Environnement et Changement climatique Canada

Environment and Climate Change Canada

Global Ensemble Prediction System (GEPS)

Update from version 4.3.0 to version 5.0.0

Canadian Meteorological Center

Technical Note

September 18, 2018

Revisions			
Version	Date	Authors	Remarks
1.0	2018/06/22	X.-X. Deng N. Gagnon	First draft
1.1	2018/08/20	X.-X. Deng N. Gagnon	Add contents to sections 3, 4, 5, 6, 7 and 11
1.2	2018/08/21	X.-X. Deng N. Gagnon P. Houtekamer	Review
1.3	2018/09/17	X.-X. Deng N. Gagnon P. Houtekamer	Add contents to sections 8 (from A&P), 9, 10, 11 and final review
1.4	2018/09/18	N. Bois D. Anselmo	Final review

Table of Contents

Improvements to the Global Ensemble Prediction System (GEPS) from version 4.3.0 to version 5.0.0

Development and Operations Divisions
Meteorological Research Division
at the Canadian Meteorological Center (CMC)
of Environment and Climate Change Canada (ECCC)

Xingxiu Deng[1], Normand Gagnon[1], Pieter L. Houtekamer[2], Stéphane Beauregard[1], Sébastien Chouinard[4], Rabah Aider[2], Martin Charron[2], Juan Sebastian Fontecilla[1], Ronald Frenette[3], and Rochdi Lahlou[4]

1- Development division at the Canadian Meteorological Center of Environment Canada
2- Meteorological Research Division of Environment Canada
3- Severe Weather Lab, Environment Canada
4-Operations division of the Canadian Meteorological Center (CMC) of Environment Canada

SUMMARY

In this implementation, the GEPS is upgraded to the version 5.0.0 which includes a newer version of the GEM model (4.8.LTS13 instead of 4.6.10) used to do the trial fields and the forecasts. For both data assimilation and prediction components, the 800 x 400 global Gaussian grid at 50 km resolution is updated to a Yin-Yang grid at 39 km resolution; the geophysical fields are now generated by using the software GenPhysX instead of the Genesis software. In addition, the assimilation component includes the following changes:

- The digital filter (DF) initialization is replaced by an Incremental Analysis Update (IAU) procedure;
- The model top is raised from 2 hPa to 0.1 hPa and the model has now 81 instead of 74 vertical levels;
- Infrared observations (AIRS, IASI and CrIS) are added;
- Less severe vertical localization is applied;
- Extra quality control with a Huber norm for all observation types is introduced;
- Further thinning is applied for satellite wind, radiance, aircraft and Scatterometer observations.
- The RTTOV operator is upgraded from version 10 to 12;
- Satellite wind observation error is now dynamic.

The assimilation component also benefits from changes in the processing of the observations made to the new GDPS (version 6.1.0) which will be implemented at the same time. For the GEPS prediction component, three additional changes are done: the model top is raised from 2 hPa to 0.1 hPa by means of 45 instead of 41 vertical model levels; a new method is used to generate Markov chains and the products are now generated on a new "user grid"..

Overall, the above changes lead to the improvement of the trial field quality, and better forecast performance of the new GEPS 5.0.0 compared to the previous GEPS 4.3.0.

1 Introduction

Since 1998, CMC has been producing medium range ensemble forecasts operationally. The Ensemble Kalman Filter (EnKF, Houtekamer et al. 2009, Houtekamer et al. 2014) has been used operationally since 2005 to supply initial conditions to the forecast component of the Global Ensemble Prediction System (GEPS, Gagnon et al. 2007, Charron et al. 2010, Gagnon et al. 2011, Gagnon et al. 2013a and b, Gagnon et al. 2014a, Gagnon et al. 2015). With this system, we are producing forecasts extending out to 16 days twice per day using the Canadian Global Environmental Multi-scale model, GEM (Côté et al. 1998a, b). Once a week (Thursday at 0000 UTC), the forecasts are extended to 32 days to produce an outlook for the next 4 weeks (see Gagnon et al. 2013b). The official forecasts of the Meteorological Service of Canada (MSC) for day 6 and day 7 lead times are produced with the outputs of the GEPS. In addition to these official forecasts, there are many products for the 1-15 day lead times which are available at: http://weather.gc.ca/ensemble/index_e.html (MSC public weather web site).

It is also noteworthy to point out that the GEPS forecasts are an integral part of the collaboration with the United States National Centers for Environmental Prediction (NCEP) in the North American Ensemble Forecasting System (NAEFS) project (Candille, 2009, Candille et al. 2010 and Cui et al. 2012). Products from this super-ensemble are available on the MSC public weather web site: http://weather.gc.ca/ensemble/naefs/index_e.html, as well as on the NCEP Environmental Modeling Center web site: http://www.emc.ncep.noaa.gov/gmb/ens/NAEFS.html and many other sites on the Internet.

Starting from the implementation of version 4.0.0 (Gagnon et al. 2014a), the GEPS system is coupled in a 2-way mode with the Global Deterministic Prediction System (GDPS). It still depends on the GDPS for: 1) initial surface fields; and 2) the processing of observations (background check, quality control and bias correction). Meanwhile, the GDPS is making use of the EnKF trial fields to calculate the background-error covariances in the 4DEnVar – data assimilation component of the GDPS (see Buehner et al. 2015 and Buehner et al. 2014).

The GEPS also provides the initial and boundary conditions for the Regional Ensemble Prediction System (REPS, see Gagnon et al. 2014a, Erfani et al. 2013, Charron et al. 2011). The

REPS is producing up to 72 h operational ensemble forecasts over North America at 15 km horizontal resolution.

It has to be noted that there were two minor upgrades to the GEPS system since last major implementation documented by Gagnon et al. 2015. Firstly, on 16 March 2016, the GEPS was upgraded from version 4.1.1 to 4.2.0 with minor changes to the GEPS assimilation component only. The changes included additional Atmospheric Motion Vector (AMV) data from Visible Infrared Imaging Radiometer Suite (VIIRS) on board the NPP satellite, additional marine winds from Rapidscat Scatterometer mounted on the International Space Station (ISS), and replacement of MTSAT-2 AMVs by those of HIMAWARI-8 (see the official note for more details at http://dd.meteo.gc.ca/doc/genots/2016/03/15/NOCN03_CWAO_151730___00994). Secondly, on 6 September 2017, the GEPS version 4.2.0 was replaced by version 4.3.0 as a result of a major migration to a new supercomputing infrastructure (see the official note for more details at http://dd.meteo.gc.ca/doc/genots/2017/08/31/NOCN03_CWAO_312003___11256).

This document describes the latest modifications to the Canadian Global Ensemble Prediction System, becoming GEPS version 5.0.0, hereafter referred to as the GEPS 5.0.0. In Sections 2, 3 and 4, we present modifications to the assimilation and prediction components of the GEPS in support of the operational implementation of September 2018 to replace GEPS version 4.3.0. Section 5 describes modifications to the reforecast procedure. Evaluation results of the final series test and the parallel run are detailed in Sections 6, 7 and 8. Information is given in Section 9 on performance of dependent systems and in Section 10 on availability of products. Finally, we summarize results in Section 11.

The current implementation represents cumulative efforts of many people not only from the GEPS team, but also from other groups such as the modelling and post-processing sections in research and development divisions, and operational sections of Environment and Climate Change Canada (ECCC) co-located at the Canadian Meteorological Center (CMC) in Dorval, Québec.

2 Modifications common to the assimilation and prediction components of the GEPS

The version of GEM model was upgraded from 4.6.10 to 4.8-LTS13. This upgrade is done for the model that is used to produce the trial fields in the assimilation part and the forecasts in the prediction component. Since the GEM version 4.7.0, a Yin-Yang grid has been introduced to obtain a more uniform grid resolution in a global model and also to increase computational efficiency with a better parallel scalability (Qaddouri and Lee 2011). A blending technique is applied to the overlapping regions of the two limited-area model domains of the Yin-Yang grid (Qaddouri 2015). The trajectory calculation method has been improved as well (Girard 2015, Qaddouri et al. 2015). With the GEM version 4.8, the Gaussian grid is no longer supported. As a

result, the 800 x 400 global Gaussian grid at 50 km resolution is updated to a Yin-Yang (YY) grid at 39 km resolution. With EnKF, the analysis increments are computed directly on the Yin-Yang grid of the model. The impact of moving to YY 39 km is slightly positive in ARCAD verification against radiosonde observations below 300 hPa (Figure 1), and also positive in the verifications against AMSU-B, Aircraft winds/temperature/dew point depression, Scatterometer winds, surface data, and GPS-RO below 4 km (figures not shown). For both data assimilation and prediction components, the geophysical fields are now generated by using the software GenPhysX instead of the Genesis software. Related to this, we note a slight degradation for the surface fields especially winds in both assimilation and forecasts. Nonetheless, GenPhysX is continuously in active development whereas Genesis is gradually phasing out.

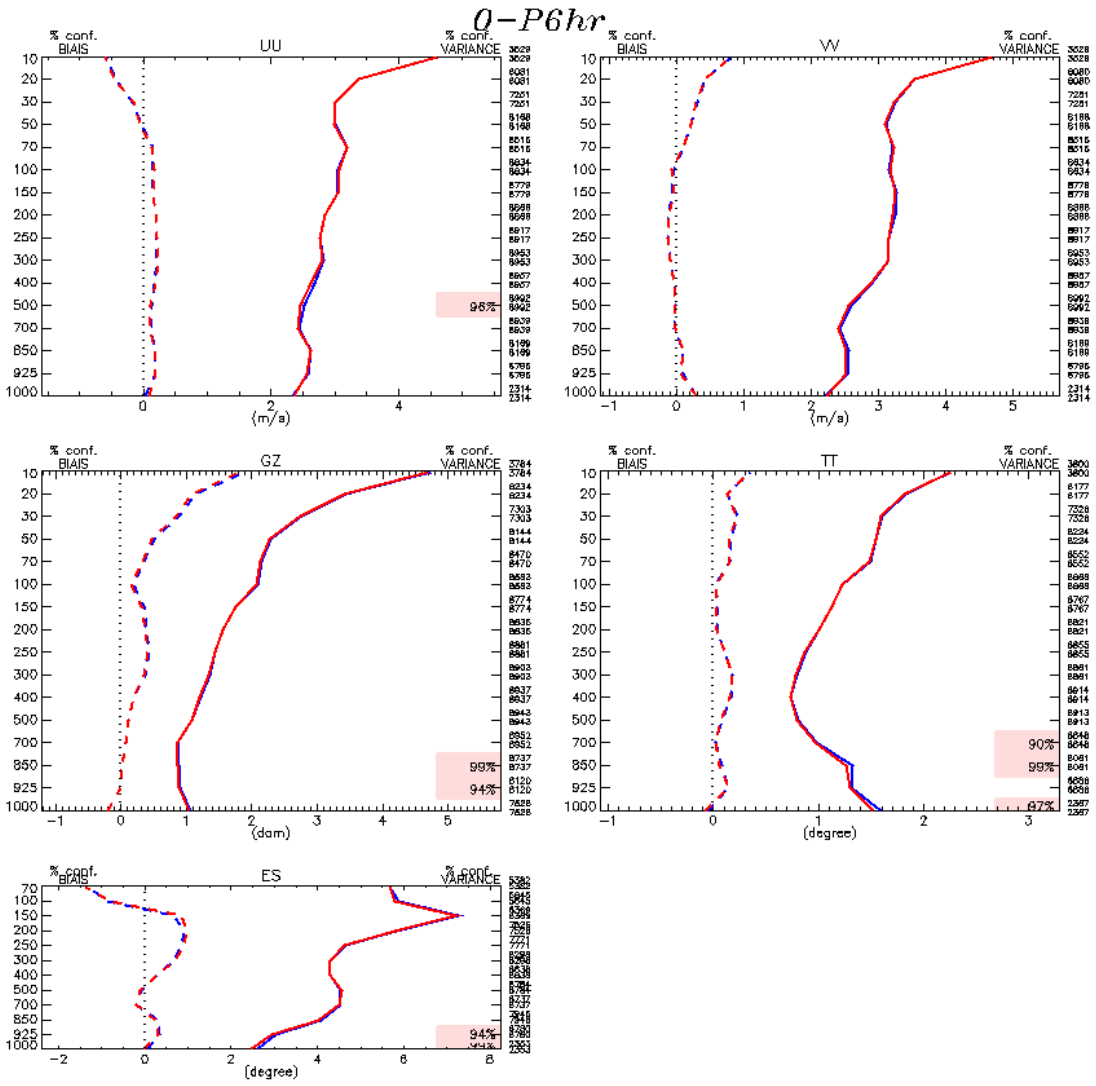


Figure 1. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosonde observations averaged globally over a 10-day period in January 2015 for EnKF experiments with GU grid at 50 km (blue) and YY grid at 39 km (red).

