Model verification and tools

C. Zingerle

ZAMG

Why verify?

The three most important reasons to verify forecasts are:

- to monitor forecast quality - how accurate are the forecasts and are they improving over time?
- to improve forecast quality - the first step toward getting better is discovering what you're doing wrong.
- to compare the quality of different forecast systems - to what extent does one forecast system give better forecasts than another, and in what ways is that system better?

What is the truth?

- The truth data comes from observational data: SYNOPs, raingauges, satellite observations, radar, analysis systems.
- Most of the time we ignore errors in observations / analysis, because the error of the observation / analysis is much smaller than the one expected by the forecasting system.

What makes a forecast good?

- **Consistency** - the degree to which the forecast corresponds to the forecaster's best judgement about the situation, based upon his/her knowledge base
- **Quality** - the degree to which the forecast corresponds to what actually happened
- **Value** - the degree to which the forecast helps a decision maker to realize some incremental economic and/or other benefit

In model verification we are mostly interested in **quality** measures, somtimes in value.

Most emphasized aspects are **accuracy** (agreement between forecast and observation) and **skill** (accuracy of a forecast over some reference).

Different specificity of forecast asks for a number of different basic verification methods

- Dichotomous (yes/no): e.g. occurrence of fog

  visual, dichotomous, probabilistic, spatial, ensemble

- Multi-category: e.g. cold, normal, or warm conditions

  visual, multi-categorical, probabilistic, spatial, ensemble

- Continuous: e.g. maximum temperature

  visual, continuous, probabilistic, spatial, ensemble

- Object- or event-oriented: e.g. cyclone motion and intensity

  visual, dichotomous, multi-categoriy, coninuous, probabilistic, spatial

Categorical or mulit-cateogrical forecasts

Contingency table

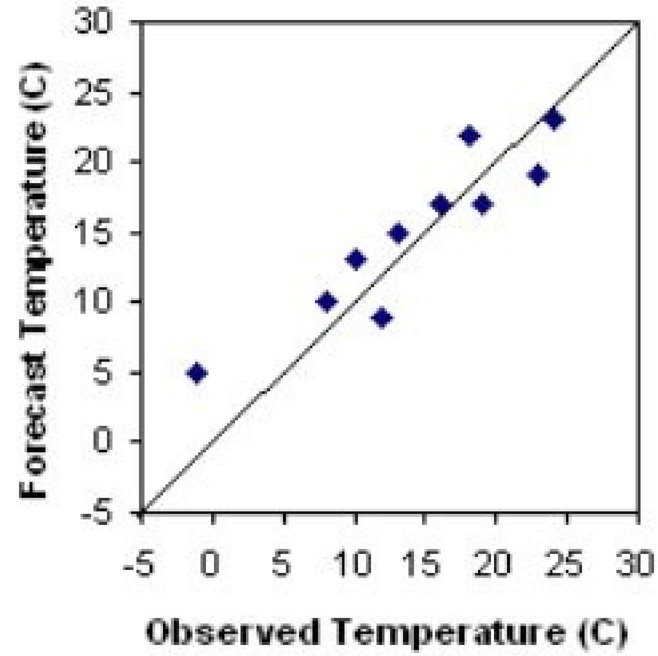
Accuracy, frequency bias, POD, FAR, POFD, SR, TS, ETS, OR, HK, HSS ...

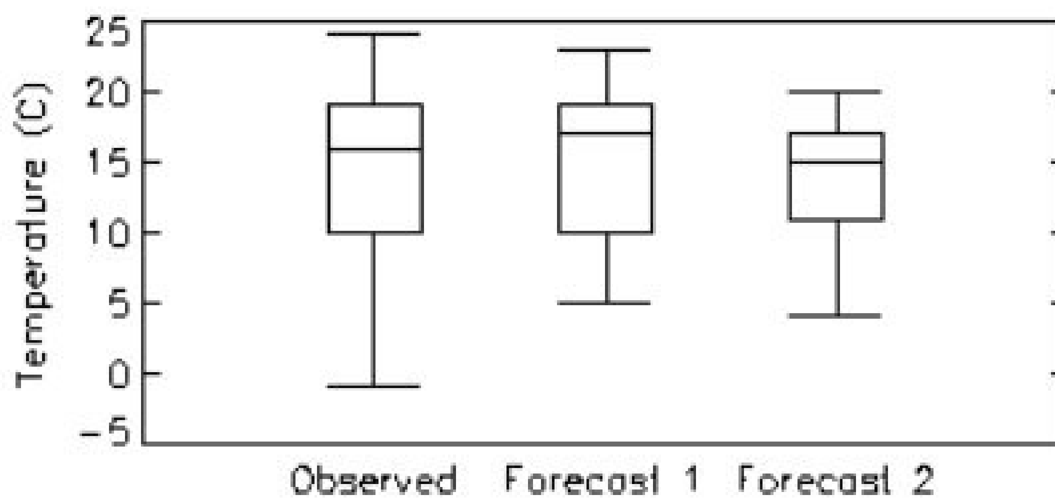
		Observed yes	Observed no	Total
Forecast	yes	*hits*	*false alarms*	*forecast yes*
	no	*misses*	*correct negatives*	*forecast no*
Total		*observed yes*	*observed no*	*total*

Conintuous forecasts

Scatterplot, boxplot

ME, BIAS, MAE, RMSE, MSE, LEPS, Skill score, correlation coefficient, anomally correlation,





