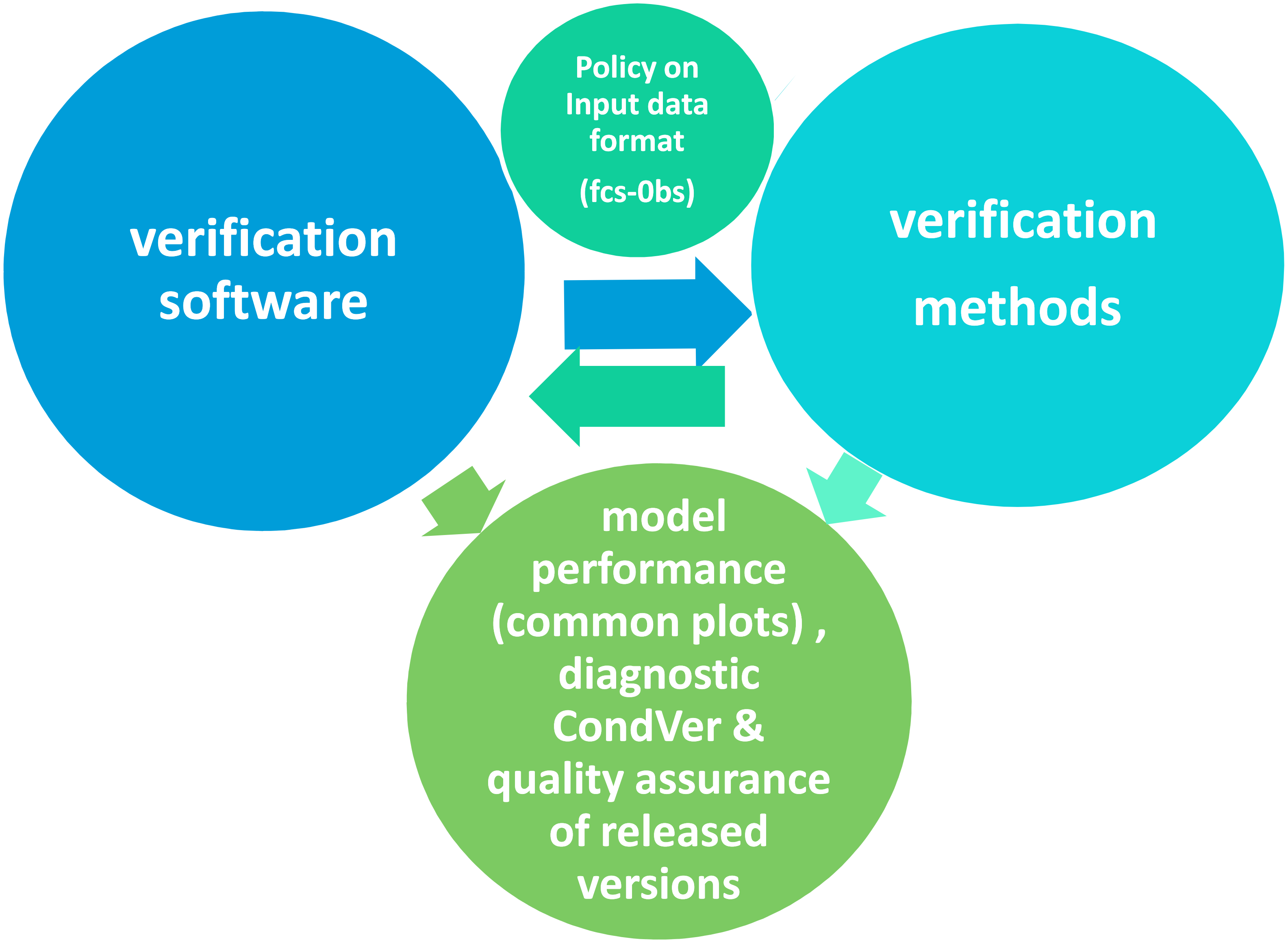
COSMO Verification Overview

Flora Gofa

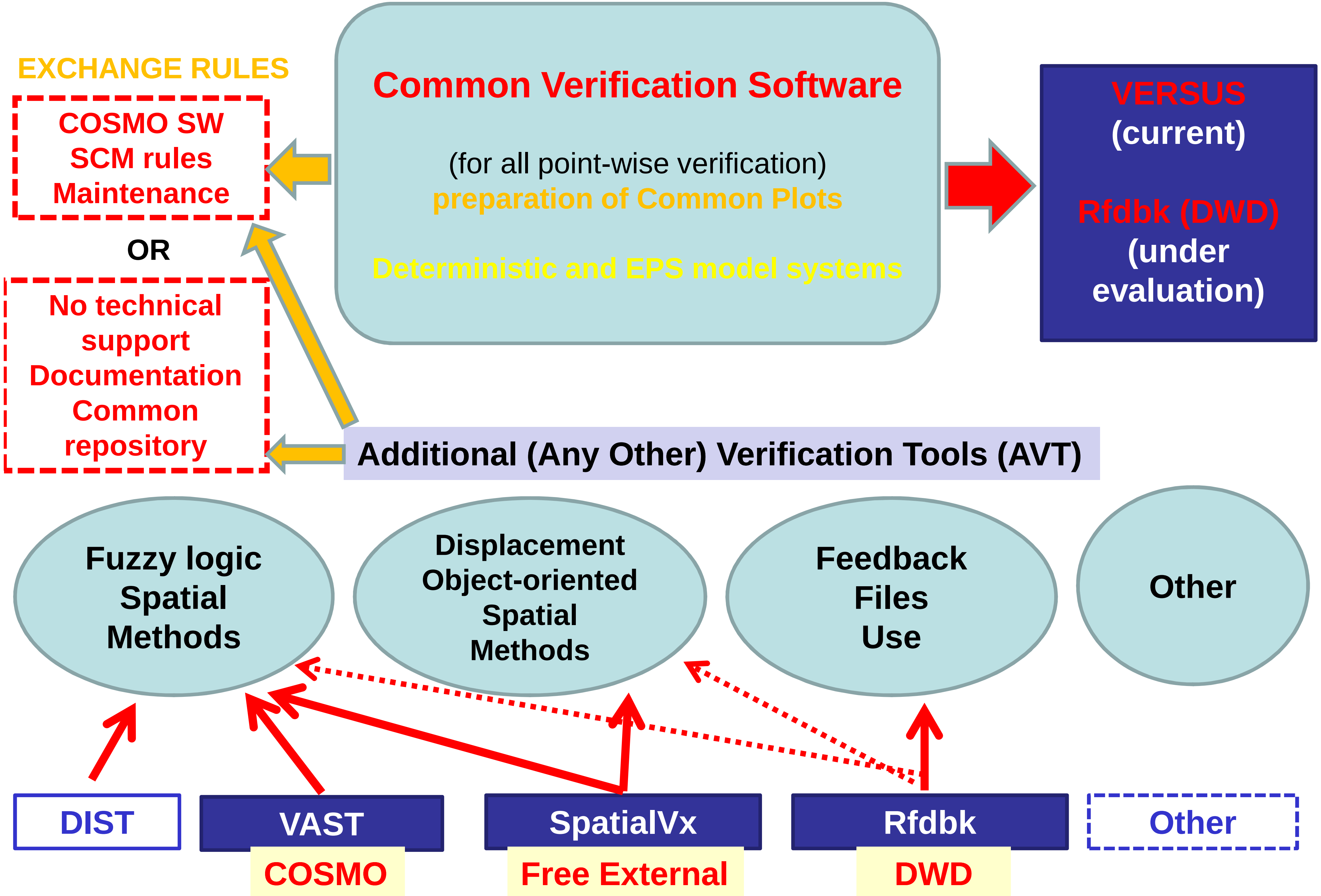
(on behalf of)

Working Group on Verification and Case studies

main activities



C V S +


EXCHANGE RULES
COSMO SW
SCM rules
Maintenance
OR
No technical support
Documentation
Common repository
Common Verification Software
(for all point-wise verification)
preparation of Common Plots
Deterministic and EPS model systems
VERSUS
(current)
Rfdbk (DWD)
(under evaluation)
Additional (Any Other) Verification Tools (AVT)
Fuzzy logic
Spatial
Methods
Displacement
Object-oriented
Spatial
Methods
Feedback
Files
Use
Other
DIST
VAST
COSMO
SpatialVx
Free External
Rfdbk
DWD
Other

Feedback File Based Verification at DWD - Rfdbk

I. Feedback Files

About

- Contain information about observations and their usage in data assimilation
- Available for each observation system used in DA (e.g. SYNOP, TEMP, AMV, AIREP, GPSRO, SCATT,…)
- Contain model analysis, first-guess and past forecast (also ensemble) for each observation
- Additional information that can be used for verification tasks (e.g. name, location, level, weight in DA, ensemble spread, talagrand index,…)
- One feedback file for each valid-time (time window), model and observation system
- Relatively small size (e.g. 10 MB ICON TEMP)
- Used for TEMP verification for a long time
- Self describing NetCDF files
- Produced by Model Equivalent Calculator MEC within the data assimilation system (3dvar, EKF, nudging) or as stand-alone

Model equivalent Calculator MEC

Installation

- Sources: Fortran 2003/2008 and some C sources from DWD
- Makefile for gfortran is provided
- NetCDF, CGRIBEX (MPI Hamburg), GRIP-API (ECMWF), (MPI recommended)
- Fortran compiler, C compiler
- Sufficient memory to hold one model state (1 ensemble state)

Required model input

- Grib or Grib2 files
- COSMO, ICON (EU Nest), IFS, HRM, ECHAM (not fully tested)
- PS, T, U, V, P, Q (mandatory, all model levels)
- T2M, TD2M, CLC, CLCT, CLCL, CLCM, CLCH, CLC, H_SNOW (optional)
- TOT_PREC, VMAX_10, TMIN_2M, TMAX_2M (optional, next release)

Required observation input

- fof/mon/cof/ekf/ver –files (existing fdbk files from nudging, LETKF or MEC)
- CDFIN (BUFR converted by bufrx2netcdf to NetCDF, BUFR in WMP-templates as used by DWD)

Output

- ver-files, NetCDF feedback files including past forecasts

I. Recap on Rfdbk Concept

- 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 is a R interface for COSMO feedback files
- Main purpose of Rfdbk is to load feedback file content with R
- Additional functionalities useful for verification is implemented as well

I. Recap on Rfdbk Concept

The idea behind Rfdbk

- Feedback file information is transformed into data table (each information related to an observation can be a table column) using R data.table package (*https://cran.r-project.org/web/packages/data.table*)

- data.table allows to perform operations on huge tables very quickly with elegant syntax

DT[i,j,by]

i : where (addresses only a set of rows)
j : select (addresses only a set of columns, column names can be used as input for R functions)
by : group (group results by instances of variables in columns)

