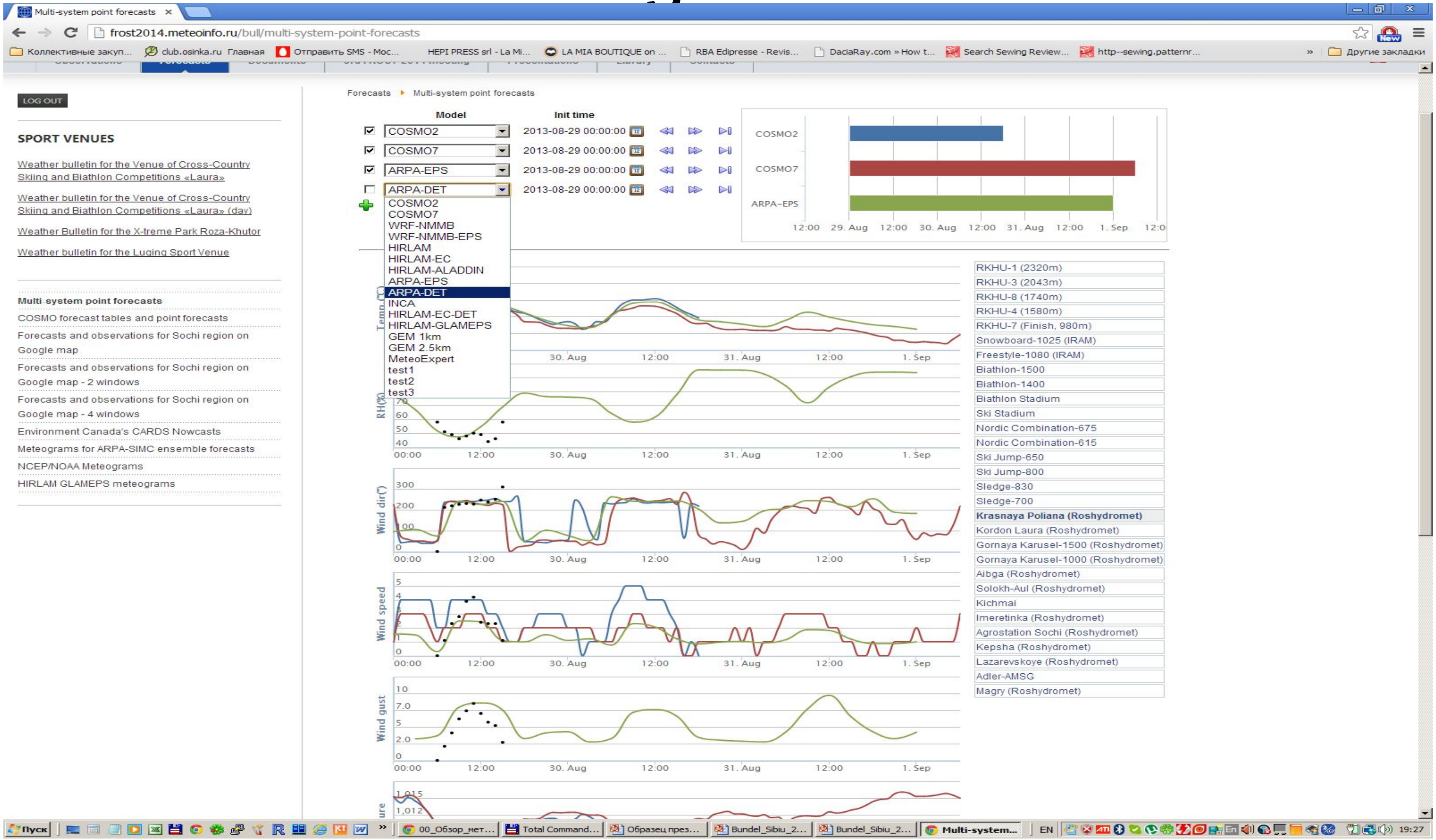
Verification in the region of Sochi-2014 Olympics, results of test periods

A.Bundel, A. Muraviev, A. Kirsanov, E. Glebova, D. Kiktev, G. Rivin, I. Rozinkina, D. Blinov, A. Smirnov, M. Zaichenko, and many others

Roshydromet

http://frost2014.meteoinfo.ru

Monitoring of forecasts



Murphy, 1997, two approaches to verification:

- *measures-oriented*
- *distributions-oriented – based on two-dimensional distribution* p(f,x) of forecasts and observations

Overview

1. Two test periods:
   1st : 01.12.2011 - 31.03.2012
   2nd : 01.01.2013 - 15.03.2013
2. Measures-oriented: Polygonal verification of COSMO and NMMB (NMMB for the 2nd test period only)
3. Distributions-oriented: COSMO station-based diagnostic verification
4. Radar-based verification: first results
5. Weather type verification
6. EPS verification in VERSUS. Beginning.

Region characteristics

- Complex relief: branches of the Big Caucasian Ridge
- The non-freezing Black Sea
- Large height variance
- Seasonal circulation features

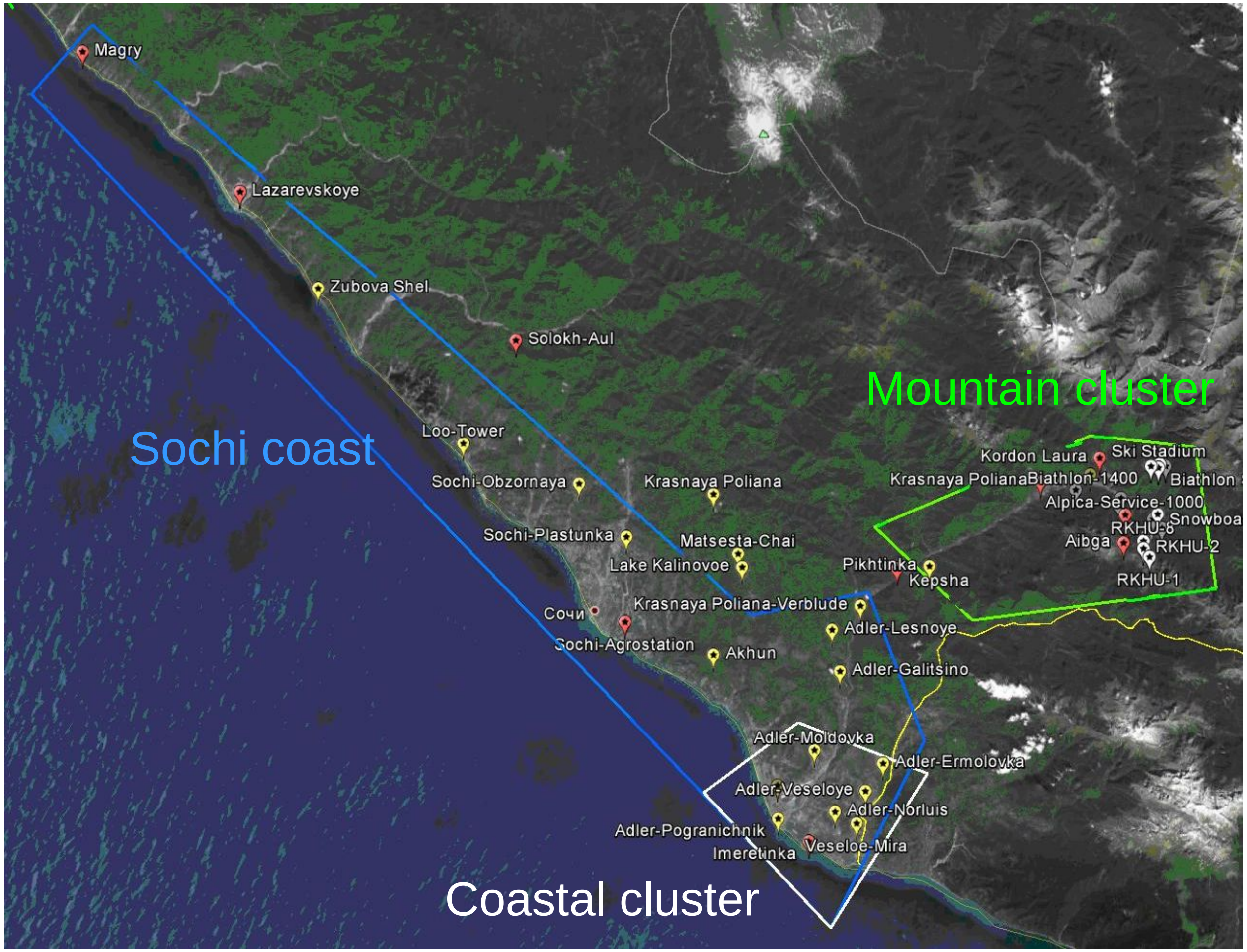
DO WE NEED SPATIAL AGGREGATION AT ALL?
We decided YES, to have a first overall picture

Areas

«***coastal***» (0 – 300 m above the sea level),
«***piedmont***» (300 – 600 m)
«***mountainous-Alpine***» (600 – 2500 m).

Because of the lack of stations in the piedmont area, piedmont and mountainous-Alpine areas were unified into one Mountain cluster.

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, 1-h lead-time step
- 4 initial times (00, 06, 12, 18)

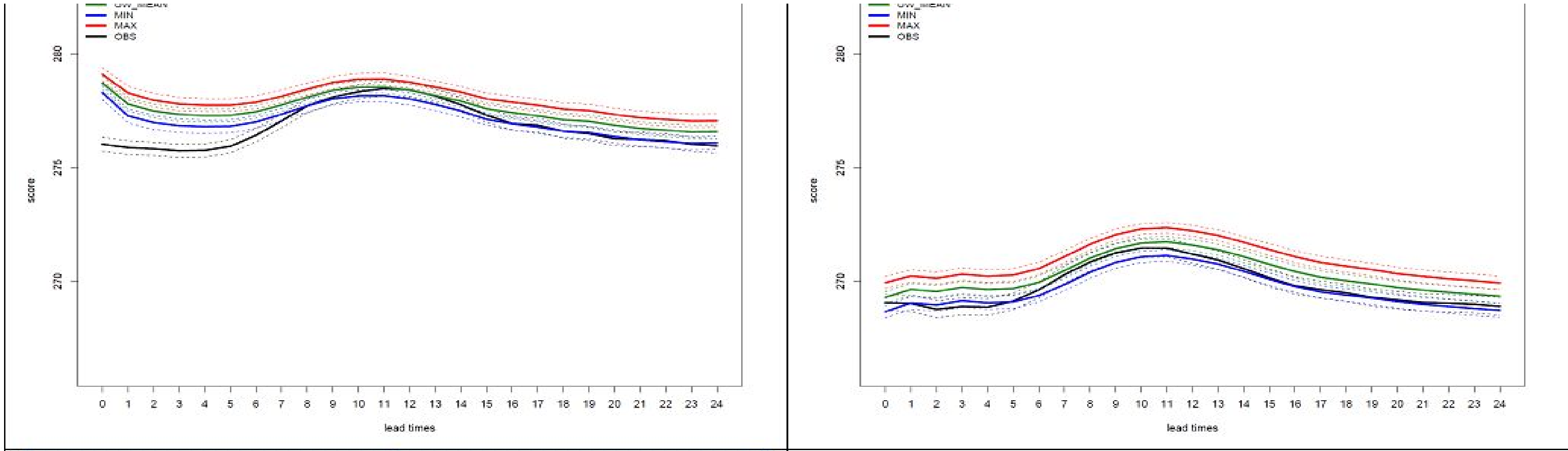
Verification setup

VERSUS and METv4 program were used.

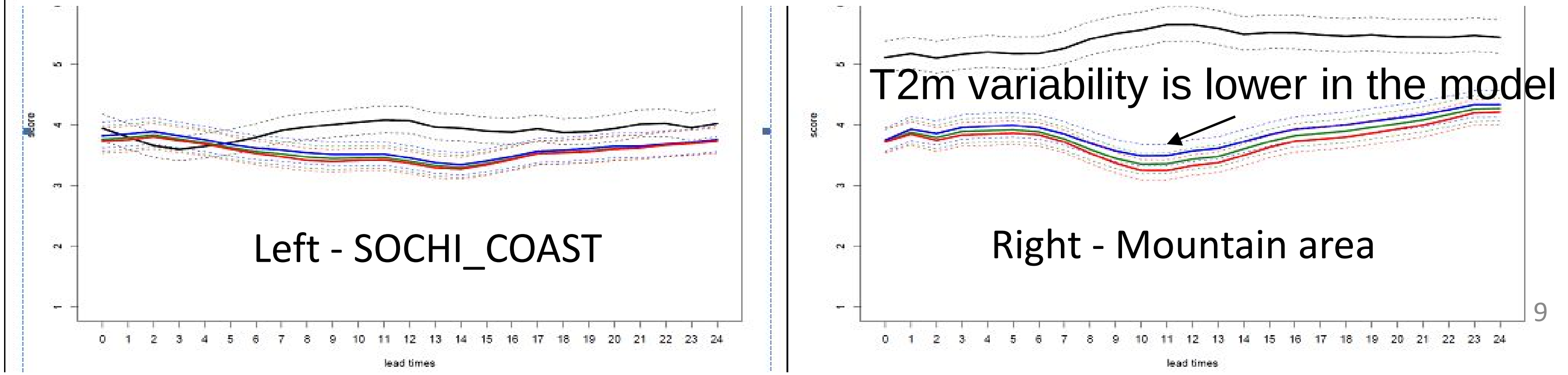
- Variables: T2m, wind speed, wind direction, MSLP
  Not yet properly analyzed: snow height – too many bad observations, precipitation – STARTED, but needs station control, visibility, total clouds, cloud base
- Thresholds (according to FROST Annex 6)
- Observation window: +/- 10 min around the forecast time
- Bootstrap and normal 95% confidence intervals in MET (300 resamplings)
- 5 interpolation methods (nearest point, distance-weighted mean of 9 nearest nodes, median, max, and min of 9 nodes) ***These methods gave similar results, on average (see next slide)***
- Obs quality control: simple thresholds (+/-50 C for T2m, MSLP < 1100 and >850 hPa, wind speed <0 and >70 m/s)