3 Modifications to the assimilation component of the GEPS (EnKF)

Besides the above-mentioned modifications common to the assimilation and prediction components of the GEPS, described below are modifications specific to the assimilation part.

3.1 New initialization method

The EnKF analyses are normally not in optimal balance between mass and motion fields. To minimize the shock due to an imbalanced initial condition for a numerical integration, an initialization procedure is usually applied. With this implementation, the digital filter (DF) initialization (Fillion et al. 1995) is replaced by an incremental analysis update (IAU) procedure (Bloom et al. 1996), which gradually incorporates analysis increments into a NWP model. The IAU technique was first introduced to the GEM model in 2015 and was implemented in the GDPS version 4.0.0 (Buehner et al. 2015). Please note that the IAU technique is used in the GEPS data-assimilation component (EnKF) only when integrating the GEM model to produce first-guess fields. The GEPS prediction component still utilizes the DF initialization method. In the EnKF, the analysis increments are introduced to the GEM model employing a sinusoidal time weighting over a 6-h window with peak at the central time of the DA window (see Figure 2). Figure 3 shows that the gravity wave activity is much lower with IAU than with DF. As in 4DEnVAR of the GDPS system, the EnKF system also recycles some key dynamics and physics variables between integrations: virtual temperature, horizontally staggered winds, cloud condensate, and a few variables related to the boundary layer parameterization. The biggest impact brought by IAU is the bias reduction of surface pressure (hence geopotential heights as seen in ARCAD verification) especially in the Tropics (figure not shown).

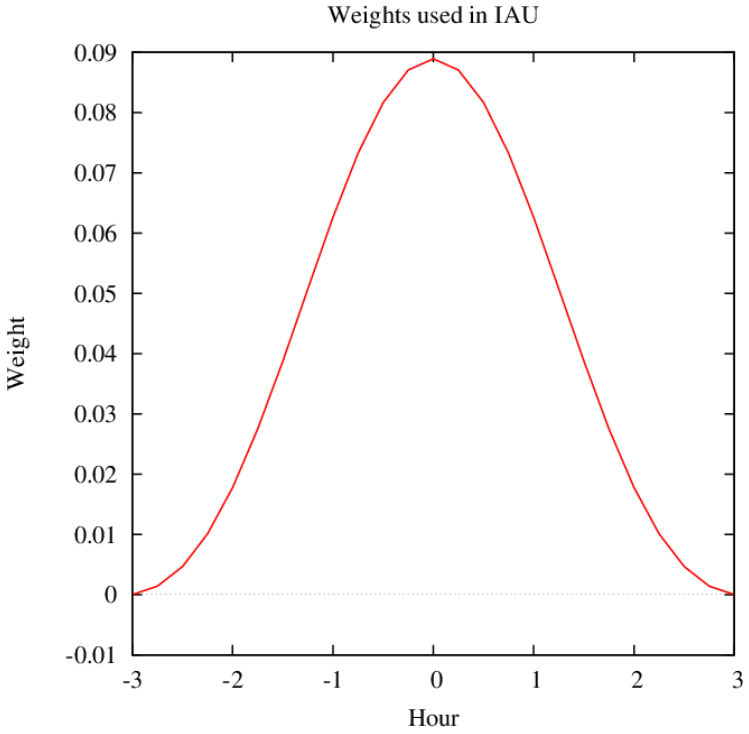


Figure 2. Weights used in IAU over a 6-h window.

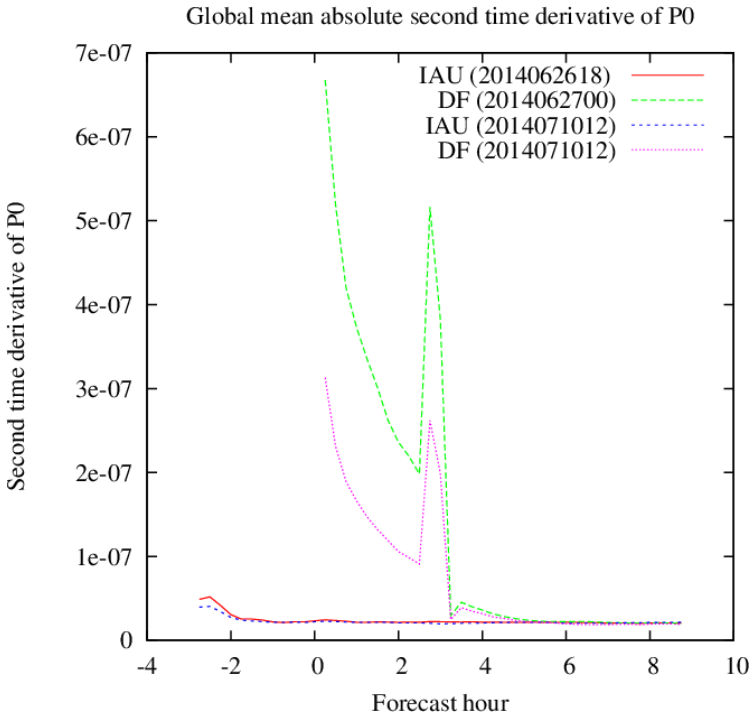


Figure 3. Global mean absolute second time derivative of surface pressure at the beginning of the EnKF spin-up (2014062618 for IAU and 2014062700 for Digital Filter) and 10 days after 4-day spin-up (2014071012 for IAU and DF).

3.2 Higher model top with more vertical levels

The model top is raised from 2 hPa to 0.1 hPa. As a result, the number of vertical levels is increased from 74 to 81. The distribution of vertical levels (Figure 4) is the same as that for the GDPS system (Qaddouri et al. 2015). As mentioned earlier, the GDPS at our center uses the trials from EnKF to calculate flow-dependent background error covariances for its 4DEnVar, which is run on a grid of 81 vertical levels with model top at 0.1 hPa. This modification facilitates a better use of the EnKF trial fields in the GDPS. Also this update allows assimilating upper channels of radiance observations (i.e., AMSU-A channels 13 - 14, and ATMS channels 14 – 15). This brings a reduction in biases and standard deviations of trial errors for AMSU-A channels 11 - 12 and ATMS channels 12 -13 (figure not shown). The impact of this update on the quality of the EnKF trials is positive in the higher troposphere/lower stratosphere (see ARCAD verifications against radiosonde observations in Figure 5 and the verifications against GPS-RO observations in Figure 6).

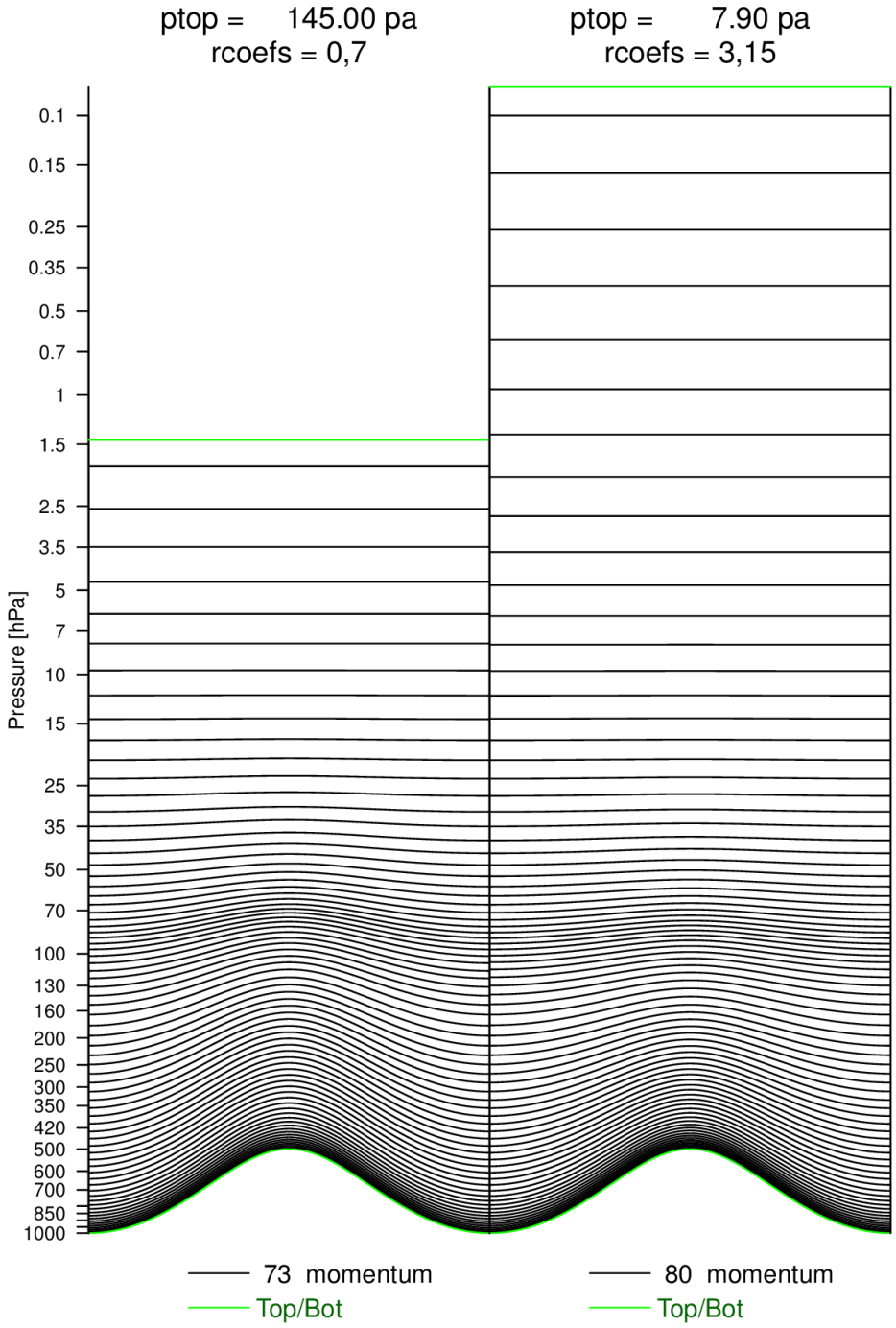


Figure 4. Distribution of vertical levels for the EnKF in the GEPS system version 4.3.0 (left) and version 5.0.0 (right).