- Based on data.tables not only scores can be calculated but also a data adjustment between experiments or conditions could be implemented

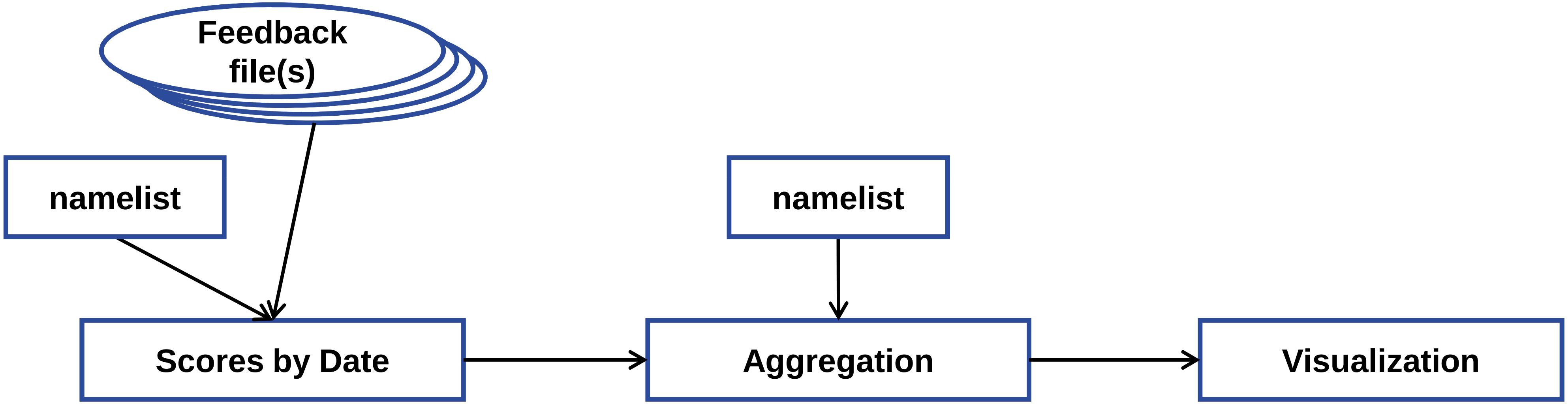
I. Recap on Rfdbk Concept

```
> DT
   obs    veri_data     veri_forecast_time
1: 291.65  291.3145                    0
2: 292.15  292.2571                    0
3: 293.65  292.7871                    0
4: 293.25  291.4255                    0
5: 301.45  300.2931                    0
---
172796: 266.45  266.2054                    18000
172797: 269.55  268.9154                    18000
172798: 272.05  271.6369                    18000
172799: 271.65  271.7378                    18000
172800: 273.65  271.7899                    18000
> DT[,list(BIAS=mean(veri_data-obs)),by="veri_forecast_time"]

veri_forecast_time            BIAS
1:                 0     0.20529260
2:               600     0.07754623
3:              1200     0.14901599
---
```

III. Recap of Verification

Deutscher Wetterdienst
Wetter und Klima aus einer Hand



Scores by Date

- 1x per valid date
- SYNOP, TEMP, etc.
- ICON, ICON-EU, COSMO-EU, IFS
- Deterministic or ensemble
- Routines or experiments
- Scores: continuous and categorical, EPS
- Conditional on: model, lead-time, valid-time, level, etc.
- All scores are based on same observation sample
- Uses R Rfdbk functionality

Aggregation

- Pool scores over period, levels, lead-time etc.
- Decide on period by setting valid date range and/or initial date range
- Pooling done correctly, considering number of observations per date, except correlations are simple averages.
- Score files transferred to visualization server
- Uses R

Visualization

- On demand interactive plot web browser application
- Plot and arrange scores, summaries, browse data
- Separate web based apps for observation and verification types
- Uses the R shiny web server

IV. Recent verification progress

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

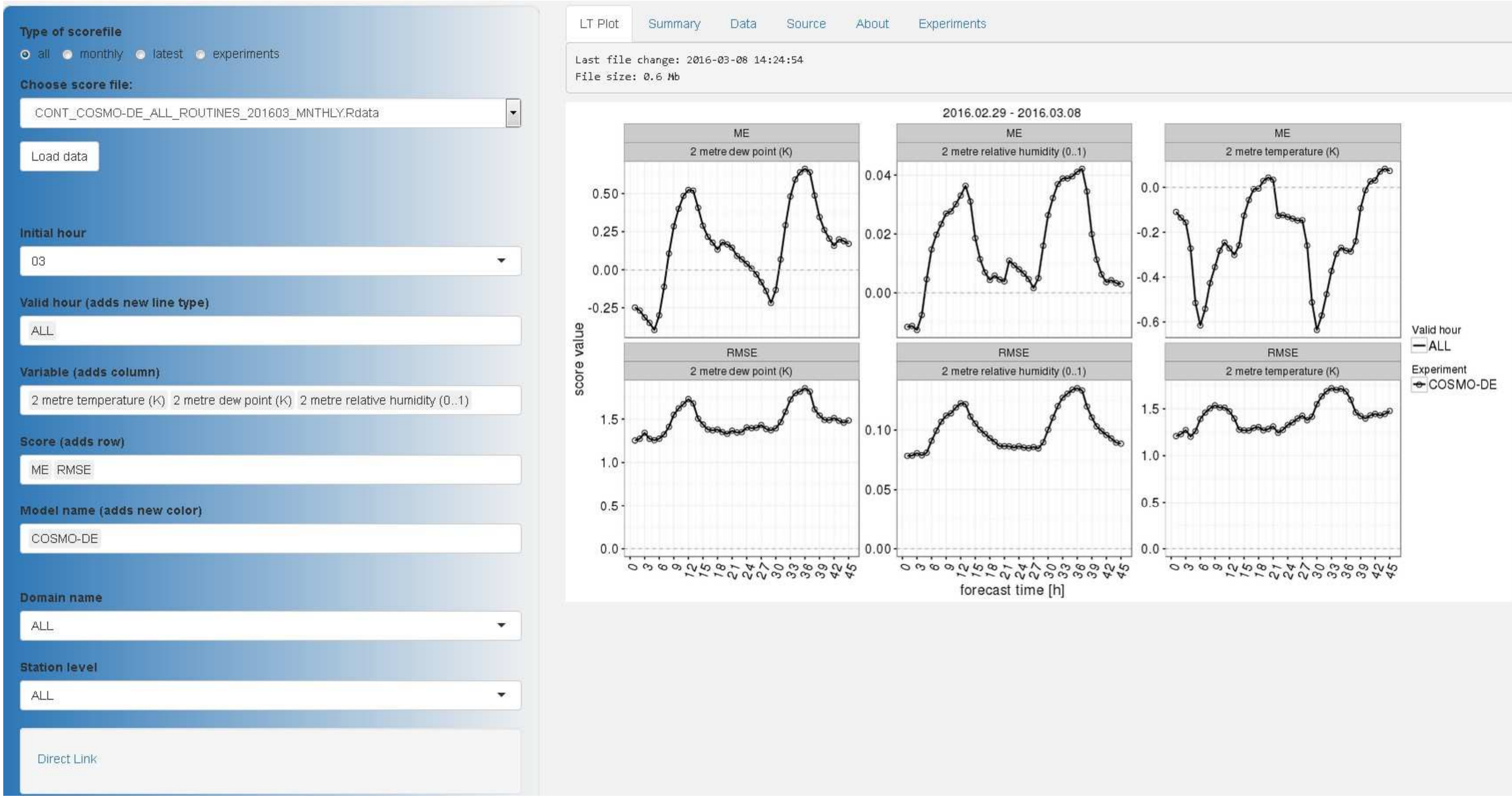
TODO

• Fill feedback files with additional observations not used in DA but required for verification (e.g. precip., gusts, T_min/max)

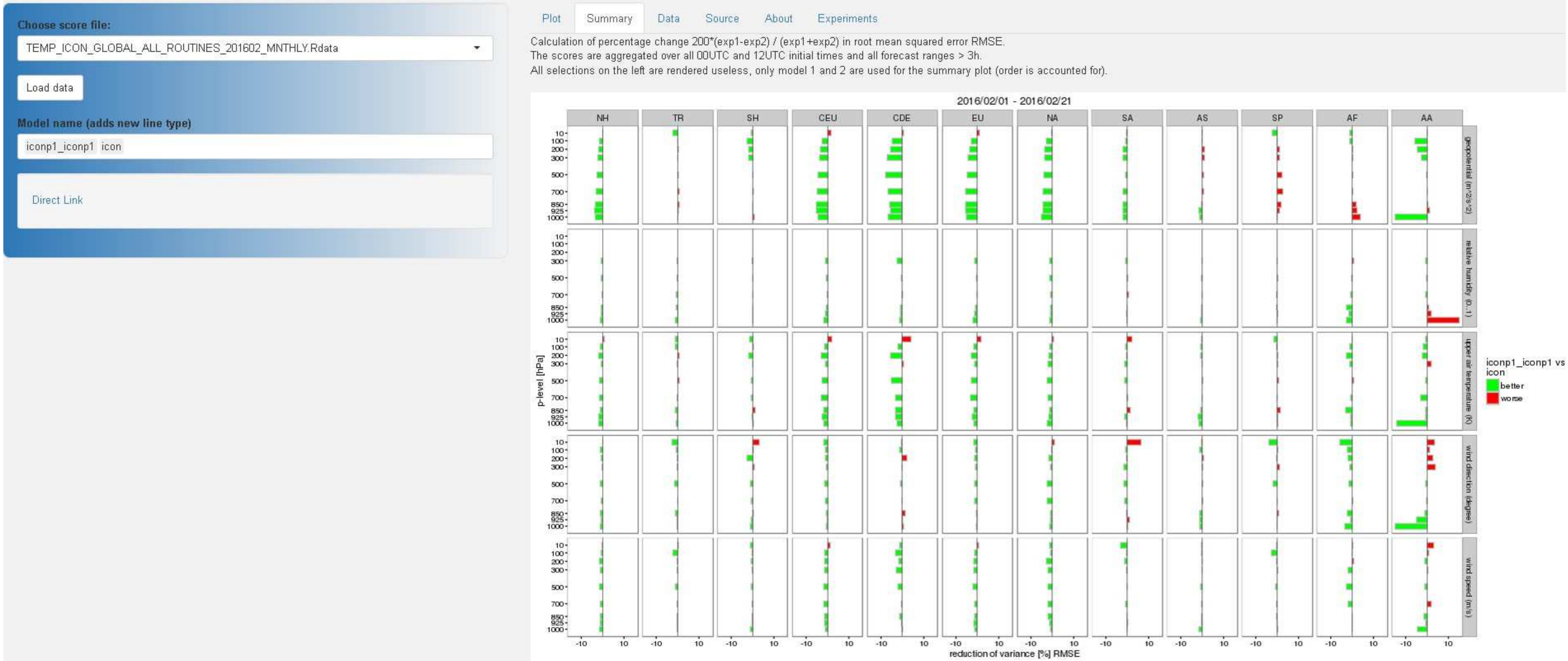
Limitations

- Spatial/object-based verification
- Conditional verification, if the required information about the observations is not in the feedback file