Different interpolation methods give similar results, on average

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



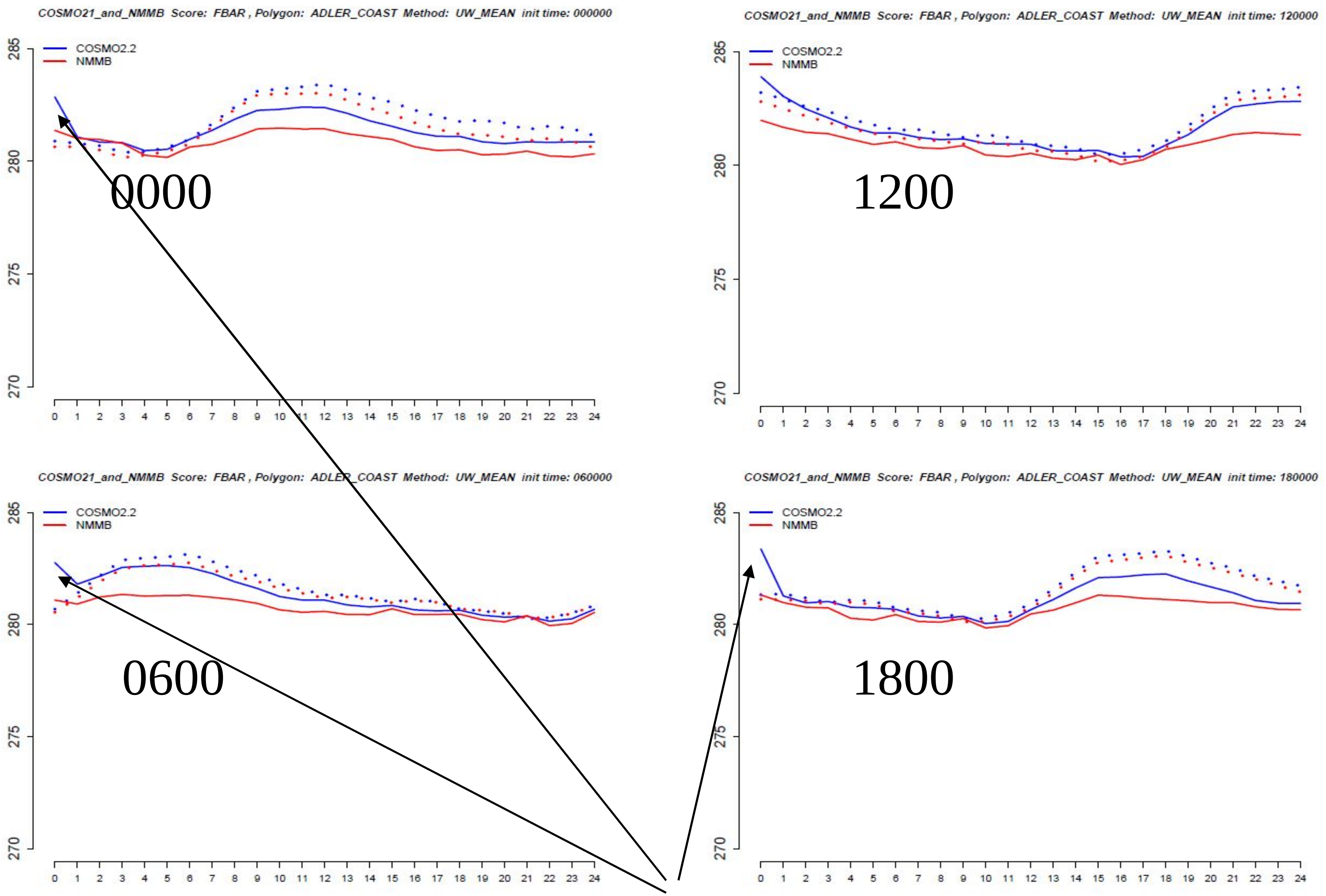
Standard deviations of forecasts from three methods (nearest point, MIN, MAX) and standard deviation of observations (black line)



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

Sochi coast

2nd test period



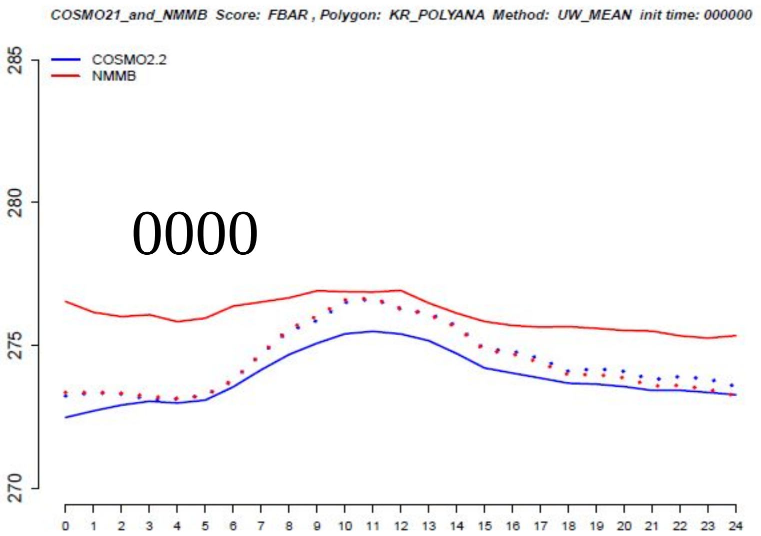
In the coastal polygons, there is a **systematic COSMO error at the initial time** that is likely due to the initial field. It is not detected in the mountain cluster.

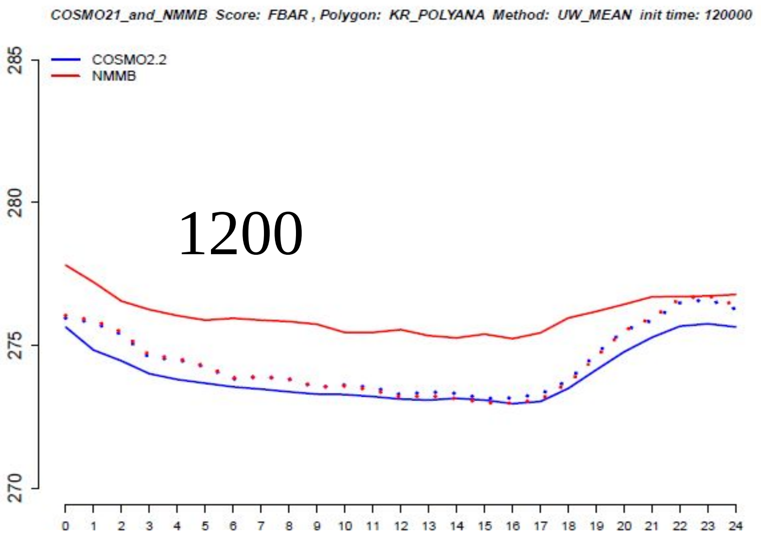
T2m (∘K) forecast and observation (dotted) means

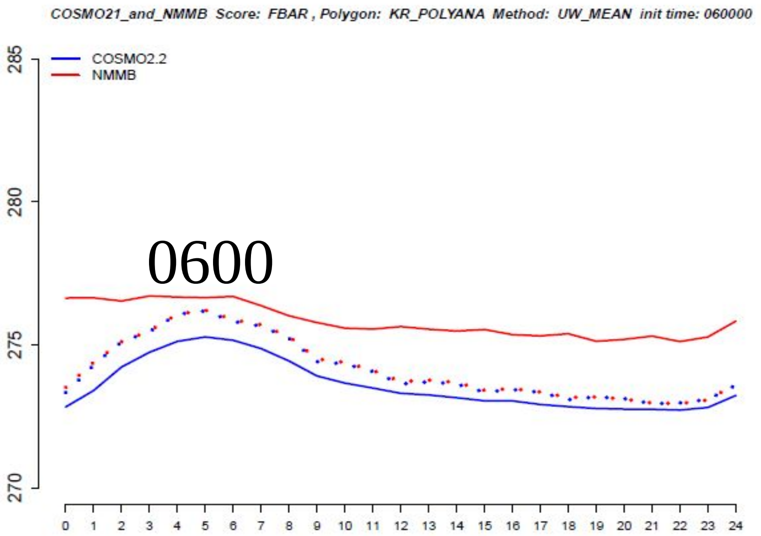
COSMO blue, NMMB red

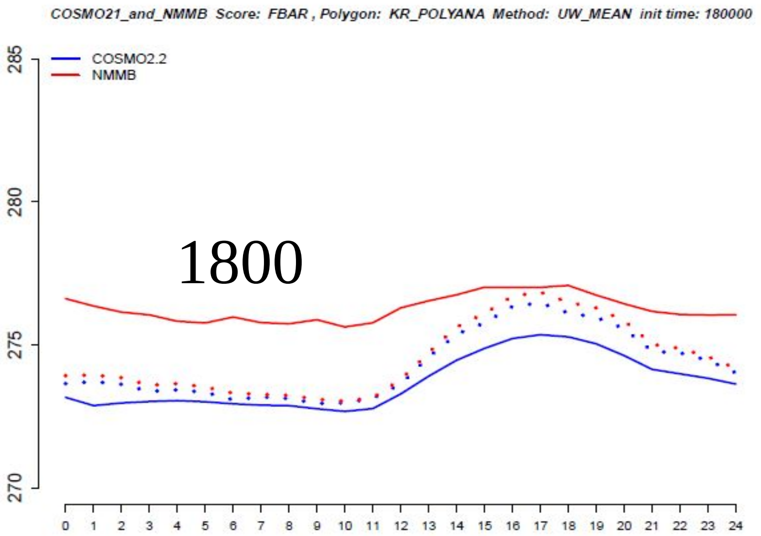
Mountain cluster

2nd test period









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

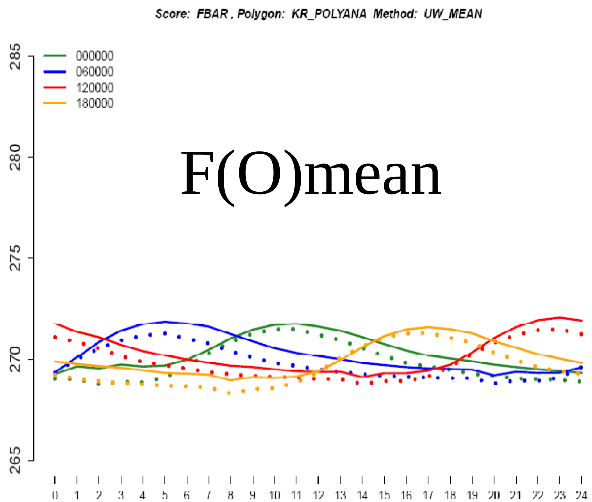
COSMO T2m Mean Error changes its sign from 2011-2012 to 2013 winter:

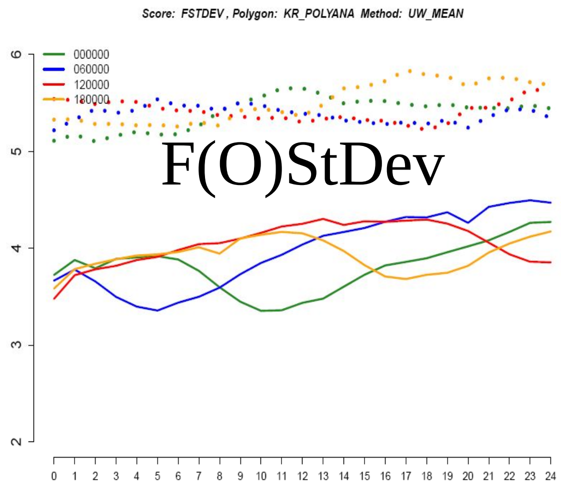
Is it due to model changes, or 2013 warm anomaly, or observation data distribution, amount, and quality?

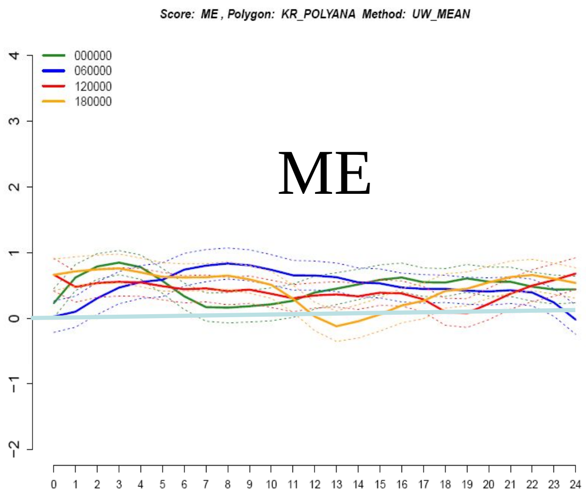
Difficulty for model calibration.