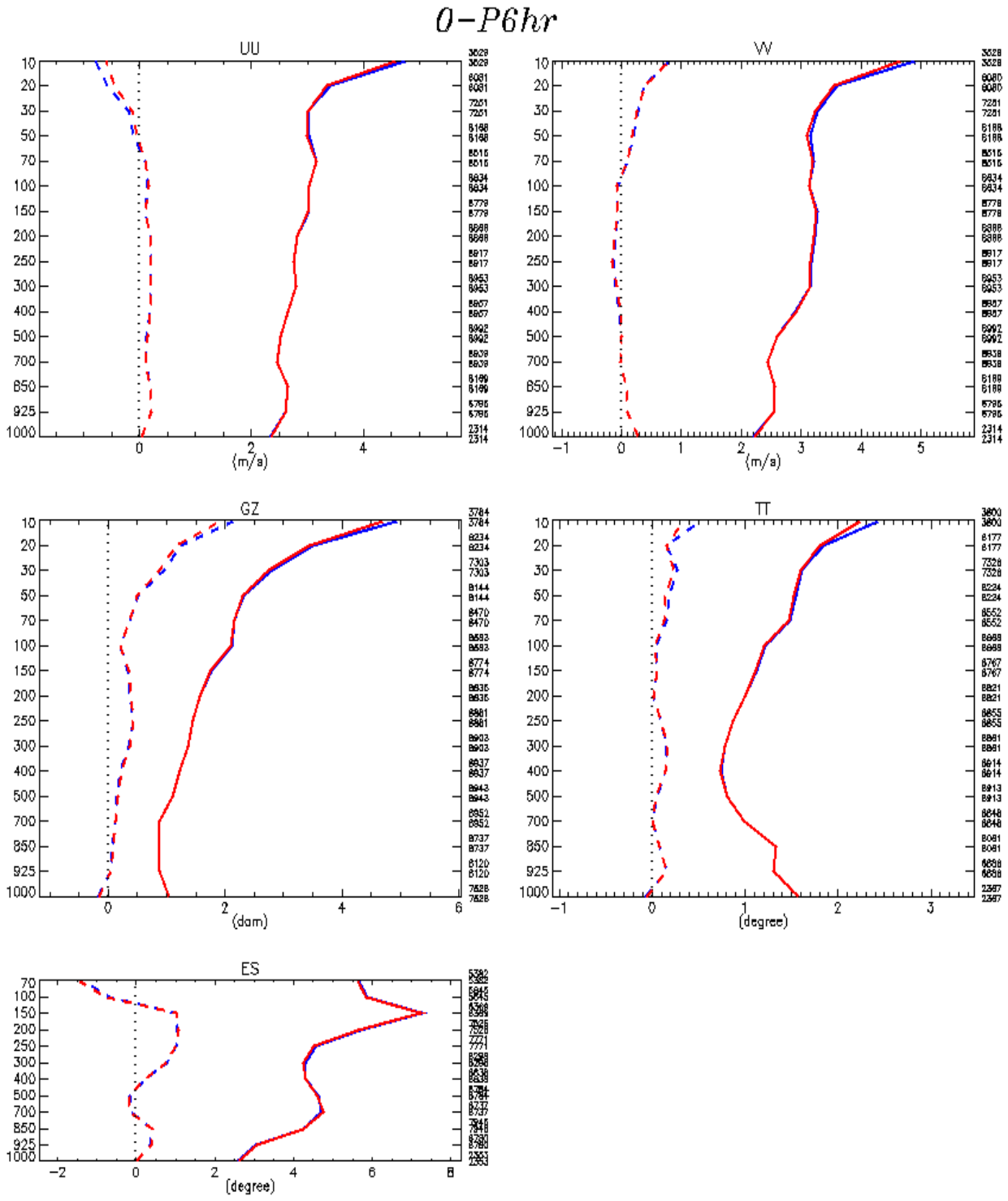


Figure 5. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosonde observations averaged globally over a 10-day period in January 2015 for the EnKF experiments with model top at 2 hPa (blue) and at 0.1 hPa (red).

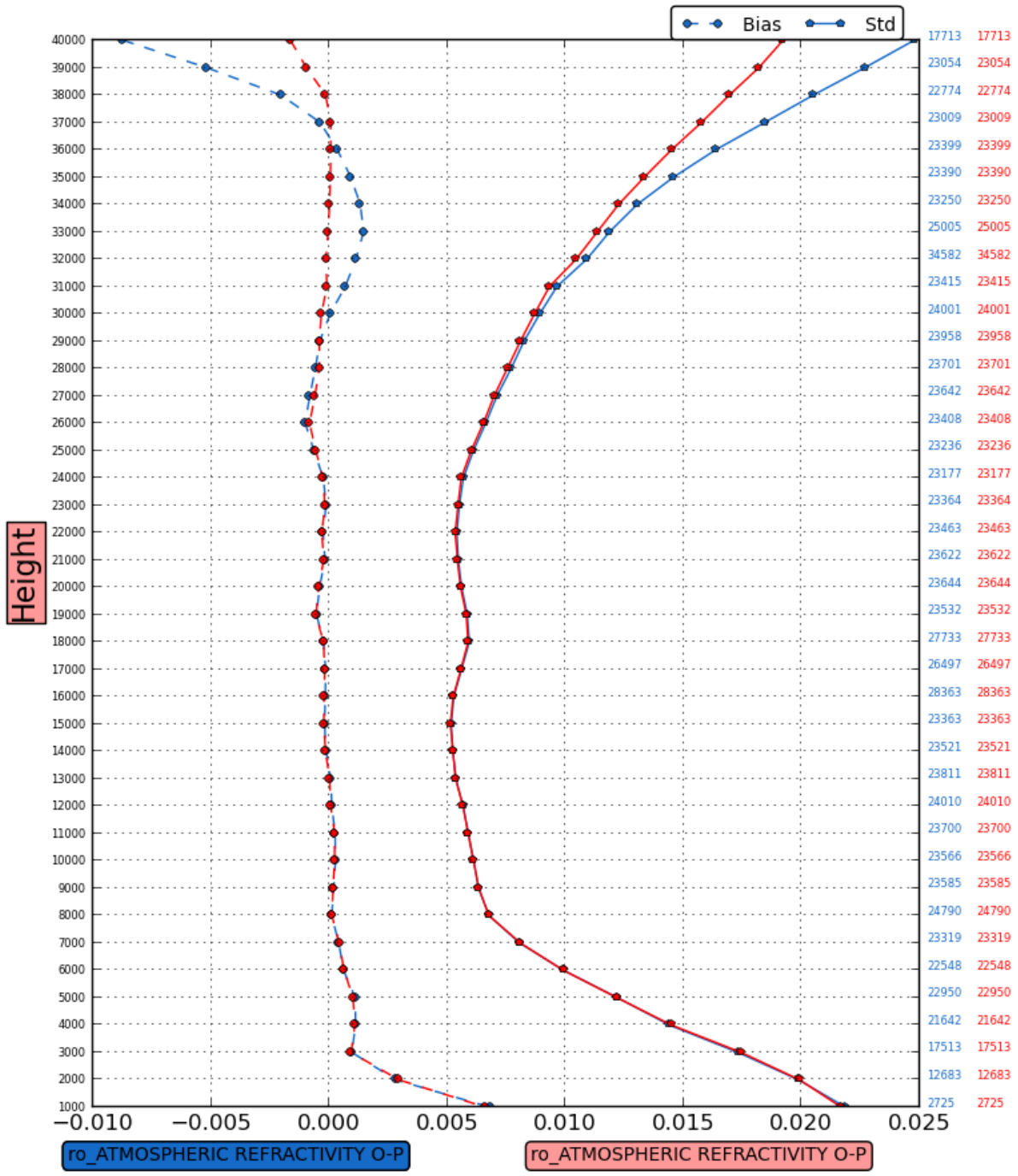


Figure 6. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against GPS-RO globally averaged over a 10-day test period in January 2015 for the EnKF experiments with model top at 2 hPa (blue) and at 0.1 hPa (red).

3.3 Infrared observations (AIRS, IASI and CrIS) are added

In the new EnKF, 14 channels of AIRS (from AQUA), IASI (from METOP-A and METOP-B) and CrIS (from NPP) are assimilated (Table 1). Observations in Arctic and Antarctic regions are rejected. Unfortunately, there are no experiments for a clean comparison for this addition.

Table 1 List of the channels of infrared observations assimilated in the new EnKF.

instrument	List of channels assimilated
AIRS	145, 150, 162, 174, 190, 198, 215, 227, 232, 1477, 1545, 1583, 1826, 1708
IASI	185, 199, 205, 226, 239, 269, 323, 379, 381, 2951, 2993, 3049, 3263, 4920
CrIS	75, 80, 85, 90, 97, 116, 136, 148, 151, 798 , 814, 842, 844, 889

3.4 Less severe vertical localization is applied

The vertical localization will relax analysis increments to zero at a vertical separation of 3 instead of 2 units of ln(P) from an observation. The modification permits making better use of radiance observations. In terms of ARCAD verification, the impact of this modification is neutral.

3.5 Extra quality control with a Huber norm for all observation types

As in the GEPS operational version (4.3.0), the EnKF still obtains bias-corrected, background-checked, and some quality-controlled observations from the GDPS, which is still coupled with the GEPS to get background error covariances for its 4DEnVar. But, because of timing, the EnKF is not using GDPS postalt files that have undergone Variational Quality Control (QC-Var). To make the EnKF more robust against outlying observations, extra quality control with a Huber Norm (Tavolato and Isaksen 2010) is introduced in the new EnKF for all observation types. In the new EnKF, the background check is performed for all data types instead of Satellite Wind only, with a less restrict reject criterion being 5 times (instead of 2 times) of the square root of the sum of the square of background error and square of observation error. For the observations with innovation between 2 times and 5 times of the square root of the sum of the squre of background error and square of observation error, the observations will be “Huberized”, namely, the observation errors will be increased thus the weight given to these observations will be decreased. With this change, we see slight improvement for winds in Troposphere (figure not shown).

3.6 Further thinning is applied for satellite wind, radiance, aircraft and Scatterometer observations.

Further thinning is performed in the new EnKF for various data families including Aircraft (AI), Scatterometer (SC), Satellite Wind (SW), and radiance observations (TO). This is done by limiting the maximum desired number of data, which is 10 for AI & SC and 6 for SW & TO, per 3-D box. The box is defined with approximately 200 km from the center to each side of the box.

Any more data than desired will be removed randomly. Figure 7 shows an example of thinning aircraft observations over North America where data is very dense. Please note that the color bar is different between the top and middle panel.

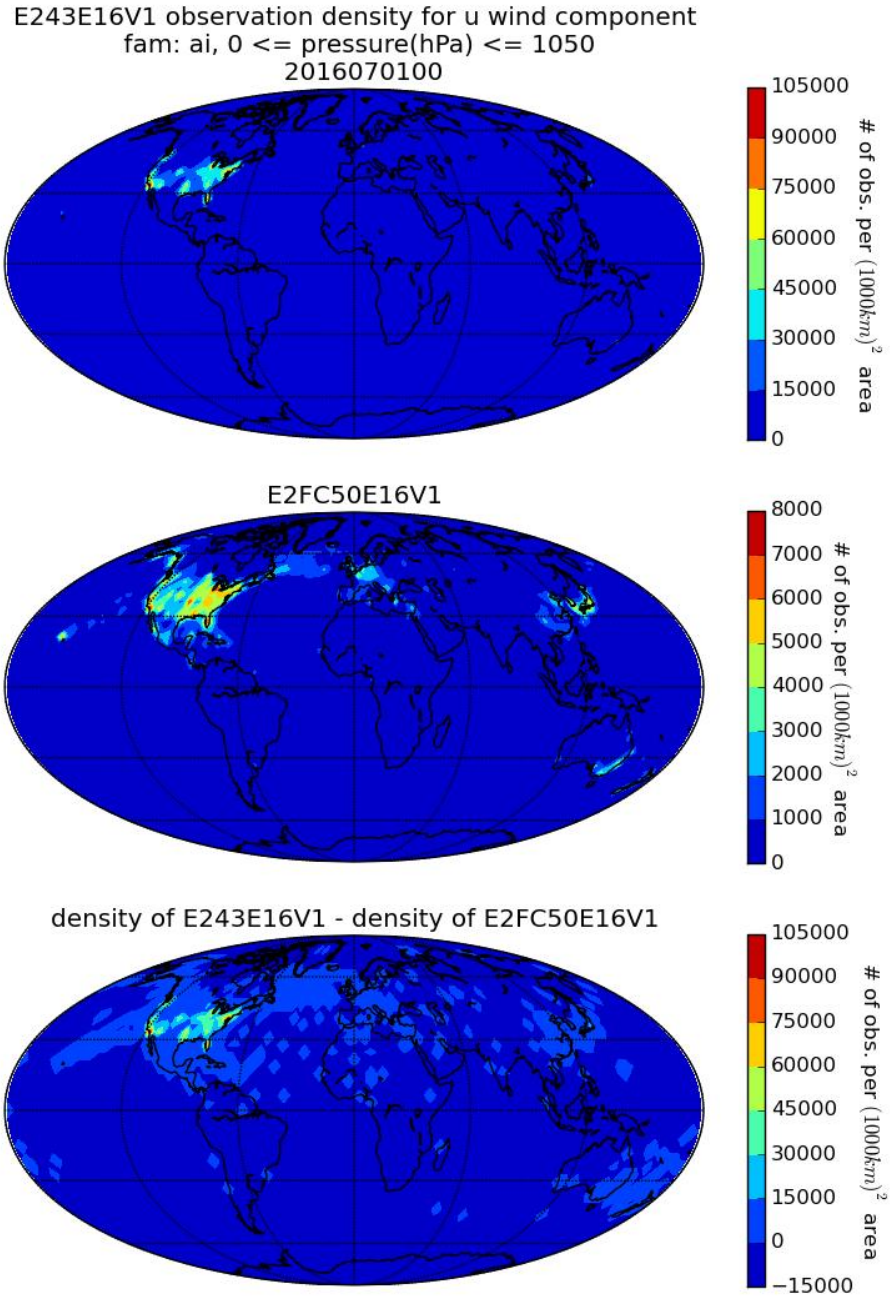


Figure 7. Number of Aircraft data assimilated in the GEPS 4.3.0 (top) and the GEPS 5.0.0 (middle) and the difference between the GEPS 4.3.0 and 5.0.0 (bottom).

