

Report on stay at ZAMG

01/10 – 11/11/2013, Vienna, Austria

New LAEF Verification Package development and LAEF
ensemble size sensitivity experiments

::Supervised by
Yong Wang (ZAMG)
yong.wang@zamg.ac.at

::Author
Martin Belluš (SHMU)
martin.bellus@shmu.sk

::Table of Contents

Acknowledgement

Foreword

I. LAEF Verification Package

II. Ensemble size sensitivity test

Conclusions

::Acknowledgement

I would like to acknowledge my romanian colleague Simona Tascu, for her continuous work on the improvement of the LAEF Verification Package. I am also grateful to our colleagues from Turkey for the billing units donation. Without them the CPU-expensive experiments with 50 EPS members would not been performed. And last but not least, many thanks to the whole local team at ZAMG for always friendly and inspiring atmosphere.

:::Foreword

This stay could be thematically divided into two parts. The first one was mostly technical, where new LAEF Verification Package driving scripts and modules in Perl and also a general script for plotting the verification scores were coded. This was done in the fruitful cooperation with Simona Tascu, who worked on the finalization of the Fortran computational part of the verification package at that time. Our stays were overlapping each-other for 2 weeks, what turned out to be very productive. During the second part of the stay, the experiments for EPS size sensitivity tests were prepared and executed at the ECMWF's computational cluster and finally verified using new version of the LAEF Verification Package at ZAMG.

:::I. LAEF Verification Package

The driving and configuration Perl scripts (5 scripts and 4 configuration/support modules, altogether over 2000 lines of code) for the new LAEF Verification Package were written and heavily tested. This was carried out in accordance with the new Fortran code for the verification scores computation prepared by Simona Tascu (please refer to her report for the detailed description of the Fortran part). Also the tool for visualization of verification scores was changed completely from the scratch. Instead of many individual shell scripts, there is now only one general Perl script for plotting whatever scores, parameters and levels. It can be run either separately on saved verification output files (arguments are given via command line) or within the verification job itself under the switch LPLOTS (when it plots in batch procedure a generic set of scores defined in the configuration file). Such solution is indeed more practical not only for the maintenance. Now one common and universal module is used for any kind of plotting instead of many copies containing the similar code for different scores, single/multiple experiments etc., which was the case before. Replacement of AWK constructions with built-in Perl string functions and regular expressions has speed up the verification processing as well.

Here is the list of new LAEF Verification Package scripts with their explanation:

MasterVerification.job

This is the only remaining shell script in the whole verification package. We decided to keep it in shell, because it is more accessible for the users and there are almost no execution parts anyway. It was strictly limited only to the minimum set of user-defined paths (application root directory with its subdirectories, data input and output directories, etc.) and basic verification settings like start date, end date, processed forecast ranges, list of verified experiments, blacklisted members (will be explained later) and similar. This is the script to be submitted by user, which subsequently launches the cascade of new Perl scripts and desired computation of the scores by Fortran executable.

DoConfSettings.pl

It's a Perl script for automatic configuration settings for the verification job. It uses *Settings.pm* and *Supports.pm* local Perl modules and exports environmental variables depending on the defined conditions and thresholds. These ENV variables are further inherited by all other verification subprocesses. This script also creates the output directories (if necessary) and unique working directory for safe processing. It performs different settings and variable cross-checks as well. Selection of appropriate ensemble members for the verification procedure is done, taken into account the user-defined blacklist from the *MasterVerification.job* script. The "blacklist" feature is especially important for our ensemble size sensitivity tests, since it allows for separate verification of various EPS members' subsets from a given experiment.

If user-defined list of EPS members, which should be skipped from the verification, is not empty, the member numbers (e.g. 5, 3, 10, 8, 11, etc.) are numerically sorted. In case the blacklisted members form a continuous series of numbers, new experiment names are created for the simplification like "*exp_name*" + "*-M₁..M_n*" (if there are more than 2 skipped members starting with M_1 and ending with M_n) or "*exp_name*" + "*-M₁-M₂*" (if less, i.e. there are just 2 skipped members M_1 and M_2). If they do not form a continuous series of numbers, the name will contain all the excluded members like "*exp_name*" + "*-M₁-M₂-M₃..M_n*". In the next step the directories with such names for the output scores are created. Scalar number of skipped ensemble members is exported as well, to be used in later processing.

Here is an example of output directory name for the upper air verification of the experiment called *LAEF50nophys* with 50 EPS members, where the last 20 members were blacklisted:
LAEF50nophys-31..50_UPPER

At the end of *DoConfSettings.pl* the verification procedure continues by sequentially calling two other Perl scripts *DoVerification.pl* and *DoPlots.pl* (if plotting is required).

DoVerification.pl

This is the main Perl driving script for the verification job. It creates symbolic links in the working directory for each input model data, observation data and reforecast data and for the whole verification period. Under the logical switch LPROC4GRIBS the verification is performed on input grib files, otherwise (if LPROC4ASCII eq True) it is done on daily ASCII files saved by the previous verification. In case the model input grib data are missing for more than 50% of verified days, the decision to continue with the verification is upon the user. If all inputs are successfully prepared, Perl script for writing the verification namelist is called (*DoNamelist.pl*) and subsequently the computation of the scores is launched (*LAEF_Verification_API.exe*). Finally the outputs are renamed according the experiment name and verified period, and are saved into the output directory.

DoNamelist.pl

This Perl script automatically creates the namelist used by the Fortran verification code. All variables are passed here by the environment exported mostly by *DoConfSettings.pl* script. Dynamically generated Fortran namelist is written into the STDOUT as well, so the user can see and check the current settings.

DoPlots.pl

This is Perl script for plotting the generic set of verification scores according to the settings in modules like *Settings.pm* and *SetupForPlots.pm*. It is not doing the plots itself, but rather runs the plotting tool (*Plotting.pl*) in batch procedure with the appropriate arguments (see below).

Plotting.pl

This is the unified tool for plotting any supported verification scores and parameters. It can be run either individually or within the verification procedure. When the script is submitted from the verification job, all arguments are provided through the environment and applied in *DoPlots.pl* script. Otherwise the arguments must be given by user via command line:

USE: *./Plotting.pl -s <score> -v <variable> -l <level> -start <start> -stop <stop> [-r <range> -t <threshold>]*

```
- score      [bias|rmse|...] (*)
- variable   [250|167|234|230|252|151] (**)
- level      [850|500|surf]
- start      [dd.mm.yyyy.HH]
- stop       [dd.mm.yyyy.HH]
- range      [000|006|...] (optional)
- threshold  [-2|0|2|...] (optional)
```