Mountain cluster

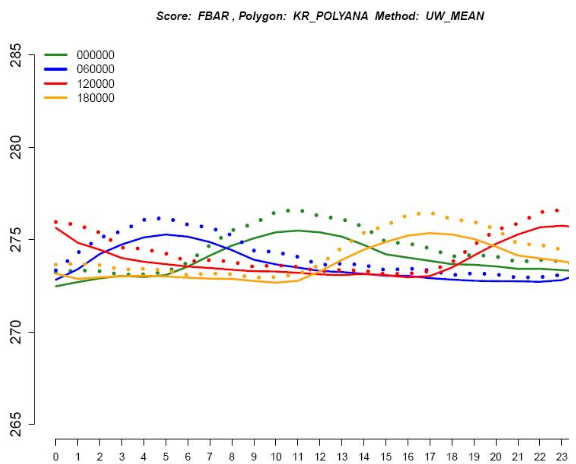
Above: 1st test period

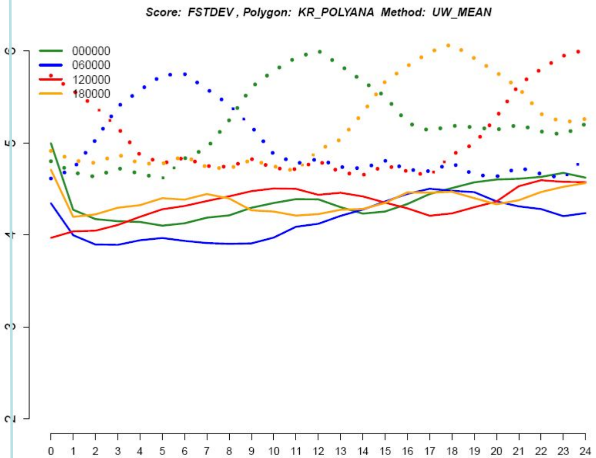


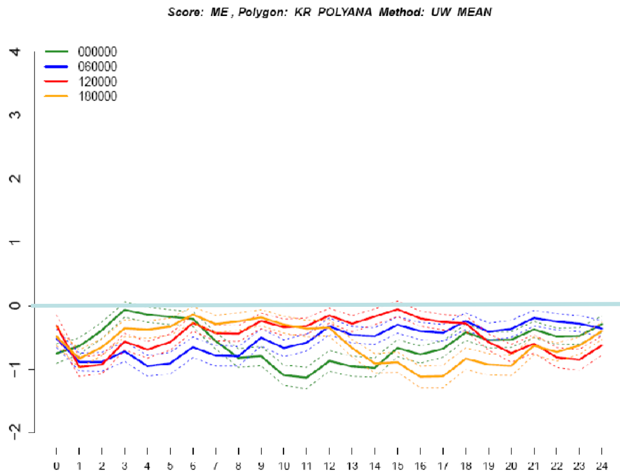




Below: 2nd test period



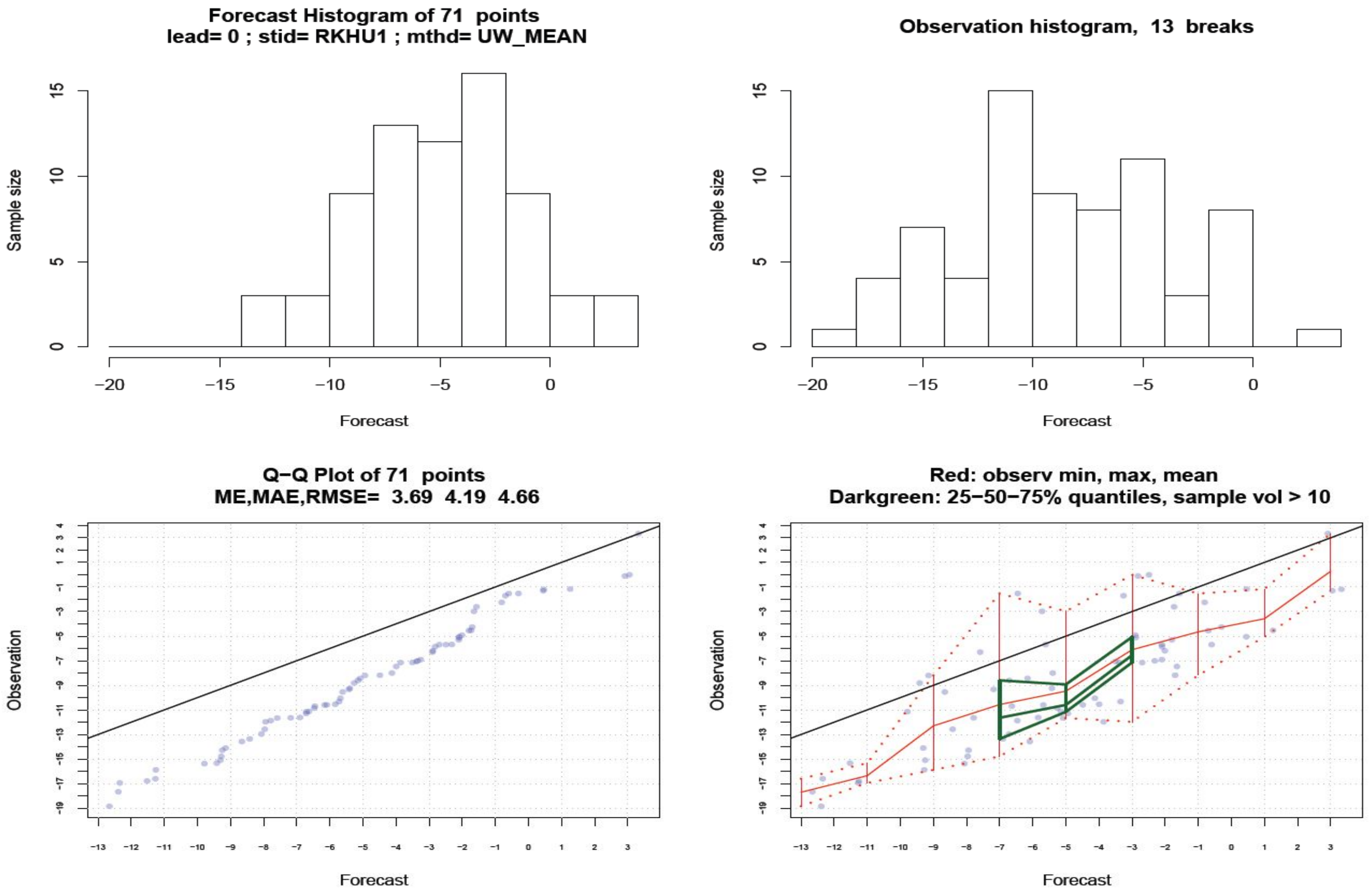




Station-based diagnostic verification

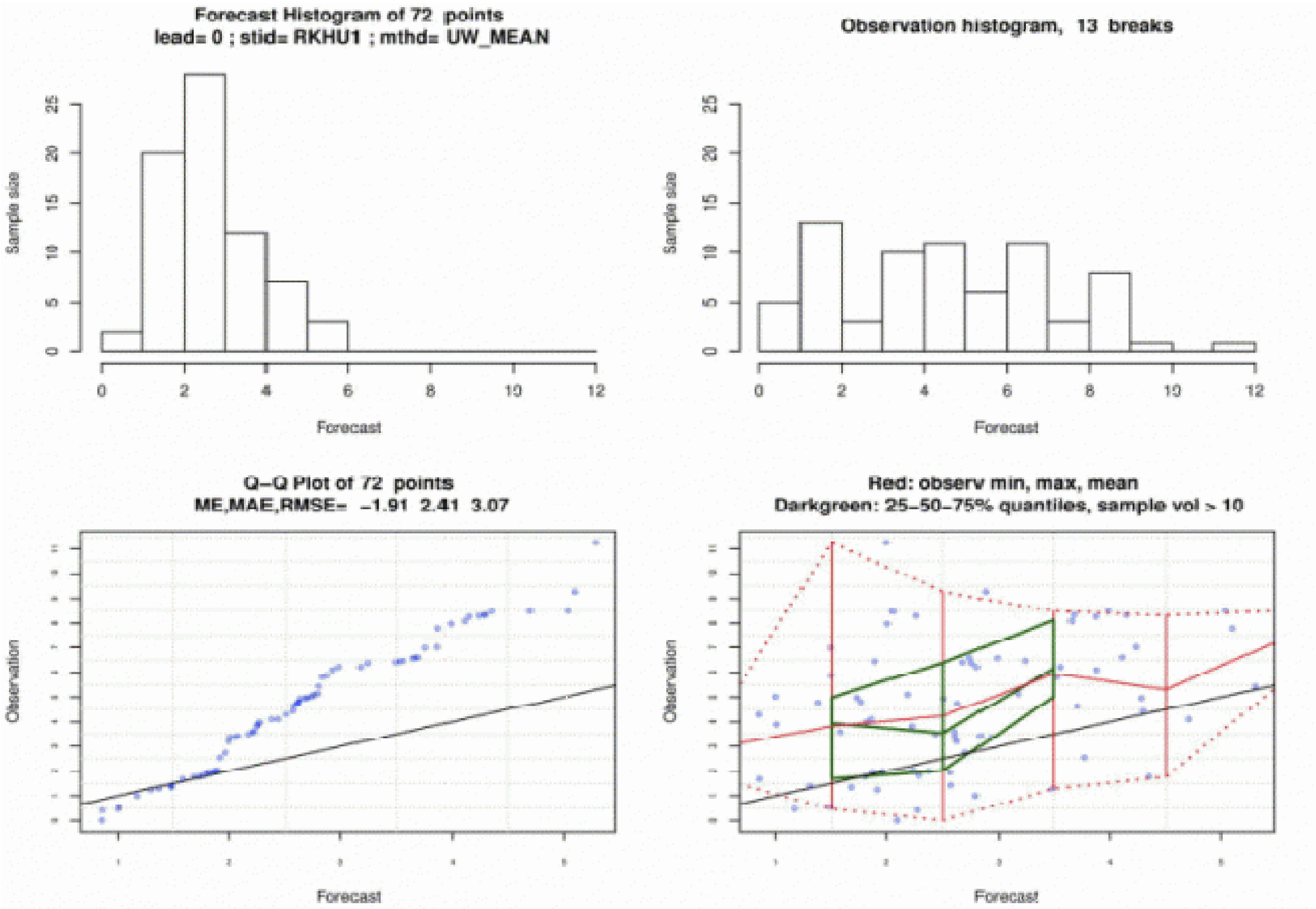
- “Diagnostic” in the sense that it focuses on the fundamental characteristics of the forecasts, the corresponding observations, and their relationships
  (A. Murphy, B. Brown, Y. Chen, 1989).
- “Station portraits” are made for each variable, station, lead time, and method (only for COSMO yet).
- They 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.



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).

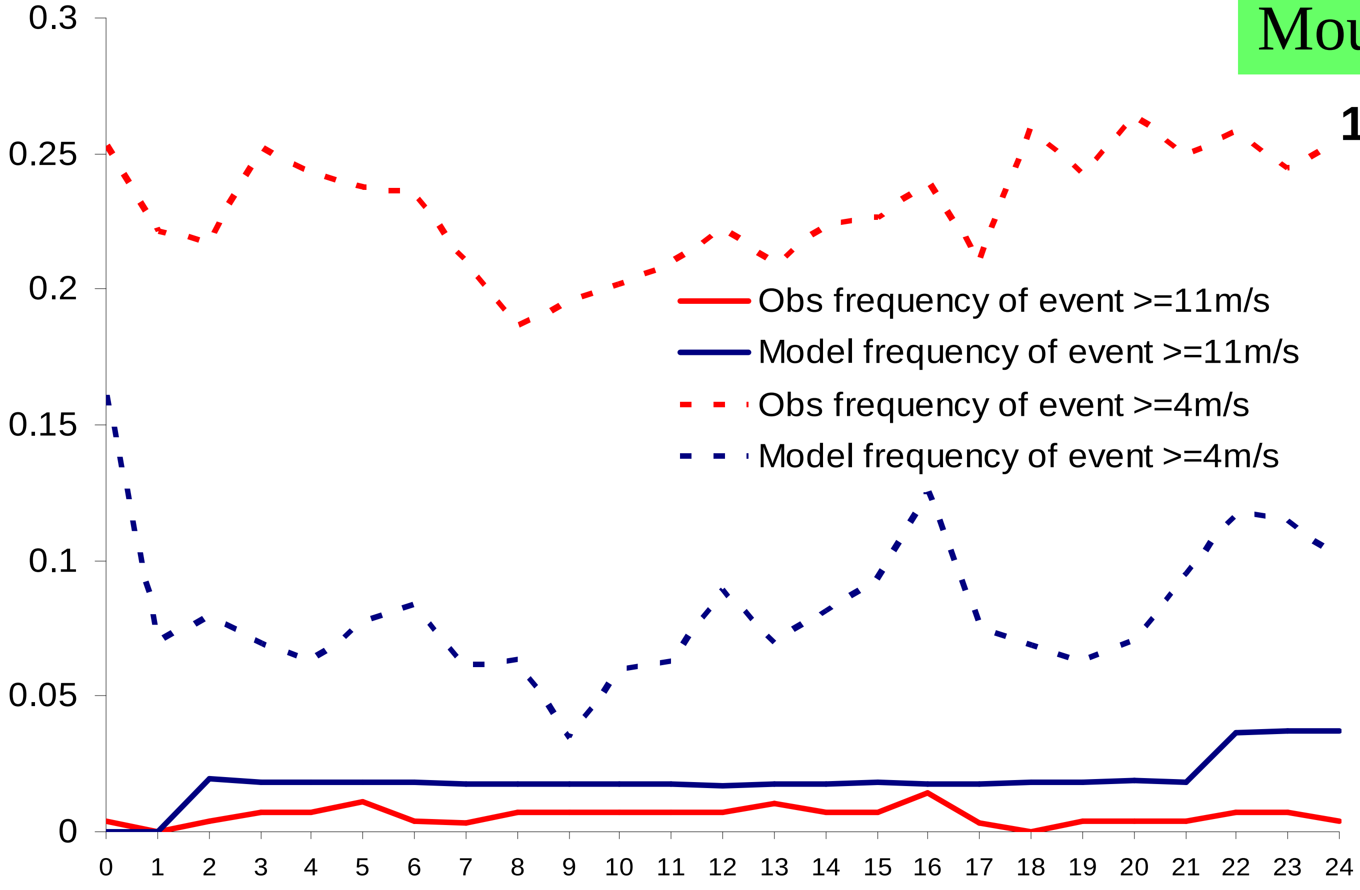
Wind speed portraits, RKHU1 stations in the mountains



Wind speed. Model and observed event frequency for two thresholds >=4m/s and >=11m/s.