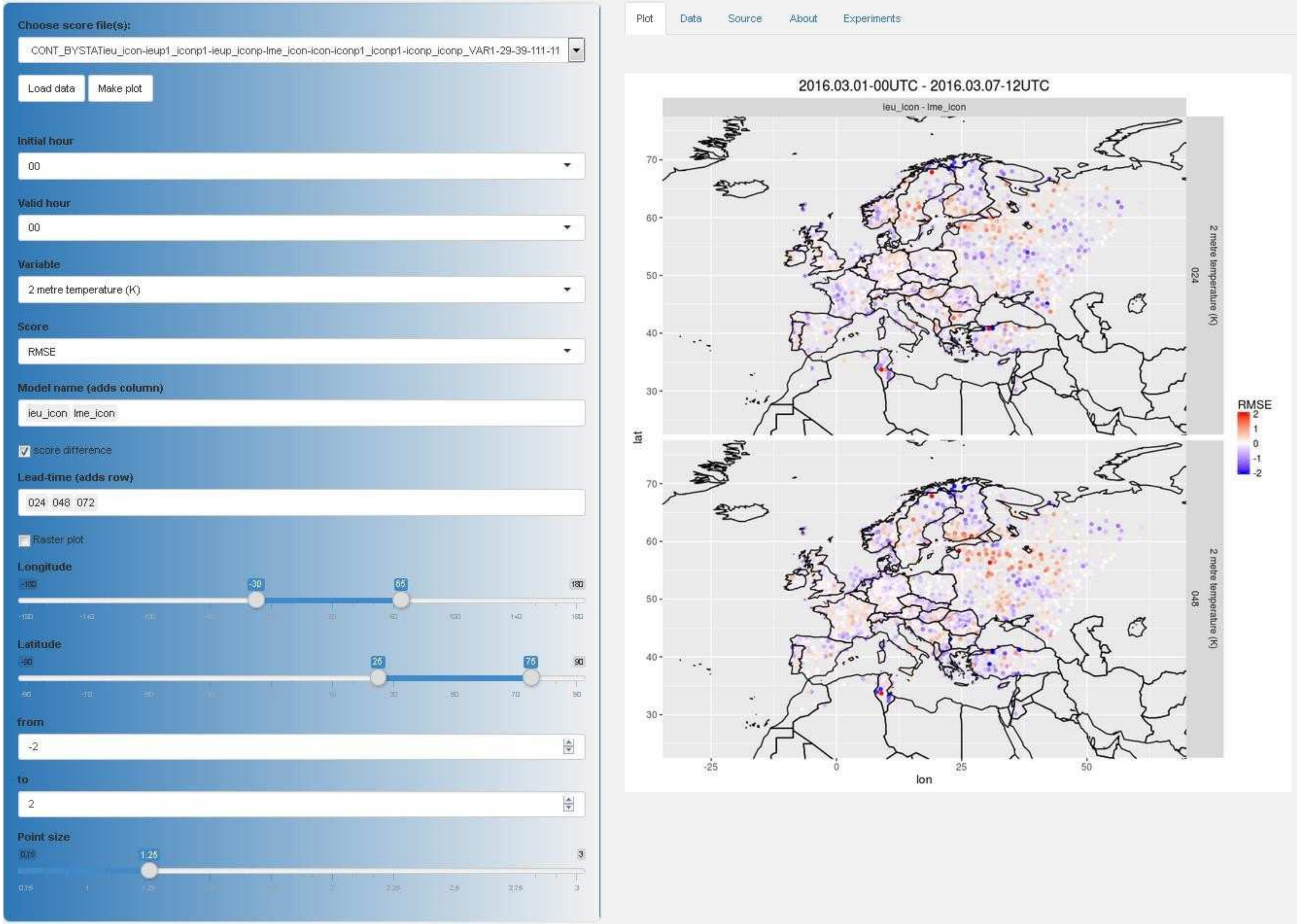
IV. Visualization



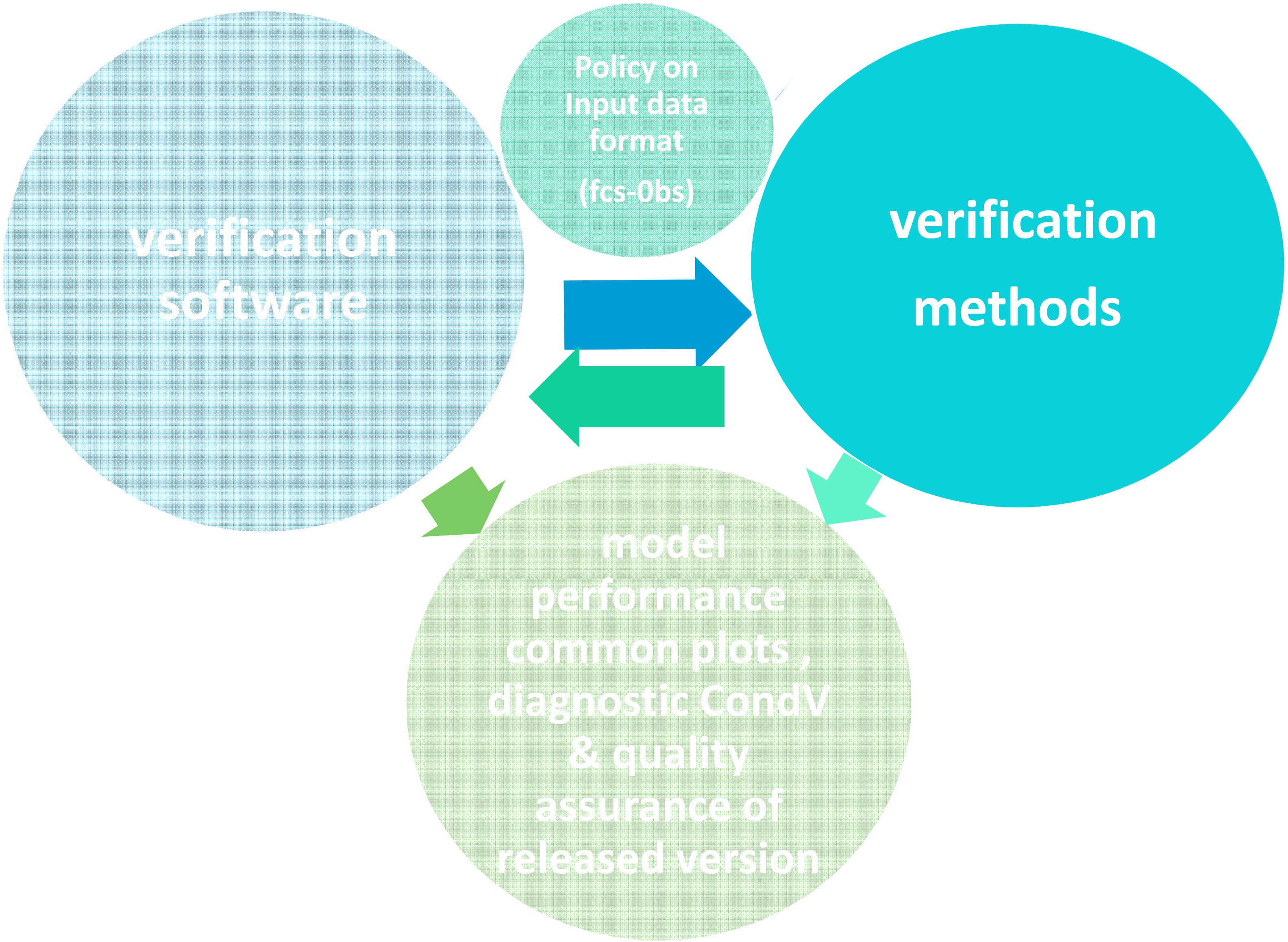
IV. Visualization



IV. Visualization



main activities



Mesoscale Verification Inter-Comparison over Complex Terrain (MesoVICT)

(https://www.ral.ucar.edu/projects/icp/)

- To investigate the ability of existing or newly developed spatial verification methods to verify fields other than deterministic precipitation forecasts, e.g., wind forecasts and ensemble forecasts.

- To demonstrate the capability of spatial verification methods over complex terrain, and gain anunderstanding of the issues that arise from this more challenging situation.

- To encourage community participation in the development and improvement of spatial verificationmethods, especially for evaluating high resolution numerical forecasts.

- To provide a community testbed where common data sets are available, but also for the sharing of data and code to assist in developing and testing spatial verification methods.

o Kick-off meeting: October 2015, Vienna (Universitat wien)
o 2nd MesoVICT meeting: September 2016, Bologna, (Arpae)

