Report on stay at ZAMG

25/05 – 26/06/2015, Vienna, Austria

Combination of IC and model uncertainties for the surface prognostic variables in ALADIN-LAEF system

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

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

::Table of Contents

::Acknowledgement

The migration from ECMWF's former IBM cluster to the new CRAY supercomputer was little bit tricky. But at the end of the day all problems were successfully resolved with the great help from Florian Weidle and Christoph Wittmann. I was also inspired by the pioneers of Hirlam group who walked the similar stubble path some months ago. I am grateful for their notes published on the internet forums. It has helped a lot.

::Foreword

The aim of this stay was to construct an experiment linking together the initial and model perturbations of the surface prognostic variables and to verify their combined effect on the regional ensemble system performance. In order to do that, firstly it was necessary to migrate the modular R&D version of ALADIN-LAEF system to the new CRAY supercomputer at ECMWF. A tricky technical task was not only the re-compilation of model and tools on new platform, but mostly the emulation of dependencies between the job steps, which is completely missing feature in Portable Batch queueing System (PBS) used on the CRAY cluster. The introduction of the appropriate running scripts was a necessity in order to run fully automated long term experiments on the new machine.

In parallel, a new version of preprocessing tool for merging and filtration of the observations was prepared, following the requests from the users. Its functionality was extended with the implementation of a time window for accepting the observations from. User can now set *$twindow* in seconds, while default value is 3600 - i.e. 1 hour. Then all records within Network Time & Date +/- 0.5 * *$twindow* will be accepted for merging. Network Time & Date is always read from the given input files. Of course, user can combine files with different validity too. The case, when some of the duplicate records has not been skipped, was fixed as well. (It used to happen when rounded LAT or LON of the station was pure zero.) New version (v03) of the obsoul_merge tool was tested and published on RC LACE forum. This new version was also used in our current experiments with the ensemble of surface data assimilations presented here.

::I. Uncertainties in NWP

Generally speaking, there are two different sources of uncertainties in NWP modelling. The first one is the uncertainty of the initial and boundary conditions and the second one is the uncertainty of the numerical models themselves. Here we do not mean the exact accuracy of the computations, but rather the approximation of the nature by half-empirical physical parameterizations or by their inevitable simplifications due to the limited computer resources.

From the very beginning of ALADIN-LAEF introduction in 2007, we have already implemented and heavily tested several perturbation methods for IC and model uncertainty simulation. These are Non-Cycling Surface Breeding (NCSB), Ensemble of Surface Data Assimilations (ESDA) and the upper-air spectral Breeding-Blending - for generation of IC perturbations; and multi-physics (MP) as well as stochastic physics approach (SPPT) - for capturing the model uncertainty.

Stochastic physics, for the perturbation of surface prognostic fields through their parameterized tendencies, was introduced into the ALADIN-LAEF R&D version last year (see RC LACE report: M. Belluš, 2014: Stochastically perturbed physics tendencies of surface fields in ALADIN-LAEF system). Nevertheless, so far it was tested only in so called dynamical adaptation mode without any other LAEF system components, nor it was put into operations. Hence, further logical step towards the operational implementation was its testing in combination with the IC perturbation. Among the other things, this comprises to run the stochastically perturbed parametrization tendencies in full assimilation cycle with the perturbed observations.

::II. Migration to CRAY supercomputer

Due to the recent changes in ECMWF's high performance computing facility (HPCF), all previous development had to be moved to the new CRAY computer. In principle, there are two source code compilers available, Intel and Cray. For fully functional assimilation cycle (model configuration c701), all used model tools like BATOR, BLEND, BLENDSUR etc., must have been compiled with the Cray compiler, likewise the program for perturbing the screen level observations ECMAPERT. All the above sources were based on model version cy36t1.

On the other hand, the SPPT development is based on a newer version of the model - cy38t1. The related source code was compiled by the Intel compiler in order to create a binary for model integration (the configuration e001 was not working properly if compiled by Cray compiler that time). It does stochastic perturbation of the physics parameterization tendencies of the surface prognostic variables (see Tab.2).

The following hints might help a lot to make CANARI (c701) running on CRAY XC30 machine:

- canari and the other tools must be compiled with the Cray compiler (not necessarily true nowadays)
- use mpiexec to invoke programs (instead of aprun)
- swap module PrgEnv-intel with PrgEnv-cray
- load module cray-snplauncher (in CMD)
- launch job in serial class "ns" with EC_total_tasks=1, EC_hyperthreads=1 and EC_threads_per_task=1
- use empty list in &NADOCK namelist for canari quality control and analysis (model was crashing with these definitions of the OBS selection criteria)
- export DR_HOOK_IGNORE_SIGNALS=-1 (don't use DR HOOK for tracebacks)

As it was already stressed, the queueing system on CRAY HPCF is PBS based and doesn't support dependencies between the job steps like IBM's Load Leveler does (in fact there are no job steps supported at all). Therefore, it was necessary to mimic this convenient feature somehow, using the functionality of PBS. A simplistic solution was to create the CMD files with the job headers and all the appropriate settings separately for each LAEF component (like canari, laeff, etc.) and submit them from a perl script where the dependencies are treated. Each CMD file can launch a given task (LAEF component) for a defined ensemble member, date and time. These are ENV variables and they are exported by the driving perl script, which allows for a parallel computation of the whole ensemble at once. One can see how the experiment launching scheme is organized on Fig.1.

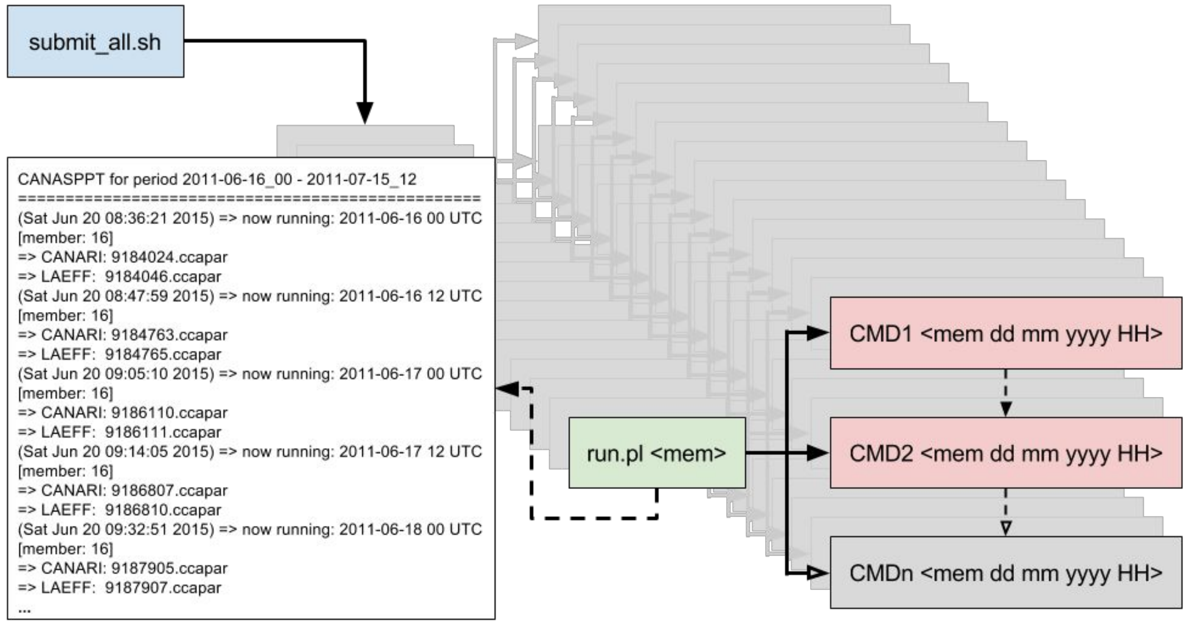


Fig.1: *An experiment scheme, where submit_all.sh is a loop through all ensemble members, run.pl does the job for the whole defined time period ($start, $stop, $step) while handling the physical dependencies between the LAEF components, and finally the jobs are launched from CMDs.*