(*) currently supported scores:

```
bias          => BIAS
brierscoreana => Brier Score ANA
brierscoreeps => Brier Score EPS
brierscoreref => Brier Score REF
brierssana    => Brier Skill Score ANA
brierssref    => Brier Skill Score REF
bsdecomp      => BS Decomposition
contingency   => ROC Score
crpsana       => Continuous Ranked Probability Score ANA
crpseps       => Continuous Ranked Probability Score EPS
crpsref       => Continuous Ranked Probability Score REF
crpssana      => Continuous Ranked Probability Skill Score ANA
crpssref      => Continuous Ranked Probability Skill Score REF
outliers      => Outliers
reliability   => Reliability
rmse          => RMSE
rpsana        => Ranked Probability Score ANA
rpseps        => Ranked Probability Score EPS
rpsref        => Ranked Probability Score REF
rpssana       => Ranked Probability Skill Score ANA
rpssref       => Ranked Probability Skill Score REF
spread        => Spread
spread_rmse   => Spread Skill
```

```

spreadskill_b => Spread Skill Bins
srrelation    => Spread RMSE Relation
talagrand     => Talagrand

```

(**) currently supported variables:

```

250 = wind speed
167 = temperature
234 = geopotential
230 = relative humidity
252 = total precipitation [mm/12h]
151 = mean sea level pressure

```

Regarding the processed experiments, if plotting is submitted within the verification job, the hash of arrays defining them is created automatically. Experiments names and number of ensemble members for each of them are provided via environment variables `$ENV{exps}` and `$ENV{neps}`. For plotting the scores individually in “offline” mode using the archived verification output files, this hash of arrays must be defined in `Exp.pm` module explicitly by user. Below is an example to plot 5 different experiments on the same chart, with the CODE names corresponding to the names of experiments used in the verification output files. Of course, these data must exist in `$ENV{ScoreDir}` or under `$data_path` if defined otherwise in `SetupForPlots.pm` module.

```

%exp = (
#-----#
# N      CODE                LEGEND                MEMBERS #
#-----#
  1 => ["LAEF50mp",          "LAEF 50 mp",    50],
  2 => ["LAEF50mp-41..50",  "LAEF 40 mp",   40],
  3 => ["LAEF50mp-31..50",  "LAEF 30 mp",   30],
  4 => ["LAEF50mp-21..50",  "LAEF 20 mp",   20],
  5 => ["LAEF50mp-11..50",  "LAEF 10 mp",   10],
);

```

Finally, the verification charts (PNG and PS files) are saved into the image directory `$ENV{PlotDir}` with the following names:

```
<variable>_<score>_<level>_<experiments>_<start>-<stop>[_<range>][_<threshold>]
```

Settings.pm

This is the main configuration module for the LAEF verification. In `%setup` hash all the verified parameters are defined with their codes used separately for the analysis, ensemble forecast, climatology and reference forecast. Also the verification thresholds for each parameter are to be set here and eventually the flags for using climatology and orography (to make altitude correction for T2m). Furthermore, the scores to be computed can be switched via `%SCORES` hash (T - compute, F - skip). The similar hash named `%PLOTS` defines the verification scores to be automatically plotted if plotting is invoked.

Supports.pm

In this Perl module there are located some support functions commonly used in the verification package. The idea is to put here all the code, which is repeatedly used by one or more driving scripts. The list of currently shared functions is following:

`&count_items(string)`

-> returns number of items in the string separated by whitespace or comma

`&shift_date(yyyymmddHH,N)`

-> returns new date (dd, mm, yyyy, HH) shifted by N hours (both directions)

`&check_date(dd,mm,yyyy)`

-> returns 0 if given calendar day is correct, otherwise dies

`&leap_year(yyyy)`

-> returns 1 if yyyy is leap year, otherwise returns 0

`&spent(start)`

-> writes to STDOUT (and also returns) the formatted time passed from the start time

SetupForPlots.pm

This is the configuration file for plotting. For each verification score the parameters defining chart type (l - lines, b - boxes, p - points), chart shape (0.5 - rectangle, 1.0 - square), X-label, Y-label, title, legend position (tr - top right, bl - bottom left, etc.) and required arguments (1 - basic set, 2 - added range, 3 - added threshold, 4 - added range and threshold) are set through `%chart_type` hash. In this module also the colors for plotting ensemble mean and individual EPS members for up to 5 experiments are specified. If it is required to plot more than 5 experiments into one chart, the color definitions must be added. Users can easily change here all predefined RGB colors in hexadecimal form (000000 to FFFFFFFF), lines/points patterns and their width/size as well. The structure of input verification data files, like the offsets for reading EPS mean and the values for individual members, is also coded in `SetupForPlots.pm` module.

Exp.pm

In this module the experiments to be plotted are defined. The hash called `%exp` with the experiment name, the name shown in legend and its number of ensemble members is created either automatically (when running within the verification job), or manually by user (when running "offline"). Number of blacklisted ensemble members is taken into account as well. Running this procedure "offline" allows even for plotting the results from the different verifications jobs into one chart, as far as the data are covering the same verification period.

Technical note:

One can find the latest version of all the verification source-files and scripts at `vvhmod:/daten2/mgruppe/bellus/verif`

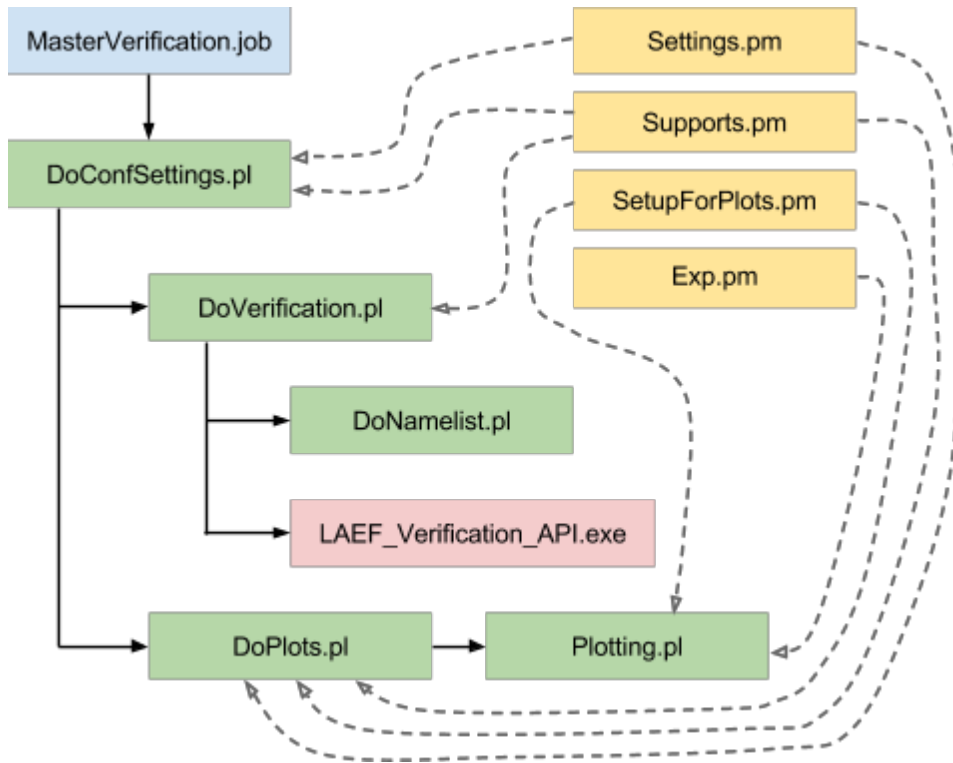


Fig 1: Scheme of the driving scripts in LAEF Verification Package

::II. Ensemble size sensitivity test

The number of ensemble members, i.e. the ensemble size, is one of the most important characteristics of the probabilistic NWP systems. Any operational production of such system is strongly limited by the available computer power. Unfortunately, the relation between the ensemble size, model resolution and EPS skill is far from linear. To get the best out of the system, while still maintain its operability, it is necessary to find some optimal configuration. This was also our main motivation to start with the ensemble size sensitivity tests for new ALADIN-LAEF system.