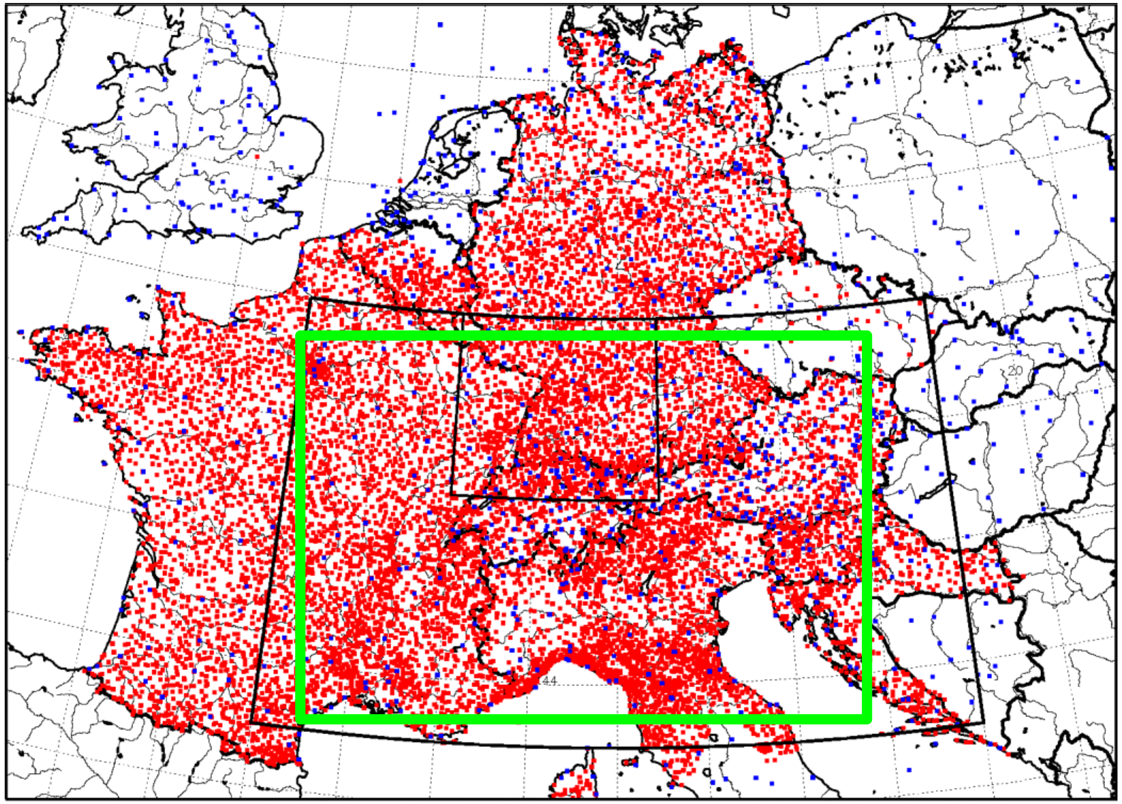
COSMO Priority Project

INSPECT: INtercomparison of SPatial vErification methods for COSMO Terrain

- runs in parallel to MesoVICT
- summarizes the COSMO experience of applying spatial verification methods to high and very-high-resolution systems
- *a wider range of **spatial** verification **methods** will become **commonly used** within the COSMO community and **Guidelines** will be proposed to ensure the correct interpretation of results of these methods.*
- Same as MesoVICT, INSPECT focuses on EPS *forecasts* and *variables besides precipitation*
- In addition to targeting the goals of MesoVICT, INSPECT provides more choice of verification domains and reference data - newer and longer periods, two **complex terrains** (the Alps and the Caucasus)
- **Share the software tools that will be developed or adapted for common use**

Observations data set

JDC-data: WWRP D-PHASE (FDP, Rotach, et al., 2009, BAMS) and
WWRP COPS (RDP, Wulfmeyer, et al., 2008, BAMS),
data available: (http://cera- www.dkrz.de/WDCC/ui/Index.jsp)



- 32 data providers
- GTS-Stations: 1232
- NGTS-Stations: > **13000**
- Mean station distance: GTS: ~36km
  GTS+Non-GTS: ~ 12km

Frames: D-PHASE (black, large)
COPS (black, small)
this study (green)

Red: Non-GTS stations
Blue: GTS stations

VERA analysis scheme: Data quality control scheme + Thin-Plate-Spline algorithm + Downscaling via the „Fingerprint“ method

Mesoscale Verification Inter-Comparison over Complex Terrain (https://www.ral.ucar.edu/projects/icp/)

Prerequisite for verification method inter-comparison:

-use of same data (Obs and FC) on the same grid and over the same area (Alpine area)

From MAP D-PHASE COPS archive

- Deterministic 2 km COSMO-2 Init-time: Initialised 06 UTC FC-range: 24h
- Deterministic 2 km CMC-GEM-H Init-time: Initialised 06 UTC FC-range: 18h
- Ensemble 10 km COSMO-LEPS Init-time: Initialised 12 UTC FC-range:132h

MCH

•Reruns COSMO-1 models for 4 cases

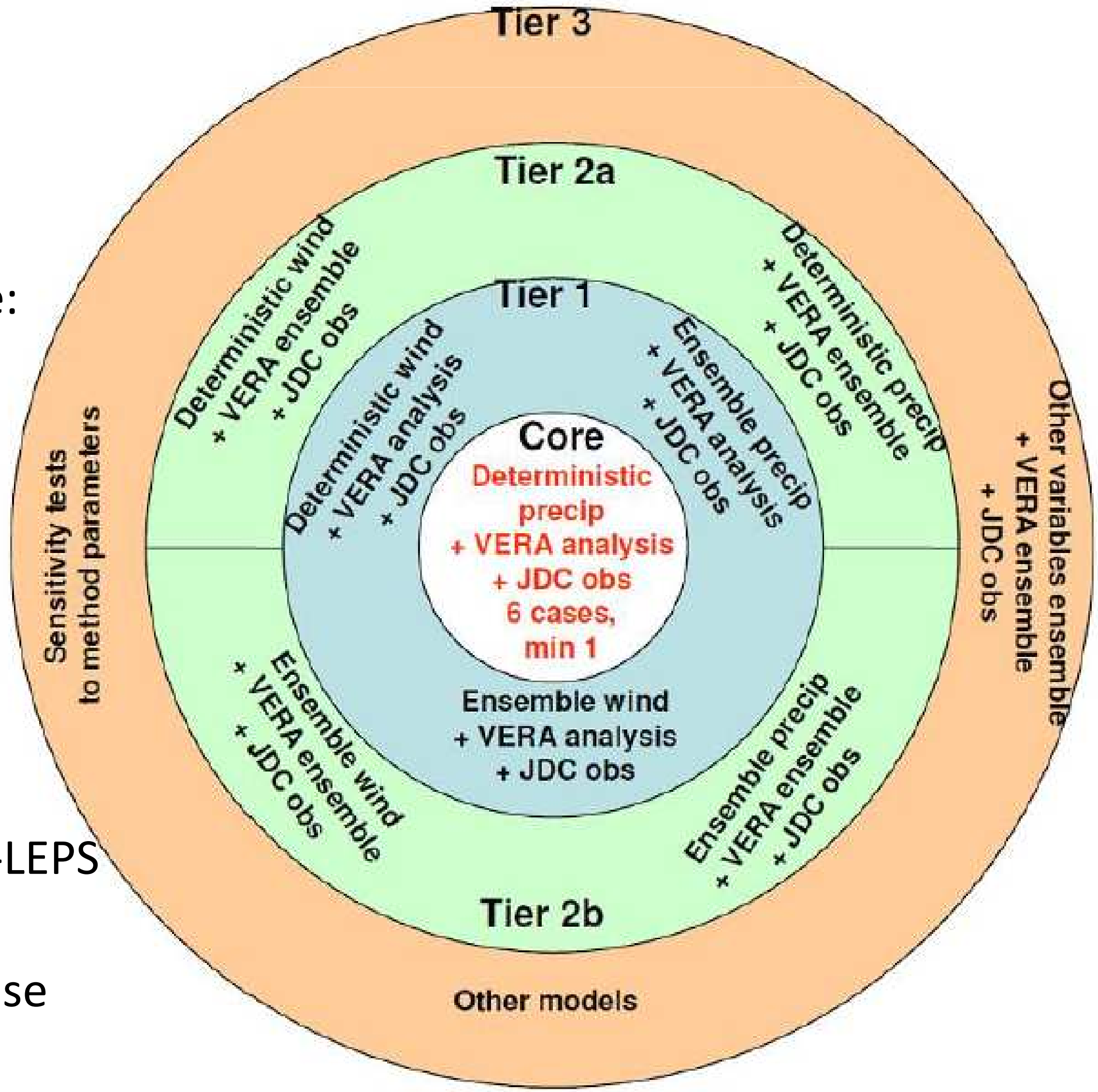
ARPAE

•ECMWF-IFS reruns for cases 1,2

•to provide boundary conditions for COSMO-LEPS

•**Roshydromet**

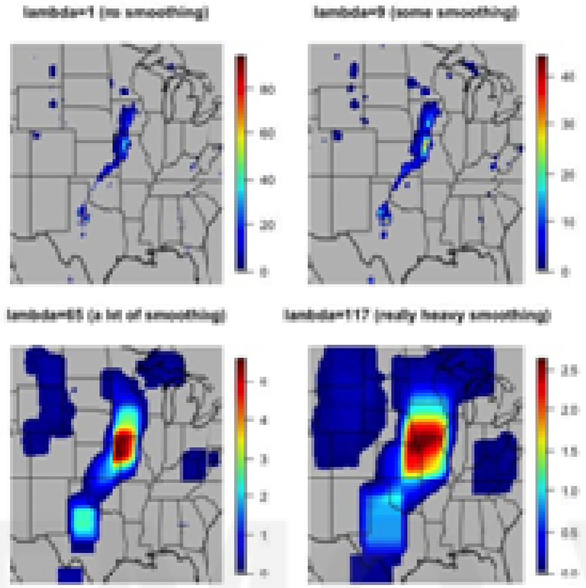
•COSMO-Ru2-EPS: rerun for 1st MesoVICT case

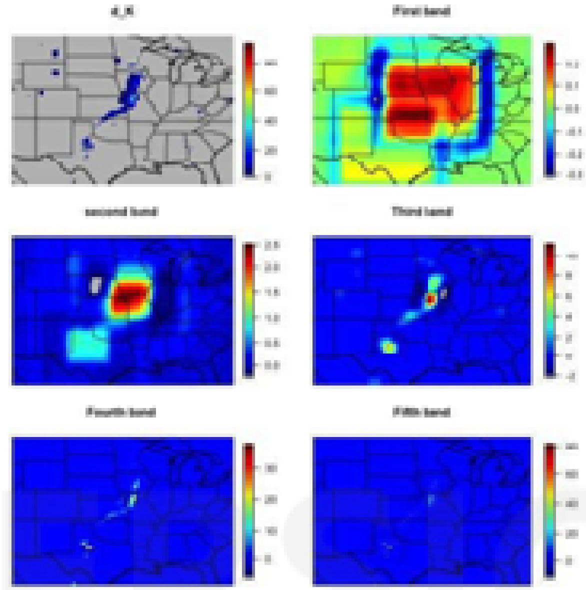


Categorisation

smoothing / neighborhood

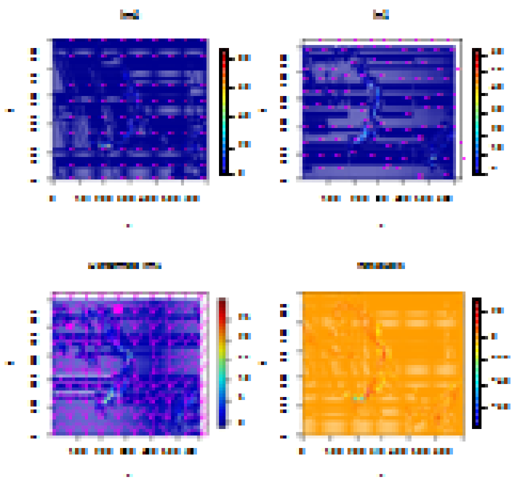
Filter



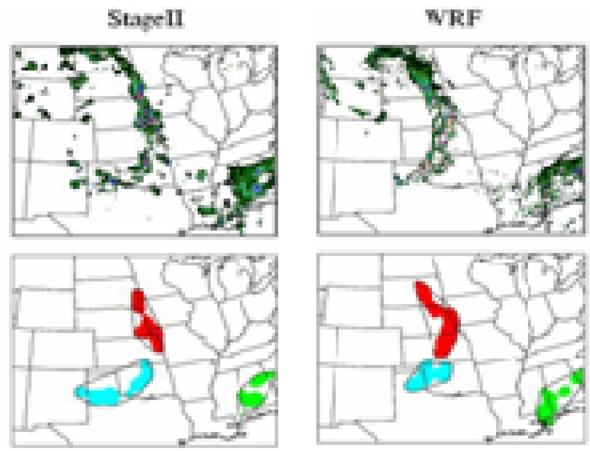


Scale Separation

Entire field



specific features



Spatial Displacement

Ca

Tr
(g

N

Sc

Fe

D

le
paration

N

Gilleland et al., Bulletin of the American Meteorological Society, 2010

Inspect **Tasks** **involving the application of spatial methods** **to MesoVICT cases**

