	0.013	0.013	0.013	0.011
0.05	Tesini	0.013	0.013	0.013	0.011
0.05	Gofa	0.013	0.013	0.013	0.011
0.05	Linkowska	0.013	0.013	0.013	0.011
0	Dumitrache	0	0	0.013	0.011
0.05	Kirsanov	0.013	0.013	0.013	0.011

4.1 Reporting

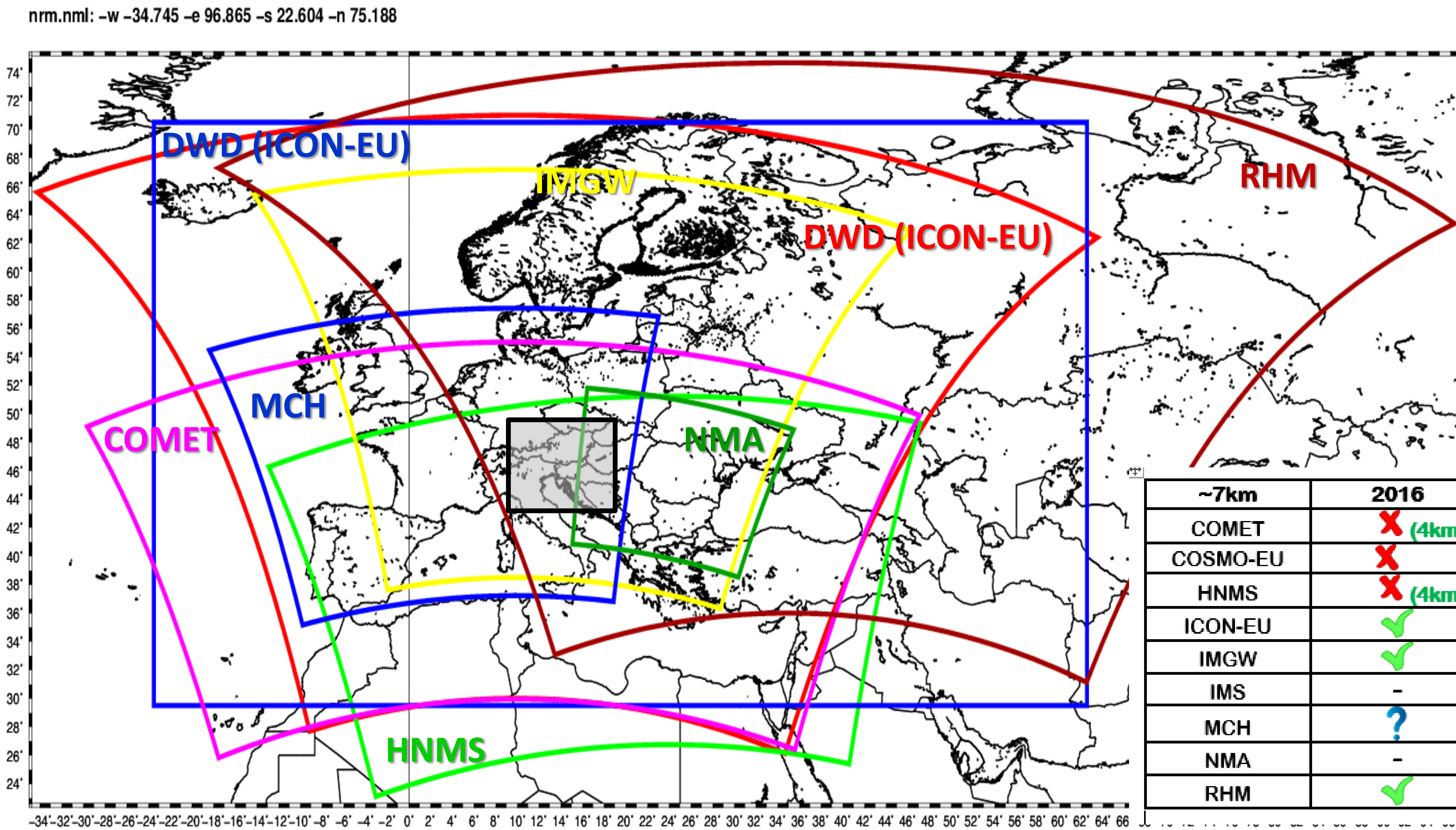