The two experiments coupled with 50 ECMWF global EPS members were carried out in order to verify the value of ensemble size for our LAM EPS:

1. ALADIN-LAEF without multiphysics (but with 12h upper-air and surface pressure breeding, surface assimilation of perturbed T2m and RH2m observations and upper-air spectral blending for the initial conditions)
2. ALADIN-LAEF with multiphysics (everything else the same as for the first experiment)

In the second experiment, the selected ten integration namelists for different physical parameterizations and model tunings (see table below) were reused for all 50 members according the key 01-10, 11-20, 21-30, 31-40, 41-50.

name	members	MPH	DPC	SHC	RAD	TRB	GUD
MP01	1/11/21/31/41	Dark Blue	Blue	Blue	Light Blue	Light Blue	Light Blue
MP02	2/12/22/32/42	Dark Blue	Blue	Blue	Light Blue	Light Blue	Light Green
MP03	3/13/23/33/43	Dark Blue	Blue	Blue	Light Blue	Light Blue	Light Green
MP04	4/14/24/34/44	Dark Blue	Blue	Blue	Light Blue	Light Blue	Light Green
MP05	5/15/25/35/45	Dark Green	Blue	Blue	Light Blue	Light Blue	Light Green
MP06	6/16/26/36/46	Dark Green	Blue	Blue	Light Blue	Light Blue	Light Green
MP07	7/17/27/37/47	Dark Red	Dark Green	Light Green	Light Green	Light Green	Light Red
MP08	8/18/28/38/48	Dark Red	Dark Green	Light Green	Light Green	Light Green	Light Green
MP09	9/19/29/39/49	Dark Red	Dark Green	Light Green	Light Green	Light Green	Light Green
MP10	10/20/30/40/50	Dark Blue	Red	Blue	Light Blue	Light Blue	Light Green

Tab 1: Description of ensemble members regarding multiphysics

MPH - microphysics

- ALARO-0 using Xu-Randall type large scale condensation
- ALARO-0 using Smith type large scale condensation
- Lopez microphysics

DPC - deep convection

- 3MT (Modular Multi-scale Microphysics and Transport)
- Bougeault and Geleyn scheme
- 3MT + cellular automaton

SHC - shallow convection

- Geleyn (1987) based shallow convection
- Kain-Fritsch-Bechtold shallow convection scheme

RAD - radiation

- Geleyn et. al 2005, Rittern and Geleyn 1992
- RRTM and Morcrette 1991 (ECMWF)

TRB - turbulence

- pseudo-prognostic TKE, Geleyn et. al 2006
- Cuxart-Bougeault-Redelsperger prognostic TKE

GUD - gust diagnostics

- classical ALADIN approach
- combination of ALADIN, Meso-NH and Brasseur
- TKE based approach (Meso-NH)

Apart from the different physical parameterizations, also various tunings are alternatively used within the same schemes to further increase the spread of the system (for more details about the multiphysics tuning please contact Christoph Wittmann). Subsequently the ensembles consisting of subsets counting first 10, 20, 30, 40 and all 50 members were verified and compared to each other.

The probabilistic forecast and hence the verification was carried out for the period of two and half months, from 1st of June till 15th of August 2011. Just for the imagination, to run such huge experiments is equal to 8 months of operational ALADIN-LAEF running with 2 suites per day. Furthermore, to warm up the surface assimilation cycle with the perturbed observations, a 20 day's integration up to +12h was performed starting on 11th of May 2011. However, these data were obviously not used for the verification itself.

Fifty ensemble members were integrated up to 54h for 12 UTC network time. For 00 UTC network time just the short 12h forecast was performed for each member to keep the assimilation cycle on.

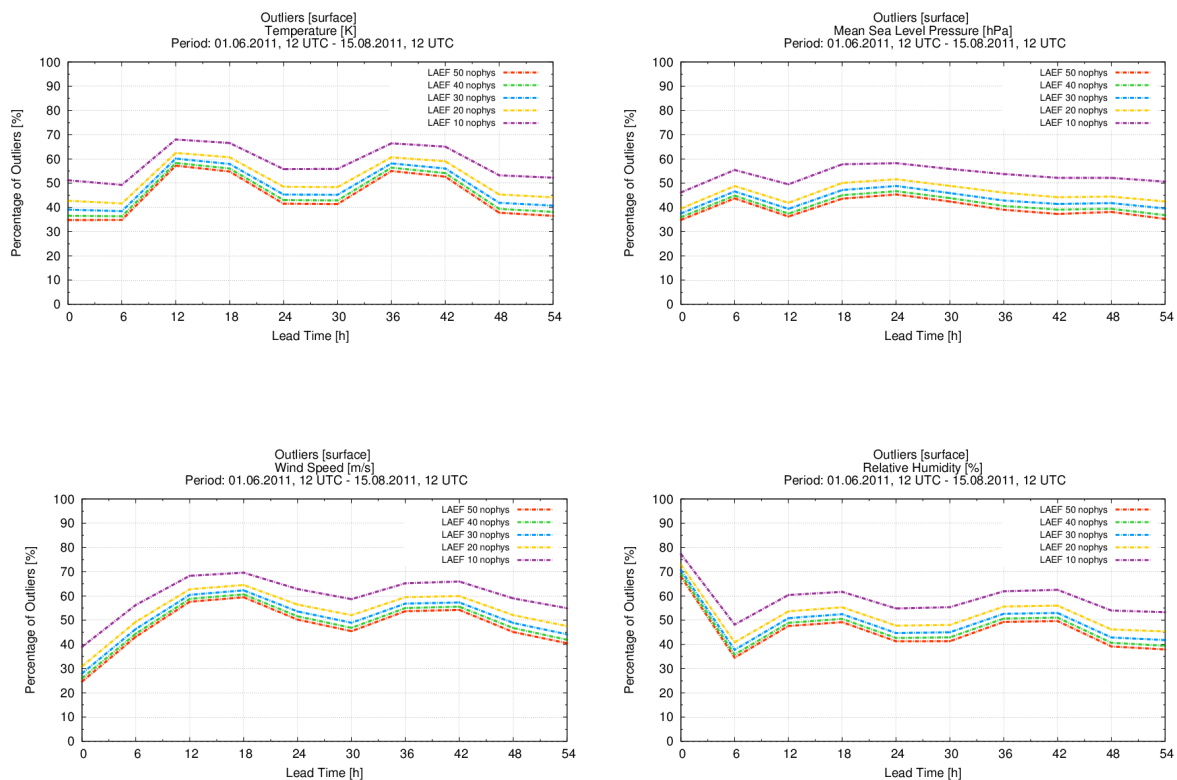


Fig 2: Percentage of outliers for different ensemble size and experiment without multiphysics for the Temperature at 2m (upper left), MSLP (upper right), Wind Speed (bottom left) and Relative Humidity (bottom right)

