Report on stay at ZAMG

Viena, Austria

13.05.2013 – 05.07.2013

Simona Tascu

NMA, Romania

Supervised by:

Yong Wang and Theresa Gorgas

Introduction

The goal of the present stay was to develop and optimize the verification package, specific for the ensemble prediction systems, available in RC-LACE consortium. The EPS outputs represent a large amount of data. Therefore, it is important to create a more flexible, easy to use, efficient and reliable verification package software.

The new verification package, which is a common framework for surface and upper levels parameters, contains two important parts: one of them is in FORTRAN language and the other one in shell and PERL programming languages. The main idea was to calculate the scores averaged over one period in FORTRAN. The section 1 summarizes the use of FORTRAN programs and subroutines. It also contains the description of the PERL scripts/programs that control and activate variables and keys used in FORTRAN computation, the description of a new score, SPREAD-SKILL relationship and future plans. Section 2 contains the experimental results of the stay.

1. The verification procedure structure

The structure of the new verification package containing shell and perl scripts, fortran programs is presented in figure 1, where the new programs/scripts are marked by red color.

1.1. Fortran routines

The most important part of the verification package is in FORTRAN. After daily scores computation (saved in specific ASCII files), the average over a given period is calculated and saved. The fortran procedures permit to easy switch from surface parameters verification to upper levels one.

In order to compute the score average over one period, new allocatable variables were introduced in fortran routines. The changes with respect to the old verification package are listed below:

- The old Module_Definevars.f90 was renamed to Module_Definevars_Namelist.f90 (specific for the namelist) and Module_Definevars.f90 contains these new variables.
- The old LAEF_Verification_Surface.f90 program is split into: Module_Definevars.f90, LAEF_Verification_API.f90, Rd_AllData.f90. These programs are now adapted for surface and upper levels parameters.
- All the score subroutines are called without arguments. Instead, MODULE_DEFINEVARS and MODULE_DEFINEVARS_NAMELIST" are used to access and control the variables.

News routines in the verification package:

- A general subroutine for reading grib data was developed using gribapi package (from ECMWF). This subroutine (Rd_Gribs_API.f90) is common for surface and altitude parameters, for different type of precipitation, for orography, etc.
- Throughout the OpenAscii.f90 subroutine, all the ASCII files are opened at the beginning, after verification dates, which are computed using the Calc_New_Date.f90 subroutine.
- Each day of the period, for every score, is summed using Calc_SumScores.f90 subroutine.
- The average computations of the scores, for one forecast range over all verification days, is realized using the Calc_AllScores.f90 subroutine

- The simplicity of the new verification package is given by the use of the variables control in all routine. The variables allocation/deallocation is carried out through inside Rd_Allocate.f90 subroutine.
- Actually, all the score subroutines are taking into account the missing model data (one member or all the members from one day).

1.2. Perl and Shell scripts

The verification package can be run using a shell script "MasterVerification.job" which offers the possibility to set: the name of EPS experiments, the number of members, the parameter names abbreviations, the directory paths (where all the data and executable file are located), beginning and ending dates, forecast ranges, number of points in latitude and longitude or the number of surface stations, type of parameter level (surface or upper level), keys to compute and/or plots the scores.

At the end of "MasterVerification.job", the perl script "DoConfSettings.job.pl" is called in order to set automatically different keys or variables based on "Settings.pm" file. The "Settings.pm" file contains settings regarding the code and name parameters, thresholds, the usage of the orography correction and climatology, etc. Also, in this file are included different keys for score compuatation and plotting (for instance CALCULATE_BIAS_RMSE or PLOT_BIAS). The DoNamelist.job script generates the namelist necessary to transfer variables from shell and perl scripts to fortran routines.

The flexibility of the new package (due to perl scripts) allows to switch, in a very simple way, from surface score computation to upper levels parameters and vice versa.

Thanks to Martin Bellus (SMHU), almost all parts of shell scripts were adapted to perl programming language.

1.3. SPREAD-SKILL relationship

SPREAD-SKILL relationship is based on the ensemble spread and root mean square error (RMSE) of the ensemble mean. For this purpose the spread and RMSE of the ensemble are computed for each forecast range, parameter, gridpoint over the whole period. For each gridpoint, one SPREAD-RMSE pair is generated. The applied method for this computation follows the Wang and Bishop, 2003 method. First, the spread values are ordered and divided in 10 subgroups of equal population (keeping in mind the RMSE associated values). For each subgroups the average is computed and saved. Figure 2 shows two examples of SPREAD-SKILL: in the left side the SPREAD and RMSE are plotted for these 10 subgroups and in the right side all the grid point pairs are plotted.

1.4. Future plans

- To compute the average over one period using daily ASCII files. In this moment it is possible only from grib data.
- To transpose the plotting shell scripts into PERL scripts.
- To adapt this package for more deterministic scores.

The new verification package proved that it is more flexible and much faster (6 times more) than the old one.