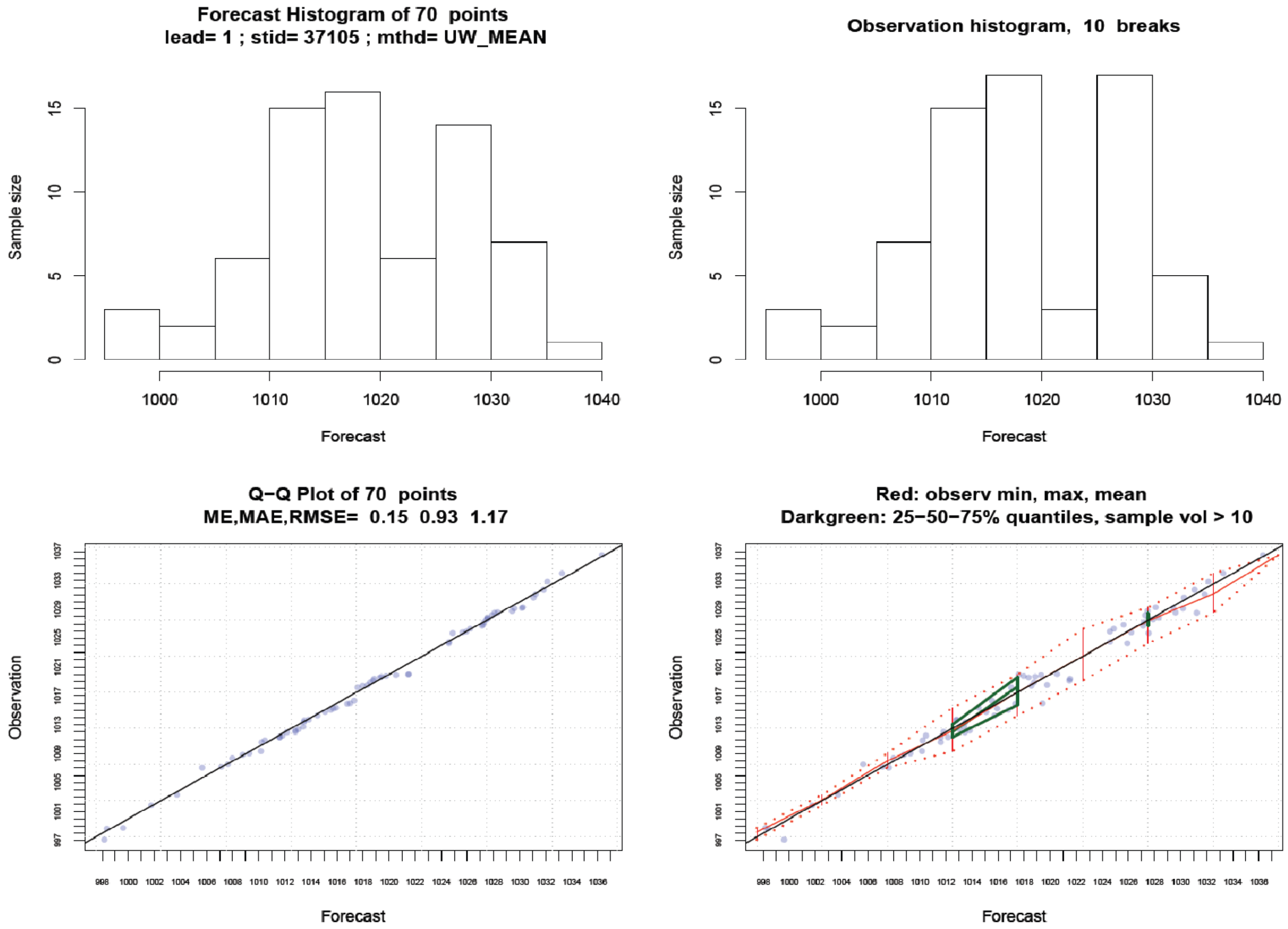
Mountain cluster

1st test period



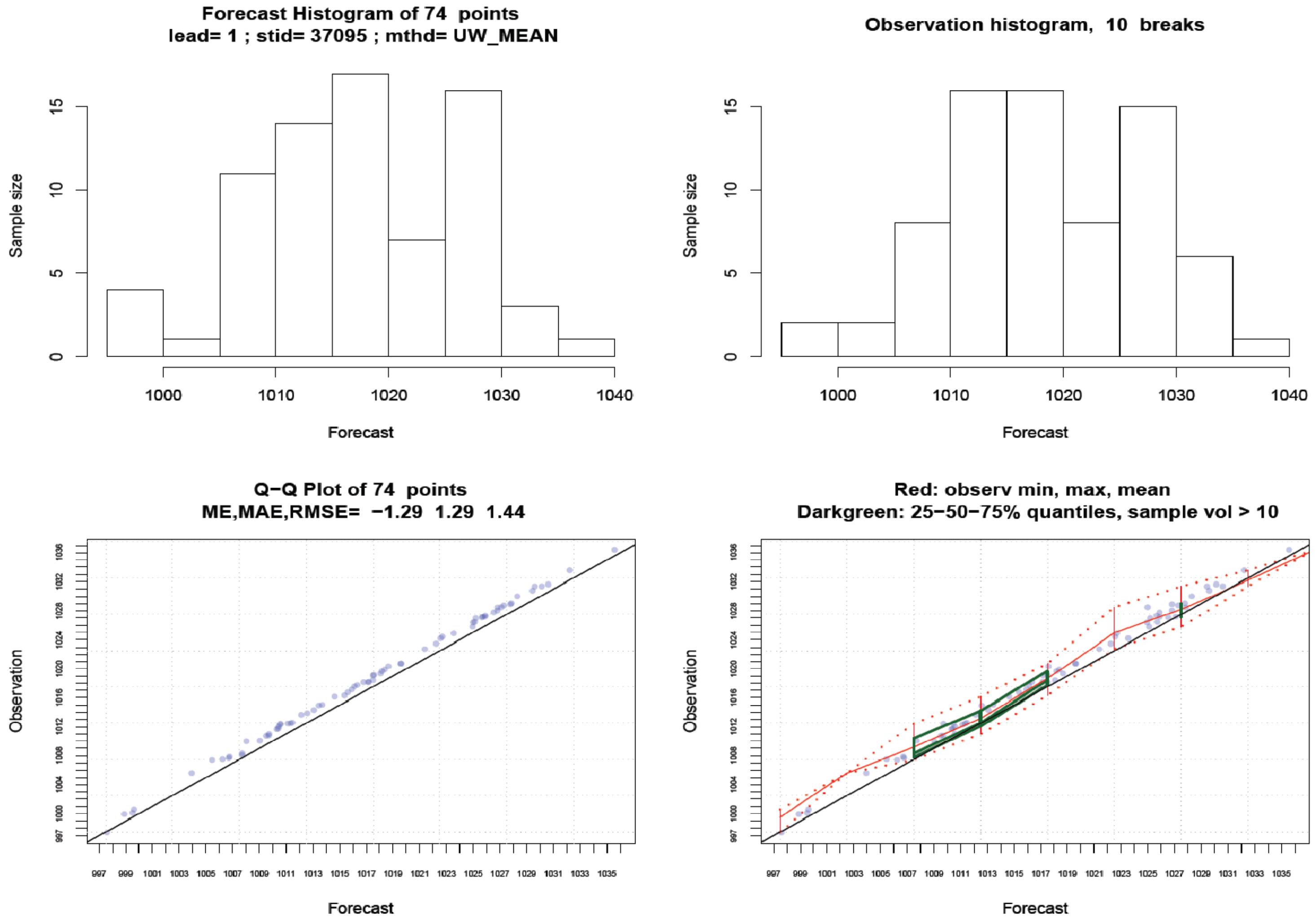
The model underestimates the event frequency for small wind thresholds, and overestimates it for great thresholds.

PMSL. 37105 Alpica-Sevice-1500 (in the mountains)



The quality is rather high and stable for all lead times

PMSL. 37095 Imeretinka (Coastal area, south)



MSLP is slightly underestimated (ME of about -1 hPa) for all lead times

INTERMEDIATE CONCLUSIONS:

- Temperature ME changes its sign from year to year. It makes more difficult the model calibration.
- COSMO yields better T2m means and the diurnal cycle, especially in the mountain cluster, compared to the NMMB model for 2012-2013 winter
- 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. The first results show that it is corrected using the data assimilation scheme
- TD is mostly overestimated
- PMSL is overall well forecasted
- Model tends to underestimate weak winds and to overestimate strong winds

Such scores are useful for the model development, BUT we still do not quite satisfy the forecasters. Among their most urgent requirements are radar measurements interpretation and verification and weather type classification.

Radar-based verification. First results

- Achievement of the scientific verification purpose (densely populated observations)
- Detection and analysis of mesoscale weather patterns (with subsequent synoptic typing)
- Usage of elaborated verification techniques - FUZZY, neighborhood, CRA (B. Ebert), Wavelet scale analysis...

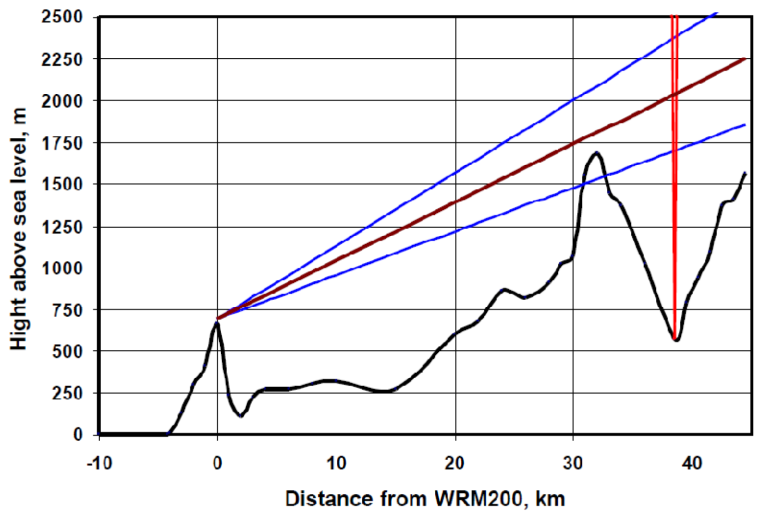
Vaisala doppler weather radar WRM200

on the Akhun mountain:

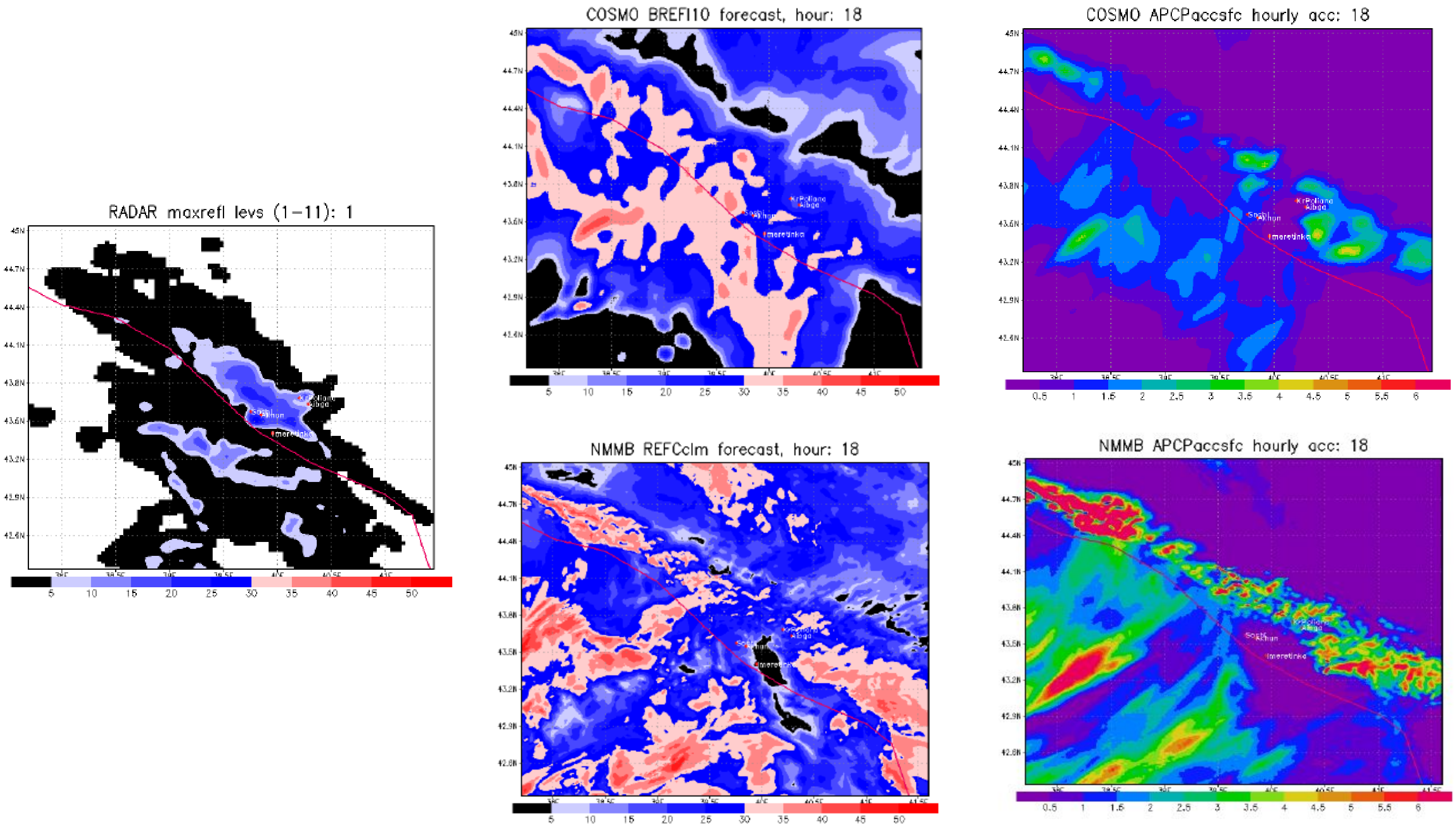
Radar shadow problem

It is recommended to use the max reflectivity in layer 1-3 km, but further calibration against gauges, satellites is needed.