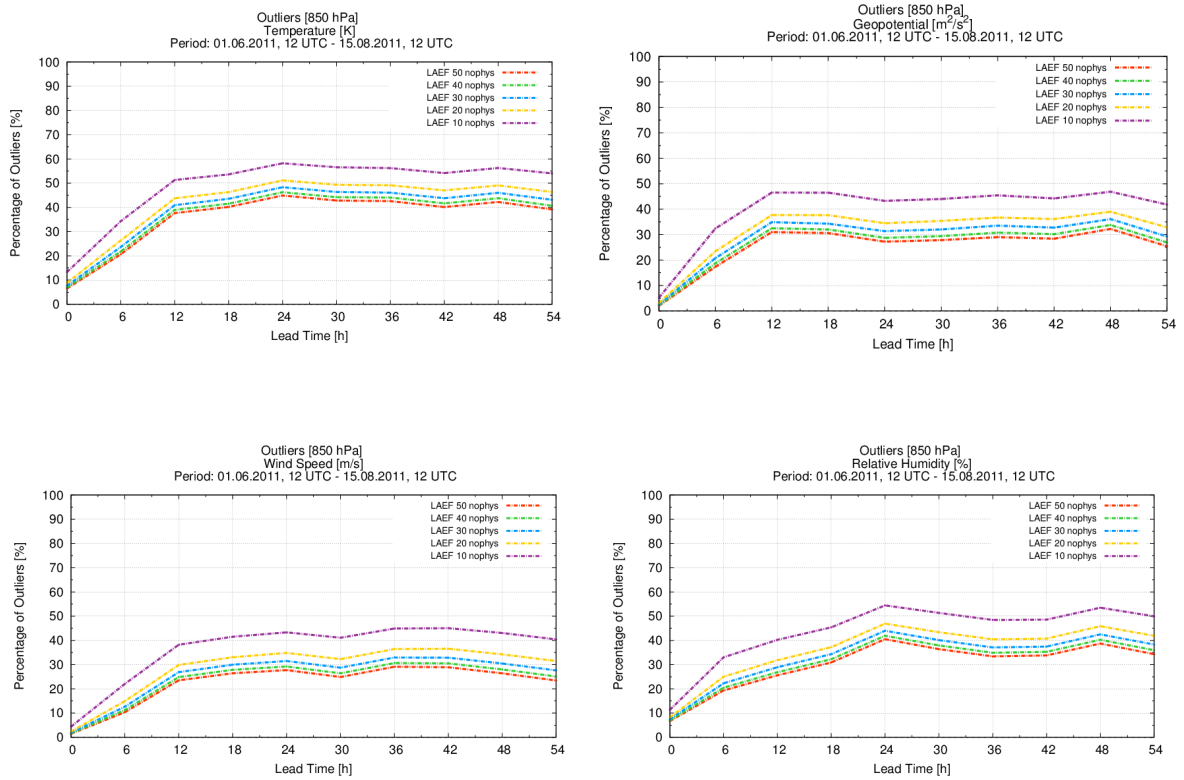


Fig 3: Percentage of outliers for different ensemble size and experiment without multiphysics for the Temperature (upper left), Geopotential (upper right), Wind Speed (bottom left) and Relative Humidity (bottom right) at 850 hPa level

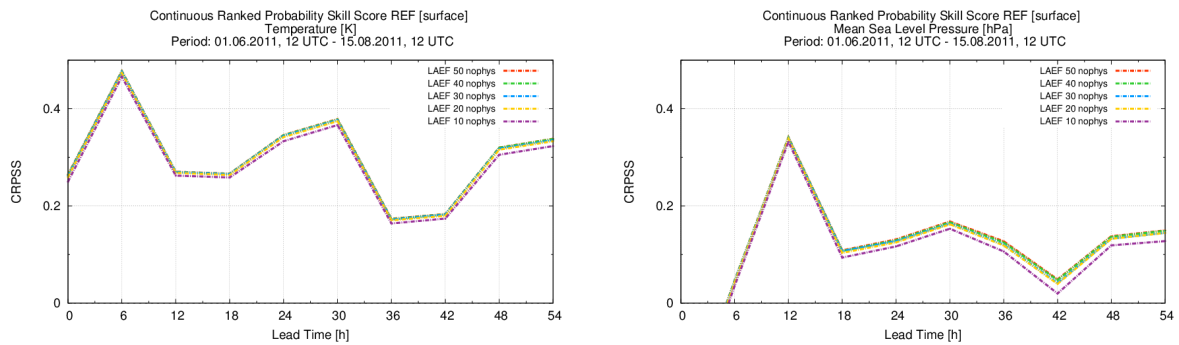


Fig 4: CRPSS with the ALARO5 as reference for different ensemble size and experiment without multiphysics for the Temperature at 2m (left) and MSLP (right)

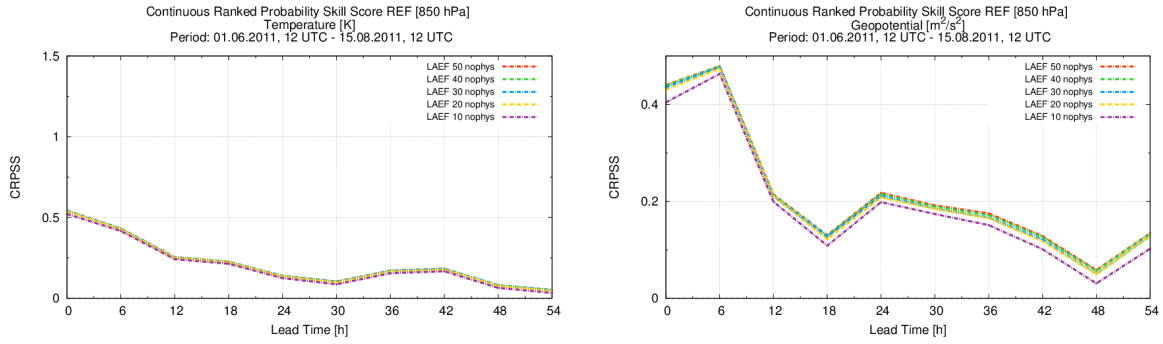


Fig 5: CRPSS with the ALARO5 as reference for different ensemble size and experiment without multiphysics for the Temperature (left) and Geopotential (right) at 850 hPa level