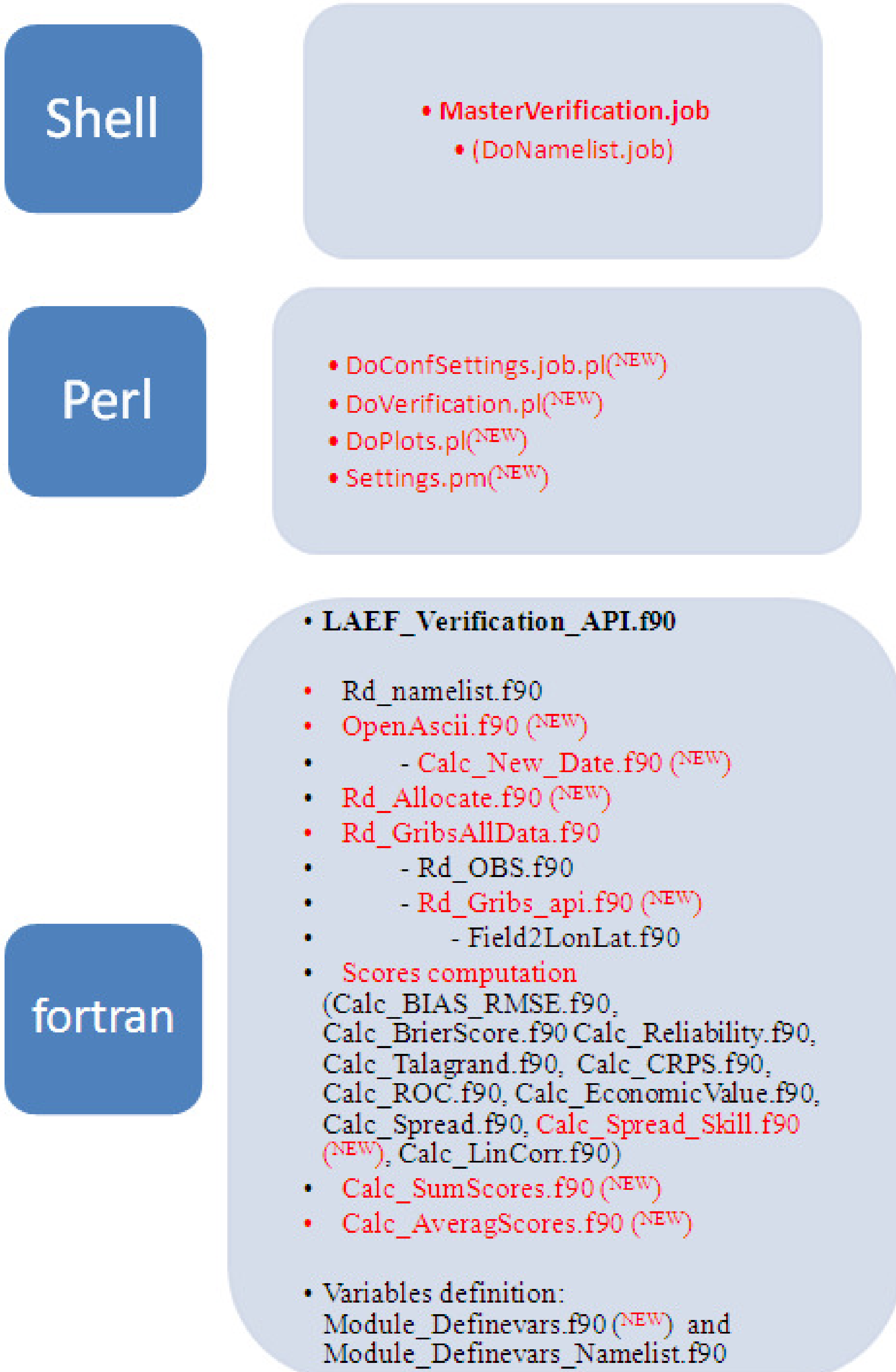


Fig. 1 Verification package structure

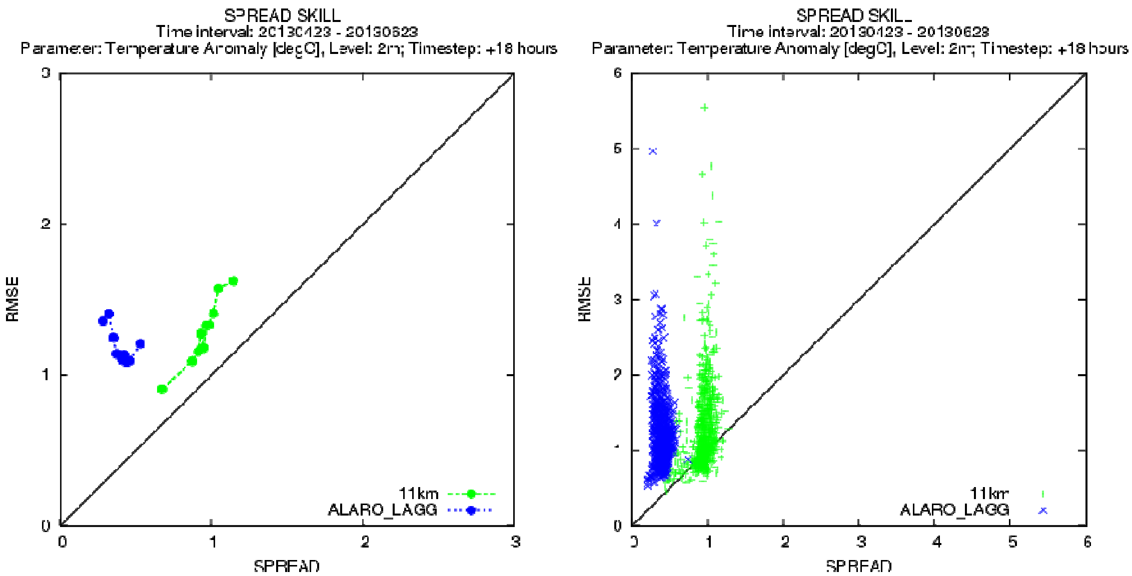


Fig. 2 SPREAD-SKILL: T2M, 23.04.2013-23.06.2013, 18 UTC for lagged ensemble (labelled ALARO_LAGG) and LAEF ensemble (labelled 11km): for 10 subgroups (left) and for all the grid points over the domain (right).

2. Experimental results
3.

Two approaches have been studied in order to verify the forecasts provided by the deterministic model and bye the ensemble system

A) The verification of a **time-lagged ensemble system in comparison with the new LAEF system**

The time-lagged ensemble system comprises the deterministic ALARO –AUSTRIA forecasts from different runs valid at the same moment, as described in table 1 (marked by green color). There are 5 members /day the time-lagged system and 17 for the new LAEF system (11 km horizontal resolution).

one day before **member5**	12 UTC	00	06	12	18	24	30	36	42	48	54	60	66	72
one day before **member4**	18 UTC		00	06	12	18	24	30	36	42	48	54	60	66
current day **member3**	00 UTC			00	06	12	18	24	30	36	42	48	54	60
current day **member2**	06 UTC				00	06	12	18	24	30	36	42	48	54
current day **member1**	12 UTC					00	06	12	18	24	30	36	42	48

Table 1: time-lagged ALARO-AUSTRIA ensemble

The validation of these two systems was realized for two months period (23 May – 23 June 2013) over the new LAEF system domain (figure 3). Few scores are shown and shortly commented below. The mean see level pressure (MSLP), temperature at 2m (T2M), wind speed at 10m (W10M), 12 hours cumulated precipitation have been verified, for 1219 surface stations. For skill score, the ALADIN-AUSTRRIA model was used as reference.