Anatoly Muraviev works now on radar-gauge merging in mountains



Radar max reflectivity in layer 1-3 km for 2013.01.25, 17:29 vs COSMO and NMMB column max reflectivity and hourly precipitation forecasts for 2013.01.25, 18:00 (18-h lead time)

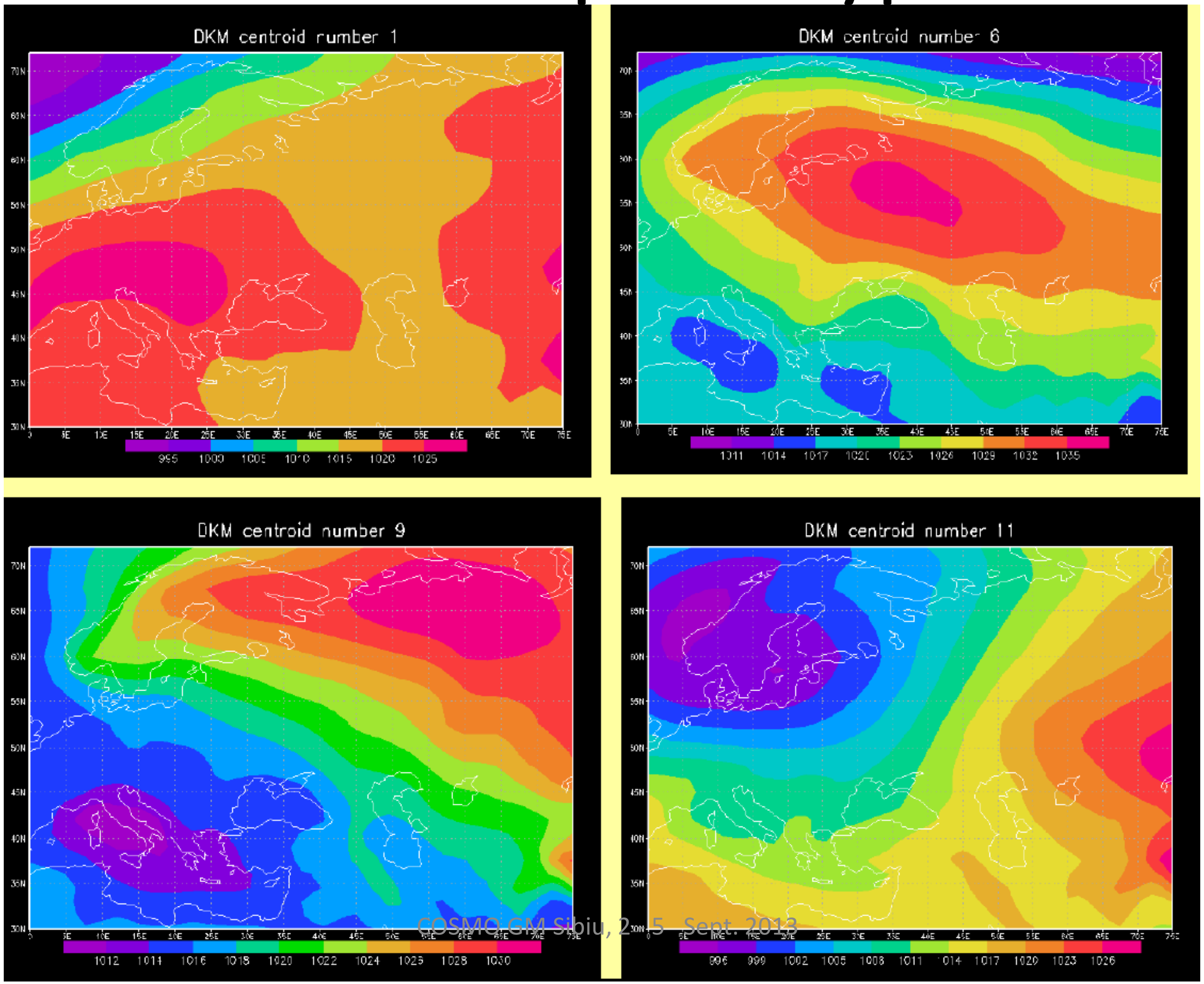


More patched NMMB pattern compared to COSMO because of smaller resolution

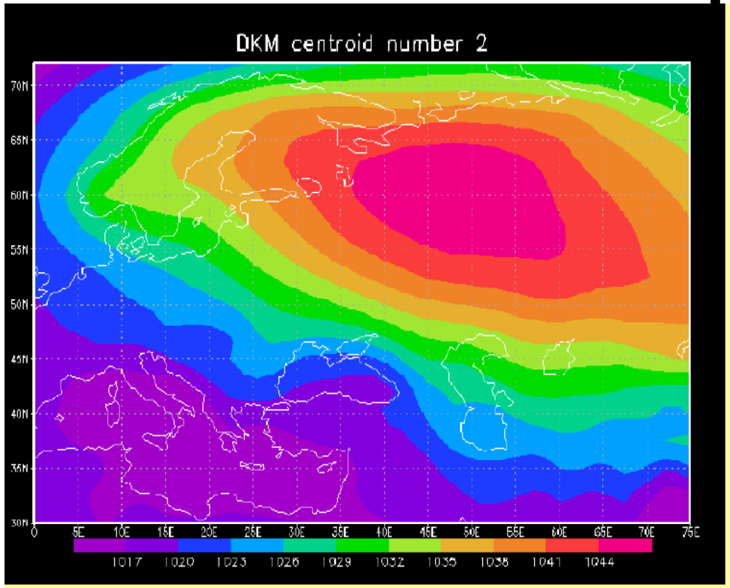
Weather type verification

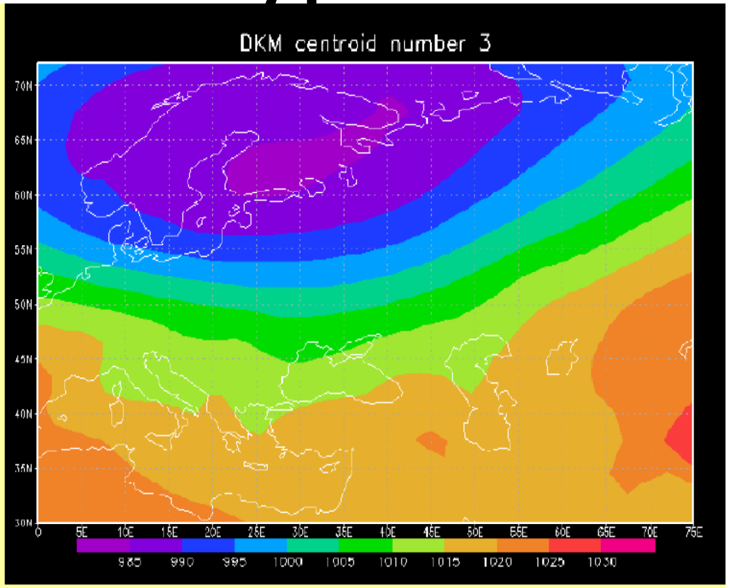
- Large number of experiments –
  - 10, 20, 30 types
  - CKM, DKM, PCT, PTT
  - three domains of different scales
  - pmsl and pmsl anomalies as classification variables (ECMWF ERA40 and interim reanalysis, DJFM 01.09.1957 – 31.01.2013)
- To evaluate "discriminative power" of classifications, Kolmogorov-Smirnov 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.

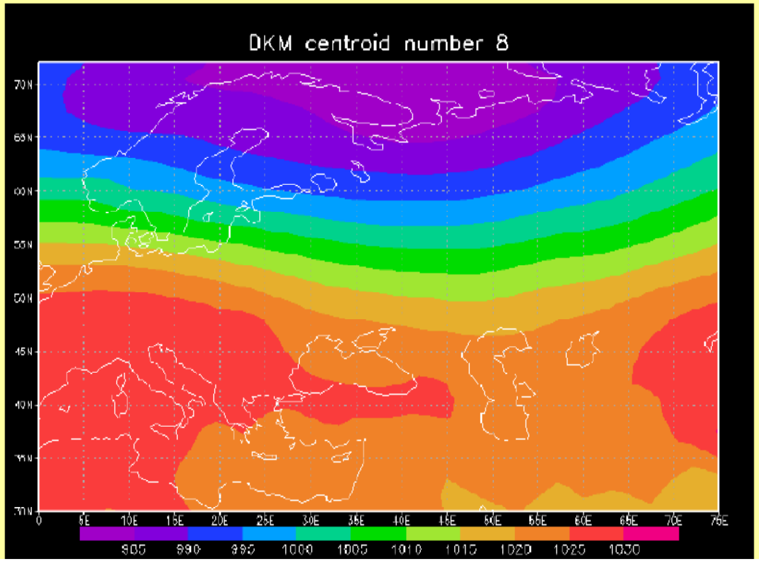
Most frequent types



Least frequent types

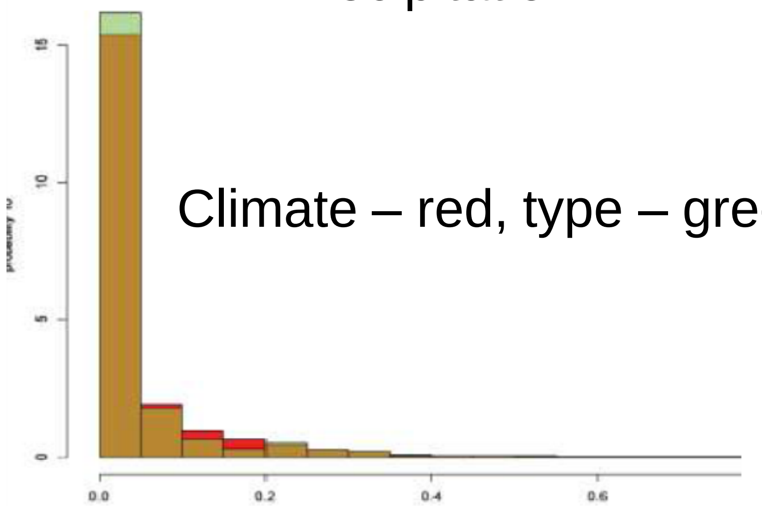






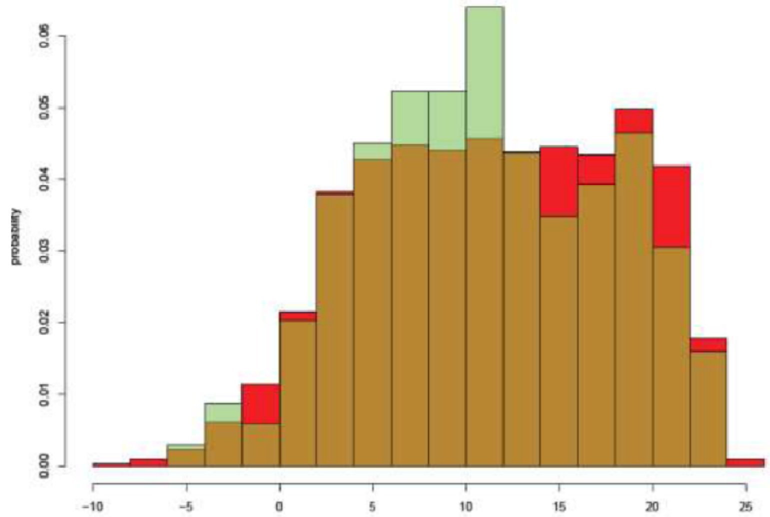
Example: Type 4 – mostly drier and cool weather is more probable

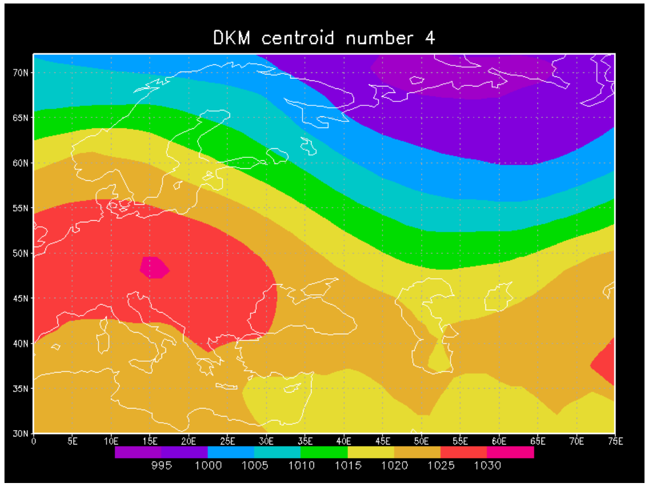
Precipitation

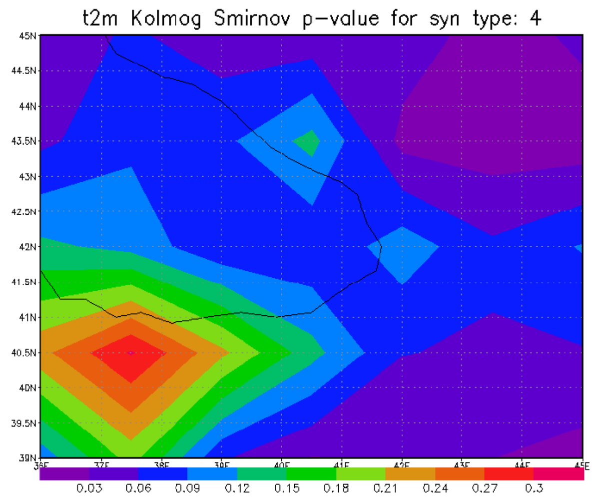


Climate – red, type – green!