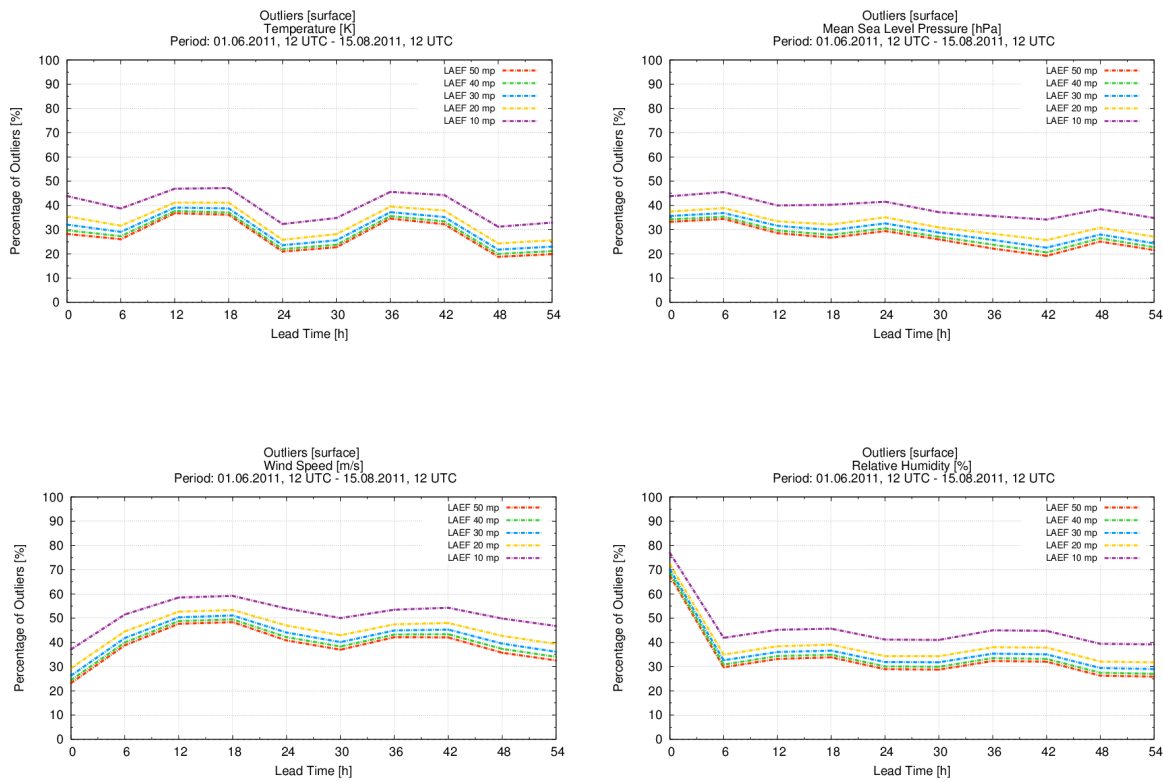


Fig 6: Percentage of outliers for different ensemble size and multiphysics experiment for the Temperature at 2m (upper left), MSLP (upper right), Wind Speed (bottom left) and Relative Humidity (bottom right)

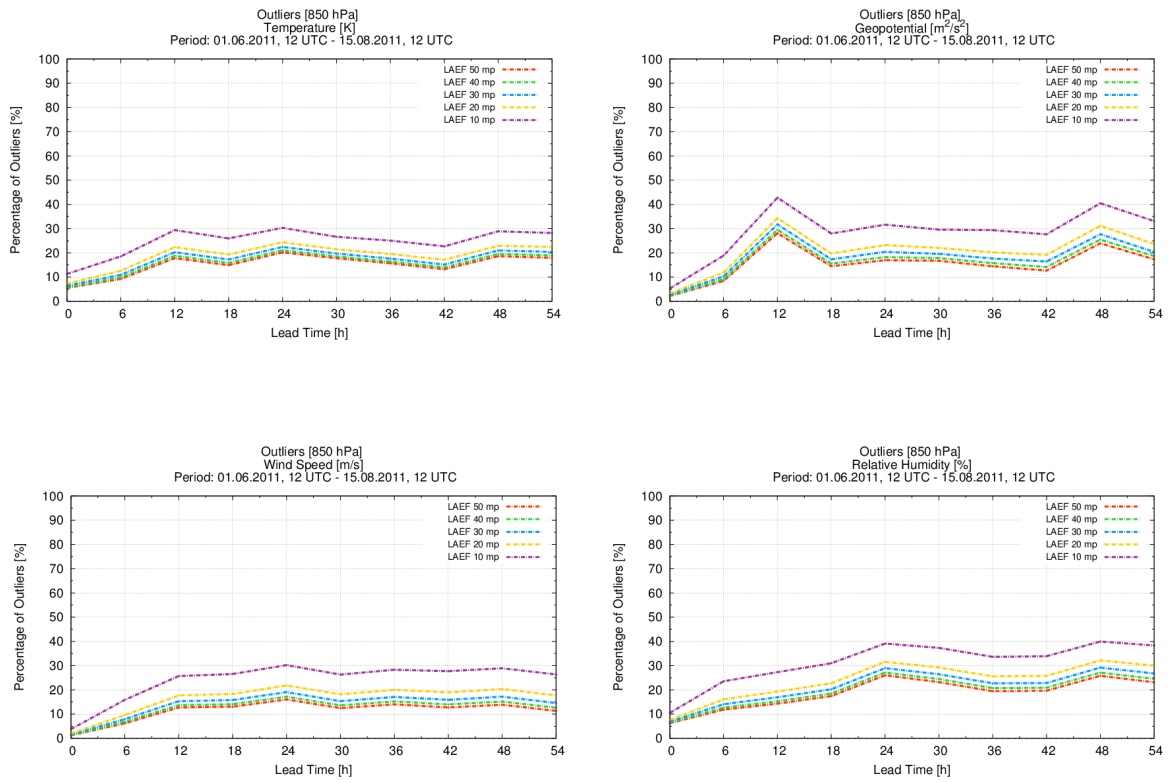


Fig 7: Percentage of outliers for different ensemble size and multiphysics experiment for the Temperature (upper left), Geopotential (upper right), Wind Speed (bottom left) and Relative Humidity (bottom right) at 850 hPa level