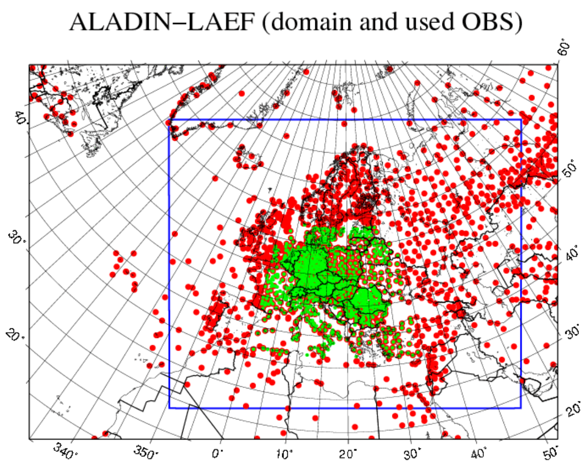


Fig. 3 New LAEF domain

Large spread indicates less predictable event, which is more difficult to be forecast. For MSLP, W10M, T2M, the spread (fig.4) is lager for LAEF ensemble (11km) those of the lagged ensemble (ALARO_LAGGED). Almost for all forecast ranges, excepting 00 range, the LAEF ensemble (11km) performs better than the time-lagged system. For 00 UTC forecast range the MSLP and T2M bias has better values for lagged ensemble which seems to be due to the assimilation cycle; it is worth to underline that the lagged ensemble contains only one analysis. The better new LAEF system performance is shown as well by the Continuous Ranked Probability Score, Continuous Ranked Probability Skill Score and Percentage of Outliers (figs 5-7).

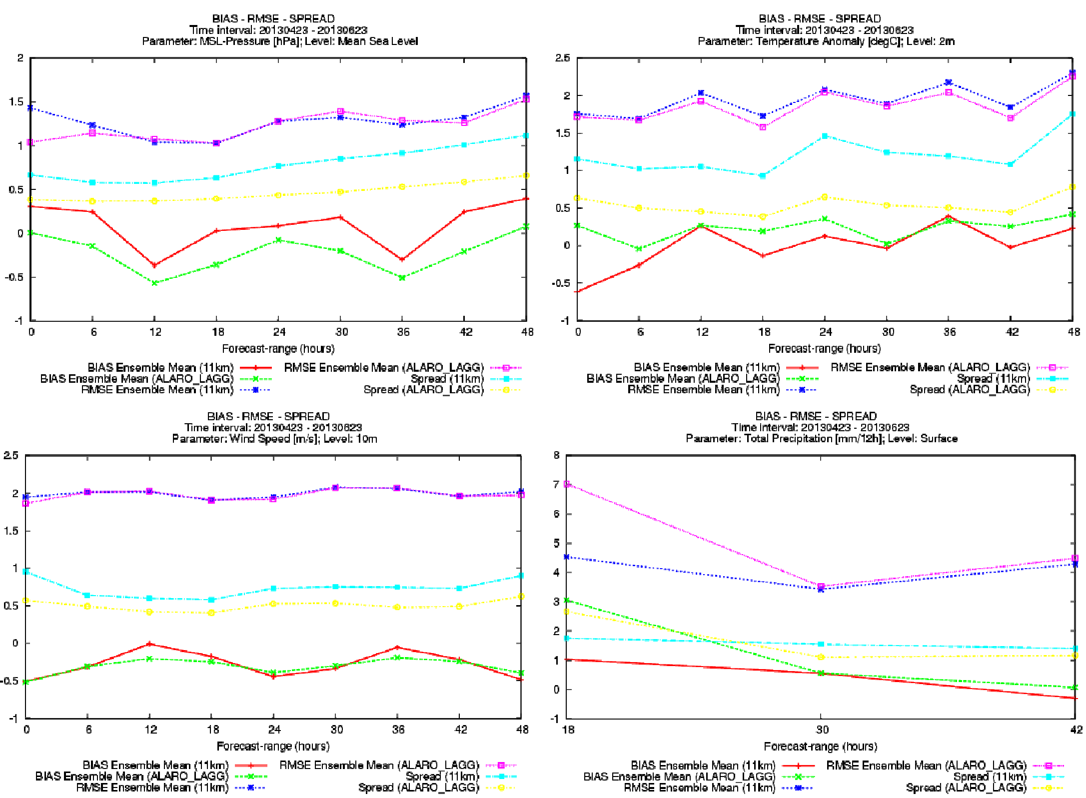


Fig. 3 Bias, RMSE and SPREAD: Mean Sea Level Pressure, temperature at 2m, wind speed at 10m and precipitation cumulated in 12hours for lagged ensemble (labelled ALARO_LAGG) and LAEF ensemble (labelled 11km) for 23.04.2013 – 23.06.2013 period.