Probability forecasts

Reliability diagram,

ROC diagram,

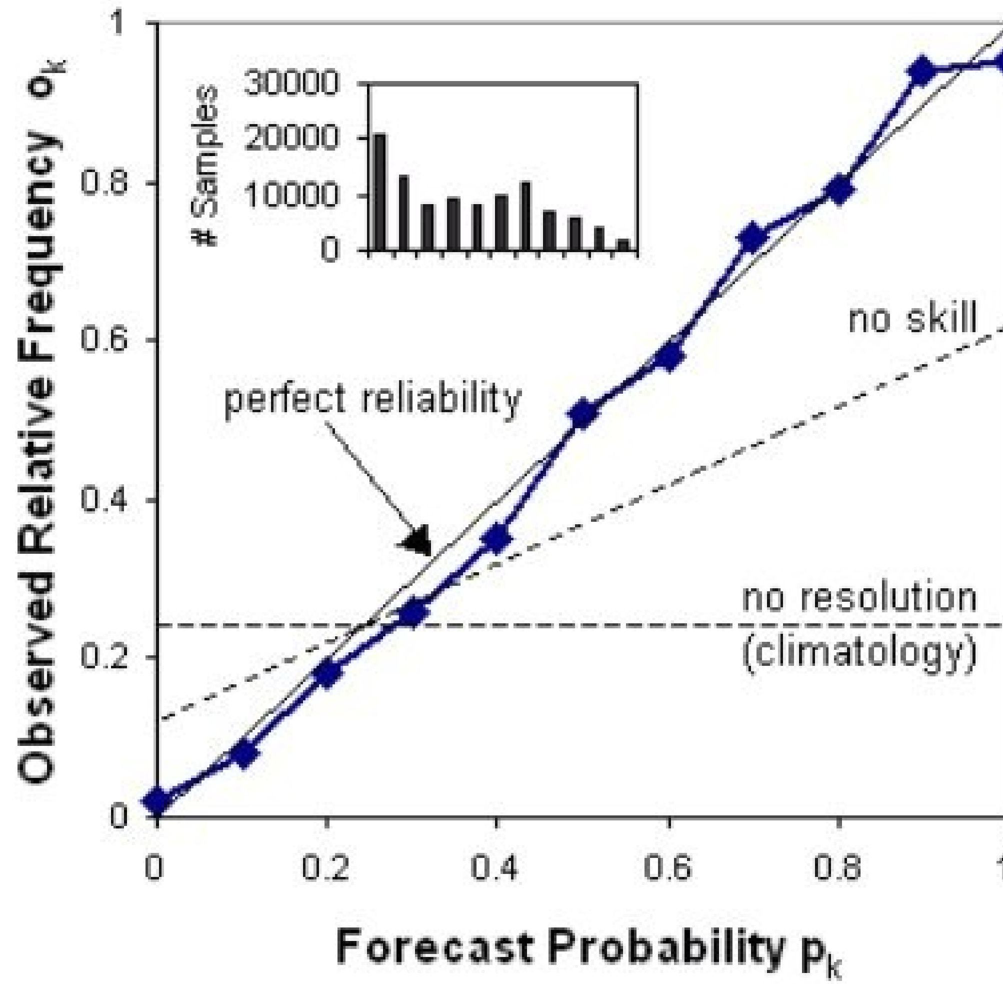
Brier score

Brier skill score

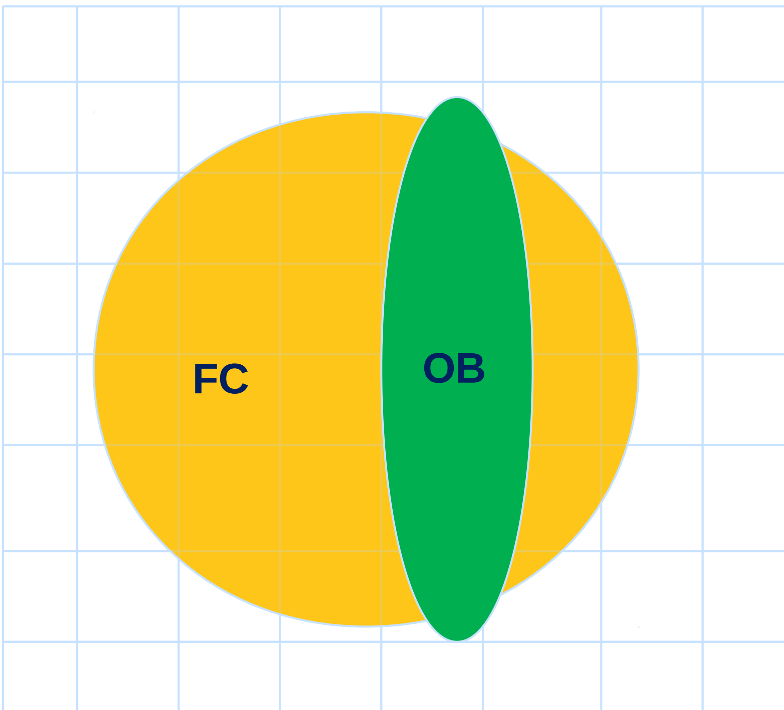
Ranked probability score

Ranked probability skill score

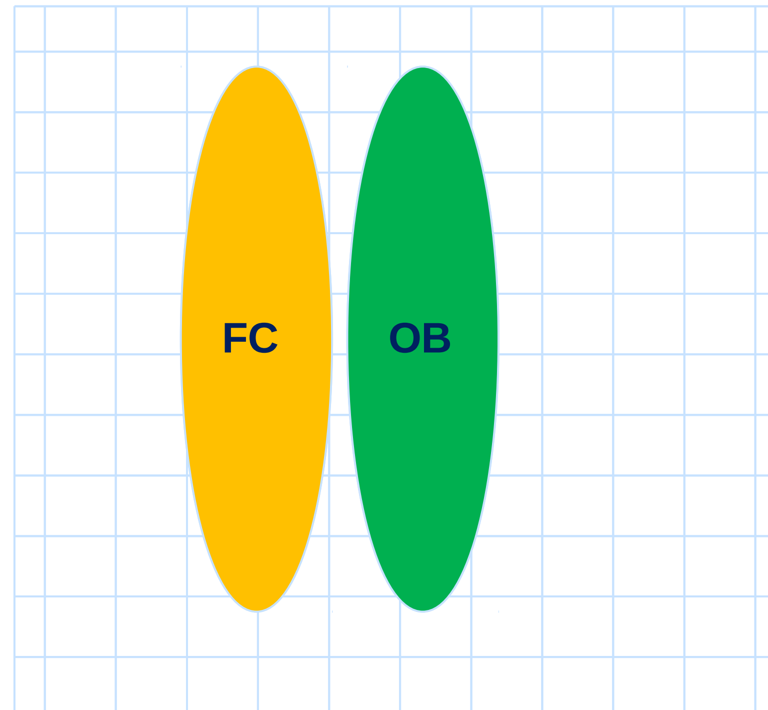
Relative value



Spatial forecasts

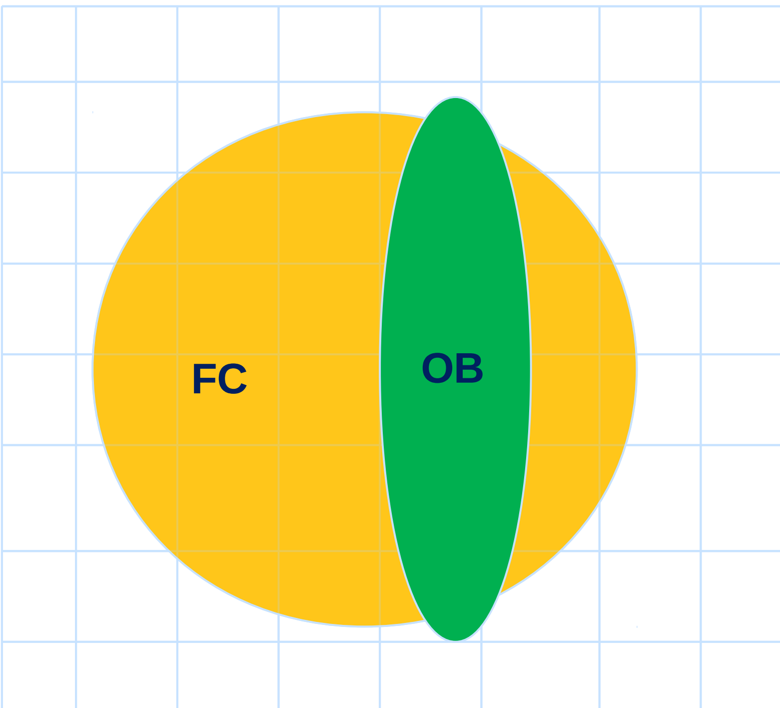


BIAS < <
RMSE <
FAR = around 0
POD = around 1
...

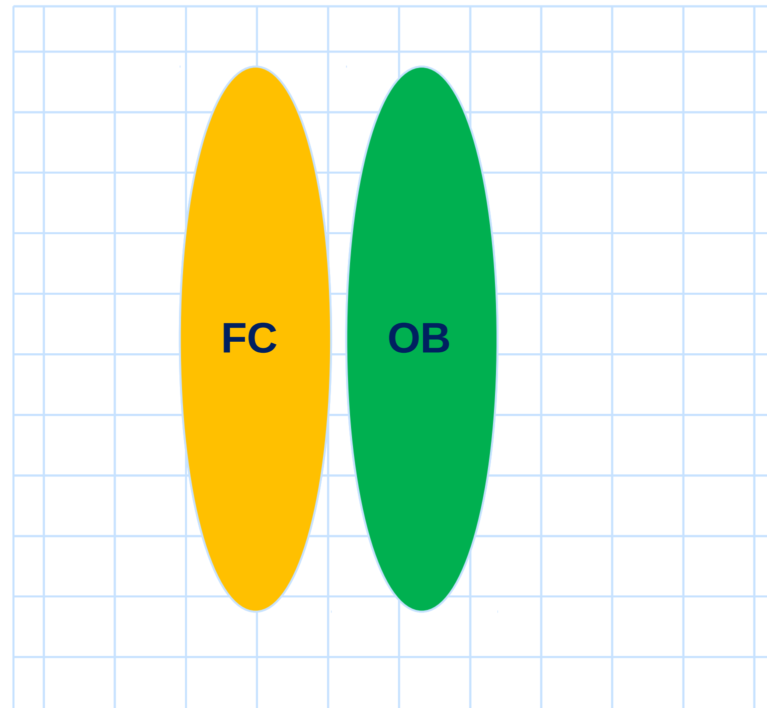


BIAS > >
RMSE > >
FAR = 1
POD = 0
...

Spatial forecasts



BIAS < <
RMSE <
FAR = around 0
POD = around 1
...



BIAS > >
RMSE > >
FAR = 1
POD = 0
...

Spatial forecasts

Double Penalty Problem: In a grid-point by grid-point verification coarse scale models scores often better than high resolution models:

Even in close fails of high resolution forecast we get worse:

BIAS goes up
POD goes down
FAR rise
RMSE rise
...