The dependencies between the individual jobs are created, where the second job waits for the first one to finish successfully. Since PBS is not supporting the job steps, each LAEF component must have its own CMD file. Then the relations between them are scripted one level above (as opposed to the IBM's Load Leveler architecture with the intrinsic dependencies between the job steps). The following is a code fraction of run.pl driving script.

```
###############################################
# run ensemble of surface data assimilations  #
###############################################

$CANARI_JOBID = `qsub $canari`;
chomp($CANARI_JOBID);
print "=> CANARI: $CANARI_JOBID\n";

###############################################
# run LAEF with SPPT after CANARI is finished #
###############################################

$LAEFF_JOBID = `qsub -W depend=afterok:$CANARI_JOBID $laeff`;
chomp($LAEFF_JOBID);
print "=> LAEFF:  $LAEFF_JOBID\n";
```

Furthermore, the new version of obsoul_merge (v03) tool was used. It applies the fixed problem of not recognised duplicity of the observations, when station LAT or LON was pure zero, and new implemented time window for the observations to be accepted from. It works well also with the local austrian TAWES files, but a small patch to account for the different terminology in naming the files ("taw" => 1, "amdar" => 2, etc.) is required as it is shown below.

```
# local files (obsoul_taw_yyyymmdd_HH0000.asc)
if ( $obs_type{$F} eq "taw" ) { $obs_type{$F} = 1 };
```

::III. Experiment settings

Our main goal was to simulate concurrently IC and model uncertainties in ALADIN-LAEF system and evaluate their common influence on the statistical scores in comparison with the particular impact of each of the components when applied separately. Therefore, we had three different experiments to carry on. Their settings are summarized in the following table (see Tab.1).

Experiment name	Upper air initial perturbation	Surface initial perturbation	Lateral boundary perturbation	Model surface perturbation
SPPT	Downscaling of ECMWF EPS	None	Coupling with ECMWF EPS	Stochastically perturbed Phys.Tend.
CANA	Downscaling of ECMWF EPS	Ensemble CANARI	Coupling with ECMWF EPS	None
CANA+SPPT	Downscaling of ECMWF EPS	Ensemble CANARI	Coupling with ECMWF EPS	Stochastically perturbed Phys.Tend.

Tab.1: *Description of ALADIN-LAEF experiments, as well as the related initial perturbation, lateral boundary perturbation and the model perturbation used in the experiments.*

The stochastic physics settings are based on the outcome of the previous work, when surface SPPT was implemented into ALADIN-LAEF system with the standard deviation of normal distribution σ=0.25, spatial correlation L=500km and time correlation τ=2h. For more details about the stochastic perturbation of physics tendencies please refer to the RC LACE report (M. Belluš, 2014).

local variable	global variable	denotation	name (tendency of)
ZTDTS	PTENTS	T_s	surface temperature
ZTDWS	PTENWS	W_s	liquid water
ZTDWSI	PTENWSI	W_{si}	frozen water
ZTDWL	PTENWL	W_r	water on leafs
ZTDSNS	PTENSNS	S_n	water in snow
ZTDALBNS	PTENALBNS	A_n	snow albedo
ZTDRHONS	PTENRHONS	ρ_n	snow density

Tab.2: *List of all perturbed surface prognostic variables in CANA and CANA+SPPT experiments with their code names.*

The ensemble of surface data assimilations (ESDA) is performed by the optimal interpolation method CANARI with the perturbed screen-level observations of temperature and relative humidity. This perturbation is done externally in the Observational DataBase (ODB), where the measured values are randomly perturbed according to the Gaussian distribution with zero mean and standard deviation proportional to the observational error.

Except the ESDA for perturbing the IC and SPPT for model uncertainty simulation (everything for the surface/soil prognostic variables only), no other ensemble techniques were used in our experiments. All ALADIN-LAEF integrations were coupled to the first 16 members of ECMWF global EPS and 54 hour forecast was performed, based on 12 UTC network time. However, an additional 12 hour integration was carried out for CANA experiments at 00 UTC network time, to maintain the assimilation cycle.

On the following figure (Fig.2) one can see the perturbation applied to the surface temperature depending on the experiment kind and forecast range. For the first ensemble member it is shown (but the same can be shown for other members too), that the surface temperature perturbation is obviously zero for SPPT experiment at +00 output, because the stochastic physics is going to be applied only since the first integration step. While CANA and CANA+SPPT experiments have their surface variables perturbed already in INIT file due to the surface assimilation based on the perturbed screen-level observations. On the other hand, the effect of perturbed IC rather quickly vanishes along the forecast range if not supported by the model uncertainty (experiment CANA). Thus the combination of stochastic physics (SPPT) and the OBS disturbance (CANA) seems to produce the most reasonable surface perturbation not only regarding the amplitude and persistence, but also providing a reasonable spatial variation for the area with coarse density of measurements (see Fig.2). A similar effect can be observed also for the other surface prognostic variables as well as for 2m diagnostic fields (not shown).

One may also notice the inconsistency between CANA and CANA+SPPT perturbation at the initial time (i.e. +00 range), which is not quite expected from the first sight. But it is logical and indeed correct, since both experiments have their own assimilation cycle and therefore the different first guess inputs (selected case was preceded by 50 days of cycling).