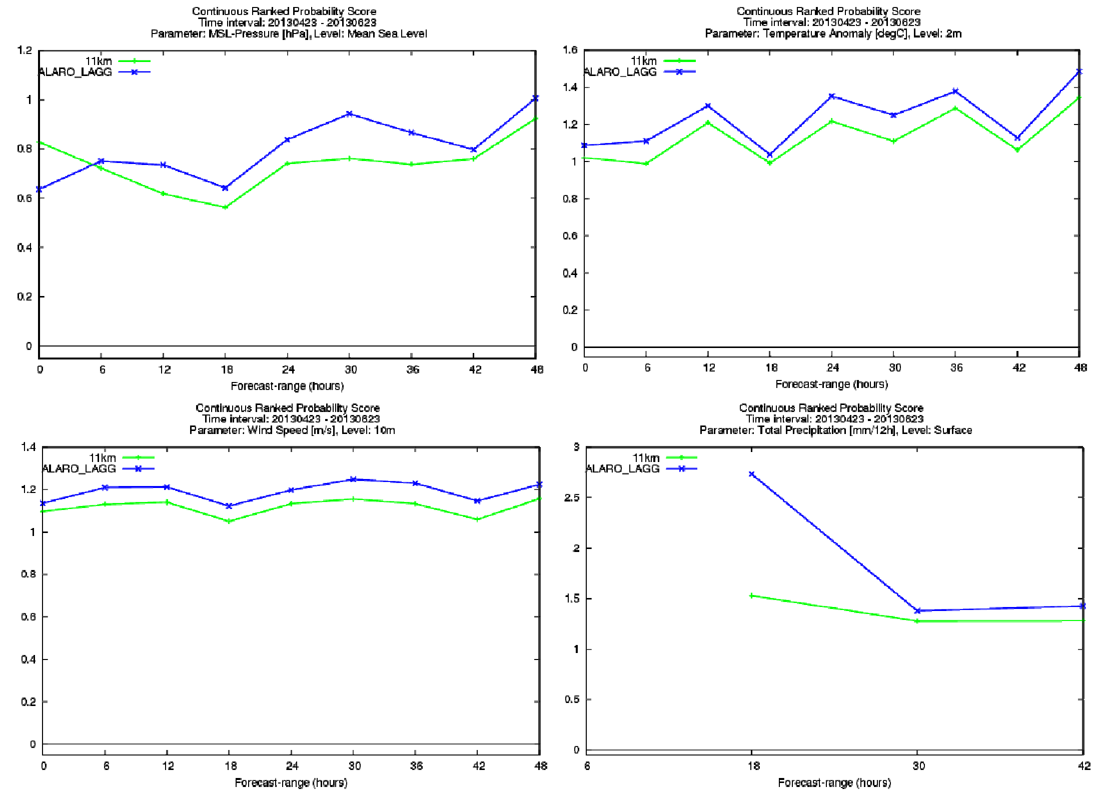


Fig. 4 Continuous Ranked Probability Score: Mean Sea Level Pressure, temperature at 2m, wind speed at 10m and precipitation cumulated in 12hours for lagged ensemble (labelled ALARO_LAGG) and LAEF ensemble (labelled 11km) for 23.04.2013 – 23.06.2013 period.

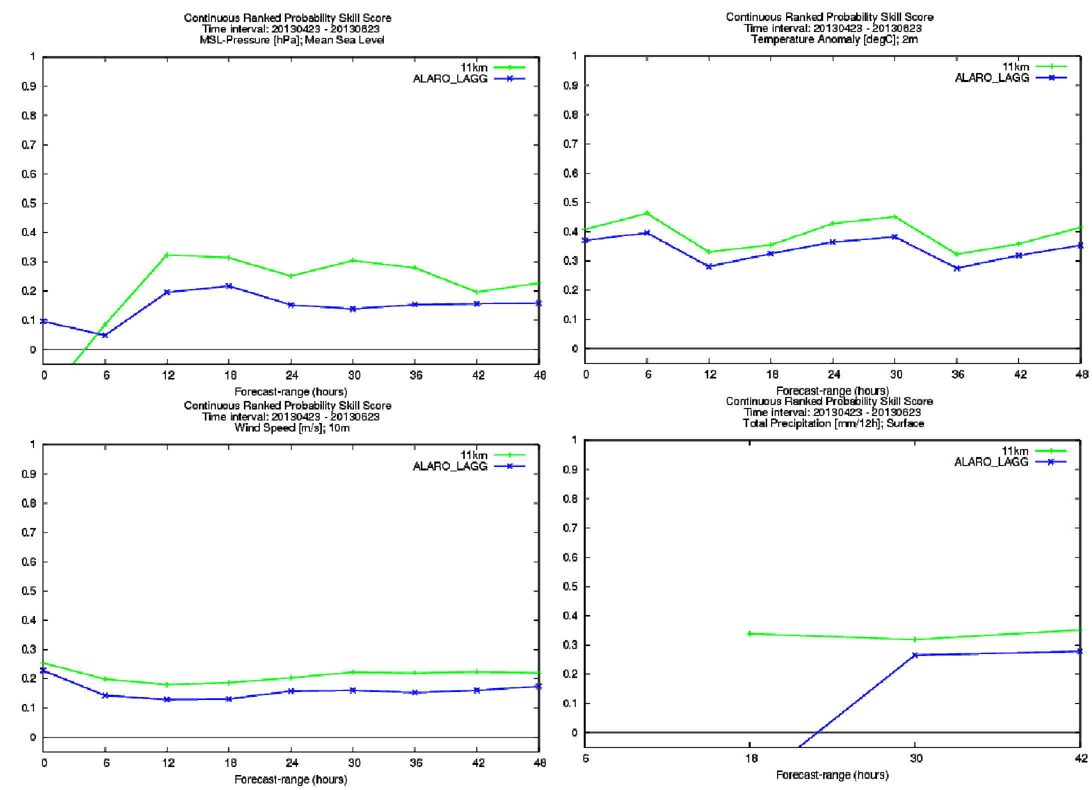


Fig. 5 Continuous Ranked Probability Skill Score: Mean Sea Level Pressure, temperature at 2m, wind speed at 10m and precipitation cumulated in 12hours for lagged ensemble (labelled ALARO_LAGG) and LAEF ensemble (labelled 11km) for 23.04.2013 – 23.06.2013 period.

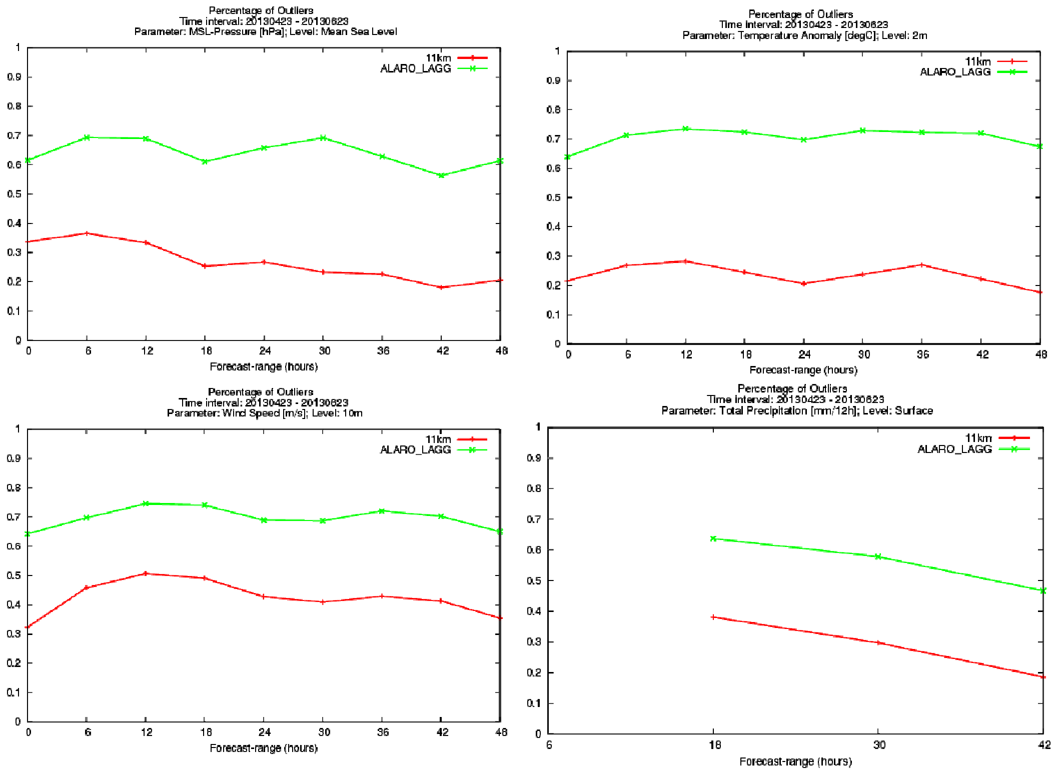


Fig. 6 Percentage of Outliers: Mean Sea Level Pressure, temperature at 2m, wind speed at 10m and precipitation cumulated in 12hours for lagged ensemble ensemble (labelled ALARO_LAGG) and LAEF ensemble (labelled 11km) for 23.04.2013 – 23.06.2013 period.