Penalized once for missing an observation and a second time for giving a false alarm

Spatial forecasts

SAL

CRA

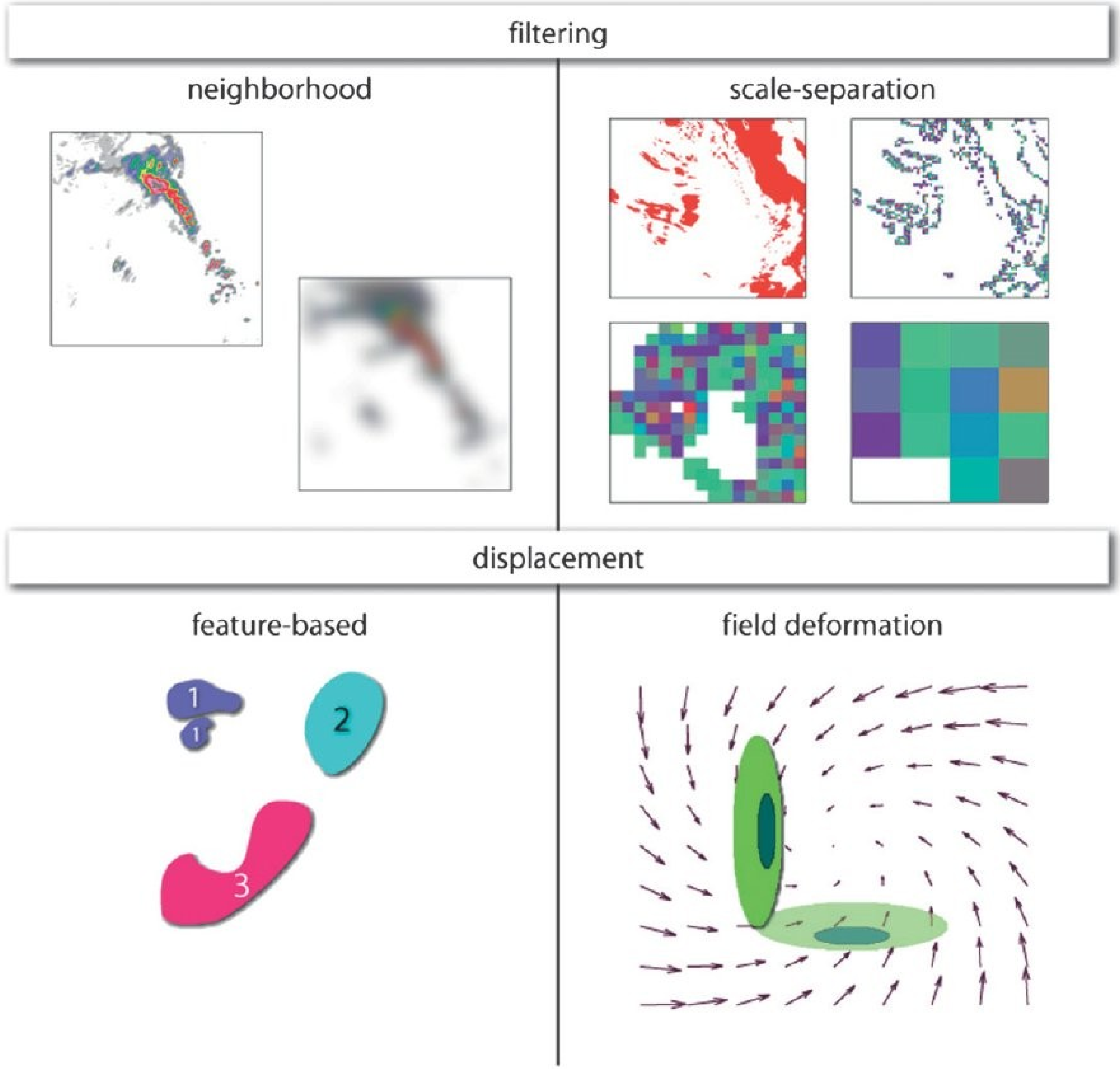
Fuzzy

FSS

Intensity scale

Morphing

...



Tools for model verification in Aladin

Aladin performance monitoring tool (Ljubljana)

Montoring of Aladin implementations in the member NMS using standard verification methods (SYNOP and TEMP, continuous and mulit-categorical variables)

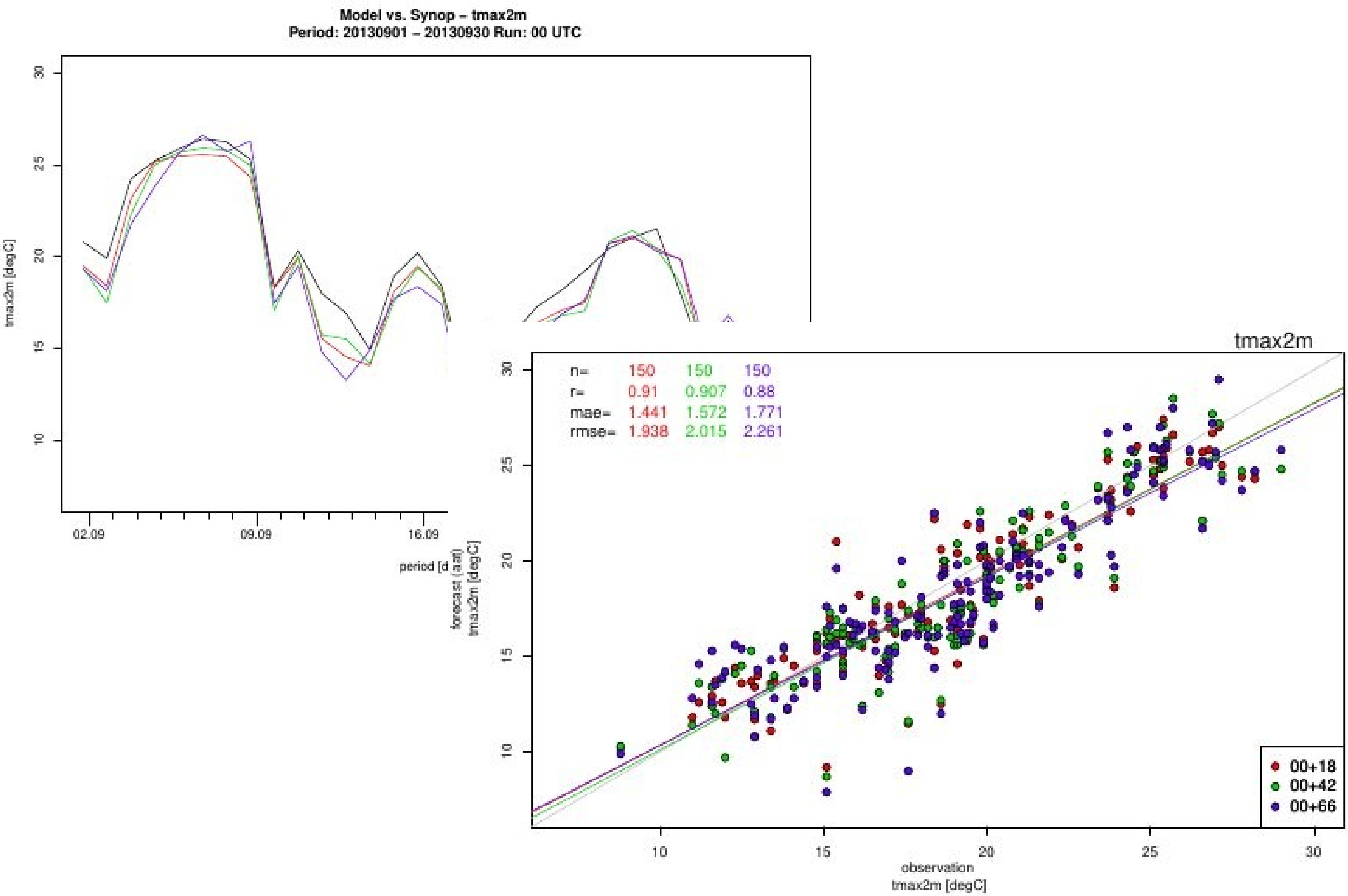
- Centralized system at Ljubljana, each member is sending his model data to the database