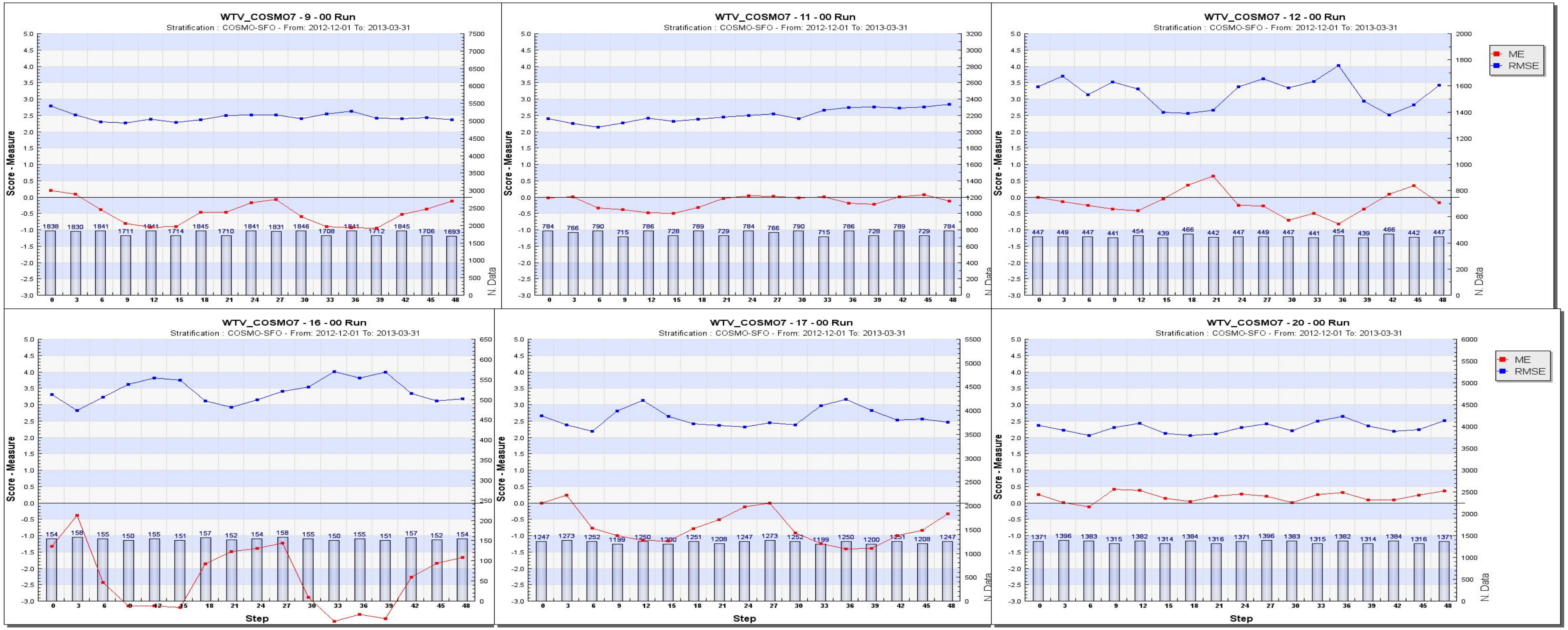
T2m



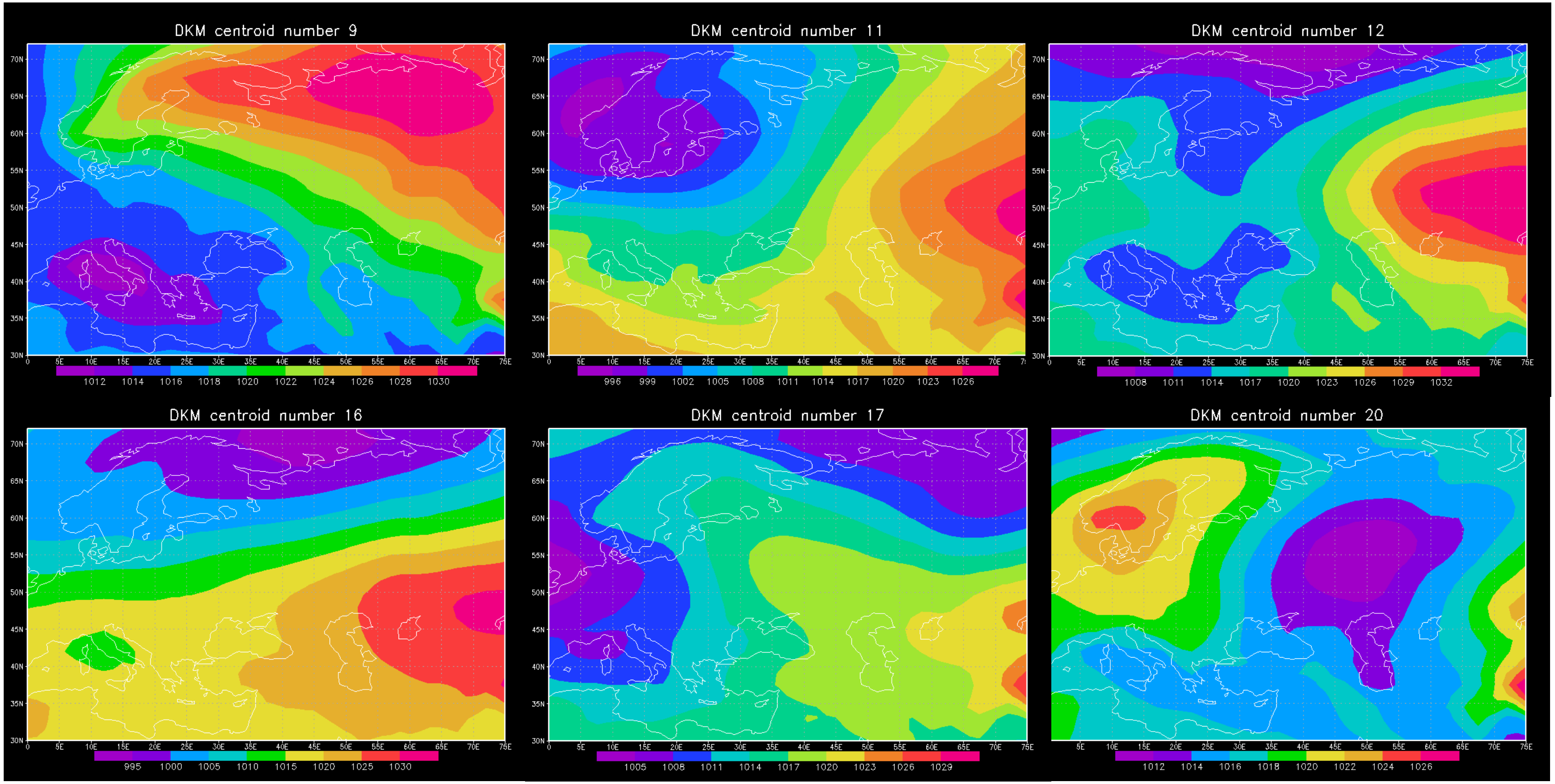




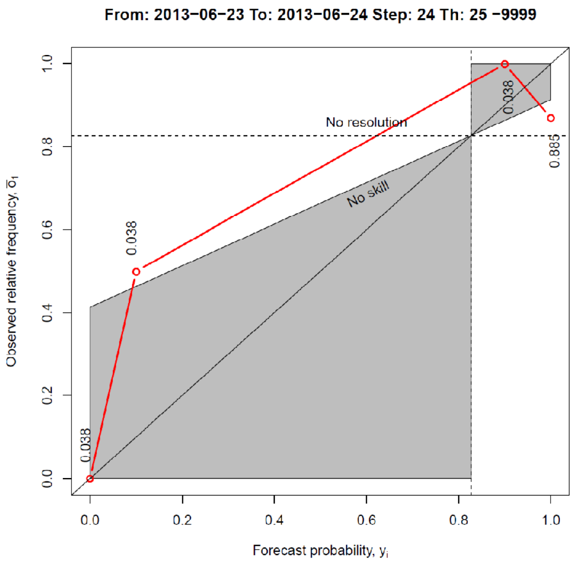
Weather type verification. COSMO-RU7, 2nd test period, whole Sochi region.

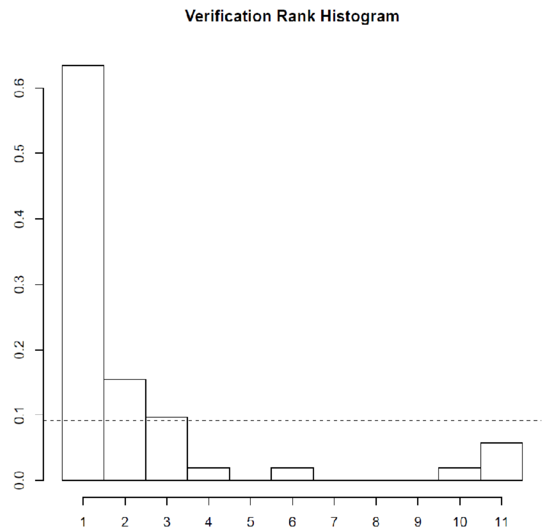


- There are differences in error cycles.
- Diurnal error cycle is most pronounced for some types.
- Type 20 – Sochi is in the rear of a cyclone with NNW flow – is the only type with mostly positive ME.
- **Such scores will be part of forecaster reference guide.**


DKM centroid number 9
1012 1014 1016 1018 1020 1022 1024 1026 1028 1030
DKM centroid number 11
996 999 1002 1005 1008 1011 1014 1017 1020 1023 1026
DKM centroid number 12
1008 1011 1014 1017 1020 1023 1026 1029 1032
DKM centroid number 16
995 1000 1005 1010 1015 1020 1025 1030
DKM centroid number 17
1005 1008 1011 1014 1017 1020 1023 1026 1029
DKM centroid number 20
1012 1014 1016 1018 1020 1022 1024 1026

EPS verification in VERSUS. Beginning.





It works, but needs testing.
Why different thresholds appear (20-25)?
We should load more data
(Italian ensembles as well as Russian 2.2-km ensembles).

Ongoing activity

1. Full variable set verification for all participating models (heavy programming efforts, unified verification complex on FROST server)
2. Ensemble forecast verification
3. Radar-based verification package for reflectivity and precipitation
4. High Impact Weather forecast strategy
5. Further efforts for weather type classification for the region
6. Template of the Refcard (guidelines) for forecaster, colorful, clear and informative
7. ONLINE VERIFICATION

Time is lacking…

Thank you for your attention!

INTERMEDIATE CONCLUSIONS:

- Temperature ME changes its sign from year to year. It makes more difficult the model calibration.
- COSMO yields better T2m means and the diurnal cycle, especially in the mountain cluster, compared to the NMMB model for 2012-2013 winter
- 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. We hope to correct it using the data assimilation scheme
- TD is mostly overestimated
- PMSL is overall well forecasted
- Model tends to underestimate weak winds and to overestimate strong winds
- Wind direction forecast quality is not satisfactory.

Such scores are useful for the model development,
BUT we still do not quite satisfy the forecasters.
Among their most urgent requirements are radar measurements interpretation and verification and weather type classification.

Forecasters say:

Problems Associated with Models

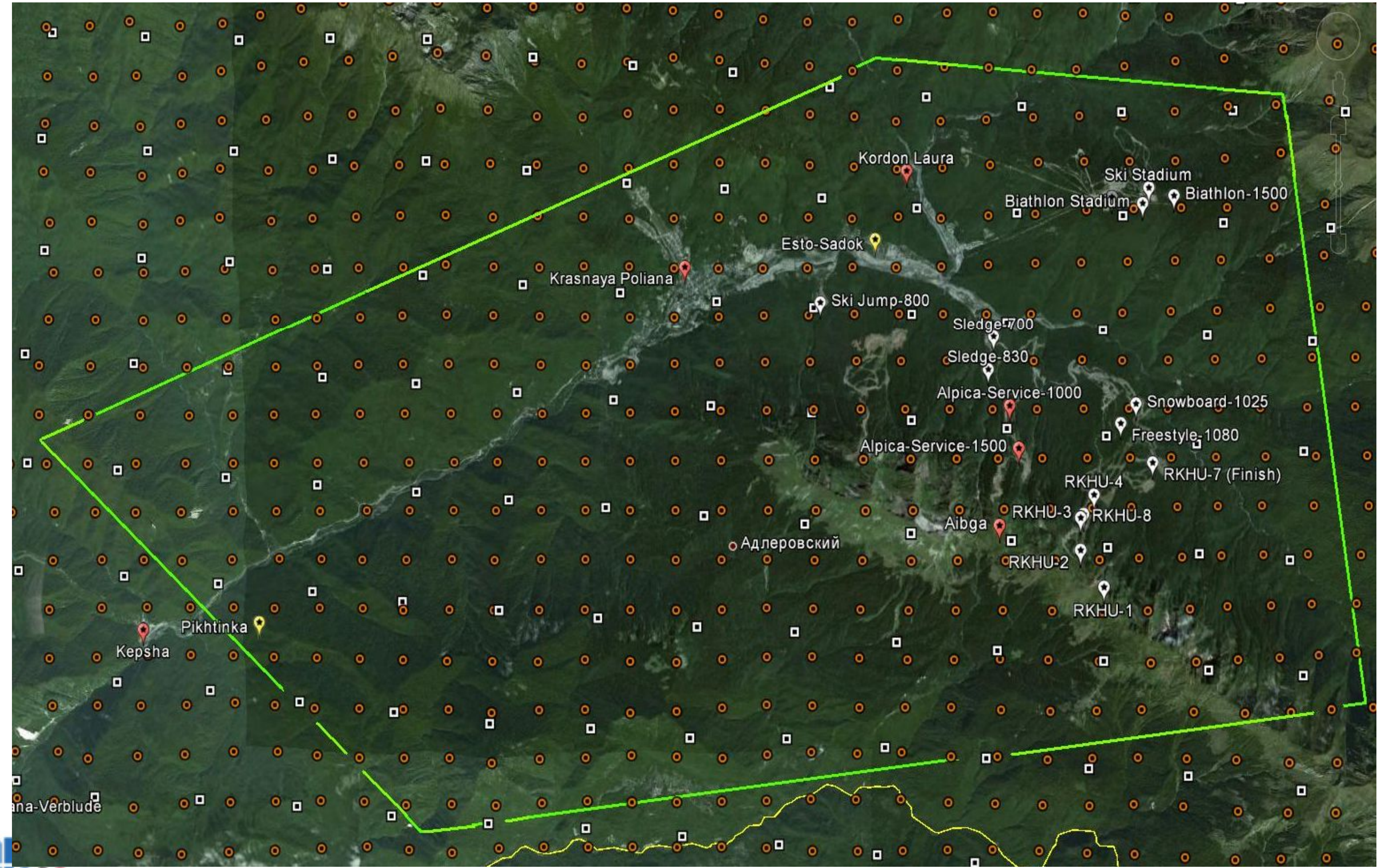
- Occasional underestimation of maximum temperature and diurnal course by the most part of models in case of fair weather and foehn
- Delayed response of models in case of the process change (e.g., from southern to northwestern)
- Difficulties with forecasting the cloudiness characteristics, visibility, and precipitation type
- Late completion of 00 UTC model runs for ARPA-SIMC
- Need in representative and not time-consuming presentation of model output data: meteograms, summary tables, and prognostic fields of meteorological parameters with the animation option

Forecasters say:

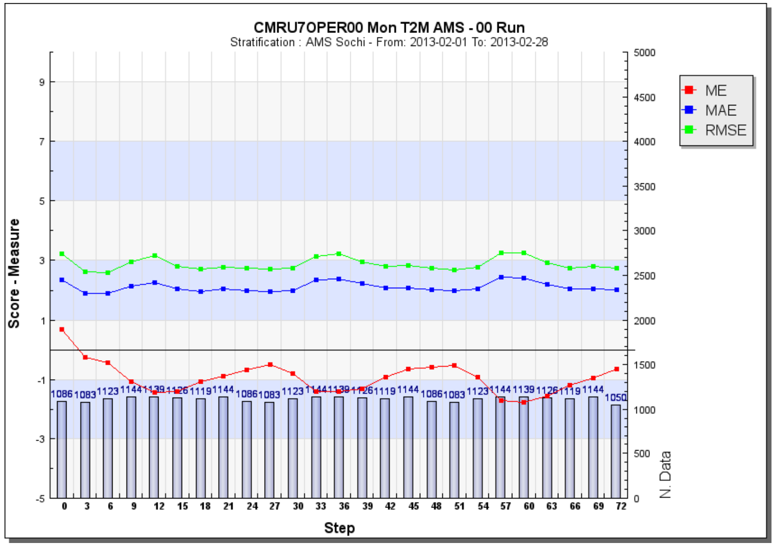
What we need

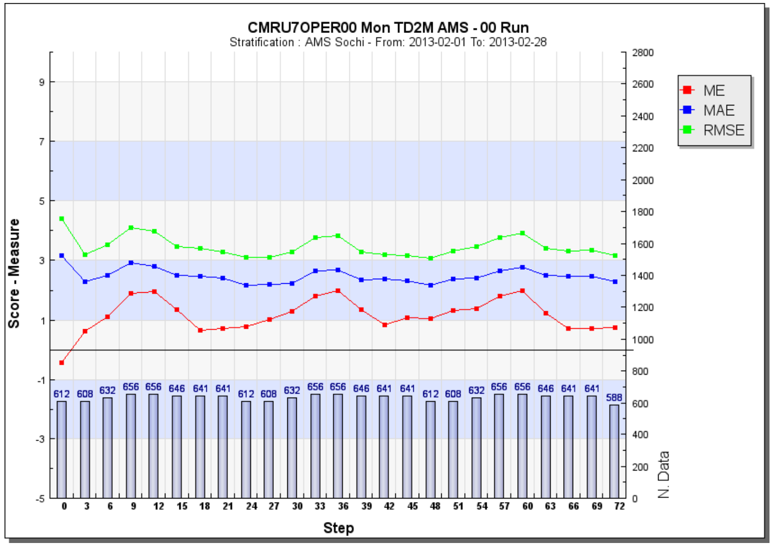
- Steady operation of automatic stations and trustworthy data;
- Installation of profiler and the use of the full potential of instruments for the monitoring and forecast of weather (MRR, Doppler radar, profiler, radio sounding, satellite products)
- Advance in forecasting such key meteorological parameters as temperature, amount and type of precipitation, wind, etc.;
- Cloudiness remains one of the major problems; in view of this, the accurate forecast of visibility and cloud base is still extremely important;
- Taking account of local processes and making corrections to the models;
- More convenient form of presentation of actual and prognostic information (charts + meteograms + summary tables).