B) The verification of ALARO -AUSTRIA forecast in comparison with the mean&median&control forecast of the new LAEF system

The mean and the median of the new LAEF system (16 members + 1 control forecast) are treated as solution of a deterministic model. Hit Rate, False Rate Alarme, Equitable Threat Score, Accuracy score, Hanssen and Kuipers discriminant, Threat Score and Correlation Coeficient were computed for T2M, W10M and 12 hours cumulated precipitation, for the same period like in the first experiment (23 May – 23 June 2013) and over the same domain.

In the next figures, the deterministic model ALARO-AUSTRIA is labelled with AUST, the mean of LAEF is labelled with LAEF_MEAN_11km, the median with LAEF_MEDIAN_11km and the control forecast with CONTROL.

The T2m bias evolution of the deterministic solutions is almost the same as ensemble scores (Figure 7). Again for 00 forecast range the bias value for the mean and median of the ensemble is worse than those of ALARO- AUSTRIA deterministic model. Starting from 12 UTC forecast range, the ensemble solutions (mean, median, control) perform better than the solution of ALARO-AUSTRIA. For temperature at 2m and wind at 10, the mean and the median converge more to the same solution. For 12 hours cumulated precipitation, the best solution is the control forecast. The accuracy score for T2m and W10m (figure 8) shows almost the same values for all solutions while for precipitation the best percentage of the correct forecasts is given by the ALARO-AUSTRIA deterministic model. The same results are obtained for correlation coefficient (figure 9) and equitable threat score (figure 10) which is sensitive to the hits.

The differences between the different solutions are more relevant for cumulated precipitation. The hit rate (figure 11) shows that the observed rain events were more correctly predicted by the mean and median of the ensemble for the 30, 42, 54 hours forecast ranges and by ALARO-AUSTRIA for 18 hours forecast range. False rate alarm (figure 12) indicates a better performance for ALARO – AUSTRIA deterministic model.

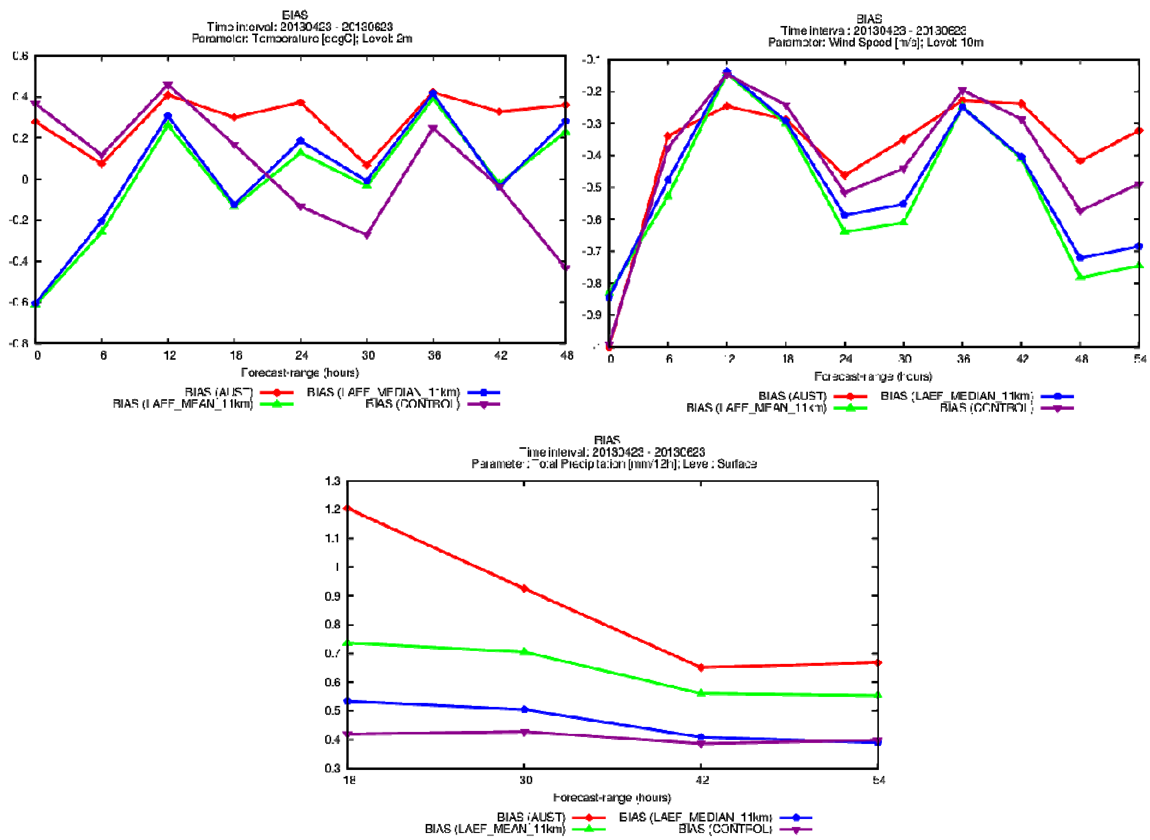
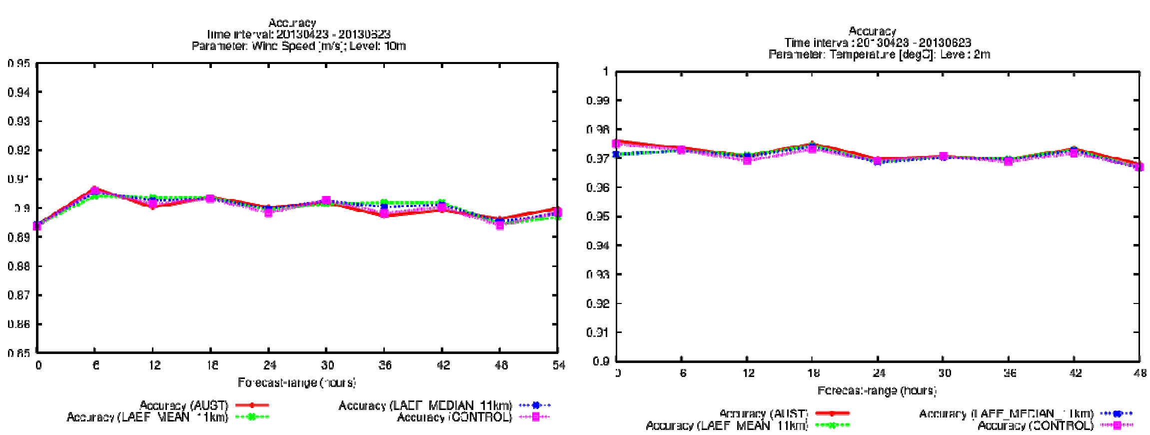