HARP

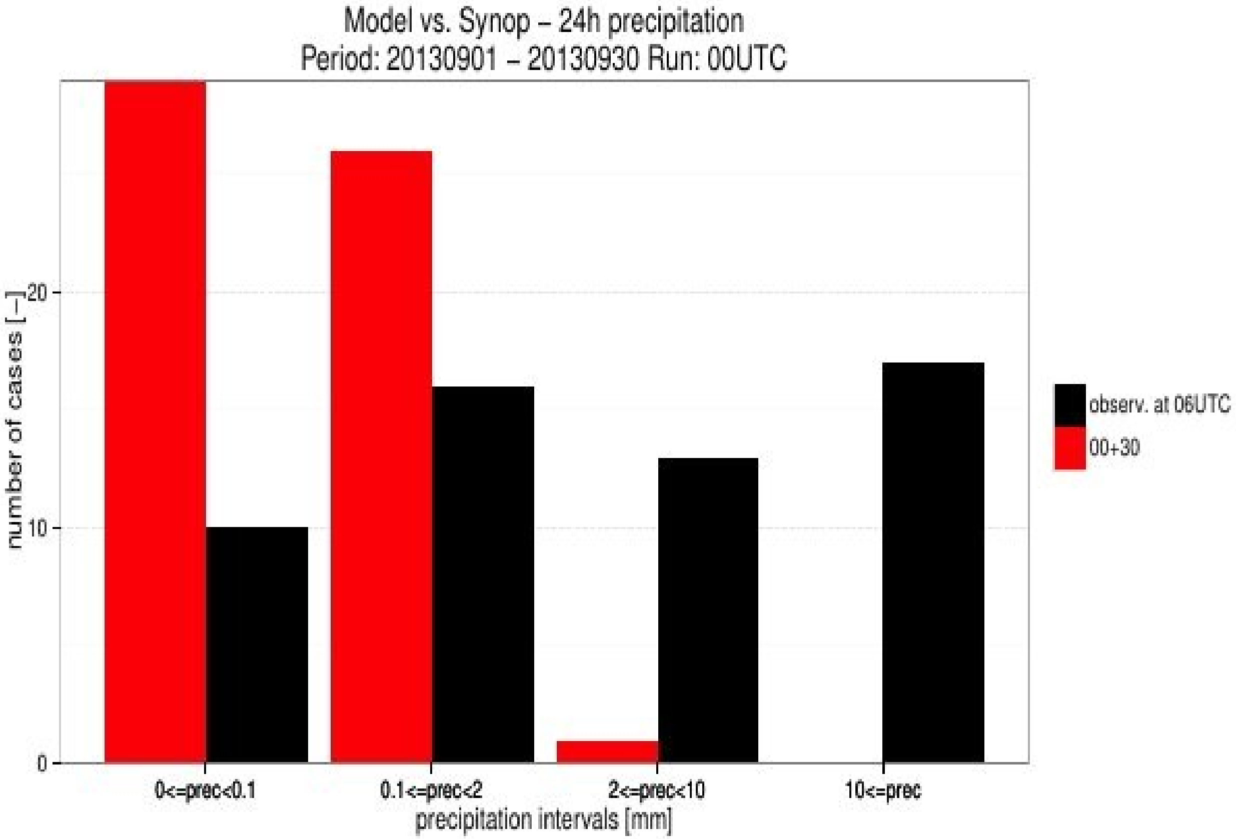
Hirlam – Aladin R-package: Adaption and development for probabilistic and spatial verification