3.7 Changes common to the assimilation components of the GEPS 5.0.0 and of the GDPS 6.1.0

As mentioned earlier, the EnKF is two-way coupled with the 4DEnVar of the GDPS, thus any change in the processing of the observations in the new GDPS will benefit the EnKF as well. In fact, many changes in the processing of the observations are made in the CMC's GDPS 6.1.0 to be implemented concurrently (see the technical note of the GDPS 6.1.0). Among all the changes, the most important having an impact on EnKF performance include significant improvement in the processing of radiosonde observations and of Atmospheric Motion Wind Vector (AMV), as well as a new rejection criterion for humidity observations. Removing humidity observations from stations (especially over Siberia) with large biases also has positive impact on the ARCAD verifications against radiosonde observations (see section 7.1). We note a decrease in the data count for AMV winds due to a new rejection of AMV data above 200 hPa (160 hPa over the Tropics) instead of 50 hPa. For radiosonde observations, we observe a significant increase in the data count for winds and temperature because of using native BUFR, which includes much higher vertical resolution and much higher accuracy for the reporting time and position. Several changes made to the new EnKF are in common with those made to the4DEnVAR in the GDPS 6.1.0. These include:

- The upgrade of RTTOV operator from version 10 to 12;
- The use of new physics functions to compute dew point depression;
- The use of dynamical satellite wind observation errors.

4 Modifications to the forecast component of GEPS

In addition to the changes described in section 2, three extra modifications were made to the forecast component of the GEPS.

4.1 Addition of 4 vertical levels

As described in section 2, the model top was raised to 0.1 hPa from 2 hPa. To cover that new extent we have added 4 additional levels at the top (now 45 instead of 41). Following this modification, no big impact was noticed in the forecast skill for altitude lower than 100 hPa (not shown). The change has allowed us to lift the model lid with minimal additional computer cost.

4.2 New method to generate the random numbers used in stochastic perturbations

A new technique to generate the random numbers used in the Physical Tendency Perturbations (PTP) and the Stochastic Kinetic Energy Back Scattering (SKEB) parameterization was introduced. Now double Markov chains are used which lead to smoother behavior in time. Also, the normalisation (stretching) of the probability density function is now done using the Error function instead of the ad hoc limitation procedure described in Charron et al. 2010. For more

details on the new normalisation, see Charron (2016). The PTP perturbation fields are now produced on a 64x32 Gaussian grid instead of a 16x8 grid, which gives smoother fields in space.

To adapt for the new normalisation procedure and the higher resolution grid, we have to reduce the stretching parameter from 1.39 to 1.10 in order to maintain the same magnitude of perturbations. All these modifications have induced very small changes in performance but the new code is simpler and it should be easier to maintain.

The use of a new type of grid (Yin-Yang instead of the lat-lon regular grid) has an impact on the SKEB parameterization near the location of the computational poles in the lat-lon grid because there is now, having no computational poles in the Yin-Yang grid, less numerical diffusion (a source term for SKEB). The grid of the previous GEPS has a convergence of points near the computational poles (along meridian) which leads to high resolution, more noise and then higher numerical diffusion. With the Yin-Yang grid, this problem has disappeared inducing a weaker forcing for SKEB in these regions. This lowering of the wind perturbations is positive because generally the wind forecast spread was too large compared to the forecast error. As will be discussed in section 6, the spread/error balance is improved. In light of these results, the SKEB alpha parameter for the perturbation amplitude remains unchanged at 0.5.

4.3 Outputs on a additional 'user grid' for product generation

The global ensemble prediction system produces output on the GEM model “native” grid. This output is now also interpolated to a “user” grid for the production of forecasting products. The main advantage of using a user grid is to allow for model updates that are independent of the generation of forecast products for users.

The user grid for the GEPS has a horizontal resolution of 0.25 degrees, with grid points at the poles and on the equator (1440x721). Note that the NAEFS exchange currently uses a resolution of 0.5 degrees for the products that are available to external users on the MSC Datamart (more details here https://weather.gc.ca/grib/index_e.html).

The GEPS user grid is associated with 3 different vertical coordinates:

1) on pressure coordinates for the following levels (hPa)

[10 20 30 50 70 100 150 175 200 225 250 275 300 350 400 450 500 550
600 650 700 750 800 850 900 925 950 970 985 1000 1015]

2) on eta coordinates for these 62 levels (internal format IP2) values:

[2000 2102 2233 2374 2508 2625 2720 2795 2852 2897 2941 2990 3044 3104 3172
3248 3334 3431 3541 3667 3812 3976 4149 4331 4522 4721 4928 5144 5369 5602
5843 6091 6348 6612 6883 7161 7446 7737 8034 8337 8646 8959 9272 9567 9845

10104 10346 10571 10780 10973 11151 11316 11467 11606 11670 11733 11792 11850 11900 11950 11975 12000]

3) on height coordinates for the following 9 levels in meters above surface (only temperature, wind components for now)

[20m 25m 40m 50m 80m 100m 120m 150m 200m]

5 Modifications to the reforecast procedure

In addition to the changes described in sections 2 and 4, the reforecast procedure period was updated from 1995-2014 to 1998-2017. It still covers 20 years with 4 members for one date per week (Thursdays) to generate a historical database of 80 reforecasts for a given date as explained in Gagnon et al. (2014b). Please note that the initial surface fields were re-generated on the new Yin-Yang 39 km grid using the Surface Prediction System (Carrera et al. 2010) driving the ISBA surface scheme forced by ERA-interim re-analyses (Dee et al. 2011).

6 Objective evaluation of the final series of tests

In this section, we present the evaluation of the performance of the new system GEPS 5.0.0 against the former operational system GEPS 4.3.0 during the final series of tests. The final series of tests were started with 4-day uncoupled EnKF runs (to allow covariances from the EnKF to have realistic flow dependence) followed by another 7-day runs coupled with GDPS 4DEnVar (to allow the coefficients of dynamic satellite bias correction to settle down on stable values) and then another two and half month runs for verification. To evaluate the impact of all the changes described above on the analyses and also on the forecasts for the first 15 days, the objective verifications were done for summer 2016 and winter 2017.

6.1 Quality of initial conditions

To measure the quality of trial fields (6-h forecasts) produced by the EnKF, the innovation statistics have been calculated using the ensemble mean trial field and the 3D radiosonde observations (the so-called ARCAD verifications). The verifications against GPS-RO and radiance data were also done. The averaged results over different regions including the globe, Northern Extratropics, Southern Extratropics and Tropics are calculated for the winter (from 21 December to 28 February 2017) and summer (from 20 June to 31 August 2016) cycle respectively.

For both the summer 2016 and the winter 2017, the ARCAD verifications globally show nice improvement for Troposphere (largely due to moving to YY 39 km) and upper Stratosphere

(mainly due to raising the model top) but neutral impact from 200 to 50 hPa (see Figure 8 for summer, figure for winter with similar features not shown). We observe an apparent degradation for dew point depression in standard deviation above 250 hPa as well as in bias from 500 hPa to 300 hPa. However, over Southern Extratropics we see improvement in standard deviation for dew point depression (Figure 9 for summer, figure for winter with similar results is not shown). Apparently, the degradation in standard deviation for dew point depression in global verification originates from the Northern Extratropics. The ARCAD verifications for the final series of tests were performed against radiosonde observations from the operational GDPS quality-controlled "postalt" files. It was discovered that radiosonde humidity observations from some stations, especially over Siberia, have large biases in the higher troposphere/lower stratosphere. This helps explain why we see degradation in dew point depression in the Northern Extratropics. We also note that the improvement for winds and Geopotential Heights are bigger in the Southern Extratropics than in the Northern Extratropics. Over Tropics, we note a very big decrease in the biases of Geopotential Heights, which is mainly associated with the use of IAU (figure not shown). The SQLITE verifications against each system's own assimilated radiosonde observations show big improvement in winds (Figure 10), temperature and dew-point depression (figure not sure). The big positive impact is due not only to the changes in the new EnKF but also to the improved quality of radiosonde observations from the new GDPS system and increased number of observations (see data count comparison in Figure 10). Figure 11 shows significant reduction in standard deviation of the surface temperature and a decrease in the data count for the new EnKF. This is mainly due to the Huber norm quality control, as well as a more restrict rejection criterion (being 200 instead of 800 meters) for the difference between the model topography and the observation height. Decreased data count and significant improvement in standard deviation are also found in both seasons for other surface variables including surface winds, surface dew-point depression, and especially surface pressure and mean sea level pressure (MSLP) (figures not shown). For surface pressure and MSLP, the biases are also greatly reduced with the use of IAU as a contributor. The verifications against radiance observations show a big improvement for AMSU-B (in both bias and standard deviation especially in Summer 2016), ATMS channels 9-13 and AMUB-B like channels (mainly in standard deviation), AMSU-A channels 9-12 in standard deviation (in summer 2016, but channels 7-12 in winter 2017, see Figure 12). The improvement could attribute to the update to YY 39 km, higher model top, upgraded RTTOV operator, Huber norm quality control and further thinning to radiance observations (confirmed by the much reduced data count in the new EnKF, as seen in Figure 12).