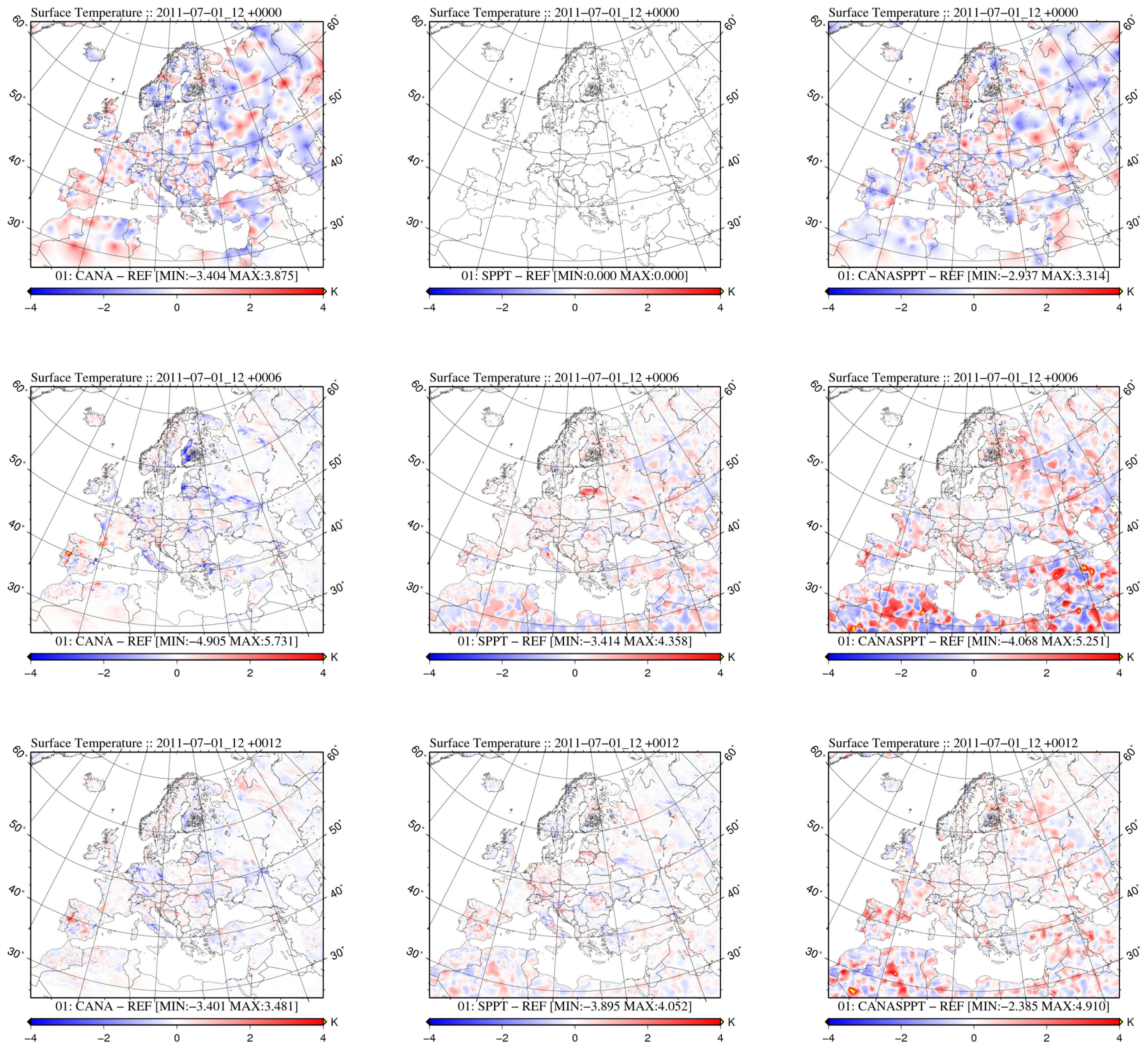


Fig.2: *The surface temperature perturbations for different experiments (from left to right: CANA, SPPT, CANA+SPPT) and forecast ranges (from top to bottom: +00, +06, +12).*

::IV. Verification

One experiment have been run to test the influence of the stochastically perturbed physics tendencies of the surface prognostic variables in combination with the perturbed initial conditions by ensemble of surface data assimilations. The results of two other previous experiments were used as a references for the impact evaluation. However, one of the reference experiments, the SPPT, has been now prolonged by one month to cover the common verification period from 15th of May to 15th of July 2011.

All the experiments were utilizing ECMWF's analysis, but CANA and CANA+SPPT have significantly reduced their surface BIAS as well as RMSE (mostly for temperature, relative humidity and also for mean sea level pressure) due to the local assimilation procedure, which was a part of the applied IC perturbation method. This is pretty obvious when looking at the scores (see Fig.3-4 and Fig.5-6). While the experiment running without the local surface assimilation procedure (SPPT), thus with the uncorrected ECMWF's set of surface fields, suffers from a notable cold BIAS. That is due to the well known discrepancy between the ALADIN and ECMWF surface treatment.

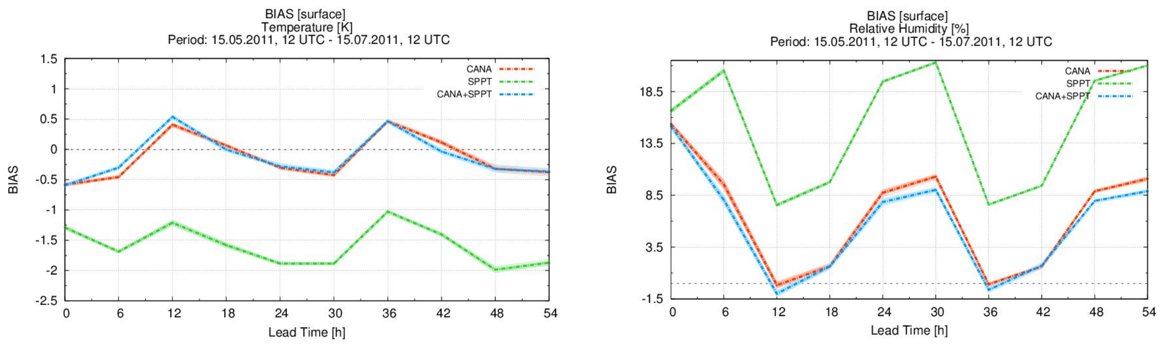


Fig.3: BIAS of the ensemble mean and the individual ensemble members by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for T2m (left) and RH2m (right).

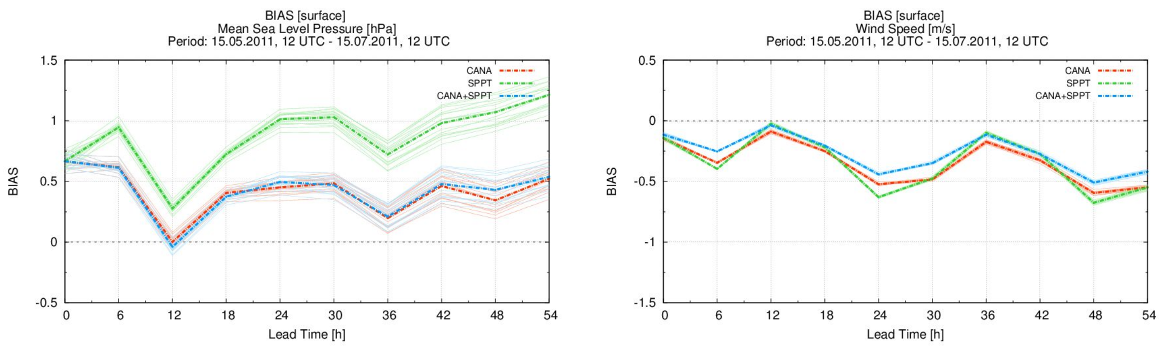


Fig.4: BIAS of the ensemble mean and the individual ensemble members by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for MSLP (left) and W10m (right).

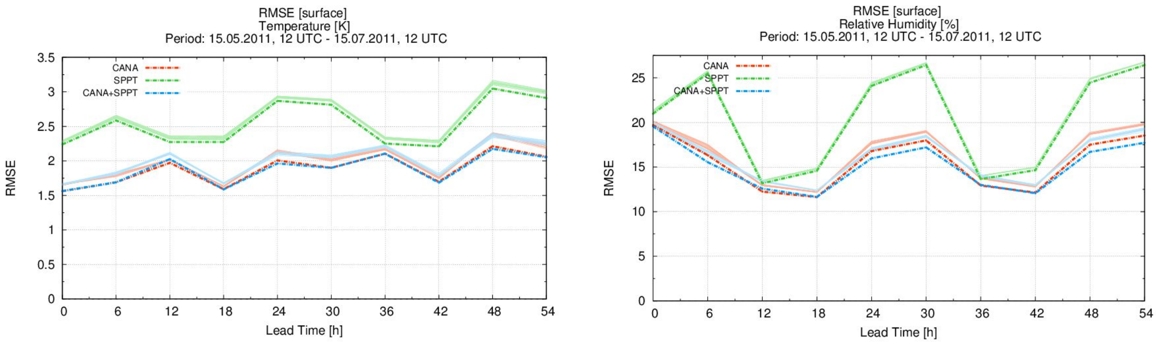


Fig.5: RMSE of the ensemble mean and the individual ensemble members by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for T2m (left) and RH2m (right).