0.2 FTEs for report preparation

D. Boucouvala, HNMS:
graphics preparation, report writing
Web page feeding

4.2 Score Production

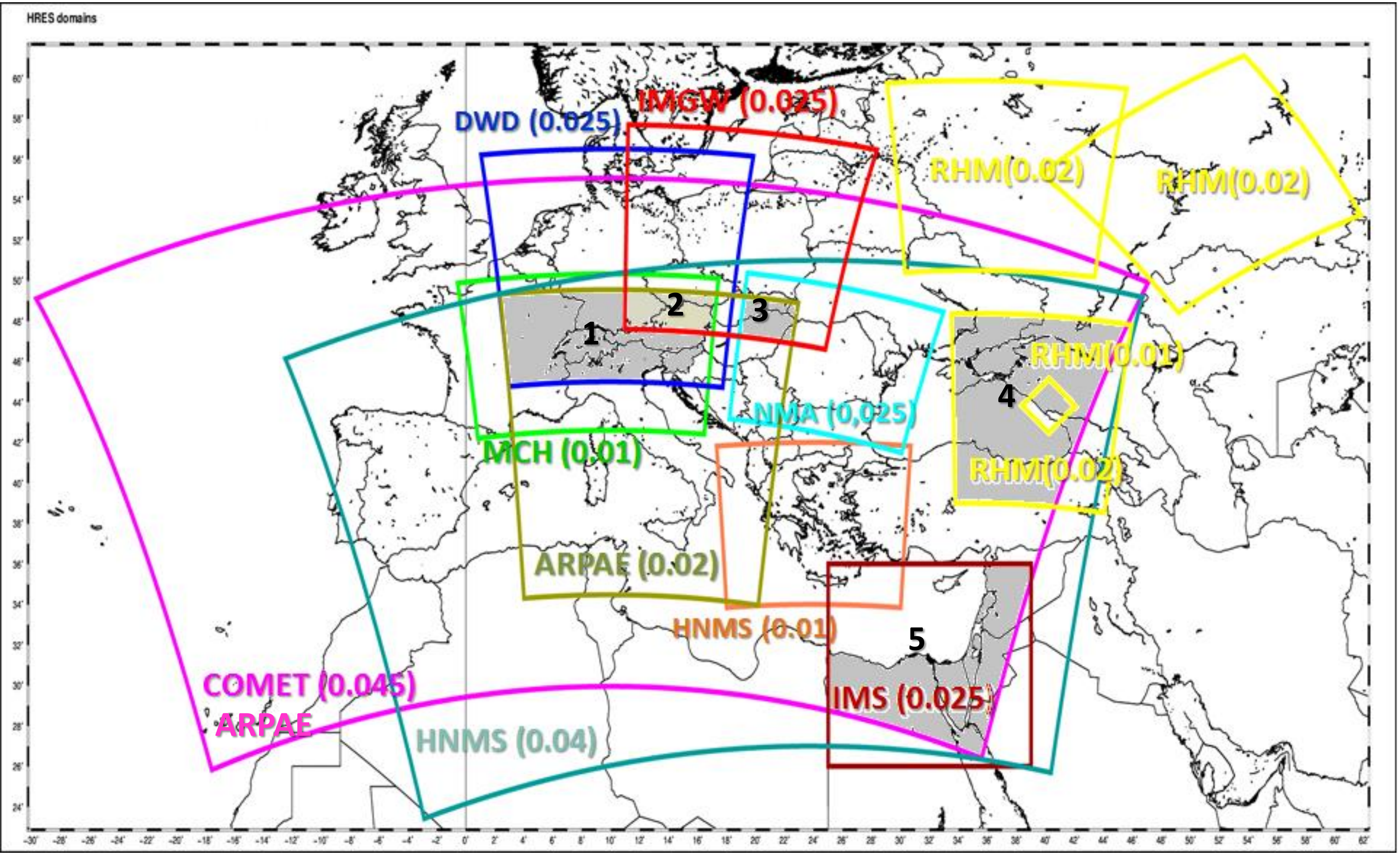
0.05 FTEs per participating service/model
+0.05FTEs for ICON global
+ECMWF/IFS