Fig. 7 BIAS: T2M, W10M and precipitation cumulated in 12hours for ALARO-AUSTRIA (AUST), the mean (LAEF_MEAN_11km), the median (LAEF_MEDIAN_11km) of the ensemble and control forecast (CONTROL) for 23.04.2013 – 23.06.2013 period.



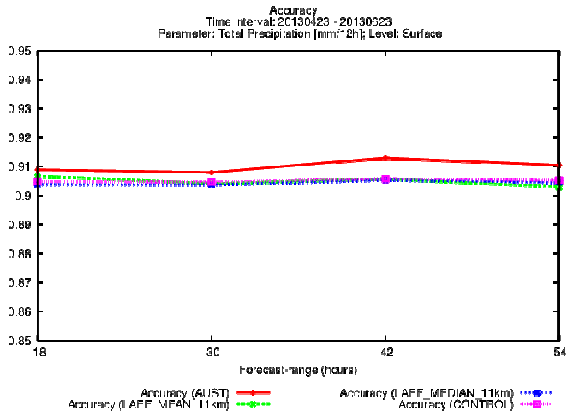


Fig. 8 Accuracy: T2M, W10M and precipitation cumulated in 12hours for ALARO-AUSTRIA (AUST), the mean (LAEF_MEAN_11km), the median (LAEF_MEDIAN_11km) of the ensemble and control forecast (CONTROL) for 23.04.2013 – 23.06.2013 period.

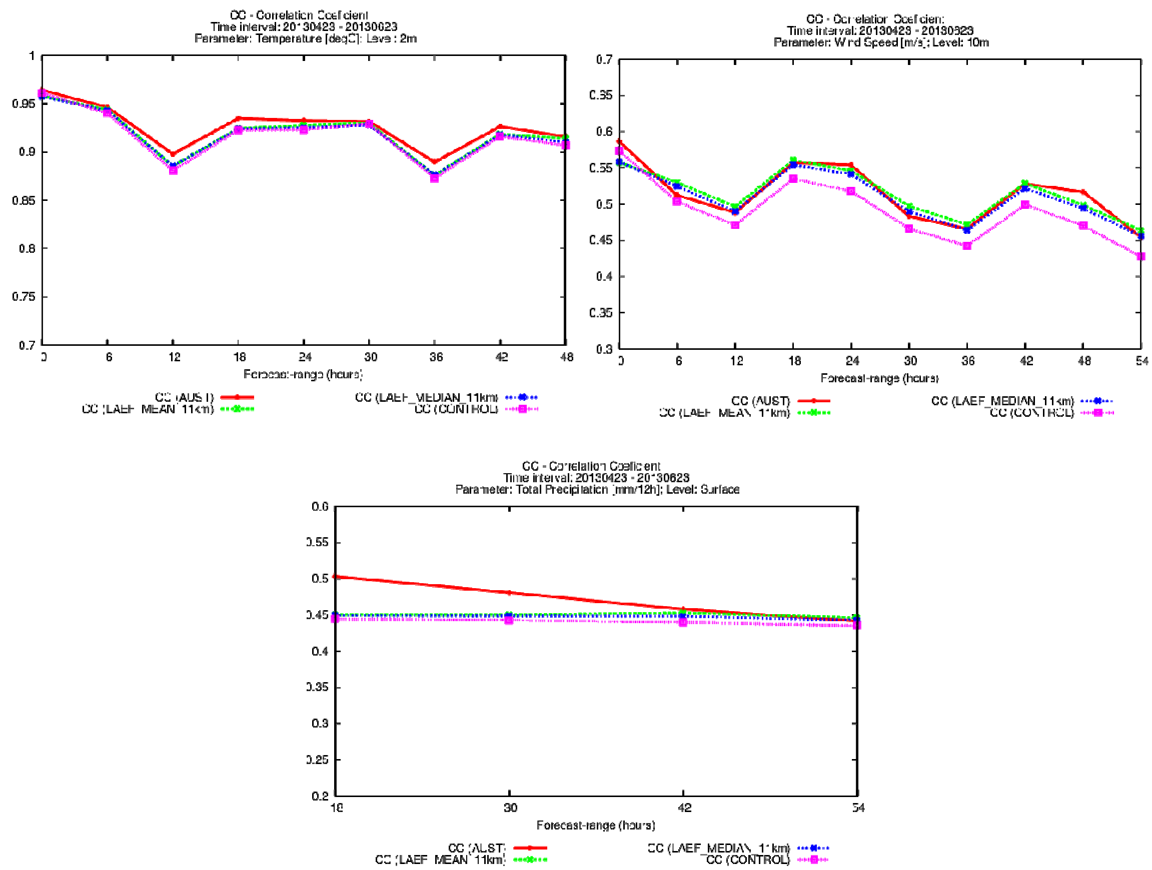


Fig. 9 Correlation coeficient: T2M, W10M and precipitation cumulated in 12hours for ALARO-AUSTRIA (AUST), the mean (LAEF_MEAN_11km), the median (LAEF_MEDIAN_11km) of the ensemble and control forecast (CONTROL) for 23.04.2013 – 23.06.2013 period.