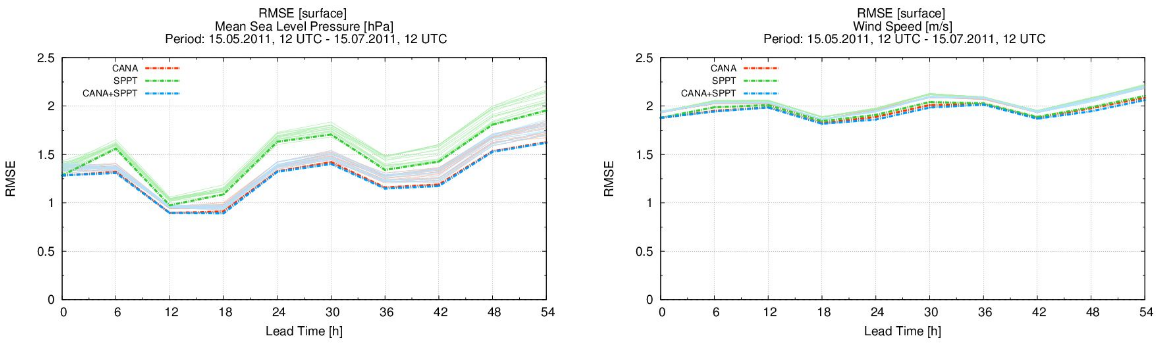


Fig.6: RMSE of the ensemble mean and the individual ensemble members by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for MSLP (left) and W10m (right).

The verification scores shown in the figures (Fig.3-10) are organized by the forecast ranges averaged over the verification period and the verification domain. The verification domain covers just the central part of Europe, with the most reliable and dense observation network. The scores like BIAS, RMSE, OUTLIERS and SPREAD mostly confirm a consistent enhancement of the tested perturbation method (the combination of IC and model uncertainties) in contrast to the references, where the above perturbation techniques were applied separately for the surface and for the model.

While this kind of statistical point verification is surely not very suitable for the evaluation of the precipitation forecast (spatially and temporally inhomogeneous fields are better to be verified using different methods like e.g. SAL), it can be actually used to compare relative improvement between the several experiments. Thus we can say, that a positive impact can be observed for the precipitation forecast as well. On the figures Fig.11-12, the spread skill by verification bins for 12h accumulated precipitation is shown. It is evident, that the perturbation of the surface model variables during the integration has some beneficial effect on the moist atmospheric processes, definitely bigger than the perturbation of initial conditions solely.

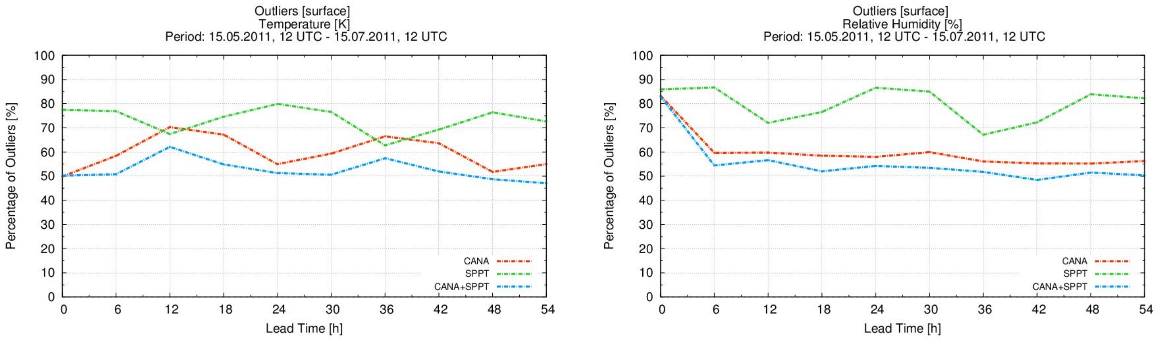


Fig.7: Percentage of OUTLIERS by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for T2m (left) and RH2m (right).

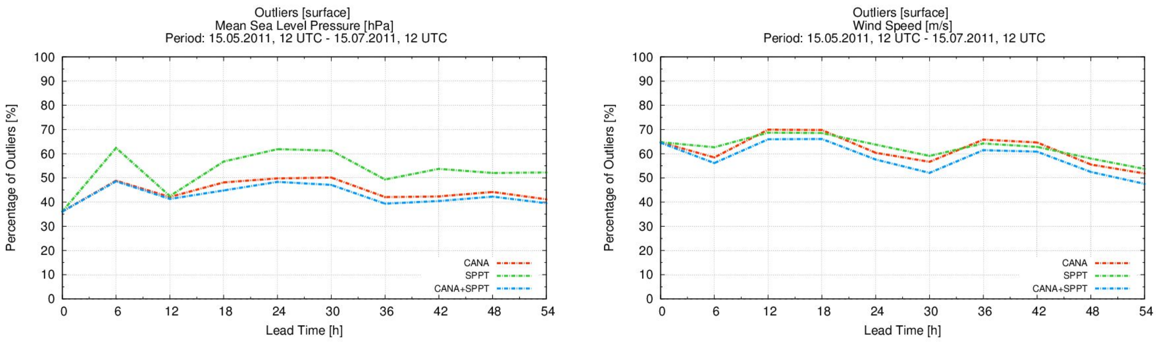


Fig.8: Percentage of OUTLIERS by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for MSLP (left) and W10m (right).

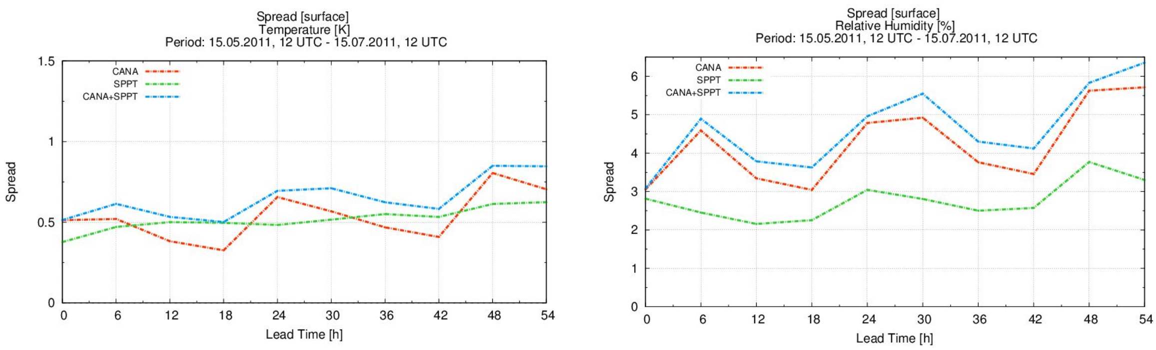


Fig.9: Ensemble SPREAD by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for T2m (left) and RH2m (right).