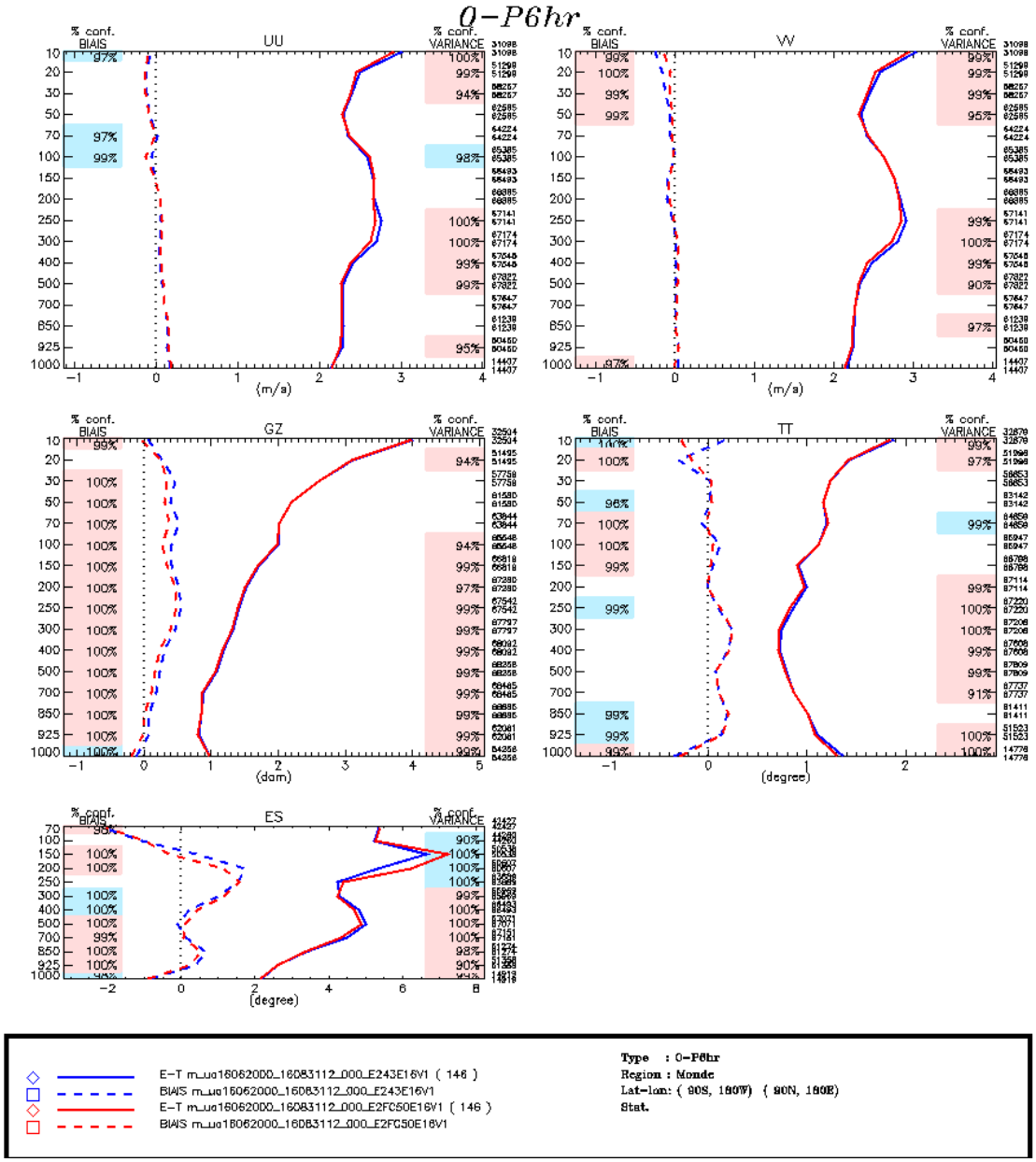


Figure 8. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the globe in the summer 2016 for the EnKF experiments with GEPS 4.3.0 (blue) and GEPS 5.0.0 (red).

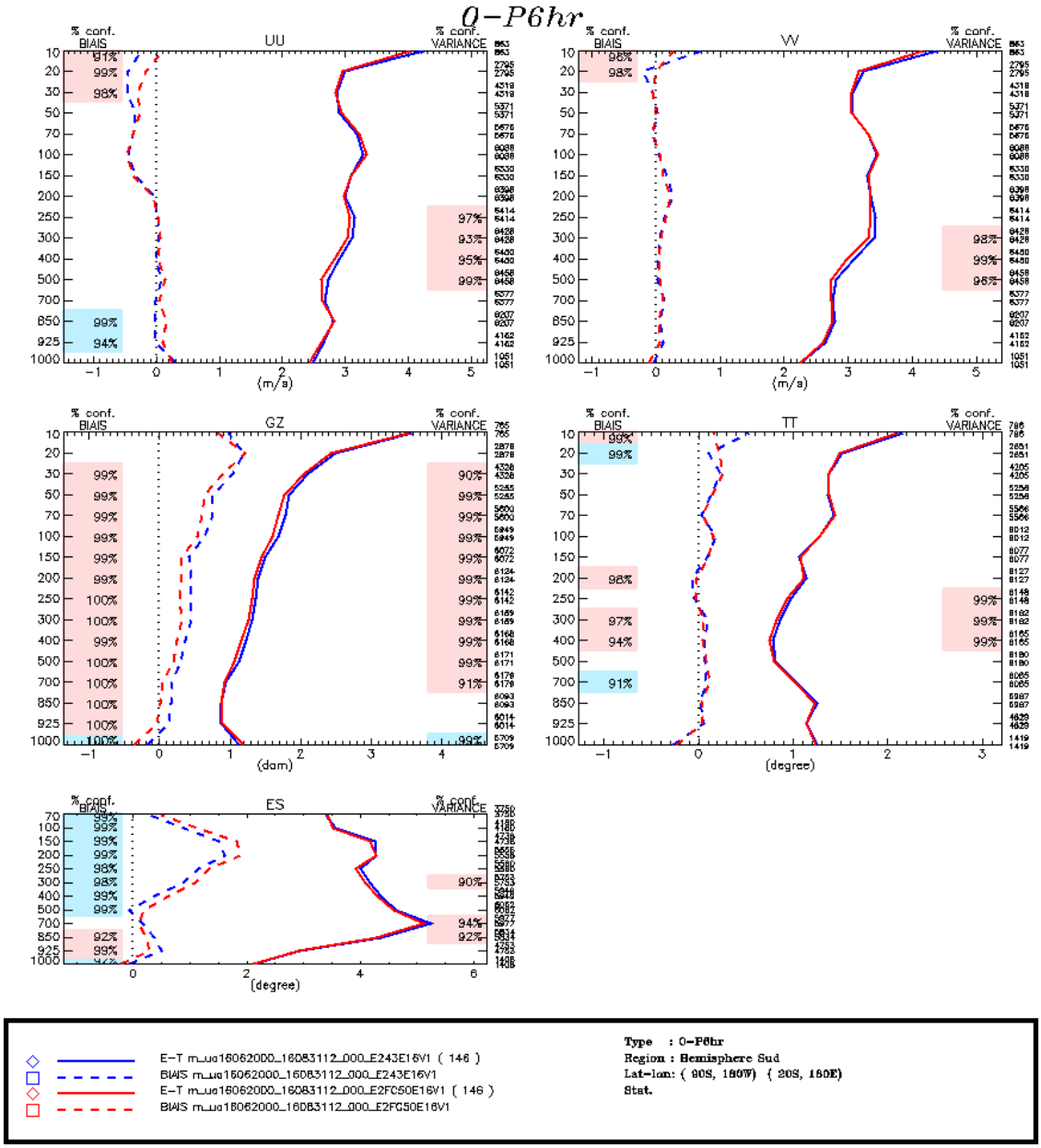


Figure 9. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the Southern Extratropics in the summer 2016 for the EnKF experiments with GEPS 4.3.0 (blue) and GEPS 5.0.0 (red).

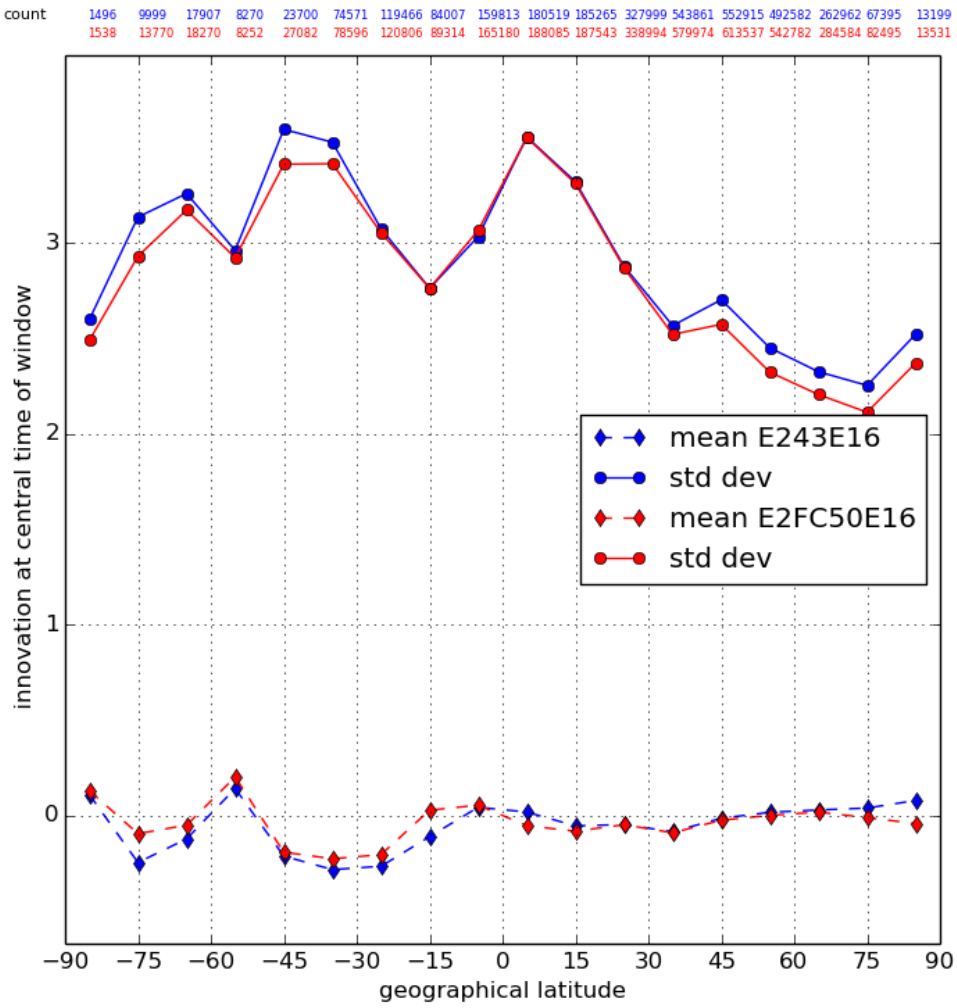


Figure 10. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes in the summer 2016 for the EnKF experiments with GEPS 4.3.0 (blue) and GEPS 5.0.0 (red).

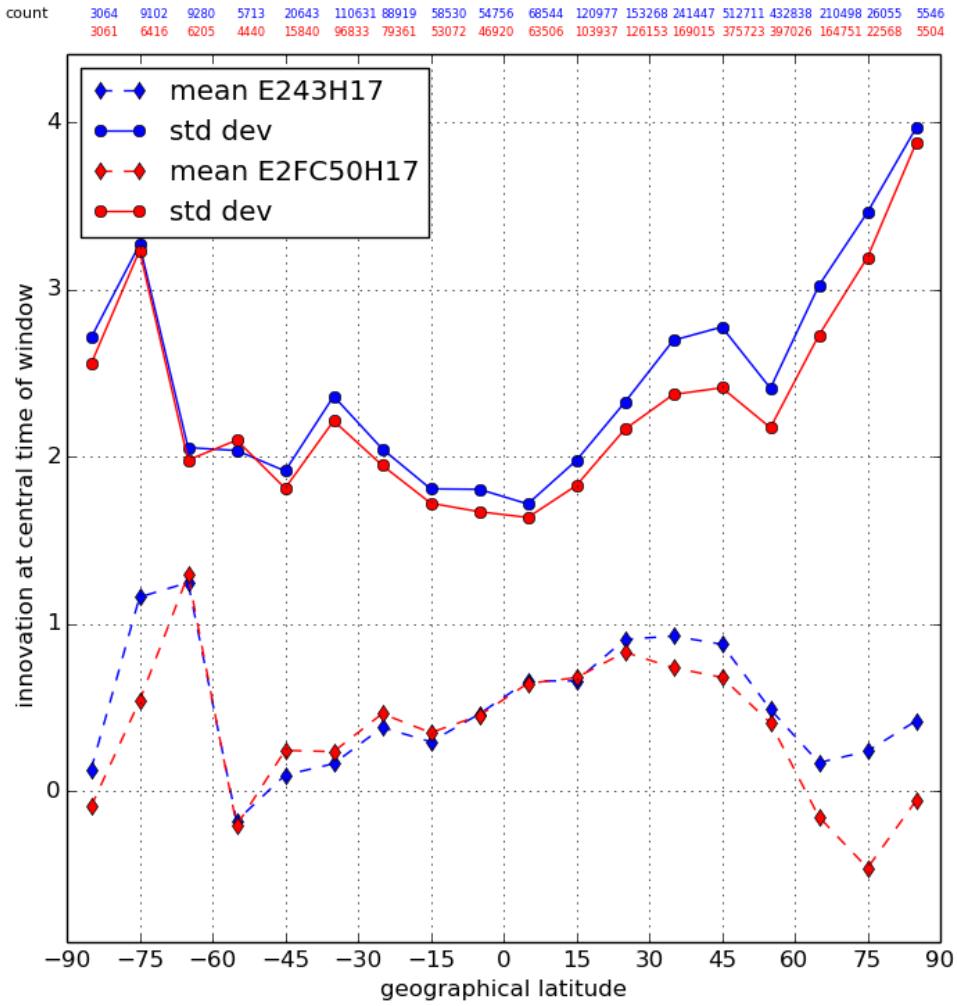


Figure 11. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against surface temperature in the summer 2016 for the EnKF experiments with GEPS 4.3.0 (blue) and GEPS 5.0.0 (red).