- Local system: Toolbox to be adapted locally at each NMS

Aladin performance monitoring tool



Aladin performance monitoring tool

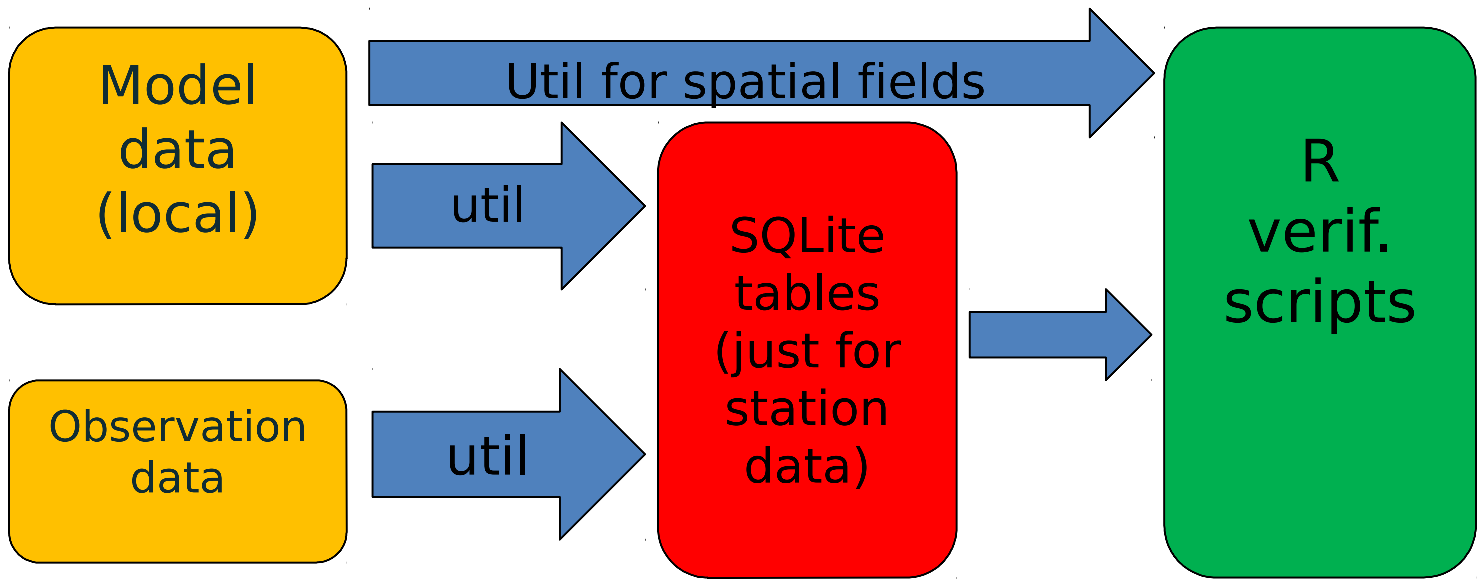


ber of cases in particular ranges of precipitation [mm/24h]

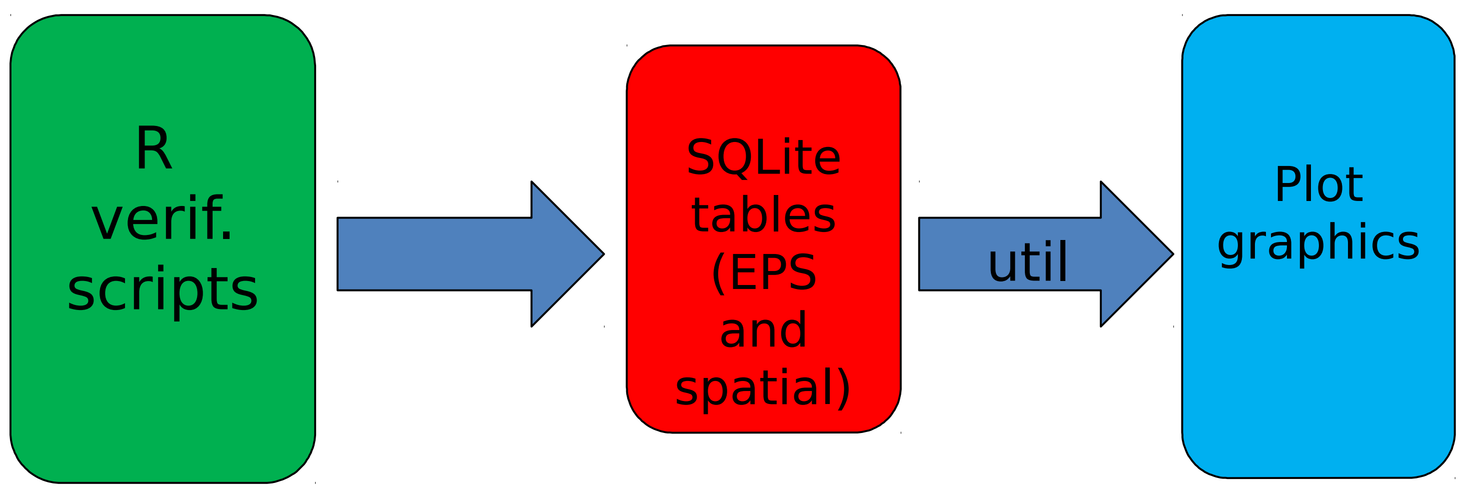
mod\obs	0<=prec<0.1	0.1<=prec<2	2<=prec<10	10<=prec	sum fc
0<=prec<0.1	13	24	18	9	64
0.1<=prec<2	7	10	10	22	49
2<=prec<10	0	0	0	1	1
10<=prec	0	0	0	0	0
sum obs	20	34	28	32	114