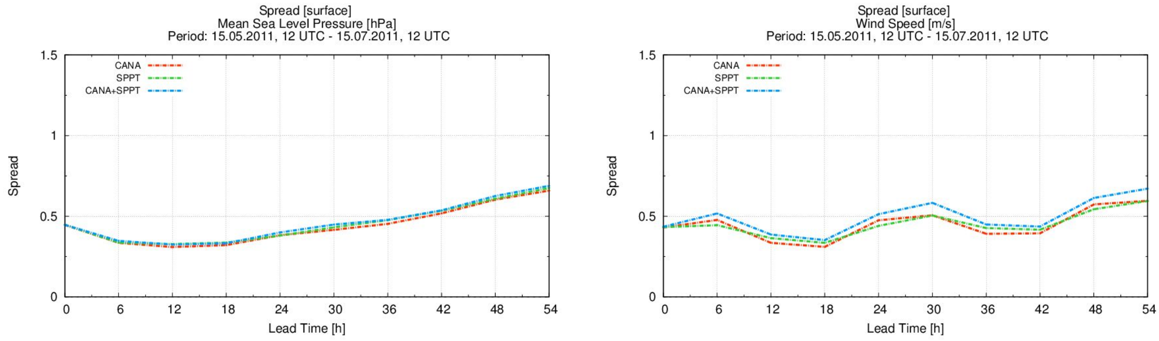


Fig.10: Ensemble SPREAD by the forecast ranges (2 months verified) for the experiments CANA, SPPT and CANA+SPPT for MSLP (left) and W10m (right).

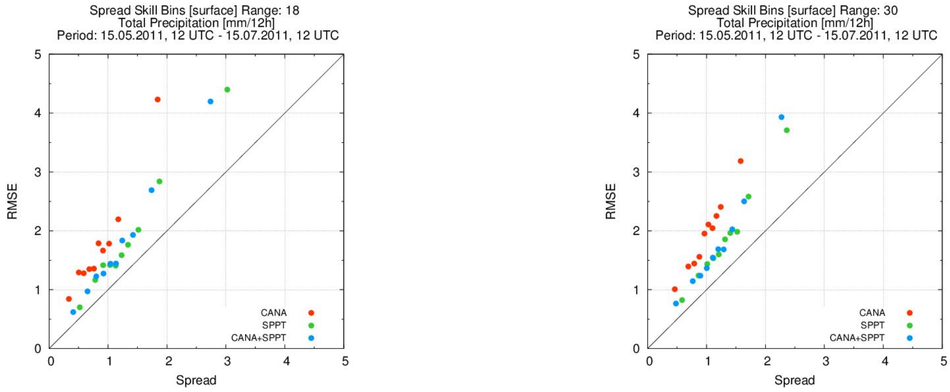


Fig.11: Spread skill bins (2 months verified) for 12h accumulated precipitation for the experiments CANA, SPPT and CANA+SPPT for the ranges +18 (left) and +30 (right).

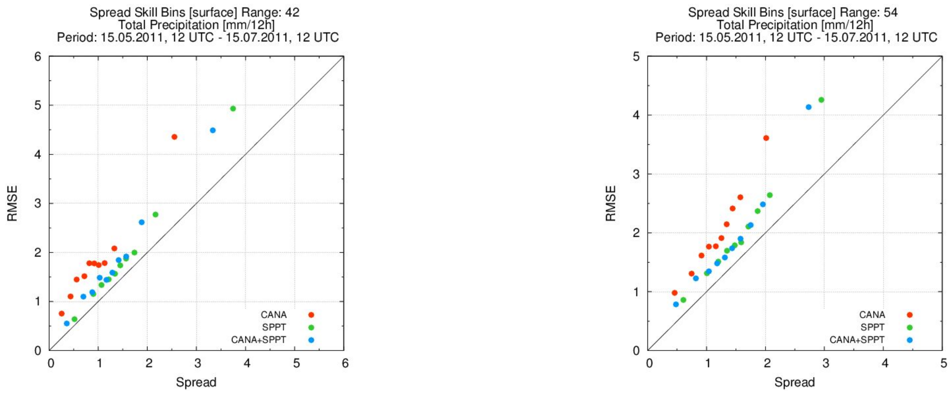
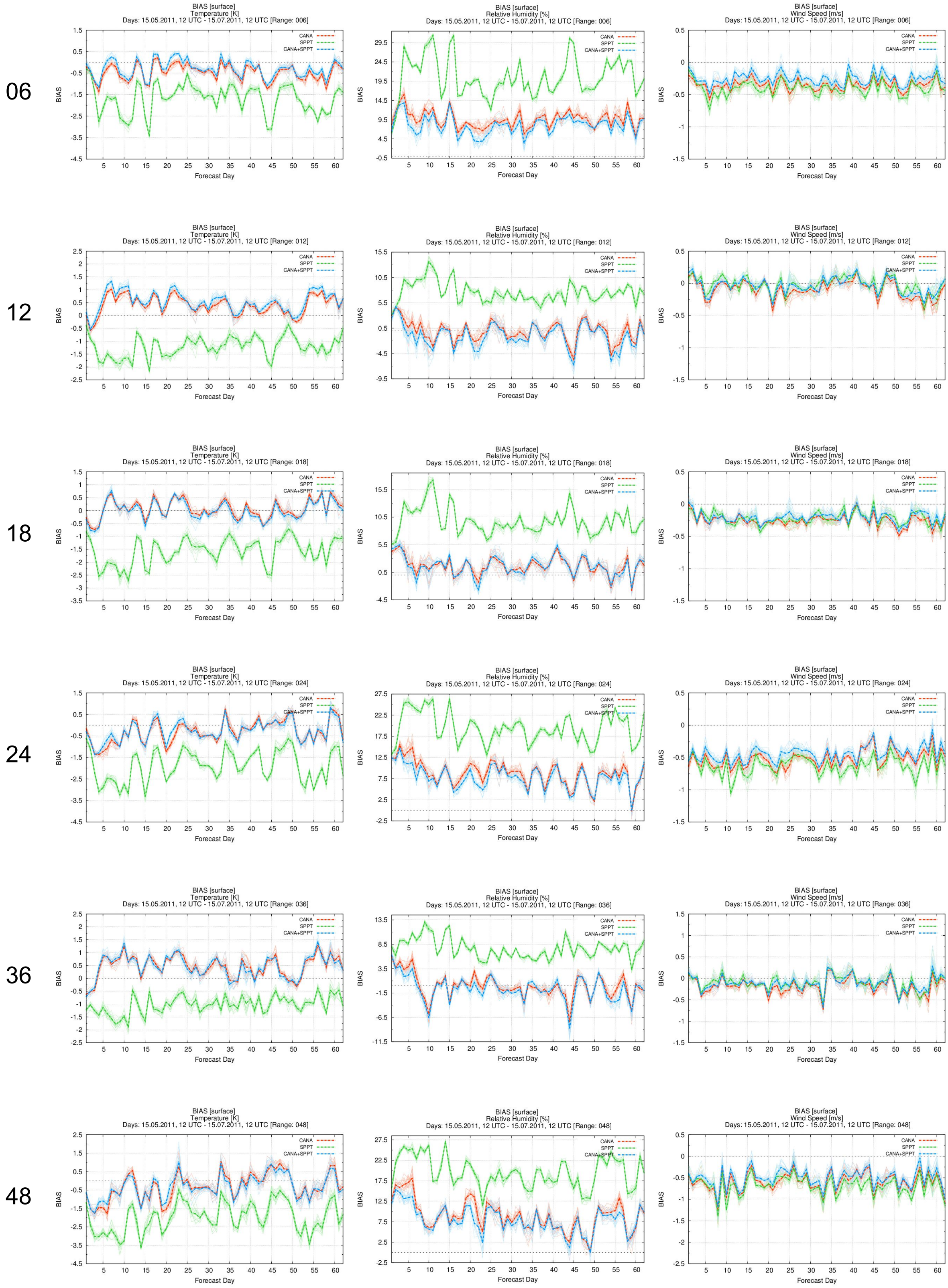


Fig.12: Spread skill bins (2 months verified) for 12h accumulated precipitation for the experiments CANA, SPPT and CANA+SPPT for the ranges +42 (left) and +54 (right).