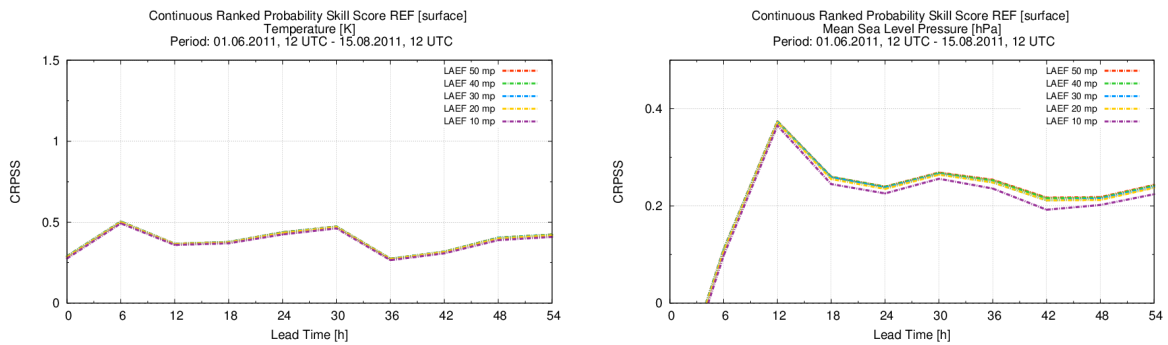


Fig 8: CRPSS with the ALARO5 as reference for different ensemble size and multiphysics experiment for the Temperature at 2m (left) and MSLP (right)

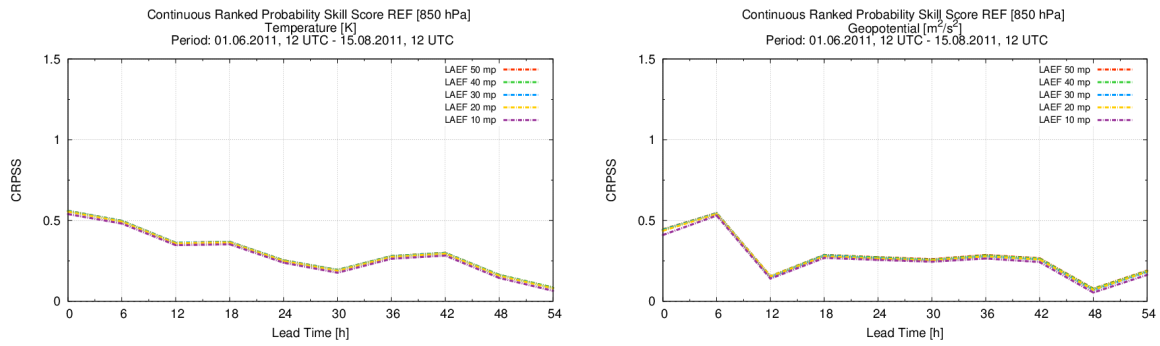


Fig 9: CRPSS with the ALARO5 as reference for different ensemble size and multiphysics experiment for the Temperature (left) and Geopotential (right) at 850 hPa level