Precipitation

- Neighborhood methods: *HNMS,MCH,DWD*
- Intensity Scale (wavelet): *HNMS*
- MODE*:* IMGW-PIB
- SAL: HNMS, IMGW-PIB
- CRA: RHM, IMGW-PIB

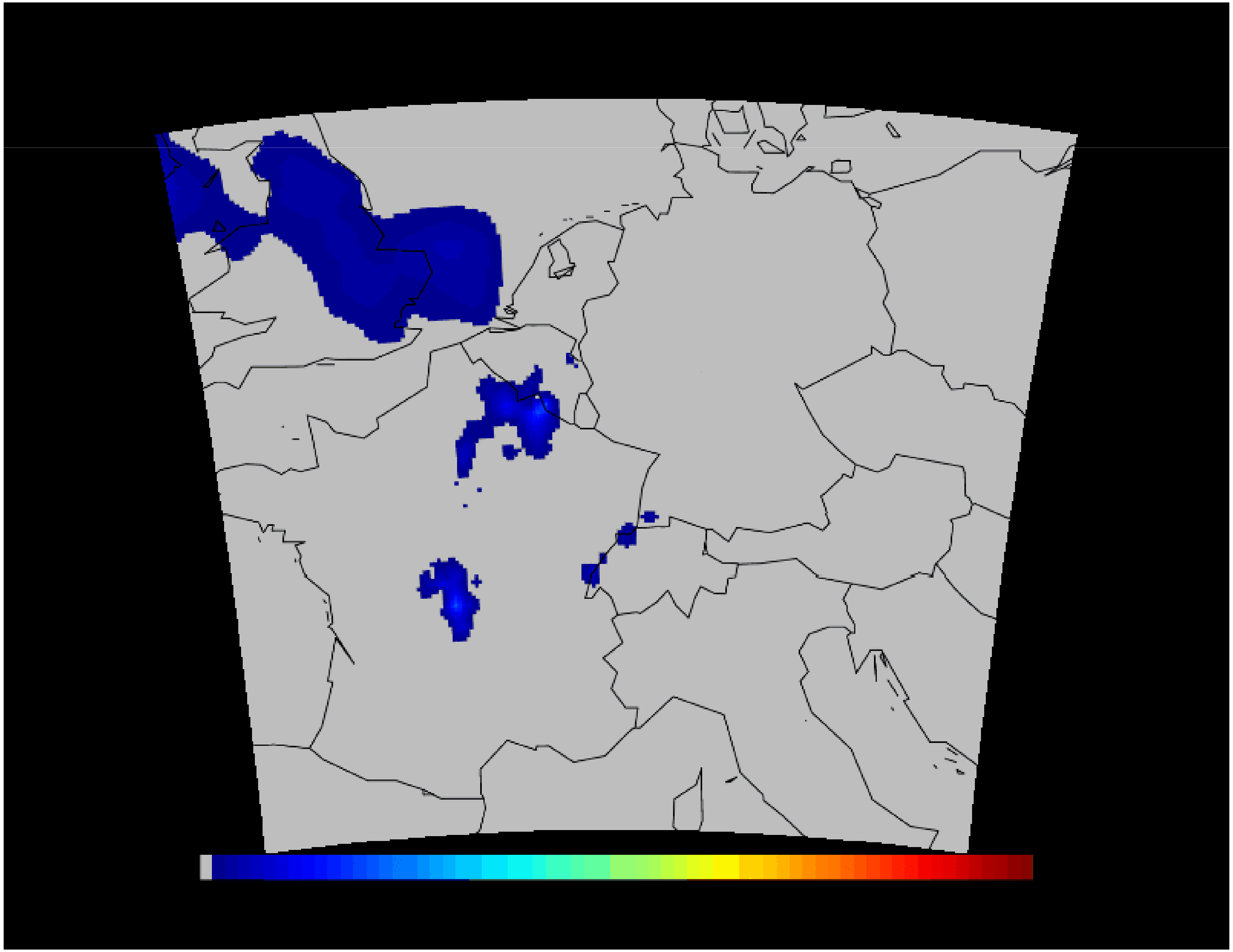
WindSpeed

- DIST filtering method (wind speed): *Arpae*

EPS

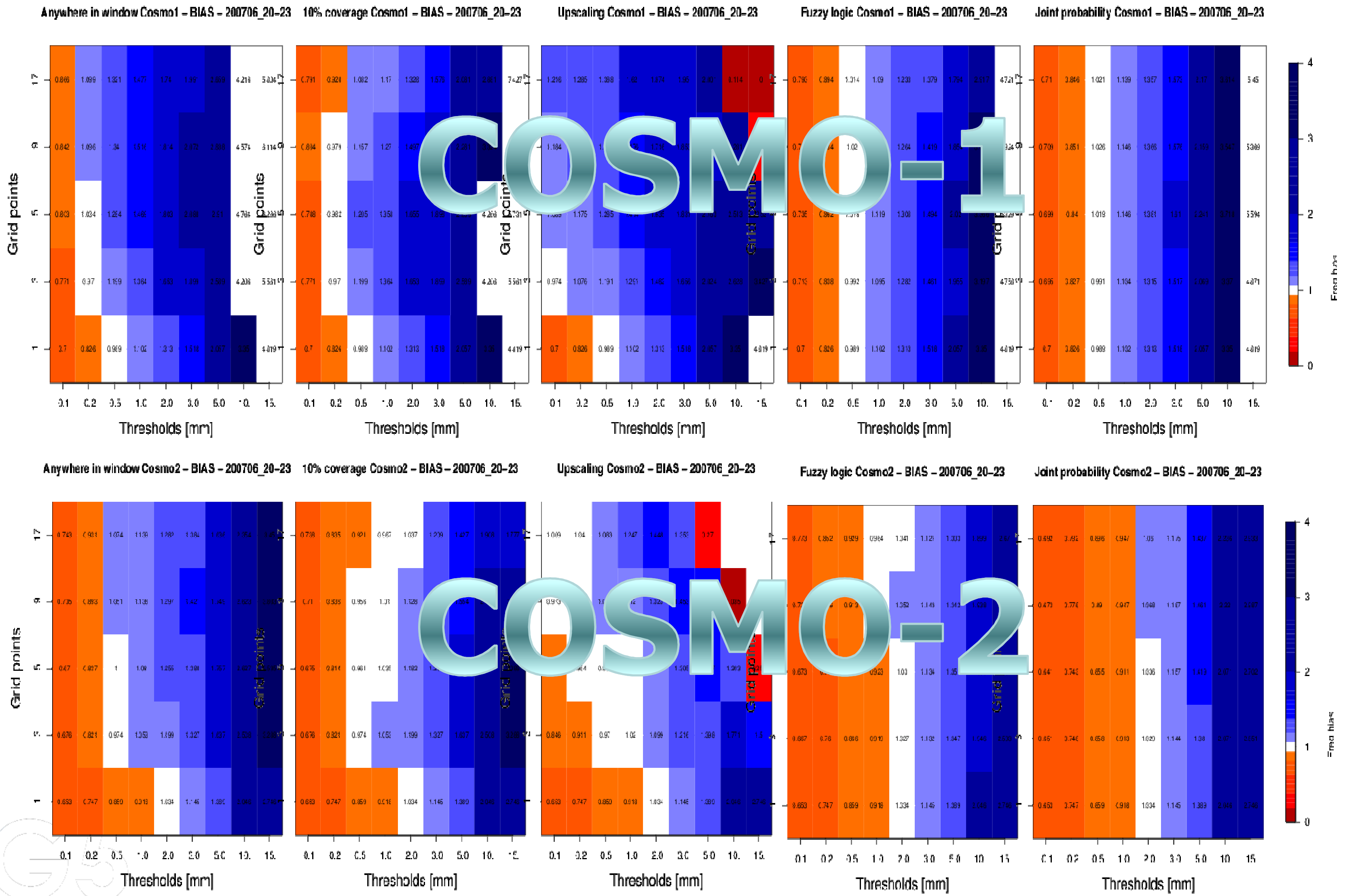
- DIST filtering method: Arpae
- SAL: HNMS
- CRA: RHM