Finally, on the following plots we show the different scores (BIAS, RMSE, spread and outliers) by the experiment days, valid for forecast ranges from +06 to +48 hours. All together we run the experiments for 62 days. This is to show the impact of the perturbation methods on daily bases for a given forecast lead time.

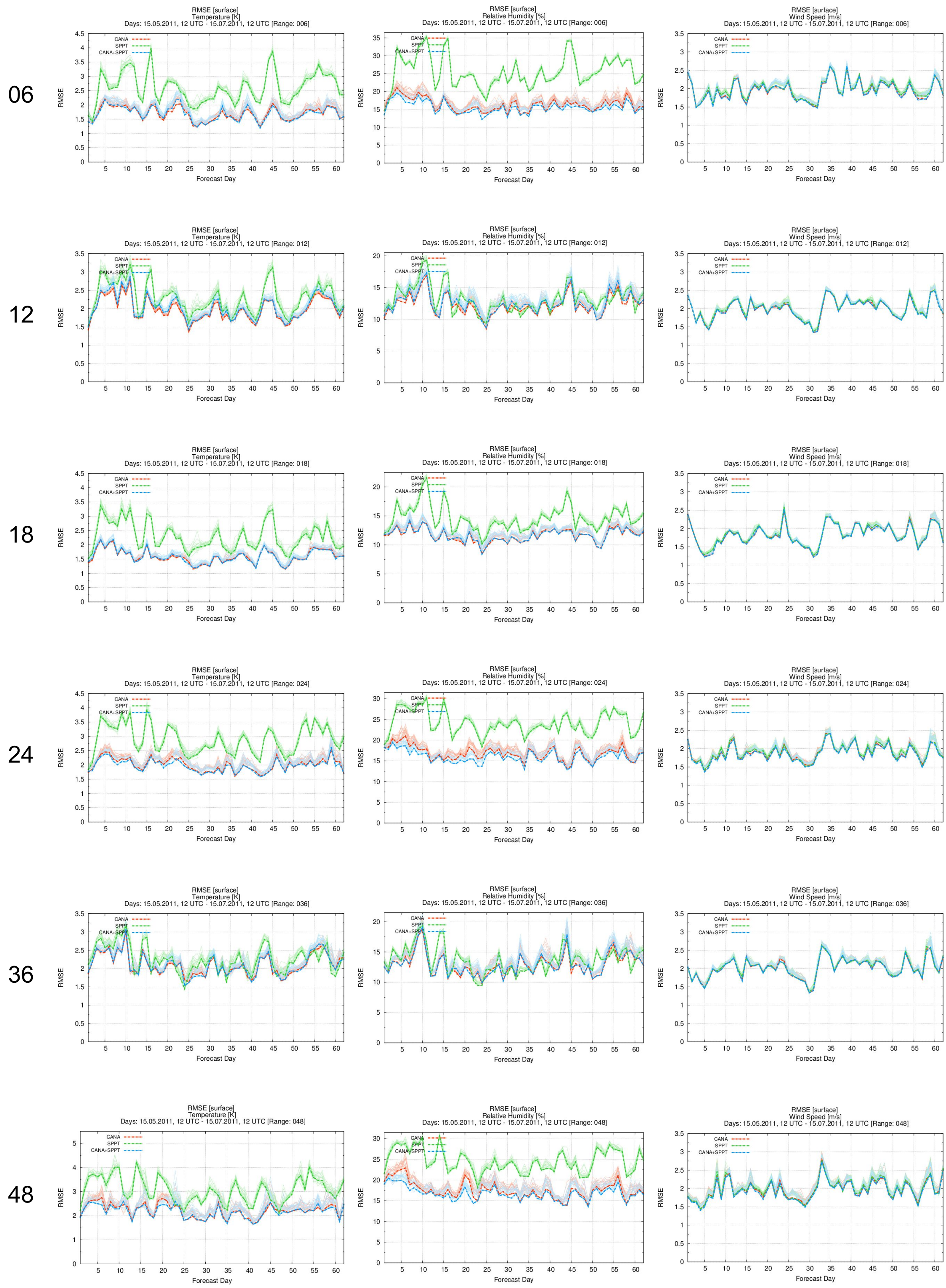
Verification period: 2011-05-15 ~ 2011-07-15   Network: 12 UTC   Score: BIAS

Temperature 2m | Relative Humidity 2m | Wind Speed 10m



Verification period: 2011-05-15 ~ 2011-07-15   Network: 12 UTC   Score: RMSE

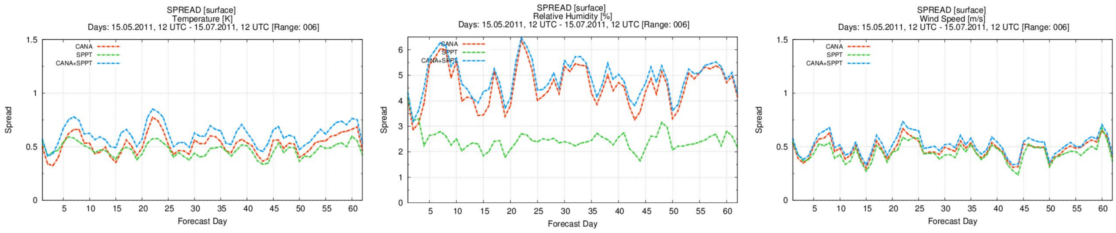
Temperature 2m   Relative Humidity 2m   Wind Speed 10m



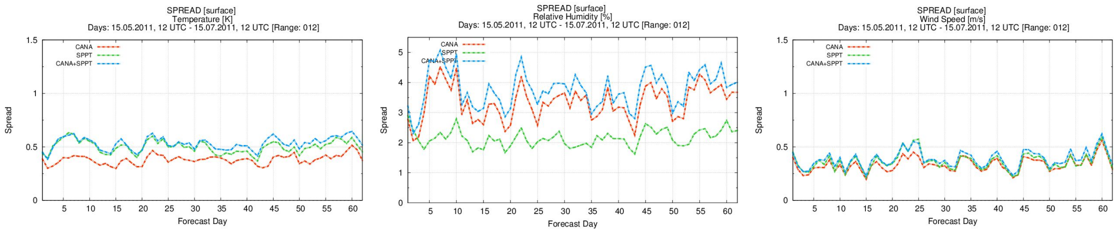
Verification period: 2011-05-15 ~ 2011-07-15 Network: 12 UTC Score: SPREAD