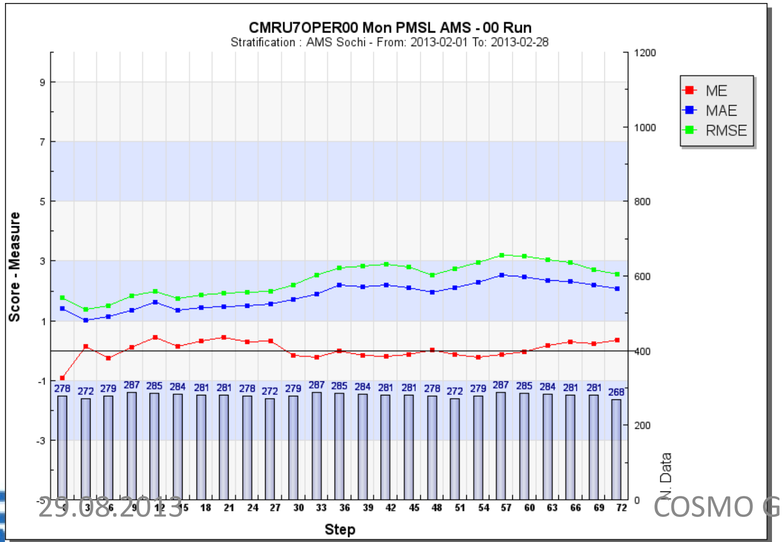
Mountain area magnified: white squares indicate the nodes of COSMO 2.2-km rotated grid, orange circles are 1-km interpolated grid, «lamps» – stations.