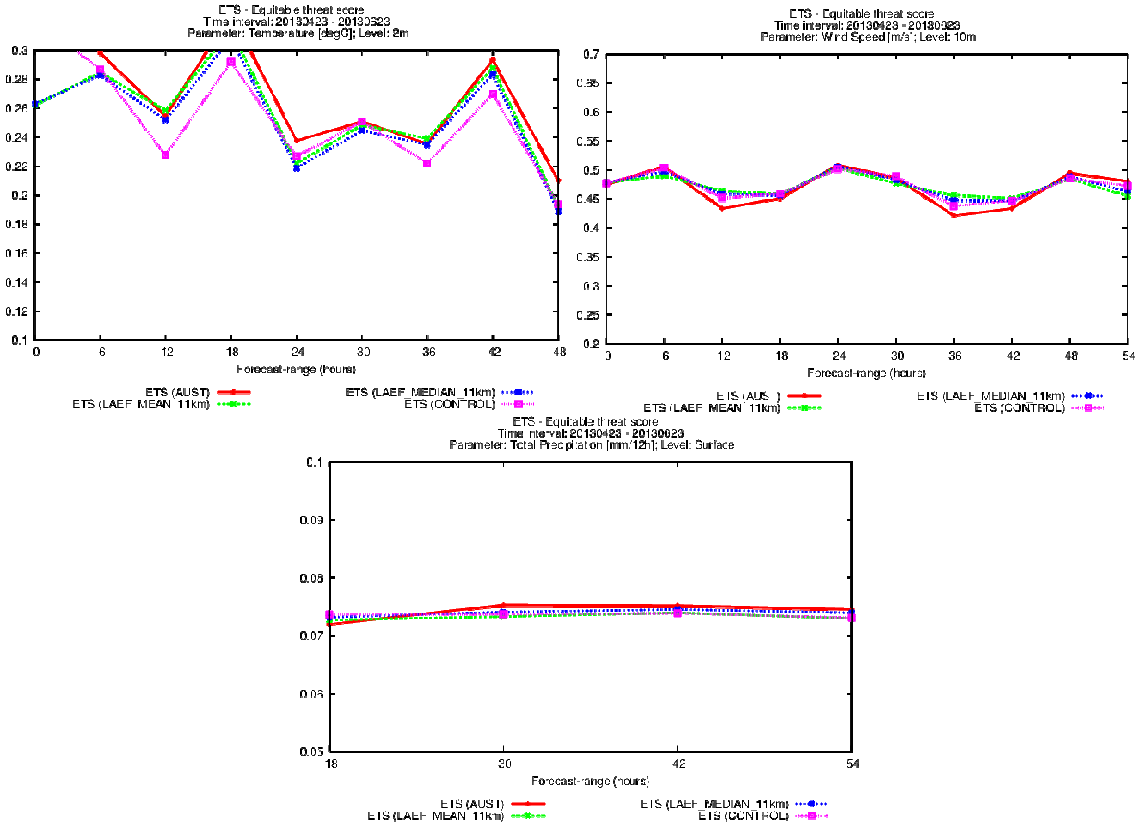


Fig. 10 Equitable Threat Score: T2M, W10M and precipitation cumulated in 12hours for ALARO-AUSTRIA (AUST), the mean (LAEF_MEAN_11km), the median (LAEF_MEDIAN_11km) of the ensemble and control forecast (CONTROL) for 23.04.2013 – 23.06.2013 period.

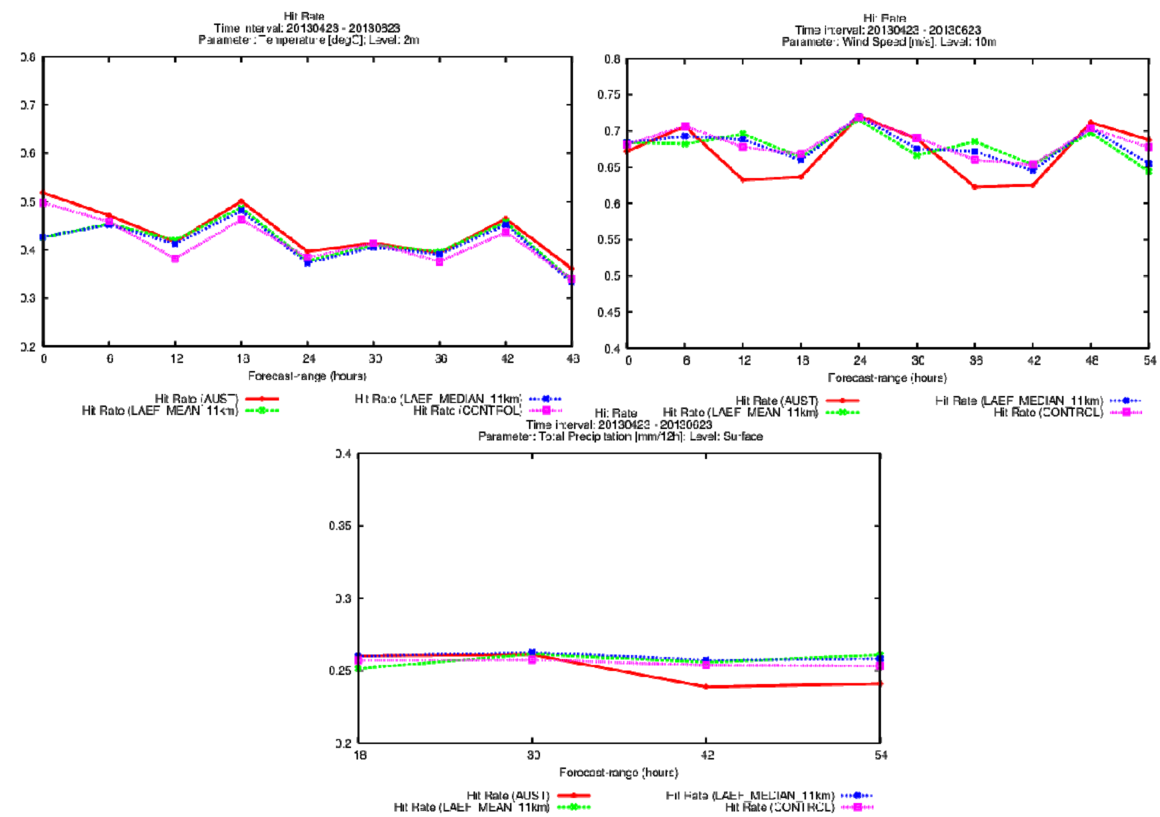


Fig. 11 Hit Rate: T2M, W10M and precipitation cumulated in 12hours for ALARO-AUSTRIA (AUST), the mean (LAEF_MEAN_11km), the median (LAEF_MEDIAN_11km) of the ensemble and control forecast (CONTROL) for 23.04.2013 – 23.06.2013 period.