Temperature 2m | Relative Humidity 2m | Wind Speed 10m

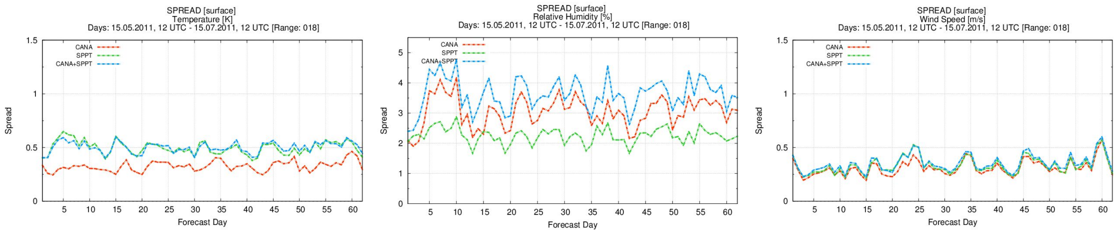
06



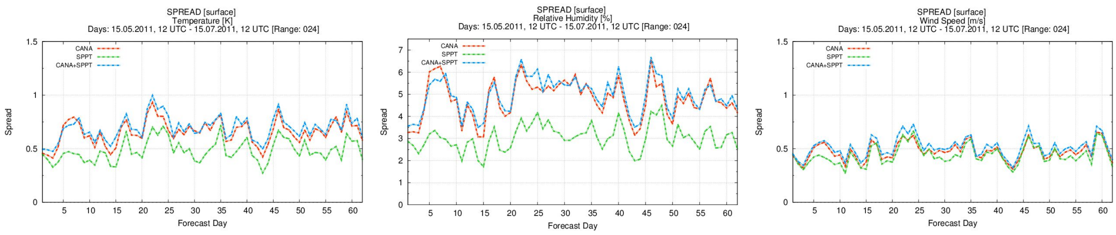
12



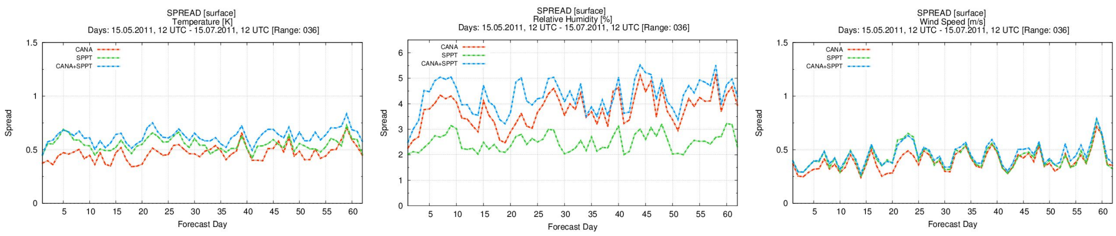
18



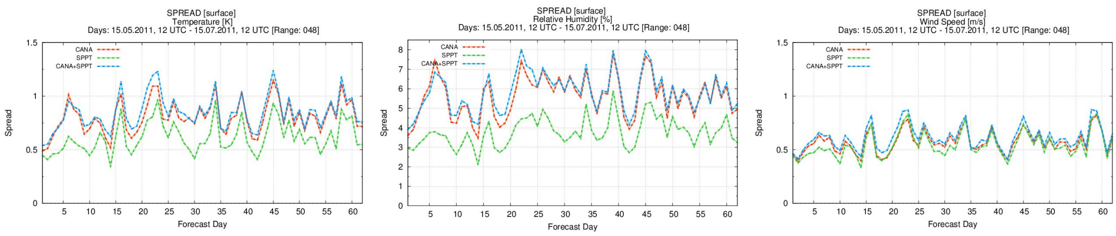
24



36



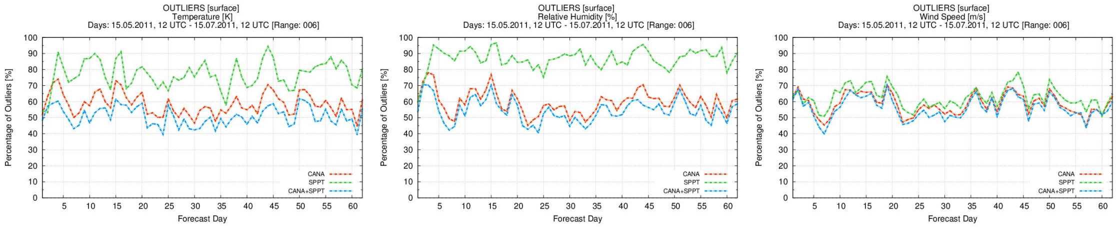
48



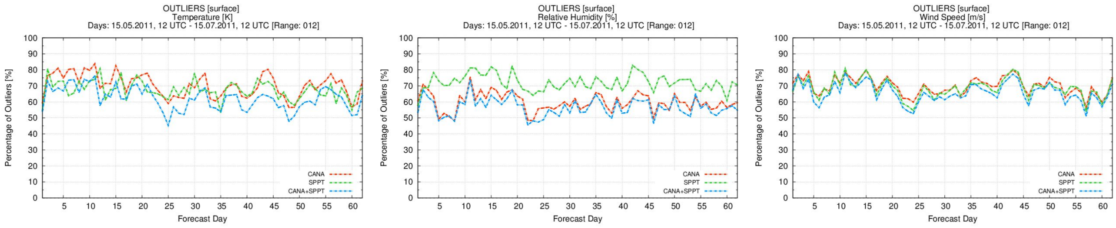
Verification period: 2011-05-15 ~ 2011-07-15 Network: 12 UTC Score: OUTLIERS