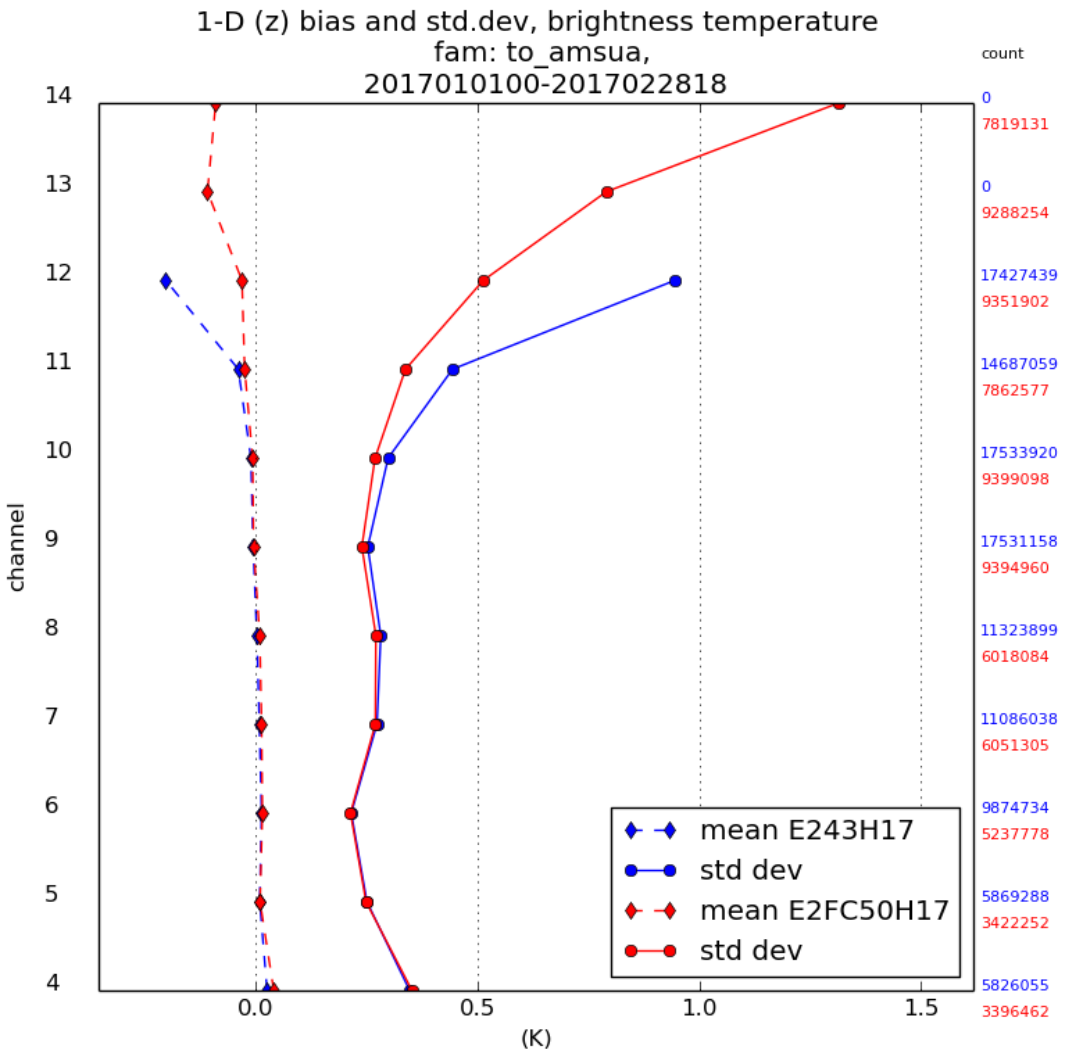


Figure 12. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) in the winter 2017 over the globe against AMSU-A for the EnKF experiments with GEPS 4.3.0 (blue) and GEPS 5.0.0 (red).

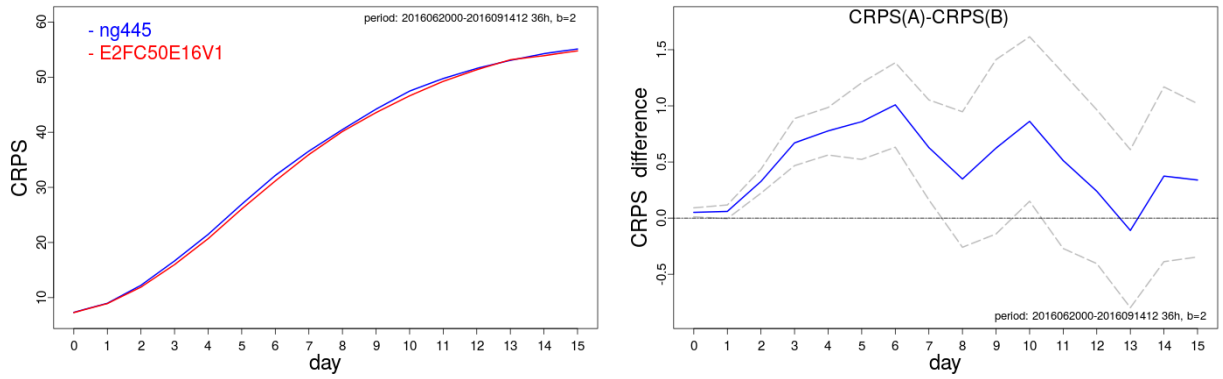
6.2 Quality of the GEPS forecasts: upper air scores

For the summer season, 15-day forecasts were generated at every 36 hours for the period from 2016061900 to 2016083112 which gives 50 cases. In winter, the period from 2016122000 to 2018022812 was used (48 cases). The forecasts were compared to the observations taken by the global radiosonde network (approximately 600 stations) at seven pressure levels: 10, 50, 100, 250, 500, 850 and 925 hPa.

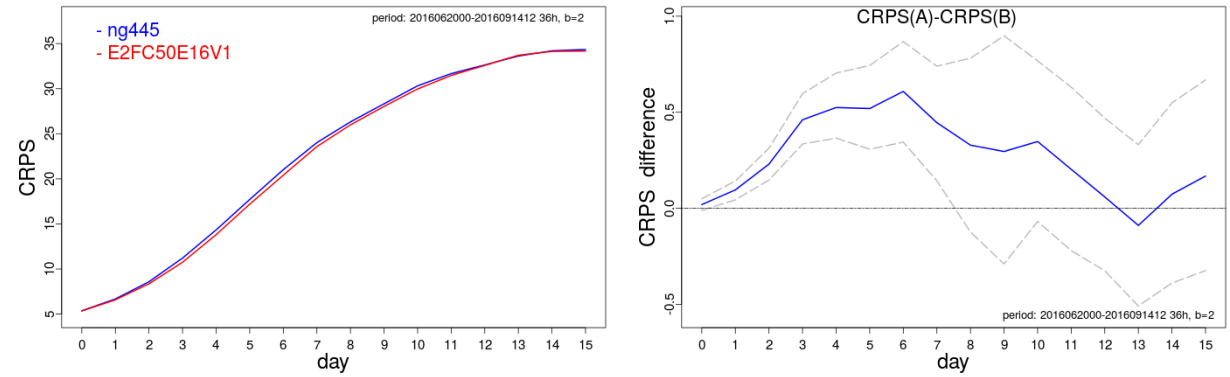
6.2.1 Results for the summer 2016 period

On the left panels of Figure 13, we are showing the Continuous Rank Probability Score (CRPS) for the geopotential heights at 3 levels in the Troposphere. The corresponding CRPS differences between the two systems with their 90% confidence intervals are shown on the right panels. These intervals were computed using bootstrapping method (Candille at al. 2007). A statistically significant improvement is noticed for the first week, whereas week 2 forecasts are of the same quality (neutral impact).

a) GZ250 *b)*



c) GZ500 *d)*



e) GZ850 *f)*

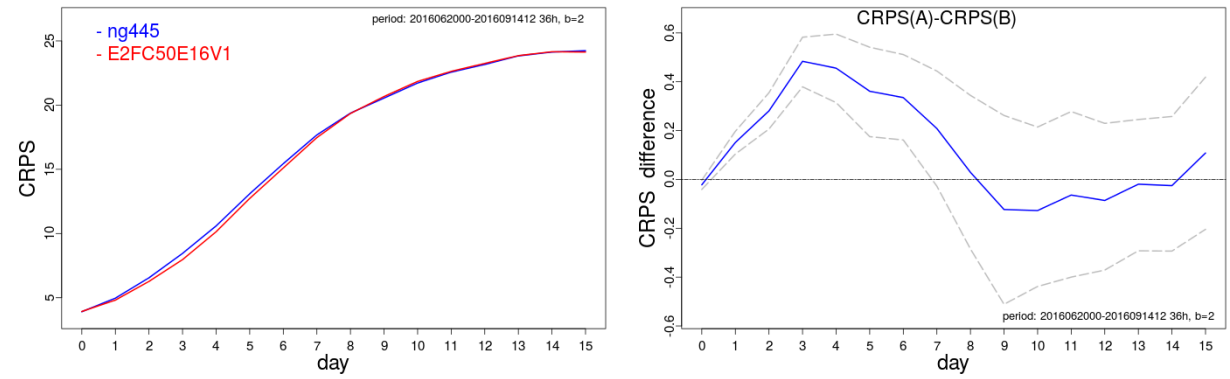


Figure 13. Left panels: CRPS of geopotential heights in Northern Hemisphere of the GEPS 4.3.0 (experiment NG445 in blue) and the GEPS 5.0.0 (experiment E2FC50E16V1 in red) during summer 2016 for three levels 250 hPa (a), 500 hPa (c) and 850 hPa (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three vertical levels.

The CRPS for the temperature is shown in Figure 14. Again, an improvement is found for the first week with a noticeable exception for temperature at 850 hPa (Figure 14 e and f) that is of higher quality for the GEPS 5.0.0 in week 2 as well (up to day 12).