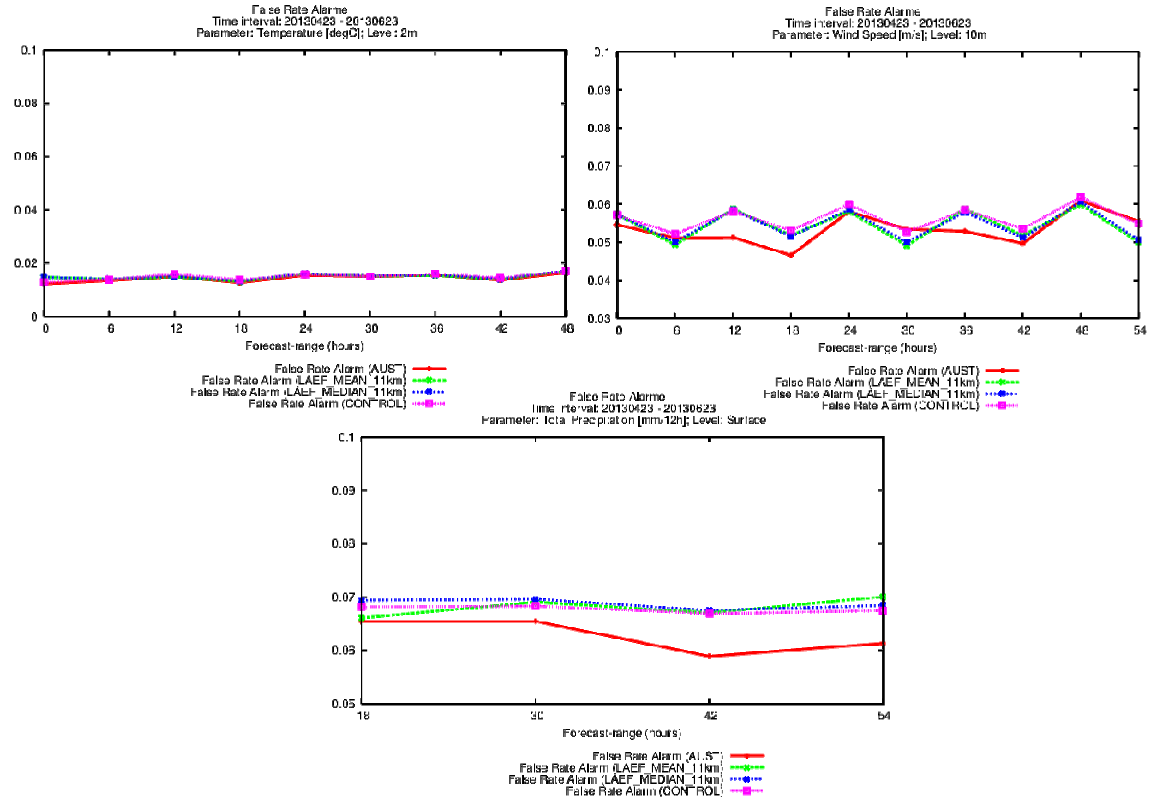


Fig. 12 False Rate Alarm: T2M, W10M and precipitation cumulated in 12hours for ALARO-AUSTRIA (AUST), the mean (LAEF_MEAN_11km), the median (LAEF_MEDIAN_11km) of the ensemble and control forecast (CONTROL) for 23.04.2013 – 23.06.2013 period

Acknowledgements:

I would like to thank to Wang Yong and Theresa Gorgas for their guidance and fruitfull discussion during my stay at ZAMG. I am very gratefully to Martin Bellus ((SMHU) for all his explanations and work regarding perl programming language. Many thanks are addressed to Florian Weidle for his ideas regarding fortran routines and constant help. The hospitality and help of Christoph Wittmann are alos acknoledged. As well many thanks to all the colleagues of ZAMG who made my stay very pleasant.

References

Wang, X., C. H. Bishop, 2003: A comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. Journal of Atm. Sci., 60, p. 1140-1158.

Appendix

Deterministic scores using the contingency table.

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

Considering the above contingency table, the scores are computed as follows:

Accuracy (fraction correct):

$$Accuracy = \frac{hits + correct\ negatives}{total}$$
, Range: 0 to 1. Perfect score: 1.

Probability of detection (hit rate):

$$POD = \frac{hits}{hits + misses}$$
, Range: 0 to 1. Perfect score: 1.

False alarm ratio:

$$FAR = \frac{false\ alarms}{hits + false\ alarms}$$
, Range: 0 to 1. Perfect score: 0.

Probability of false detection (false alarm rate):

$$POFD = \frac{false\ alarms}{correct\ negatives + false\ alarms}$$
, Range: 0 to 1. Perfect score: 0.

Threat score (critical success index):

$$TS = CSI = \frac{hits}{hits + misses + false\ alarms}$$

Range: 0 to 1, 0 indicates no skill. Perfect score: 1.

Equitable threat score (Gilbert skill score):

$$ETS = \frac{hits - hits_{random}}{hits + misses + false\ alarms - hits_{random}}$$

where $hits_{random}$ ... $\frac{\ldots se\ alarms)}{total}$

Range: -1/3 to 1, 0 indicates no skill. Perfect score: 1.

Hanssen and Kuipers discriminant (true skill statistic, Peirces's skill score):

$$HK = \frac{hits}{hits + misses} - \frac{false\ alarms}{false\ alarms + correct\ \ negatives}$$

Range: -1 to 1, 0 indicates no skill. Perfect score: 1.

Heidke skill score (Cohen's □):

$$HSS = \frac{(hits + correct\ \ negatives\ ) - (exp\ ected\ correct\ )_{random}}{N - (exp\ ected\ \ correct\ )_{random}}$$

where

$$(expected\ correct\ )_{random} = \frac{1}{N}\left[\begin{matrix}(hits + misses)(\ hits + false\ alarms) \quad + \\ (\ correct\ negatives + \ misses)(\ correct\ negatives + false\ alarms)\end{matrix}\right]$$

Range: -∞ to 1, 0 indicates no skill. Perfect score: 1.