Temperature 2m | Relative Humidity 2m | Wind Speed 10m

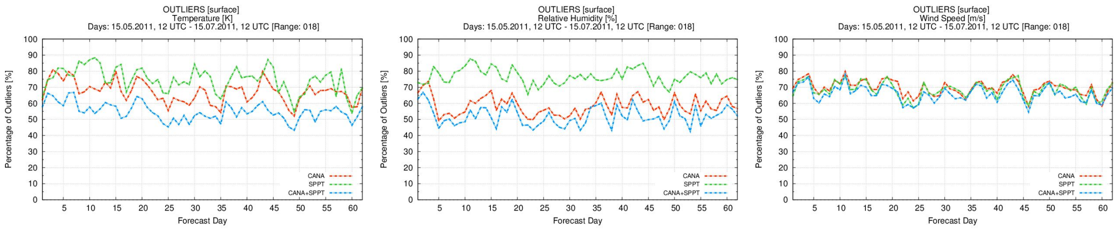
06



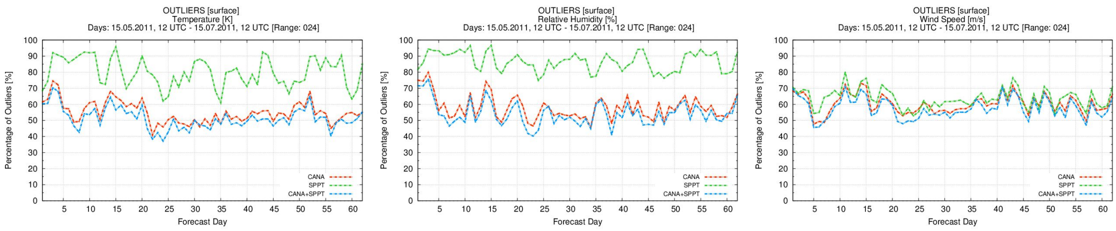
12



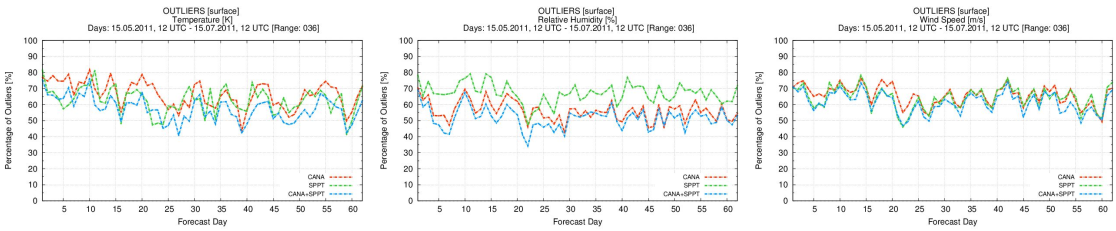
18



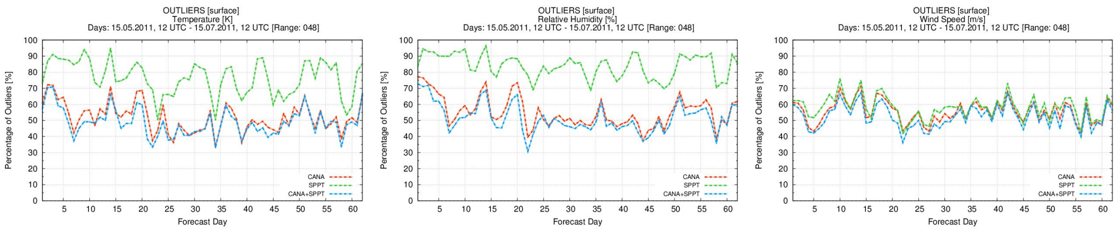
24



36



48



::Conclusions

The SPPT scheme was implemented into ALADIN-LAEF system in order to add a stochastic perturbation into the surface model fields and thus simulate the model uncertainty. Similarly, the ensemble of surface data assimilations by CANARI through the perturbed screen-level observations has to simulate the uncertainty of the initial conditions. The main objective of this study was to confirm the theory, that by the combination of the above uncertainties one can achieve even further enhancement of the regional ensemble system.

Positive impact gained by the combination of IC and model uncertainties, applying both the ensemble of surface data assimilations and the stochastic perturbation of physics tendencies for the surface prognostic variables, can be observed for the screen level temperature, relative humidity as well as for wind speed and mean sea level pressure.

ALADIN-LAEF system with implemented surface SPPT scheme (in cy38t1) in combination with the ESDA proved to be stable and reliable. This was tested for 2 months verification period. The statistical scores have shown some satisfying results. The ensemble system with combined IC and model uncertainties for the surface prognostic fields has bigger spread, less outliers and also smaller BIAS (and in most cases also slightly enhanced RMSE) in comparison with the references, where IC and model were perturbed separately. Thus our conclusion is, that the overall impact is clearly positive for all the monitored surface fields.

::Appendix

Some technical notes - source code and output data location:

(**ECMWF CRAY**)

ALADIN-LAEF perl applications (source code):

```
/home/ms/at/kah/bellus/app_CANASPPT/
drwxr-x--- 4 kah at 4096 Jun 19 11:07 canari/
drwxr-x--- 4 kah at 4096 Jun 16 17:38 laeff/
drwxr-x--- 2 kah at 4096 Jun 22 14:56 setup/

/home/ms/at/kah/bellus/app_SPPT025/
drwxr-x--- 4 kah at 4096 Jun 23 12:16 laeff/
drwxr-x--- 2 kah at 4096 Jun 23 12:19 setup/
```

Perl and Shell scripts for running the experiments:

```
/home/ms/at/kah/bellus/exp/CANASPPT/
-rw-r----- 1 kah at    2094 Jun 19 11:12 canari.cmd
-rw-r----- 1 kah at    3607 Jun 19 11:15 laeff.cmd
-rwxr-x--- 1 kah at    4043 Jun 20 08:34 run.pl
```

-rwxr-x--- 1 kah at   119 Jun 19 11:27 **submit_all.sh**

/home/ms/at/kah/bellus/exp/**SPPT025**/
-rw-r----- 1 kah at 3606 Jun 23 12:31 **laeff.cmd**
-rwxr-x--- 1 kah at 2805 Jun 23 12:38 **run.pl**
-rwxr-x--- 1 kah at 119 Jun 23 12:26 **submit_all.sh**

<u>ICMSH, PF, GRIB data (the results):</u>
ec:/kah/mbell/**CANASPPT**/
drwxrwxr-x   4 kah at 512 Jun 19 11:32 SCC/lae
2011051512/..**2011071512**/
 ICMSHCC<mb>+00<rr>
 PFLAEFCC<mb>+00<rr>

drwxr-x---   2 kah at 512 Jun 23 09:52 **grib**/
-rw-r-----   1 kah at 2374458620 Jun 23 09:51
CANASPPT.2011051512-2011053112.tar.gz
-rw-r-----   1 kah at 4234338062 Jun 23 09:52
CANASPPT.2011060112-2011063012.tar.gz
-rw-r-----   1 kah at 2106705227 Jun 23 09:52
CANASPPT.2011070112-2011071512.tar.gz

drwxr-x---   2 kah at 512 Jun 23 12:00 **seeds**/
-rw-r-----   1 kah at 204800 Jun 23 12:00
CANASPPT.2011051100-2011071512.seeds.tar

ec:/kah/mbell/**SPPT025**/TCC/lae/
2011051512/..**2011071512**/
 ICMSHDW<mb>+00<rr>
 PFLAEFDW<mb>+00<rr>

ec:/kah/mbell/**CANA**/lae/
2011051512/..**2011071512**/
 ICMSHCC<mb>+00<rr>
 PFLAEFCC<mb>+00<rr>

<u>GMK-pack (aladin source code for surface SPPT):</u>
/perm/ms/at/kah/mbell/packs/
drwxr-x--- 7 kah at 4096 Jun 12 09:28 **38t1_sppt**/
algor/module/**spectral_arp_mod.F90** (bugfix)
arpifs/phys_dmn/**mf_phys.F90** (call sppten_isba)
arpifs/phys_dmn/**sppten_isba.F90** (perturb phys.tendencies of ISBA fields)

(**ZAMG archive**)
<u>GRIB data for the verification domain:</u>
 /ment_arch/mproj/bellus/
CANA: ensemble of surface assimilations by CANARI (perturbed

T2m/RH2m OBS)
SPPT_025: run with stochastically perturbed phys.tend. of surface fields: Ts, Ws, Wsi, Wr, Sn, An, Rn (TCC, SPPT sigma=0.25, L=500km, t=2h, DADA)
CANASPPT: ensemble of surface assimilations by CANARI + stochastically perturbed phys.tend. of surface fields: Ts, Ws, Wsi, Wr, Sn, An, Rn (TCC, SPPT sigma=0.25, L=500km, t=2h)

<u>Verification period:</u>
15.05.2011 ~ 15.07.2011, **12 UTC** run - 2 months (CANA, SPPT_025, CANASPPT)