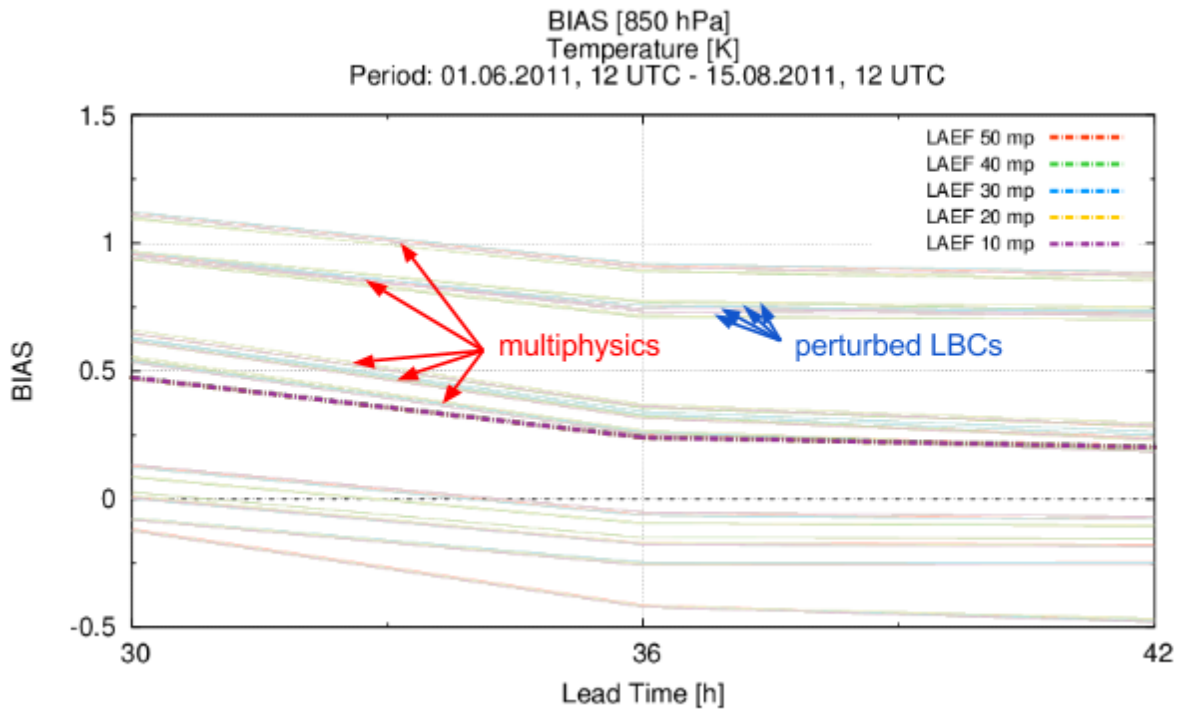


Fig 10: Clustering of Temperature errors by different physical parameterizations

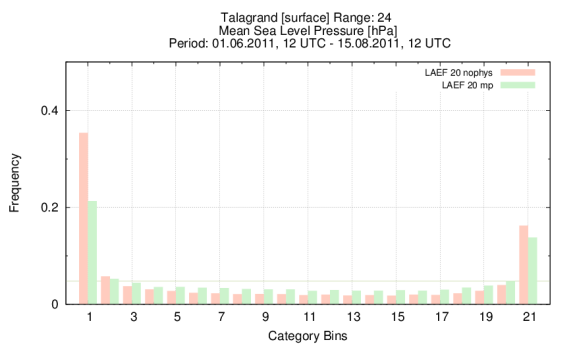
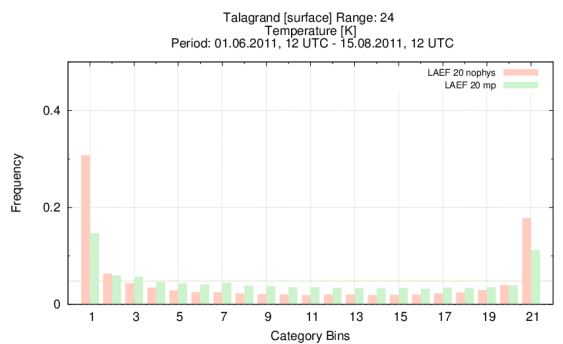
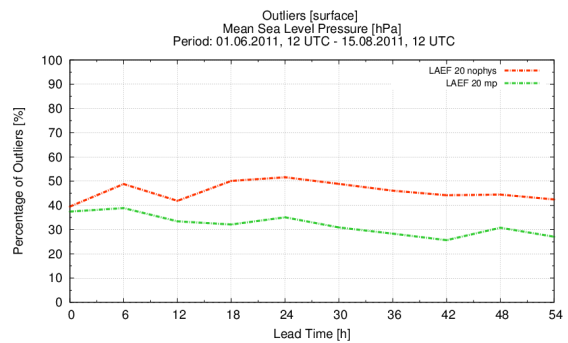
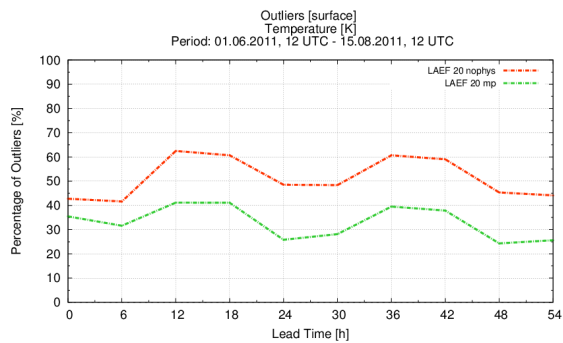
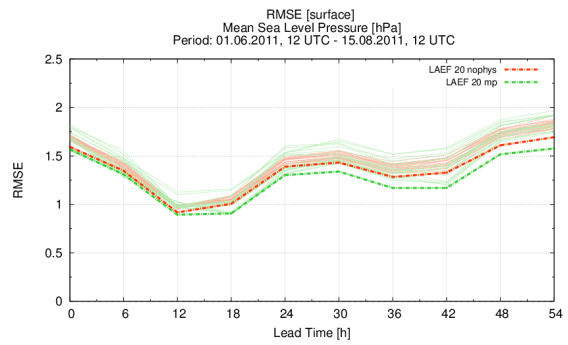
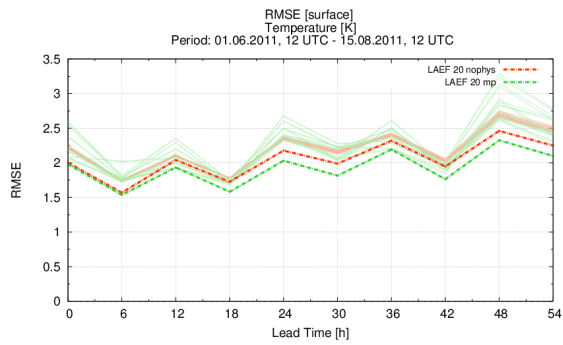
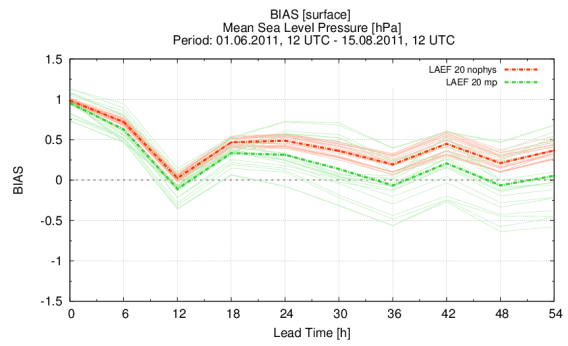
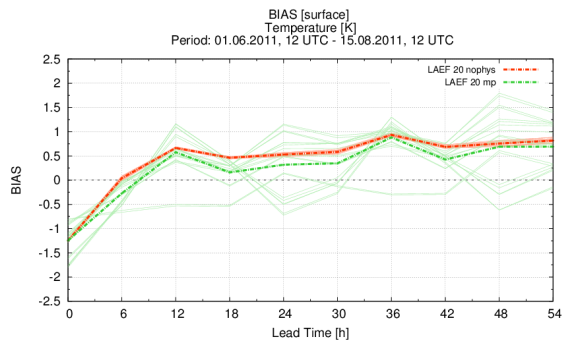


Fig 11: BIAS (first row), RMSE (second row), Outliers (third row) and Talagrand diagram (fourth row) for Temperature at 2m (left) and MSLP (right)