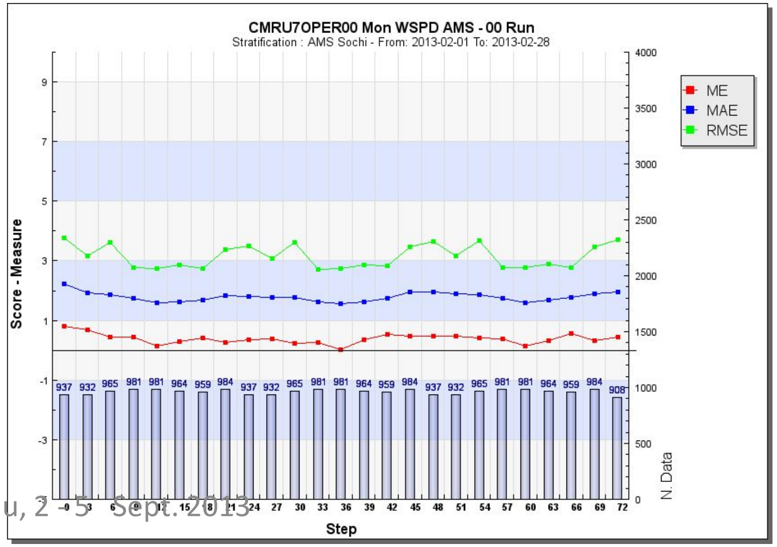


COSMO_RU7 February 2013



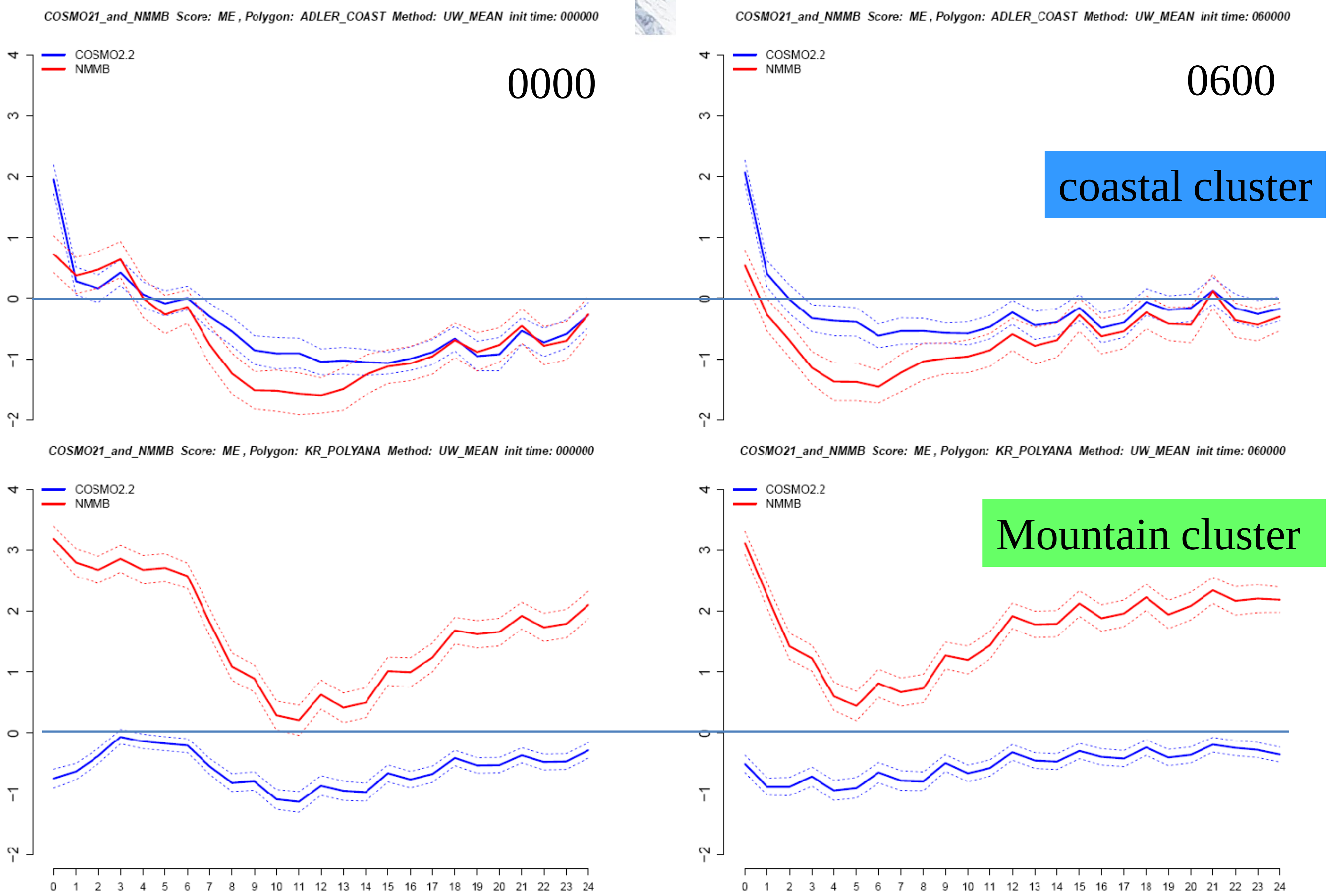


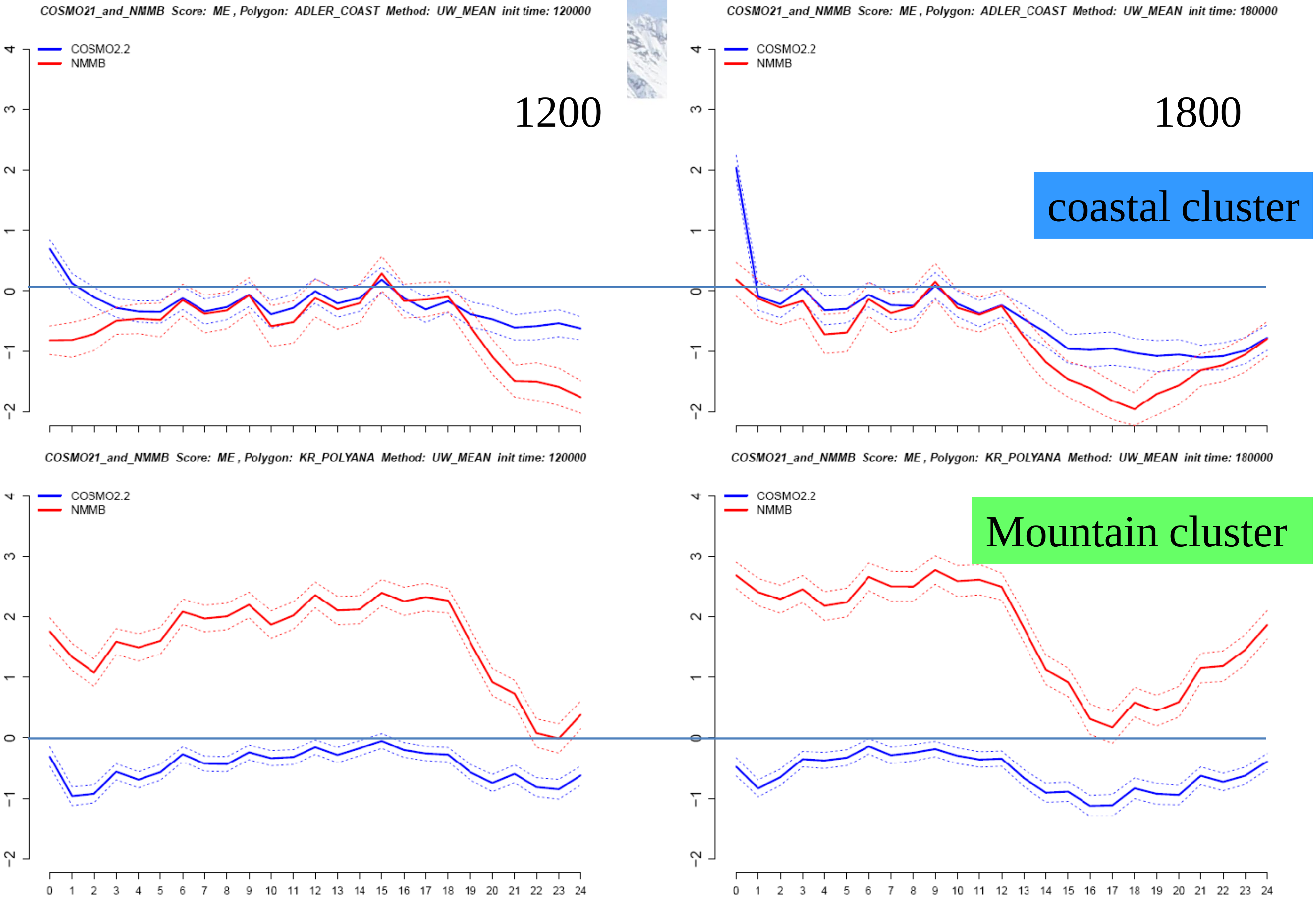




Mean error (ME) for coastal and mountain clusters

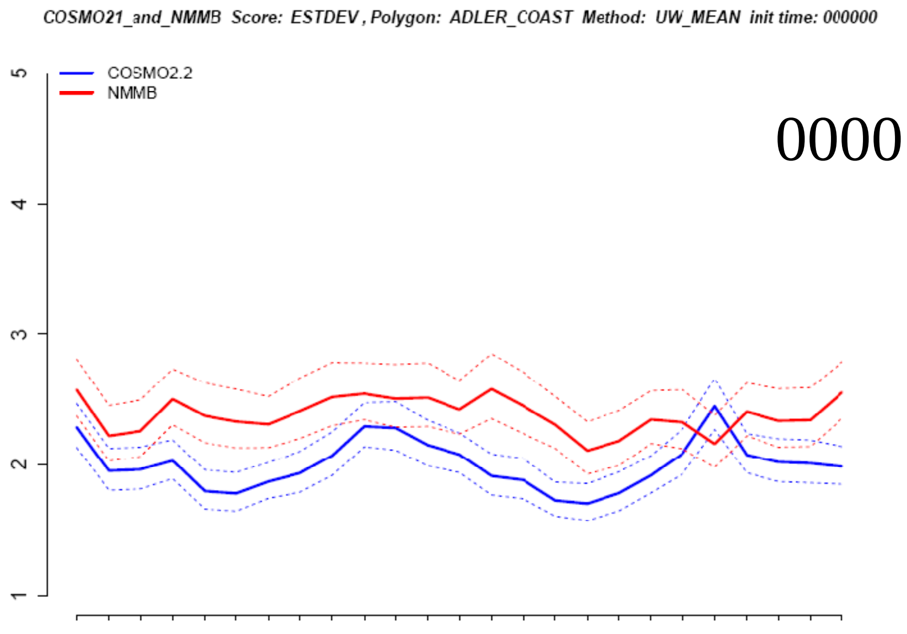
01.01.2013-15.03.2013

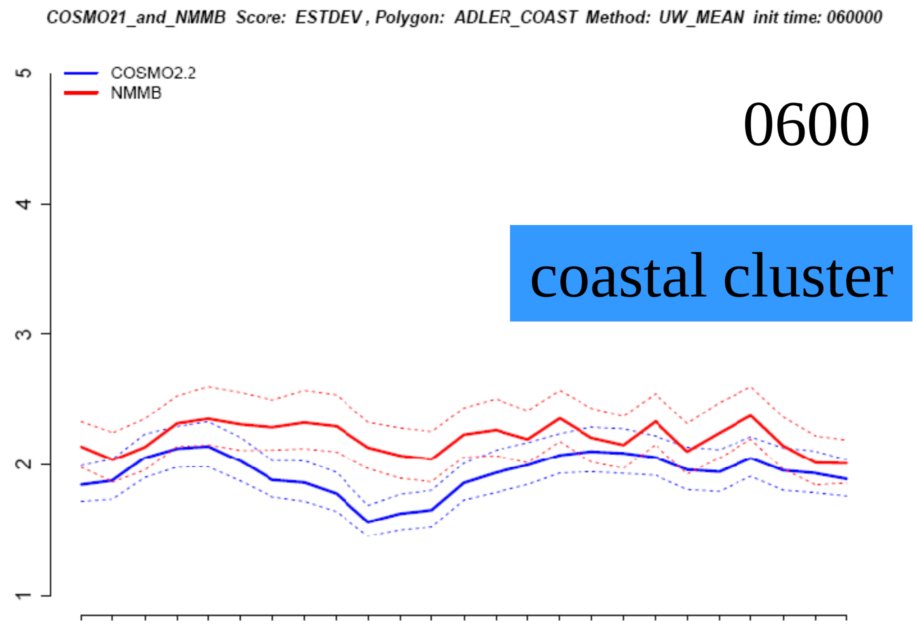


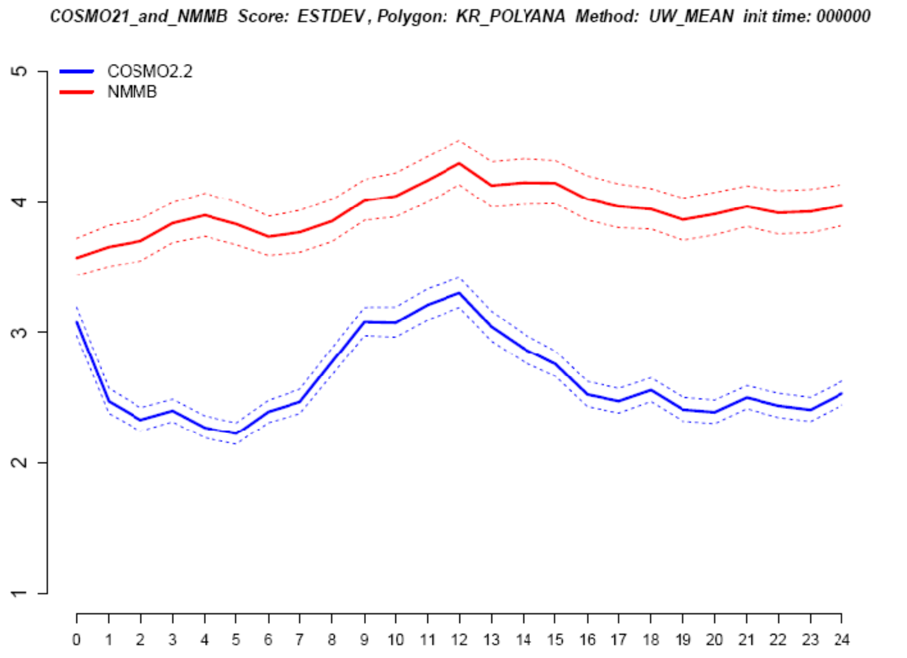


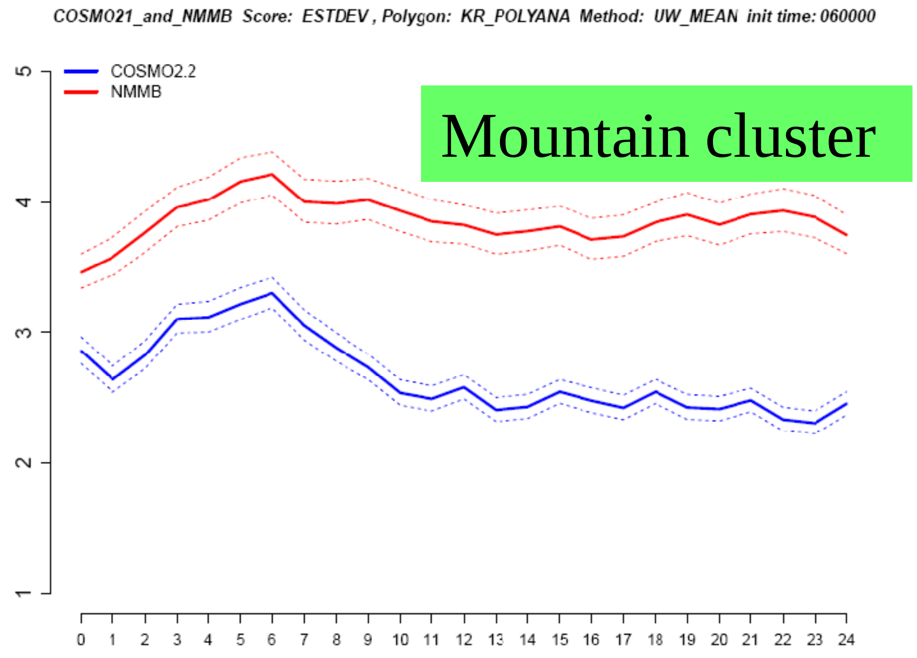
!Better COSMO performance in mountains, but ME < 0 almost everywhere

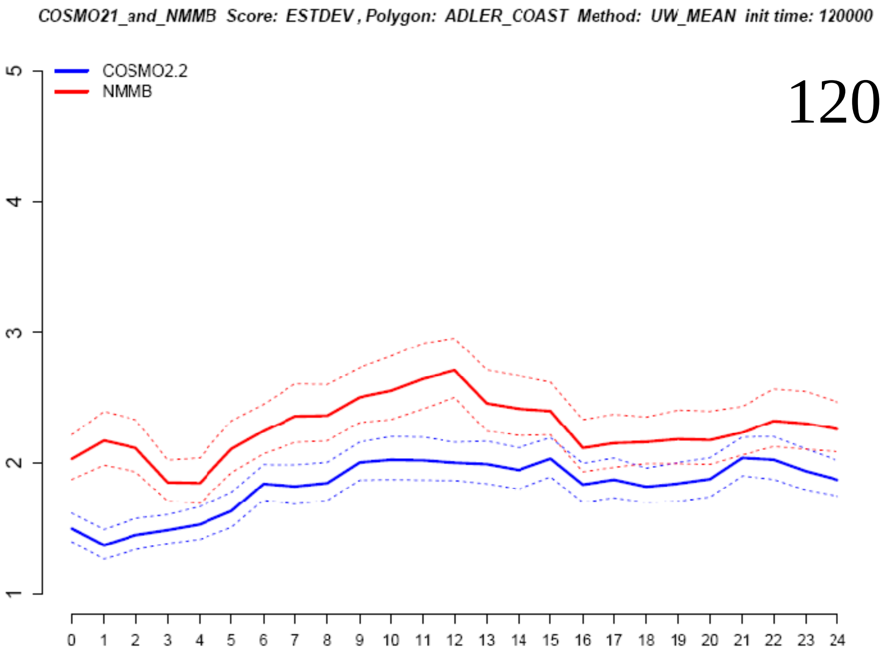
Error standard deviation (bias corrected), ESTDEV=√BCMSE

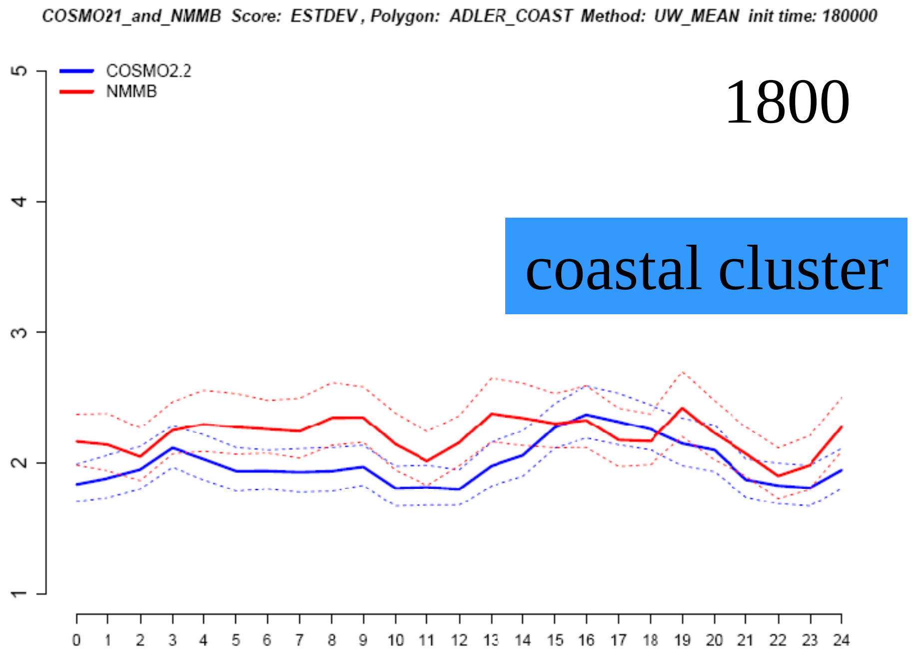


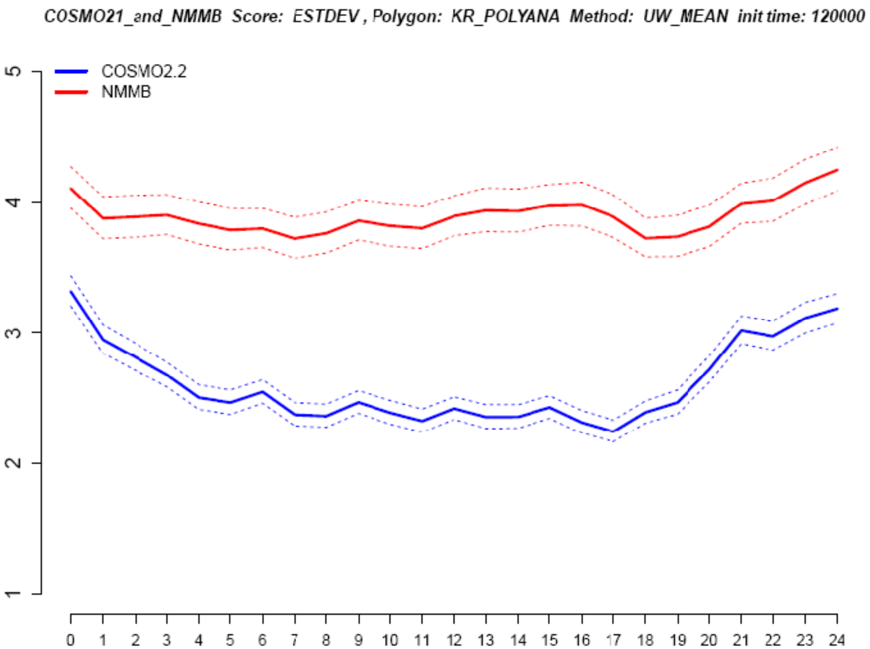


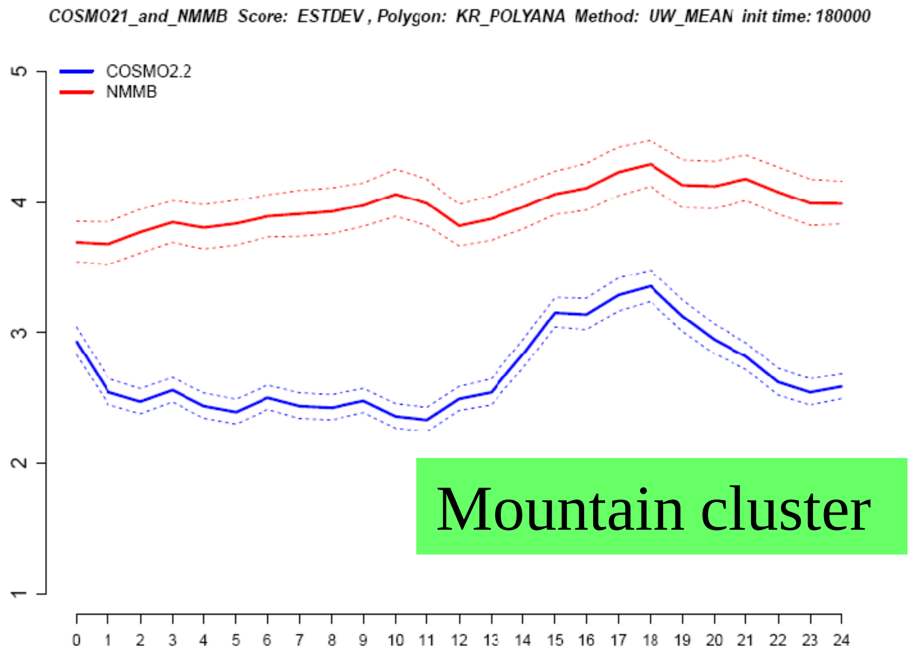





COSMO21_and_NMMB Score: ESTDEV, Polygon: ADLER_COAST Method: UW_MEAN init time: 120000
COSMO2.2
NMMB
1200
1 2 3 4 5
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


COSMO21_and_NMMB Score: ESTDEV, Polygon: ADLER_COAST Method: UW_MEAN init time: 180000
COSMO2.2
NMMB
1800
coastal cluster
1 2 3 4 5
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


COSMO21_and_NMMB Score: ESTDEV, Polygon: KR_POLYANA Method: UW_MEAN init time: 120000
COSMO2.2
NMMB
1 2 3 4 5
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


COSMO21_and_NMMB Score: ESTDEV, Polygon: KR_POLYANA Method: UW_MEAN init time: 180000
COSMO2.2
NMMB
Mountain cluster
1 2 3 4 5
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Types

- Temperature
  mostly cold types: 1, 4, 11, 13, 19
  mostly warm types: 2, 6, 7, 8
- Precipitation
  mostly dry types: 4, 9
  mostly wet types: 16