operational coarse (-7km) res models

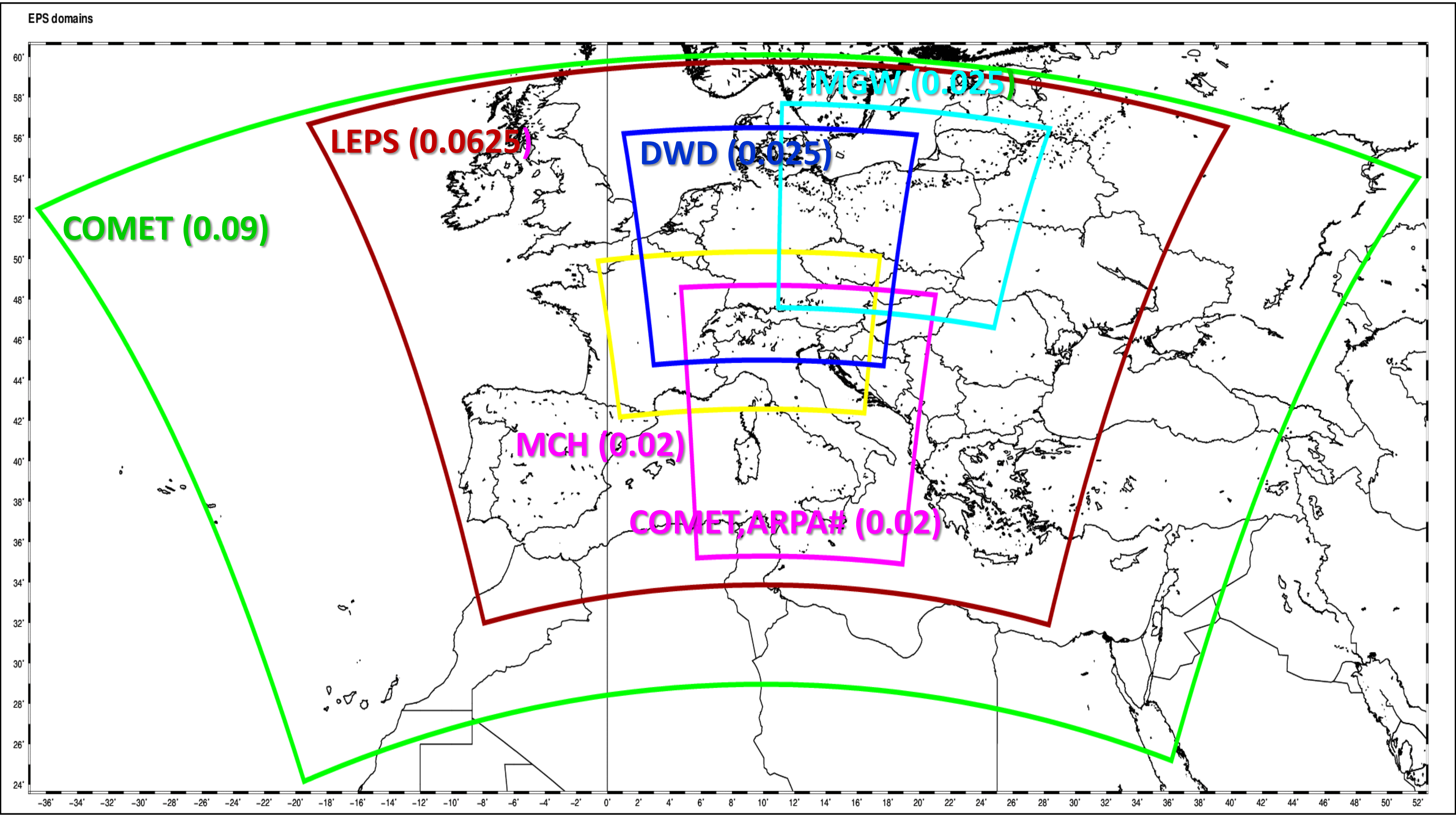


~7km	2016
COMET	✗ (4km)
COSMO-EU	✗
HNMS	✗ (4km)
ICON-EU	✓
IMGW	✓
IMS	-
MCH	?
NMA	-
RHM	✓

CP COSMO HRES: all scenarios



operational COSMO EPS models



Common Plot Reports
2015-2016

Presentation of Verification Overview (D. Boucouvala)

- Add 12UTC run
- Keep the coarser resolution comparison (~5-7km) for one year (trend since 2011)
- Add high res model comparison on a common area with restricted model participation
- Motivation that models can predict extreme values associated with dangerous weather (rare binary events). Extremal dependence scores will be added on seasonal reports
- Operational EPS verification is postponed to be discussed next year

WG5 Contributing Scientists

Dimitra Boucouvala, HNMS
Roberto Bove, COMET
Anastasia Bundel, RHM
Rodica Dumitrache, NMA
Felix Fundel, DWD
Flora Gofa, HNMS
Amalia Iriza, NMA
Pirmin Kaufmann, MCH
Alexander Kirsanov, RHM
Xavier Lapillonne, MCH
Joanna Linkowska, IMGW
Elena Oberto, ARPA-PT
Ulrich Pflüger, DWD
Maria Stefania Tesini, ARPAE
Naima Vela, ARPA-PT
Antonio Vocino, COMET