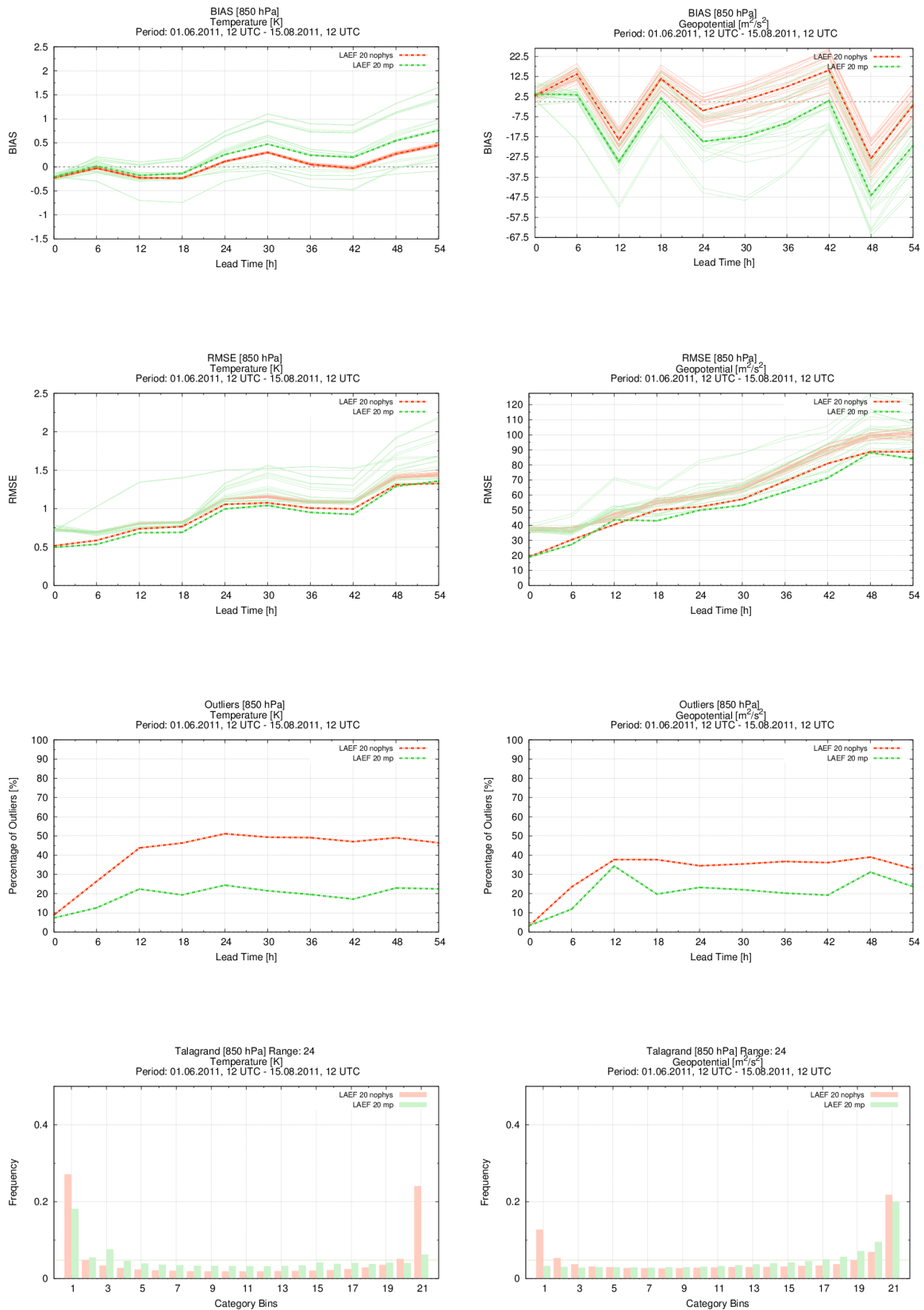


Fig 12: BIAS (first row), RMSE (second row), Outliers (third row) and Talagrand diagram (fourth row) for Temperature (left) and Geopotential (right) at 850 hPa level

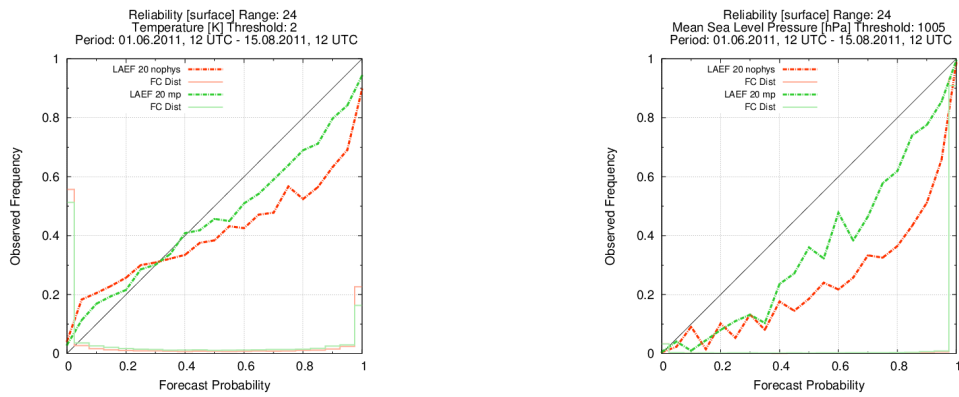


Fig 13: Reliability diagram at +24h for Temperature at 2m (left) and MSLP (right)

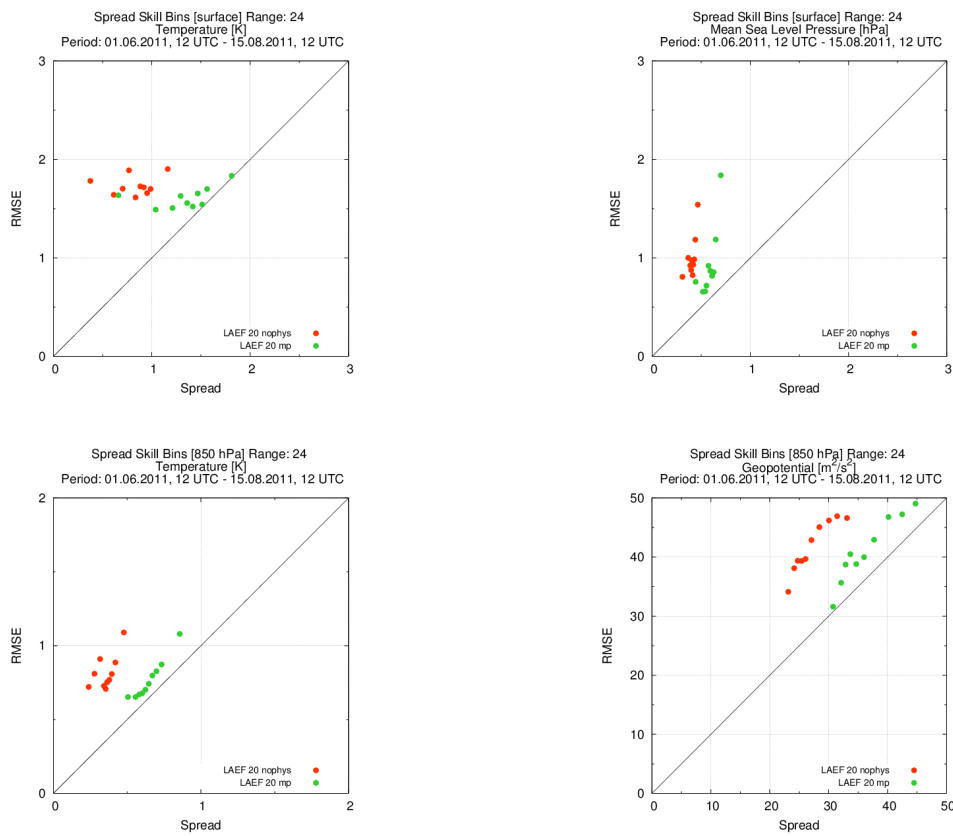


Fig 14: Spread Skill relation for the category bins for Temperature (left) and MSLP/Geopotential (right), while upper row is for surface and bottom row for 850 hPa level

::Conclusions

We have shown, that the biggest qualitative gap in relevant statistical scores is between the ensembles containing 10 and 20 members, while for more populated ensembles there is already very little improvement of the system performance observed (see Fig. 2-9). This can be applied equally to both experiments, the one without multiphysics (where the perturbation is coming from the initial and boundary conditions solely) and the other one with multiphysics (with LAM model uncertainty included as well).

BIAS and RMSE scores for the EPS means are almost independent on the ensemble size, which was rather expected. On the other side, CRPSS for instance indicates slightly better ensemble performance (skill relative to ALARO5 forecast) for increasing lead time and more than 20 members (see Fig. 8-9). The clustering of errors by the different physical parameterizations is obvious as well. It can be clearly observed from the BIAS charts with the individual members' errors plotted (see Fig. 10). Such effect is more pronounced for the Temperature, while for the Geopotential at upper levels it is not so evident (not shown).

Based on the verification results, our proposal for the optimal ensemble size of current ALADIN-LAEF configuration would be around the 20 members. For bigger ensembles there is still some slight enhancement, but the CPU cost is growing rapidly. Having all the data, we did finally also the intercomparison of such big ensemble (20 members) and our two experiments with and without the model perturbation. The significant improvement of the ensemble system is obviously gained here only by the added model uncertainty simulated by the different physical parameterizations and settings (see Fig. 11-14).