HARP Hirlam – Aladin R-package

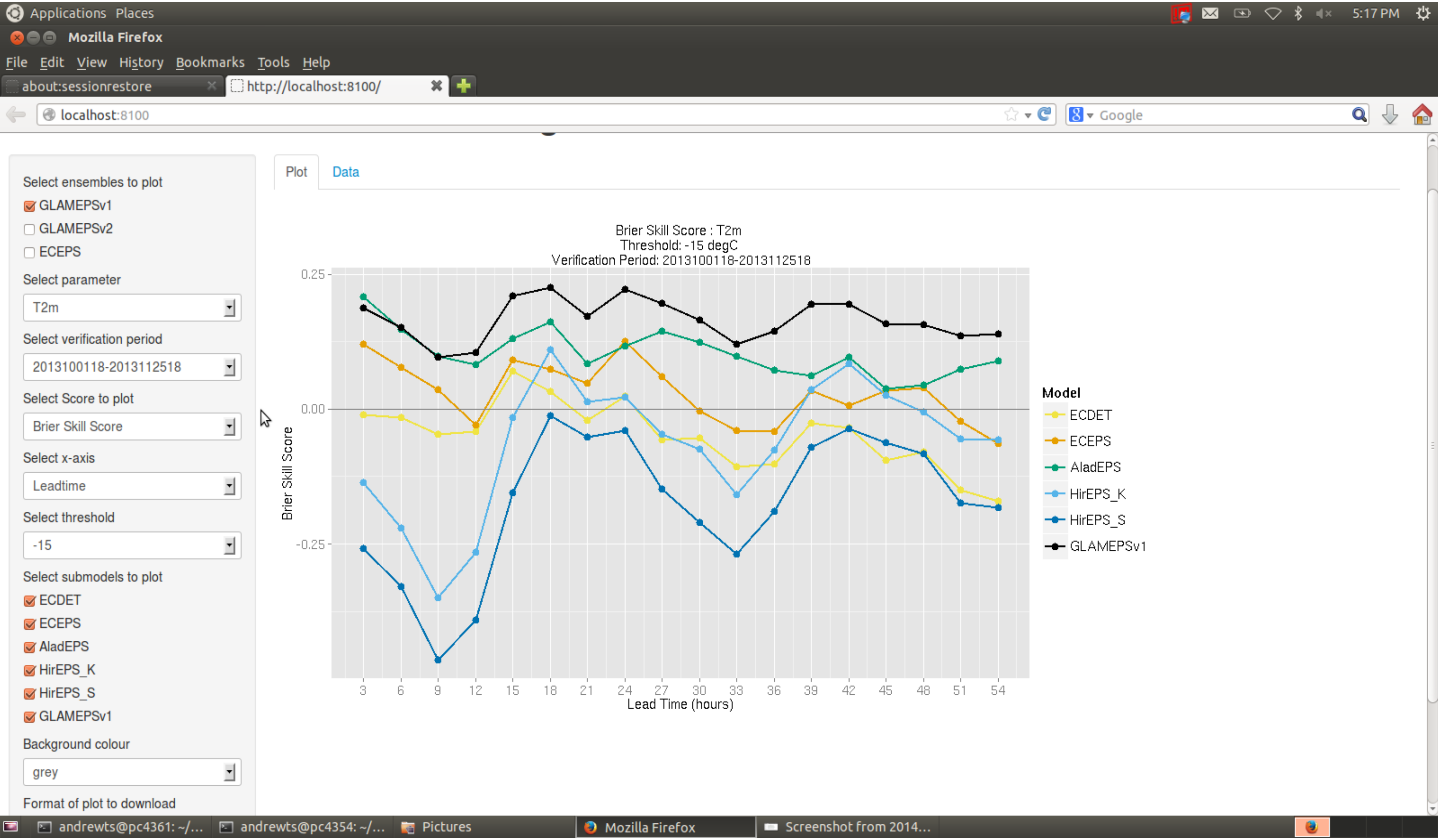
- Read from SQLite data files or from spatial fields
- Calculate scores
- Write to SQlite results files