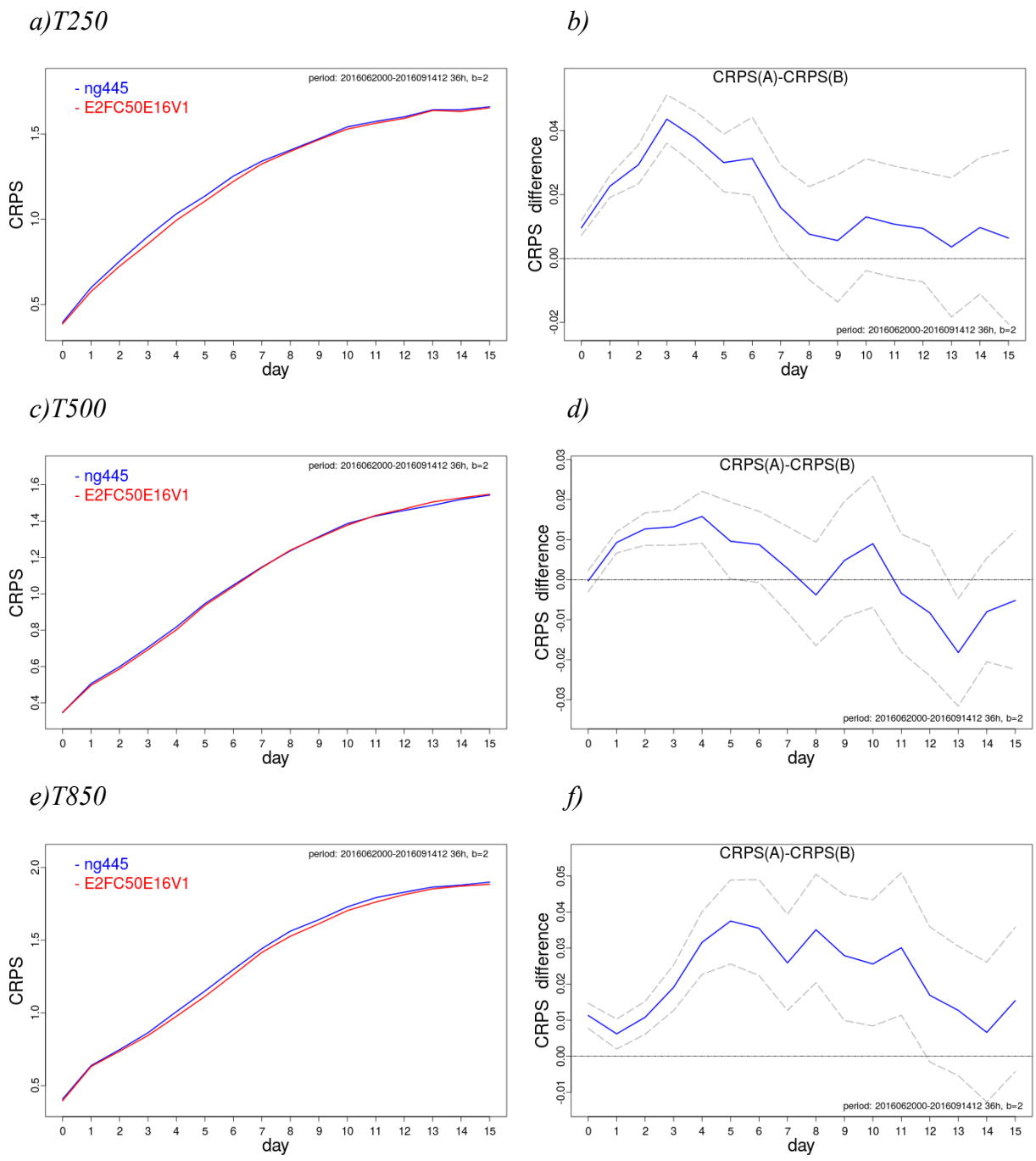
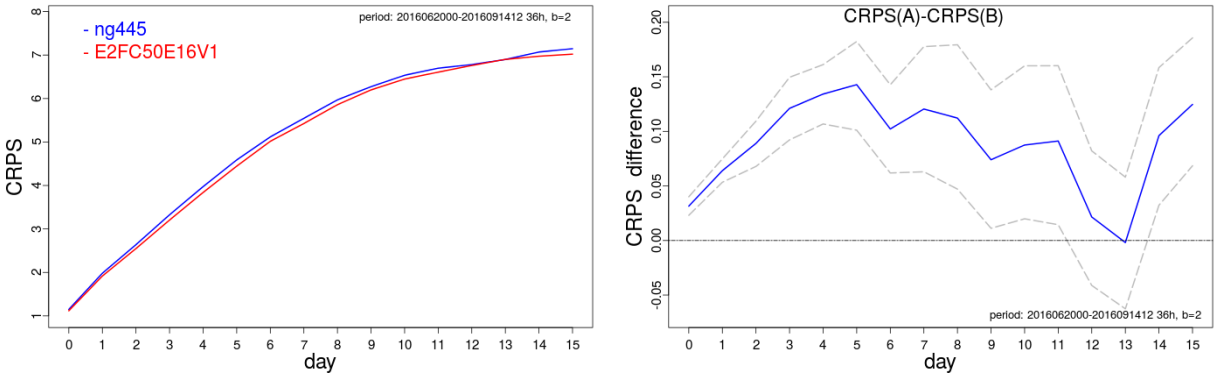


Figure 14. Idem as Figure 13 but for temperature.

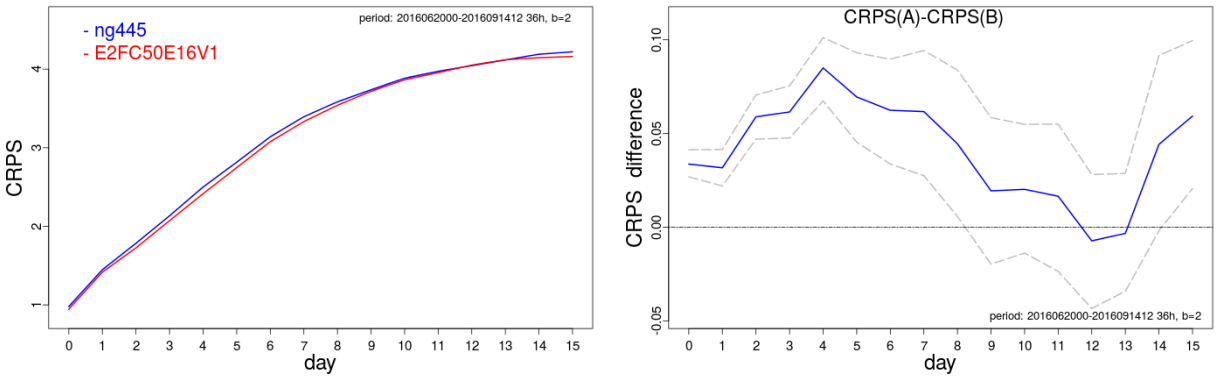
The zonal wind forecast performance is illustrated in Figure 15. A nice improvement is noticed in week 1 and even further at 250 hPa. Similar results are found for the meridional component

(not shown). As explained in section 4.2, the change to a Yin-Yang grid has eliminated most of the noise around the former computational poles, which leads to a weaker forcing for the SKEB parameterization which is a major contributor to wind forecast spread. This behavior can be seen in Figure 16 where Root Mean Square Error (minus observational error) and spread in the Arctic region are displayed. There is a computational pole in this region in the GEPS 4.3.0. We can see that the spread is smaller for levels close to the surface with the new GEPS 5.0.0 which is a better match with the forecast error (solid and dashed lines are closer together). Higher up at 250 hPa where SKEB is less active, there is in fact a larger spread with the new system past day 5, which is also in better agreement with the forecast error because the spread is too low (or the errors are too large).

a)U250 *b)*



c)U500 *d)*



e)U850 *f)*

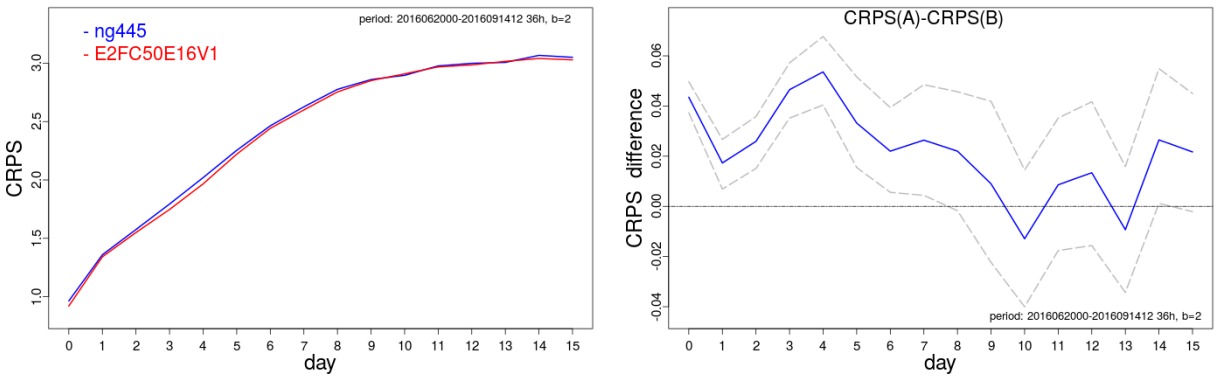
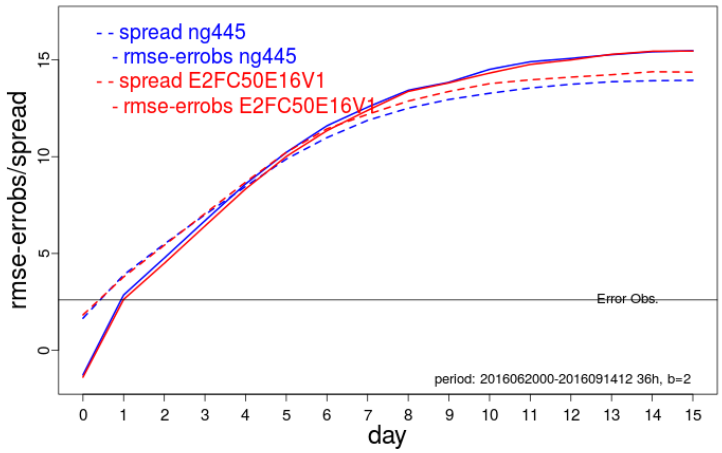
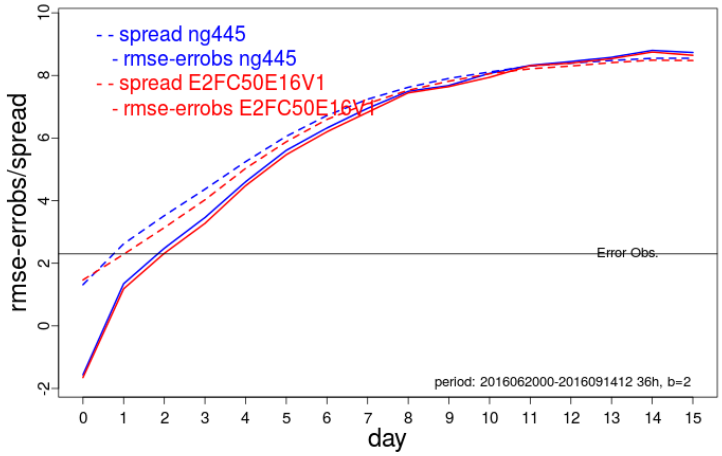


Figure 15. Idem as Figure 13 but for the zonal wind component.

a) V250



b) V500



c) V850

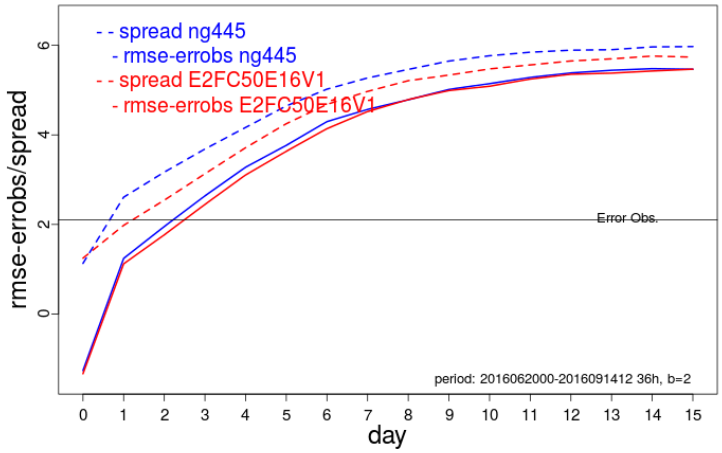
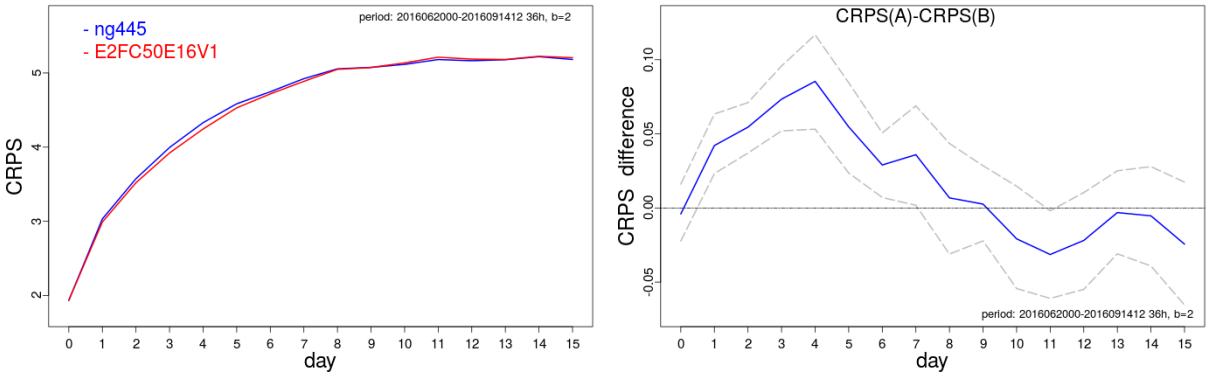