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



Neighborhood method scores COSMO-1 vs. COSMO-2

- **BIAS**

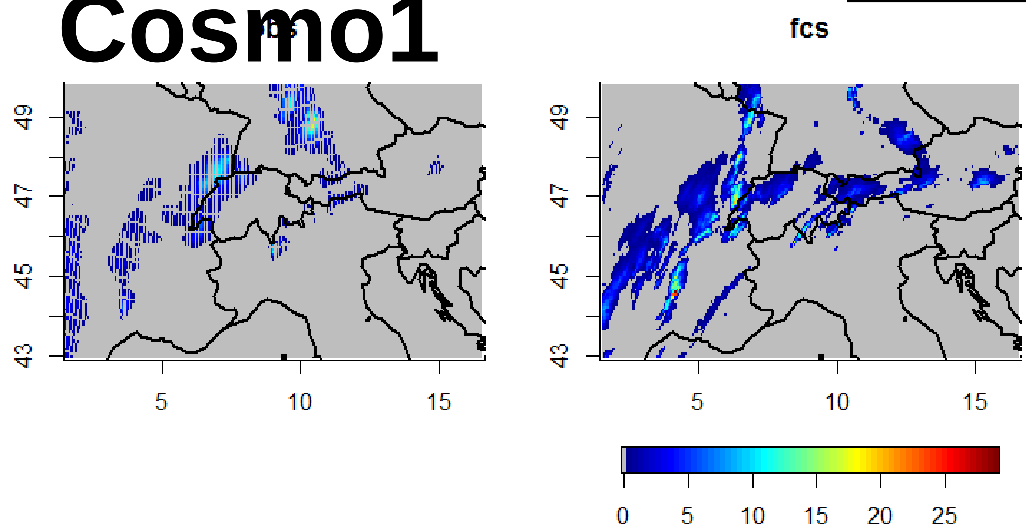


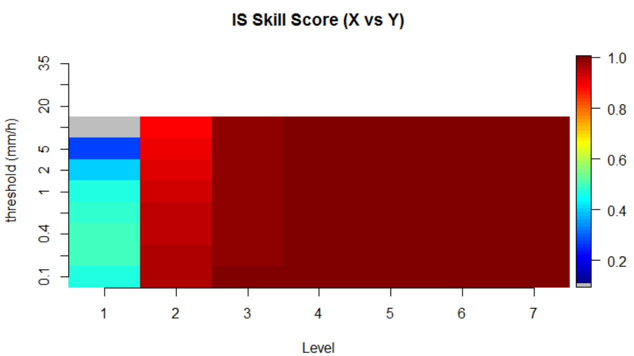
Intensity-scale verification

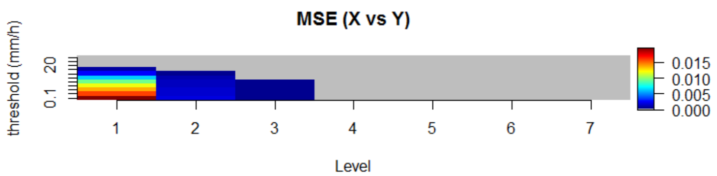
Haaer decomposition wavelet

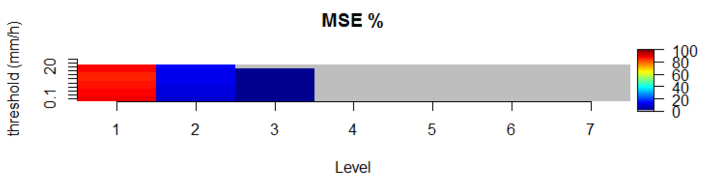
20070621-01:map

Cosmo1

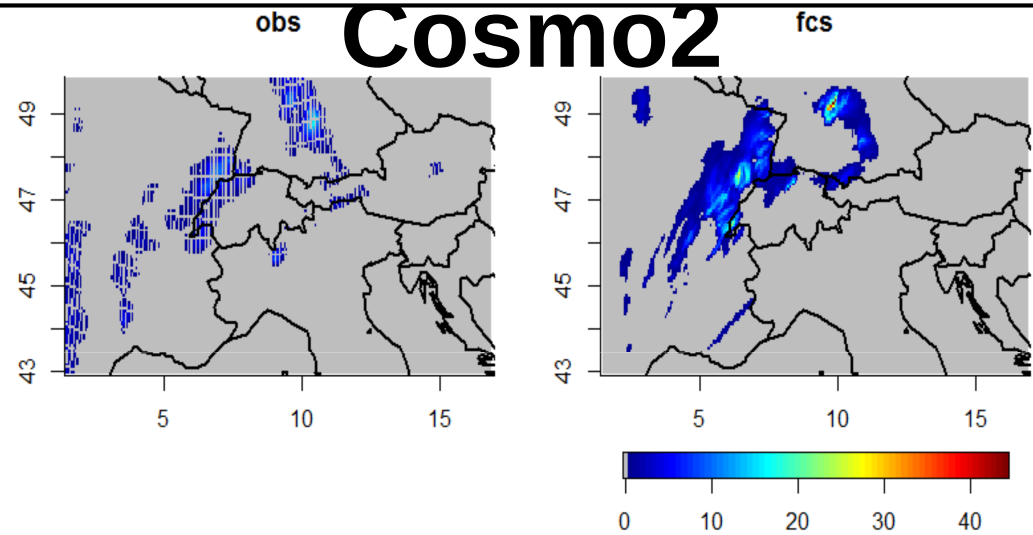


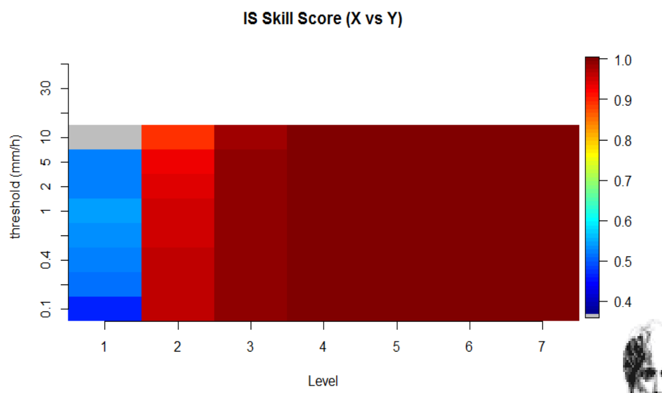


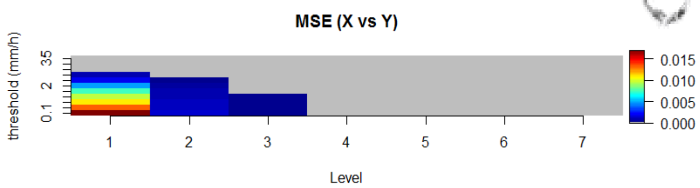


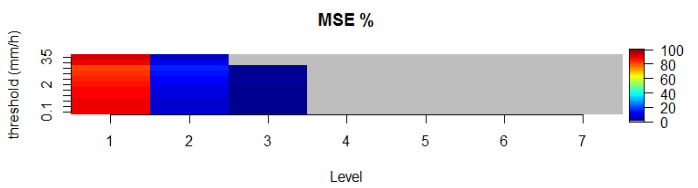


Cosmo2









ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

CRA – Contiguous Rain Area (E.E. Ebert, J.L. McBride 2000)

http://www.cawcr.gov.au/projects/verification/CRA/CRA_verification.html

MSEtotal = MSEdisplacement + MSEvolume + MSEpattern

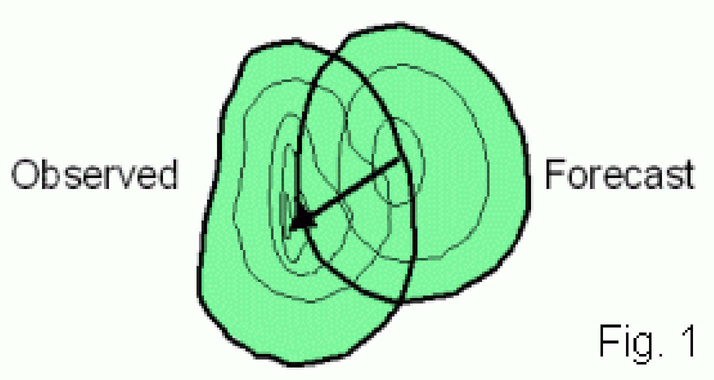


Fig. 1

MSEdisplacement = MSEtotal – MSEshifted

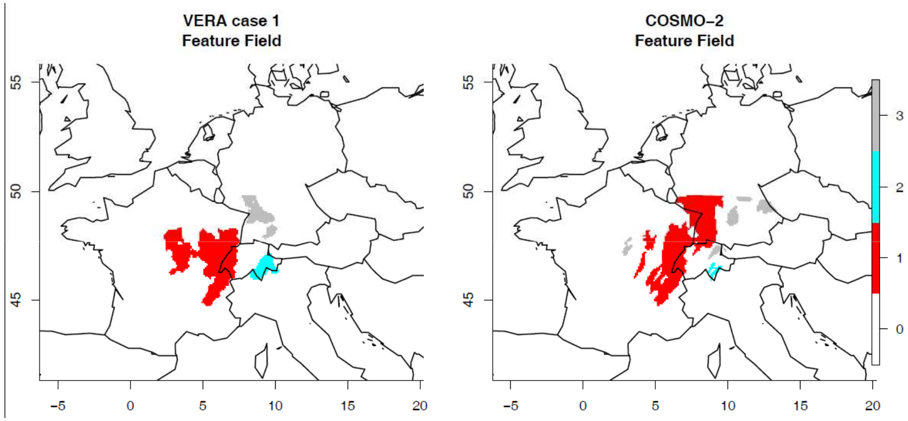
MSEvolume = (***F*** - ***X***)2

where ***F*** and ***X*** are the CRA mean forecast and observed values after the shift.

MSEpattern = MSEshifted – MSEvolume

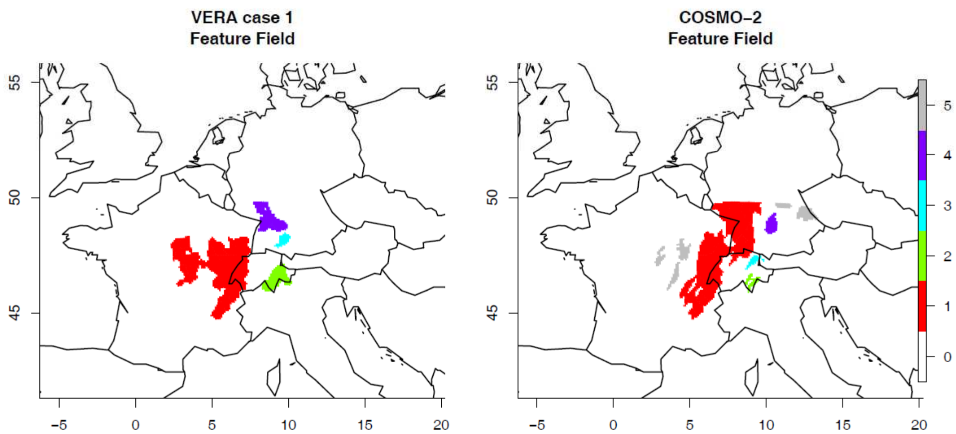
CRA scores, 2007062106

Centmatch 2



ir	MSE.total	MSE.shift	MSE.displace	MSE.volume	MSE.pattern
1	0.0347	0.0404	-0.0057	0.0000	0.0404
2	0.0030	0.0030	0.0000	0.0000	0.0030

Minboundmatch



ir	MSE.total	MSE.shift	MSE.displace	MSE.volume	MSE.pattern
1	0.0352	0.0404	-0.0051	0.0000	0.0403
2	0.0030	0.0030	0.0000	0.0000	0.0030
3	0.0081	0.0049	0.0032	0.0000	0.0049