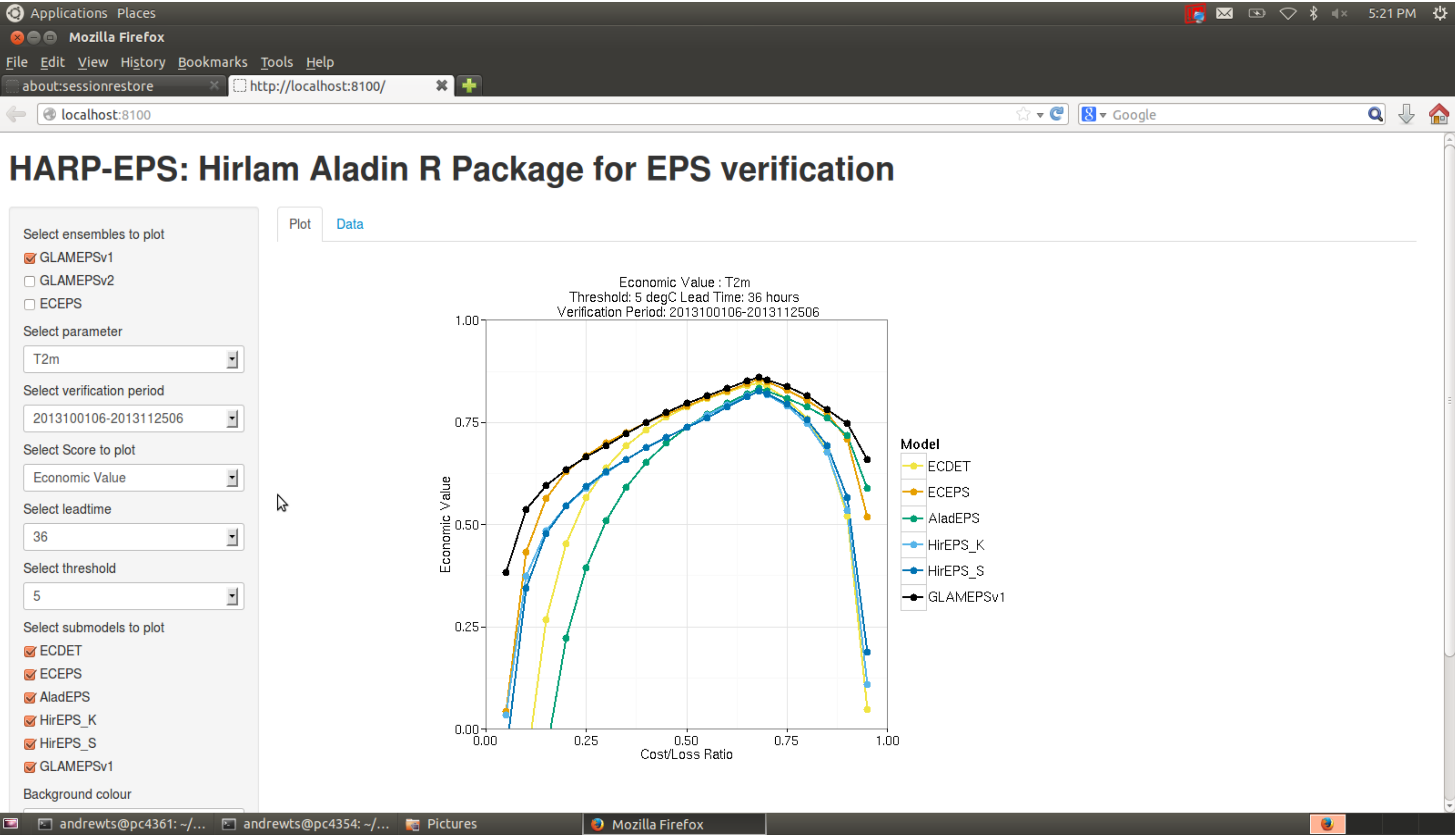
- Read from SQLite results files
- Plot



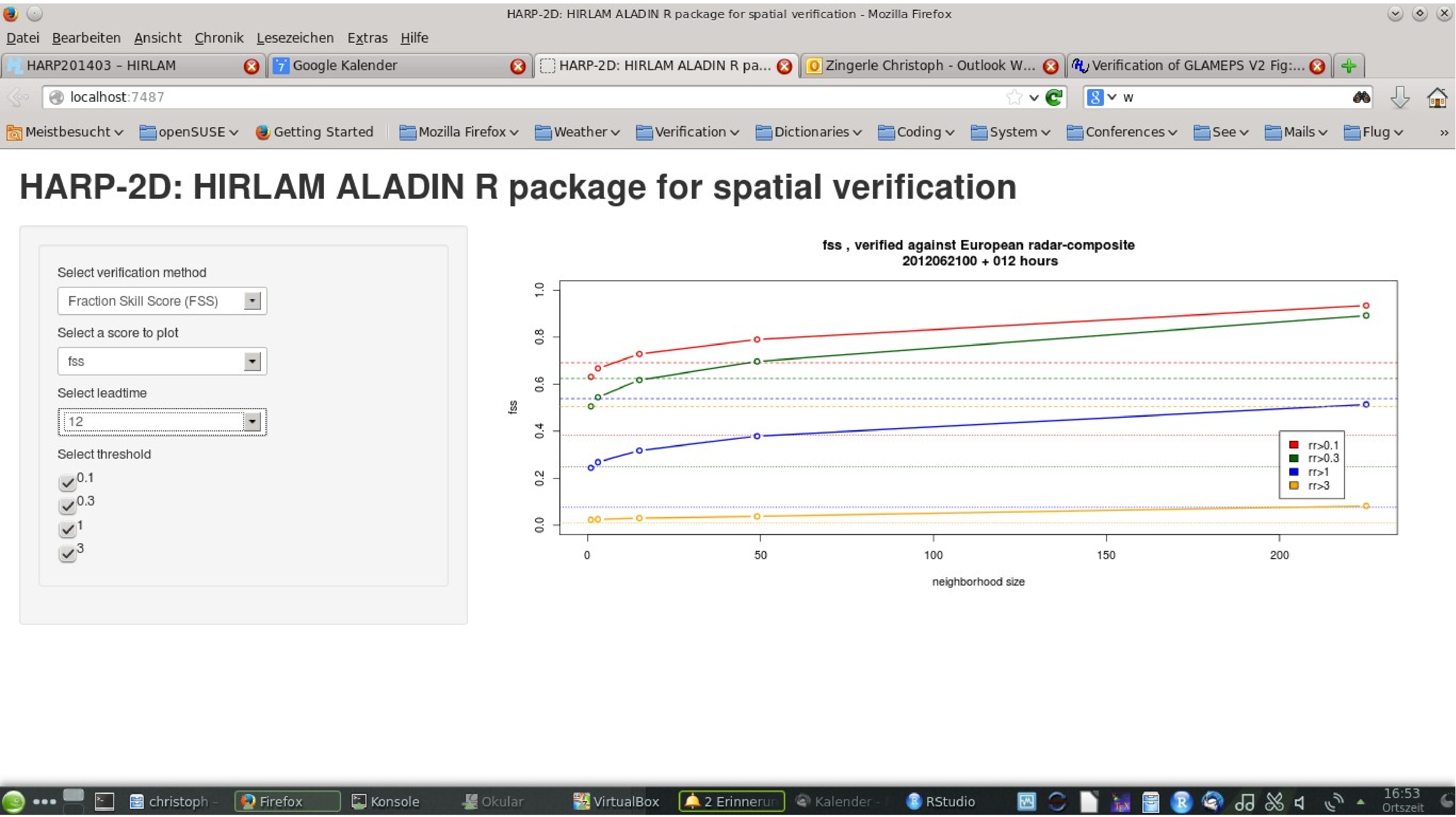
HARP Hirlam: EPS verification



HARP Hirlam: EPS verification



HARP Hirlam: spatial verification



APMT:

- Point verification
- Centralized
- Monthly report (pdf) for each country
- Beeing currently re-fitted

HARP:

- EPS & spatial verification
- Locally
- Operational visualization locally (visualization utility)
- Still under development