CRA scores, 2007062021, precip > 0.5 mm/h

All methods are acceptable

Minboundmatch

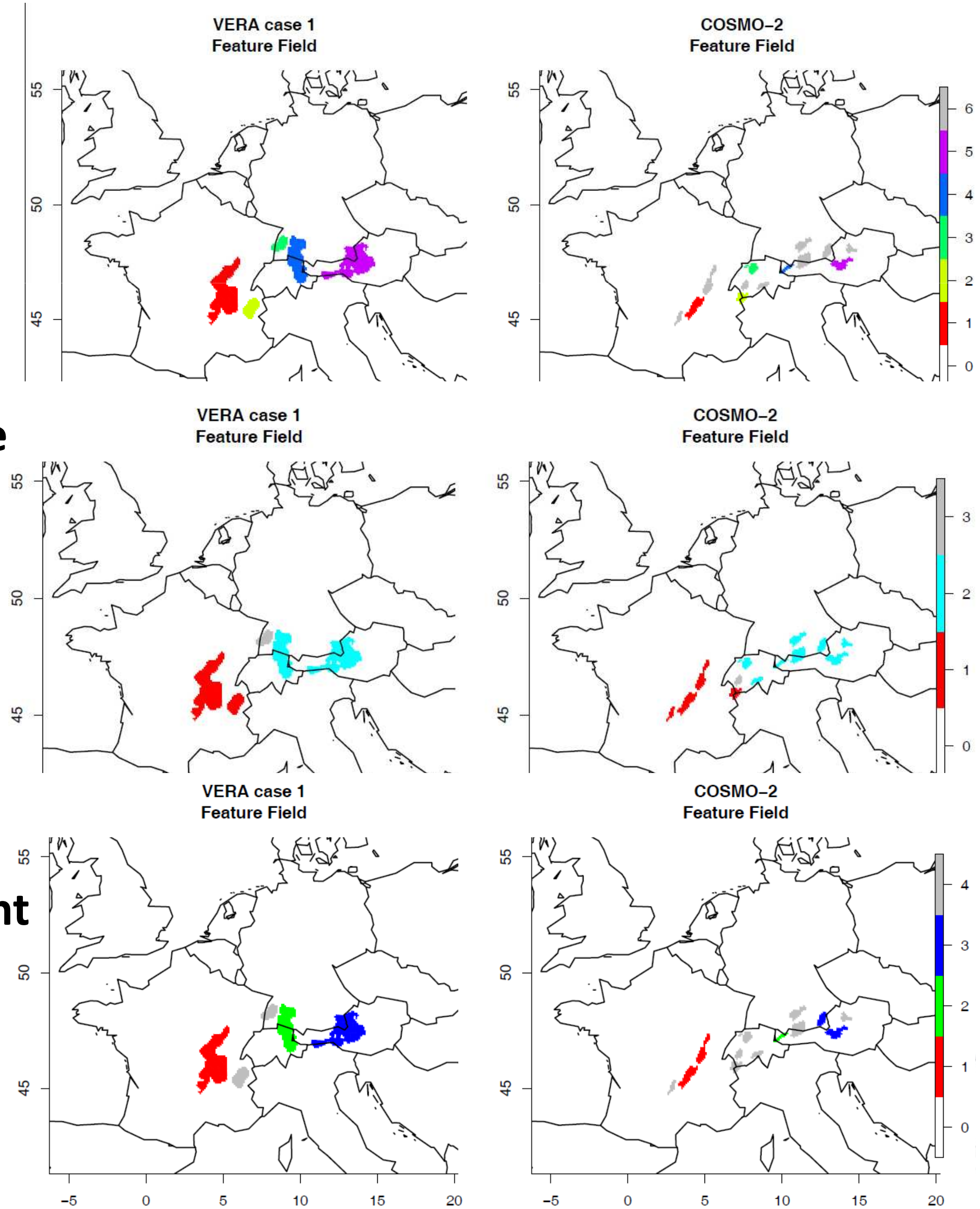
Minboundmatch **more promising, but with a minimum boundary separation distance beyond which features should not be matched**

Centmatch 1

Centmatch 1 makes implicit mergings

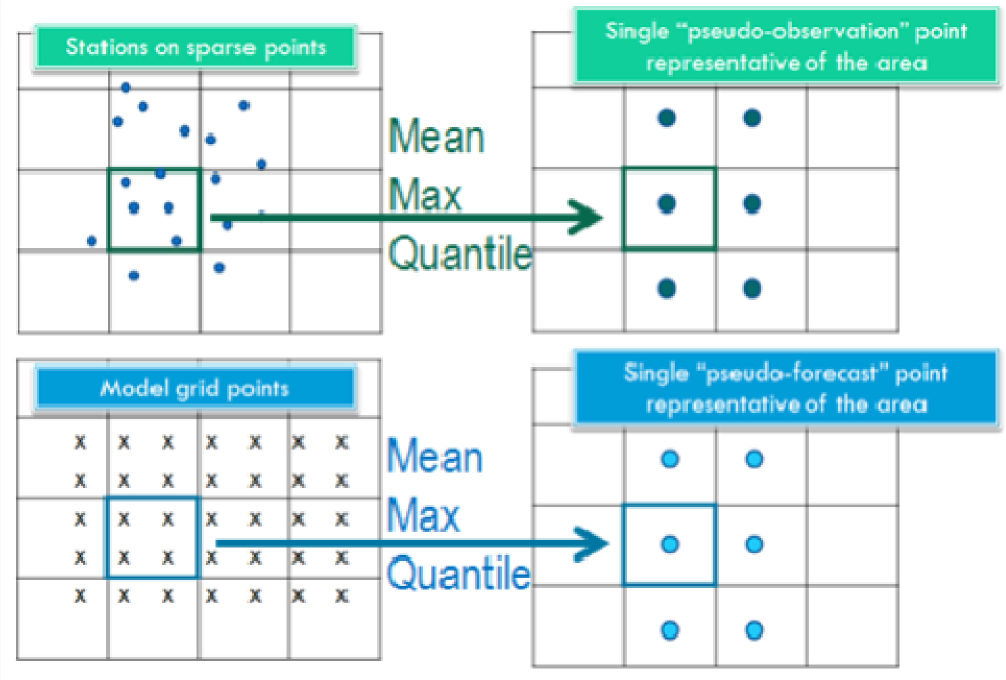
Error comes mainly from fine structure (MSE.pattern) for lower precipitation thresholds. For higher thresholds, displacement error contribution increases

Centmatch 2



The "distributional method (DIST)"

- The verification domain is subdivided into a number of "boxes", each of them containing a certain number of observed and forecast values.
- For each box, several parameters of the distribution of both the observed and forecast values falling in the box can be computed (mean, median, percentiles, maximum).
- Verification is then performed using a categorical approach, by comparing for each box one or more parameters of the forecast distribution against the corresponding parameters of the observed distribution, using a set of indices.

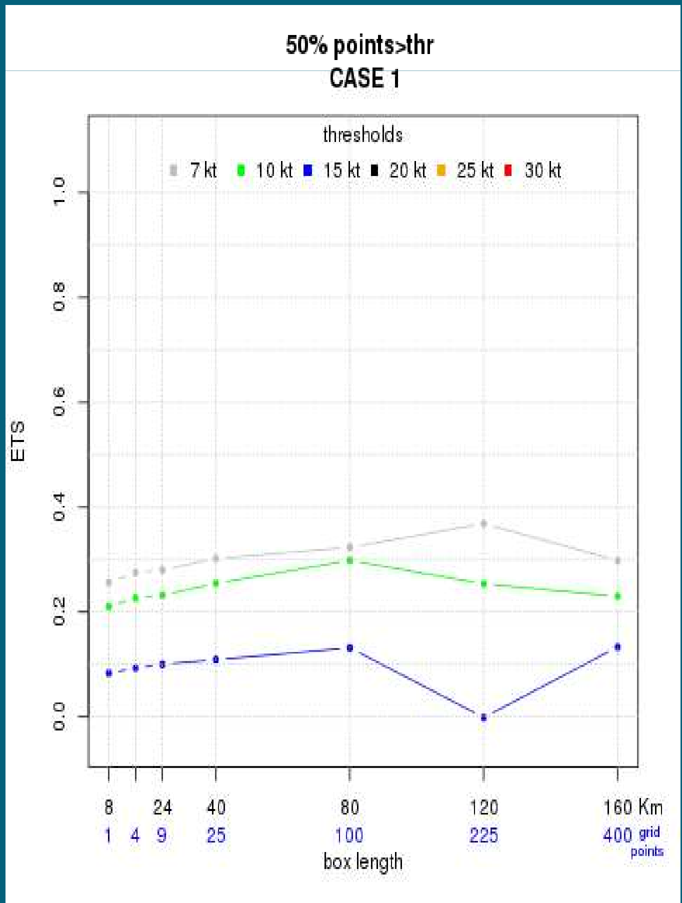


arpae agenzia prevenzione ambiente energia emilia-romagna

Marsigli, C., Montani, A. and Paccangnella, T. (2008), A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. Met. Apps, 15: 125–143. doi: 10.1002/met.65

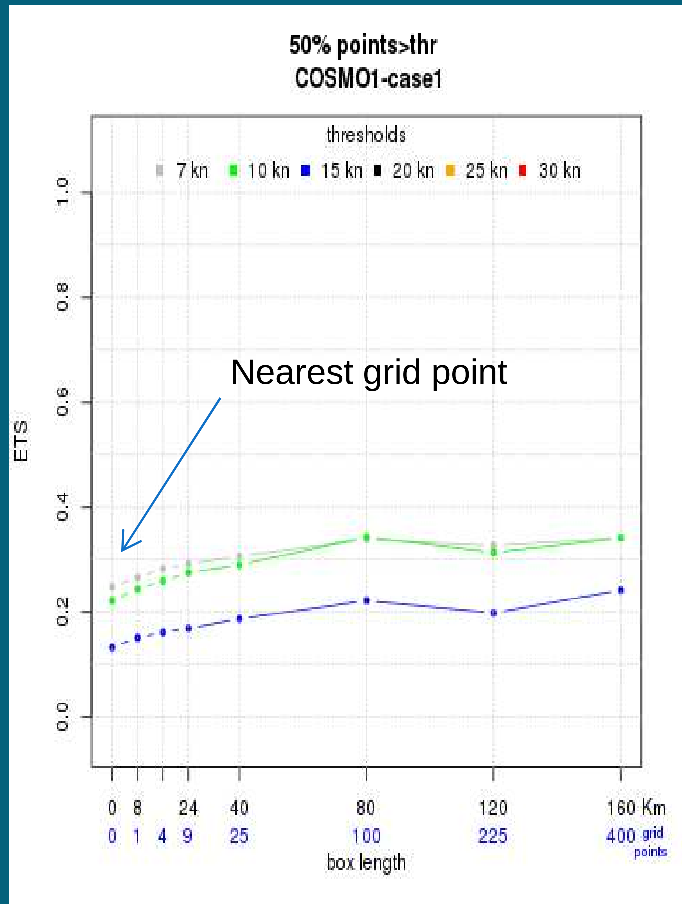
Impact of boxes size: Case 1

COSMO_2 - 1 hour



m/s	Knots
3.4	7
5.4	10
7.9	15

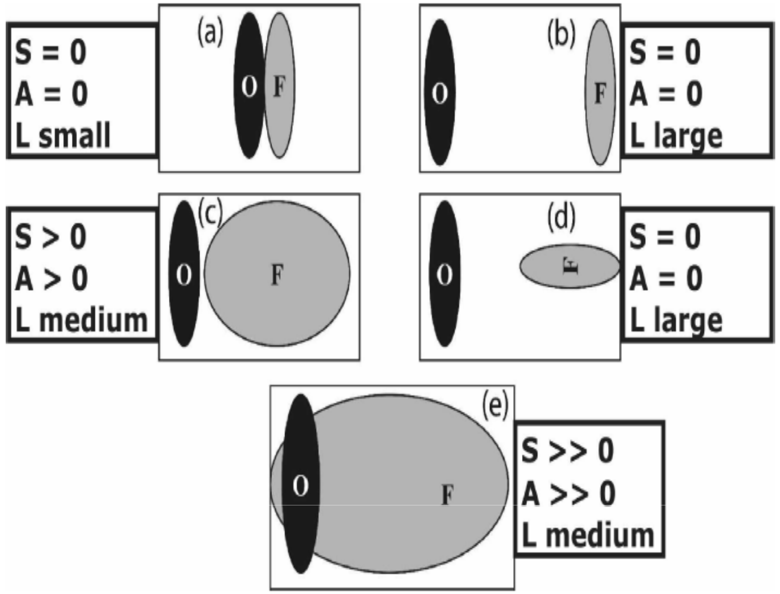
COSMO_1 - 1 hour



- The event is defined as "median exceeding a predefined threshold"
- The scores are plotted as a function of the box dimension

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

SAL:
Feature based verification measure



SAL Method (Wernli et al. 2008, 2009) For each pair of gridded observations/forecast field 3 indexes are calculated.

S: Structure Component (Compares Total Volume of Normalized Objects of obs/fcst . Captures size and shape of objects) (Values from -2 to 2) S=0 perfect, S >> 0 forecast predicts more widespread pcp , S<< 0 forecast predicts more peaked objects

A: Amplitude Component (Normalized difference of domain-averaged values of forecast and obs field) (Values from-2 to 2) A=0 perfect, A >> 0 forecast overpredicts pcp A<< 0 forecast underpredicts

L: Location Component (Consists of L1+L2) (L Values from 0 to 2) (0 perfect)

L1 : normalized distance between centers of mass of the obs/fcst fields (not sensitive to rotation around center of mass)

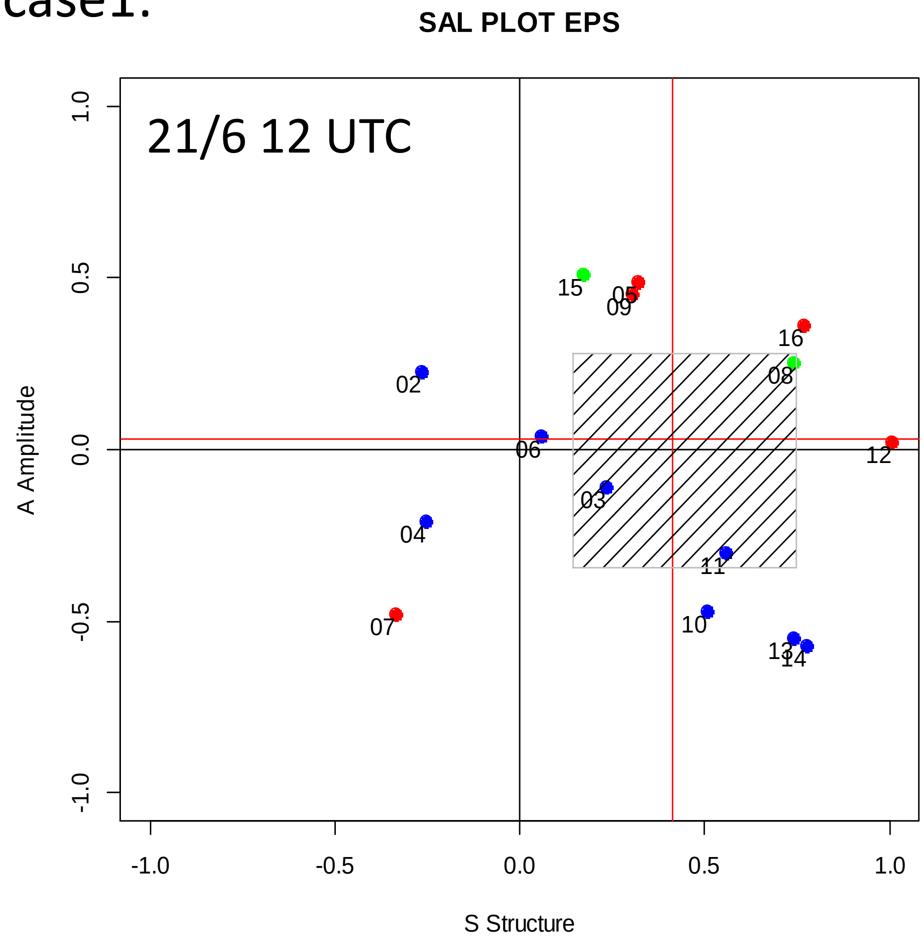
L2: difference of normalized distance between center of mass and individual objects over observed and forecast field.

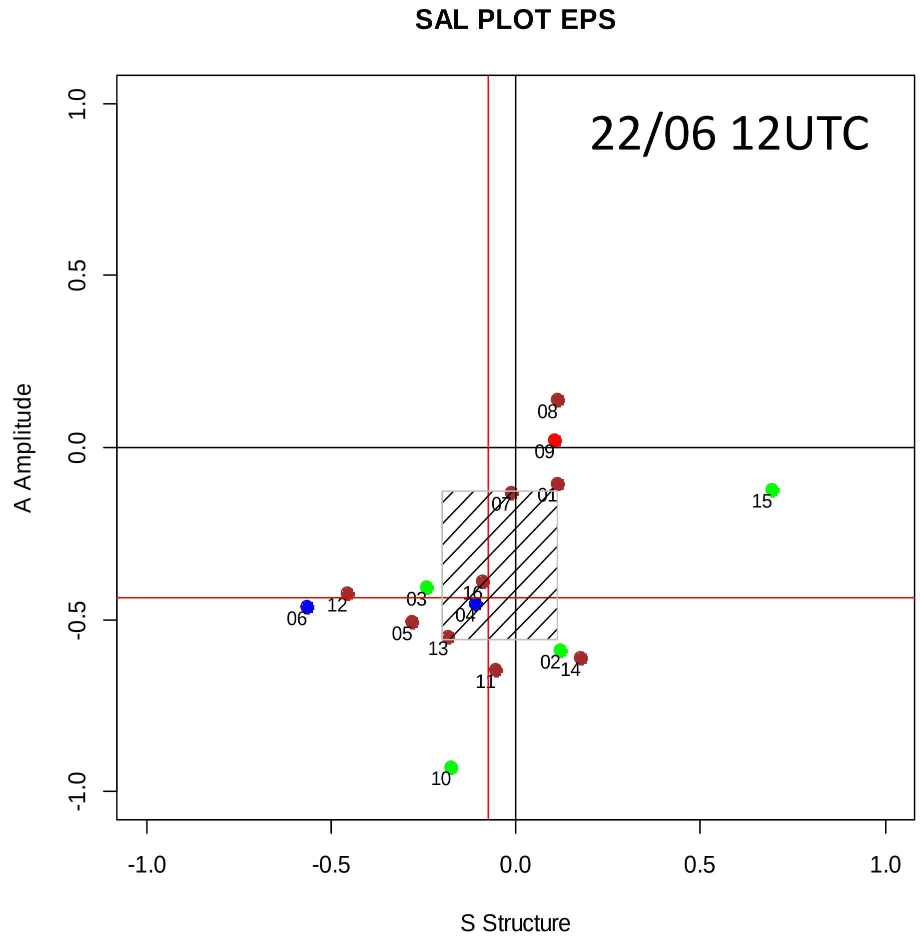
SAL PLOTS use for EPS evaluation

ΕΘΝΙΚΗ ΜΕΤΕΩΡΟΛΟΓΙΚΗ ΥΠΗΡΕΣΙΑ
HELLENIC NATIONAL METEOROLOGICAL SERVICE

SAL method can be applied in order to estimate the performance of an ensemble forecast ***(Barrett et al. 2015).*** Each point of the SAL plot represents one member, and an ensemble performance can be estimated.

An example case shown here used data for COSMO2 LEPS 16 members MesoVICT case1:





S >0 larger objects predicted

A < 0 domain values underestimated

Analysis of the usefulness of various spatial methods - Guidelines

- ✓ efficiency in calculation time
- ✓ ability to deal with different density of observations
- ✓ stability against observation errors
- ✓ proving added value of high-resolution models
- ✓ ability to address specific issues of interest (e.g. location errors, intensity errors, performance at different scales) etc.

- ***IS method:* interpretation of scores is not straight forward*. ***Not suitable for operational verification*** *as it is not concentrated in the average behavior of the model over areas but on single forecast.*
- ***Neighborhood methods:*** *with the right choice of decision model and aggregation on several timesteps/runs can provide a more* ***operationally*** *"useful" type of information. Suitable for other parameters than precipitation.*
- ***CRA method.*** *error usually comes from the fine structure of the field for lower precipitation thresholds. For higher thresholds, displacement error contribution increases.* ***Matching is tricky, important to consider each case before application of particular matching function. Aggregation of results with attention***
- ***SAL method*** *gives information on three attributes of a forecasted field. It has to be applied to relatively small domains. Further investigation on the object identification by thresholds according to each analyzed case is needed. Comparison of Wernli method and SpatialVx exhibits small differences on L parameter.* ***Application of SAL to an EPS forecast with all members could be a useful tool for EPS forecasts evaluation****.*

highlights of presentation

- Common verification software restricted to common plots and CV diagnostic applications
- Additional verification tools (software) to be developed/adopted for supplementary verification needs – avoid duplication of efforts
- Policy of observation and forecast data as input for verification tools
- Importance of interaction with international community for “key” verification issues (MesoVICT) – Spatial methods
  - **no single method can address all the errors**. One has first to decide on the properties that make his of **forecast “useful**” and then apply the method that focuses on the errors that are most important. Adoption of methods that can provide **comprehensive** results to the user

COSMO verification contributing team

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
Andrea Montani, Arpae
Elena Oberto, ARPA-PT
Ulrich Pflüger, DWD
Maria Stefania Tesini, ARPAE
Naima Vela, ARPA-PT
Antonio Vocino, COMET

Ευχαριστώ

Grazie

Thank you

III. Verification Environment

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Run time

- Generation of a score file (valid for a single date): ~30 sec
- Aggregation of one month of single score files: ~10 min

Memory Consumption

- Single score file: < 4Gb
- Aggregation: < 25GB

Example

- Monthly, global, deterministic TEMP verification for 3 models, from scratch (feedback files are on file system) takes < 1h