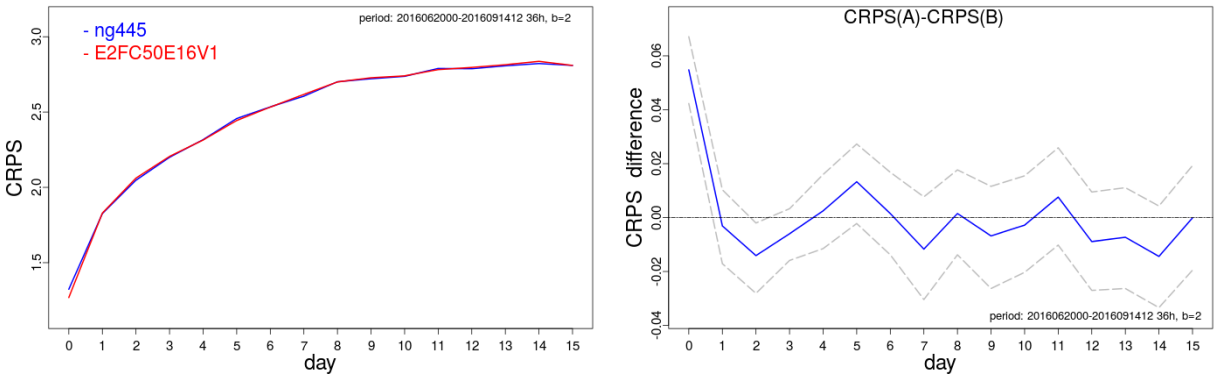
Figure 16. Root mean square error minus observational error (solid line) and forecast spread (dashed line) for the meridional wind component in the Arctic of the GEPS 4.3.0 (experiment NG445 in blue) and the GEPS 5.0.0 (experiment E2FC50E16V1 in red) during summer 2016 for three levels 250 hPa (a), 500 hPa (c) and 850 hPa (e).

The moisture (dew-point depression) forecasts were also evaluated against radiosondes data and the CRPS is shown in Figure 17. Significant improvement is noticed in the very first days of week 1, after which the differences are neutral.

a)ES500 *b)*



c)ES850 *d)*



e)ES925 *f)*

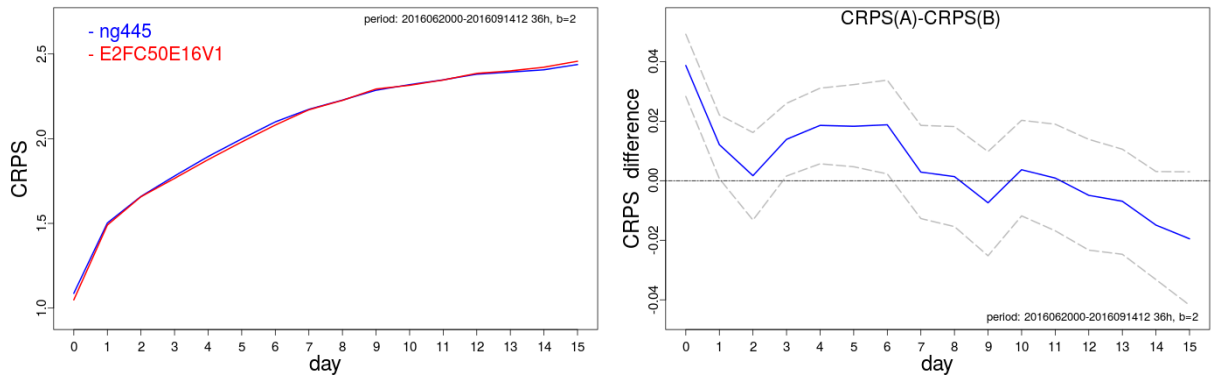


Figure 17. Left panels: CRPS of dew-point depression in Northern Hemisphere of the GEPS 4.3.0 (experiment NG445 in blue) and the GEPS 5.0.0 (experiment E2FC50E16V1 in red) during summer 2016 for three levels 500 hPa (a), 850 hPa (c) and 925 hPa (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three vertical levels.

6.2.2 Results for the winter 2017 period

The difference in forecast performance in winter 2017 is smaller between the two systems compared to the ones just discussed for the summer 2016 in Northern Hemisphere.. This is true for all fields except dew-point depression which is more improved in winter (not shown). Generally, for all other variables, a small improvement is seen in the first 4 days and neutral results for the other lead times (not shown).

6.3 Quality of GEPS forecasts: surface scores

The forecasts of surface fields were also evaluated using the global network of surface synoptic station data (approximately 10000 stations). This data is used to produce the objective surface analyses at CMC using a long cut-off (named G6). The forecasts of the 1.5-m temperature (T1.5m), the mean sea-level pressure (MSLP), the 10-m wind speed (UV10m), and the 1.5-m dew-point depression (ES2m) are verified.

6.3.1 Results for the summer 2016 period

The CRPS of the 1.5-m temperature, 1.5-m dew-point depression (ES1.5m) and mean sea-level pressure (MSLP) for the North American region are shown at the Figure 18. The temperature forecasts are clearly better (significant up to day 7). Also, the MSLP is significantly better in the medium-range (day 4 - 8). The ES1.5m is not improved over North America. However, for week 1, there is significant improvement in Asia (not shown).

a) T1.5m *b)*

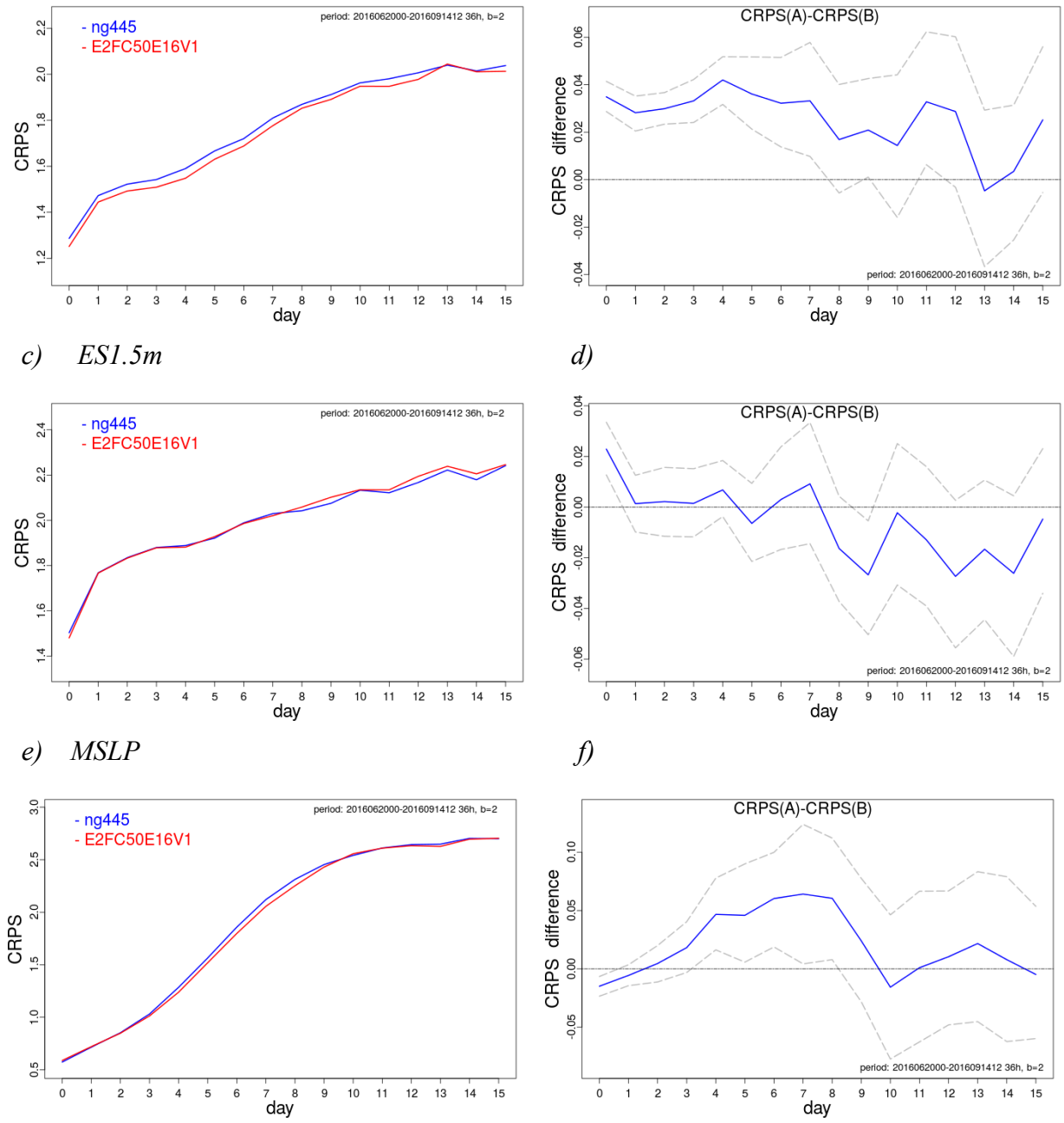


Figure 18. Left panels: CRPS in the North America of the GEPS 4.0.0 (experiment NG445 in blue) and the GEPS 5.0.0 (experiment E2FC50E16V1 in red) during summer 2016 for the 1.5-m temperature (a), the 1.5-m dew-point depression (c) and mean sea-level pressure (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three fields.

The results for the 10-m wind speed are worse as shown in Figure 19. Both CRPS components are degraded (reliability and resolution, not shown). The change of source for the model orography has generated this problem. The bias of the wind varies in function of the bins. The model over-forecasts low wind (< 5 knots) events while under-forecasts the higher wind categories (R. Frenette, personal communication). Overall the bias is positive because the number of cases with weak wind dominates the sample. Clearly, there is room for improvement for this field at 10 meters. However, the wind components in the lower troposphere are improved as shown for the zonal wind at 850 hPa in Figure 15 (also true at 925 hPa, figure not shown). So, the problem is localized near the model surface. In the next version of the model a less aggressive orographic filter will be used and it should lead to improvement of winds near the surface.

a) CRPS UV10m *b)*

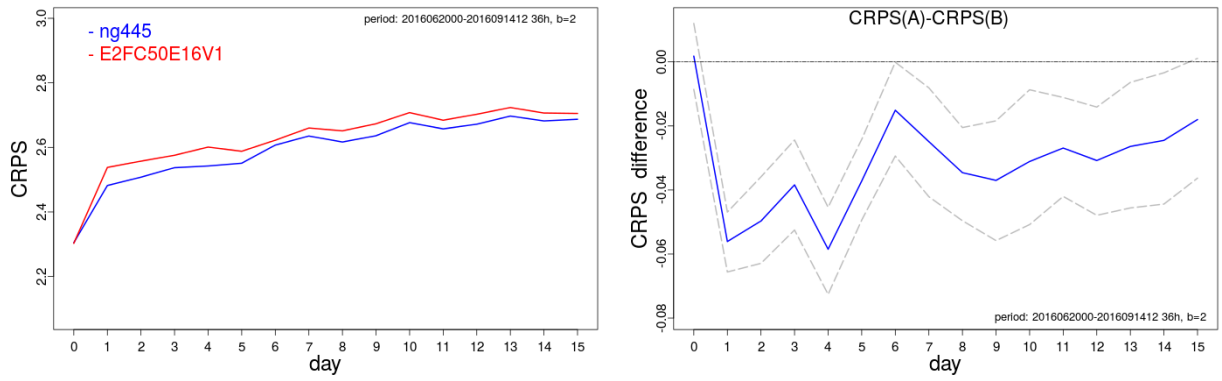


Figure 19. Idem as Figure 18 but as 10-m wind speed.

6.3.2 Results for the winter 2017 period

The CRPS of the 1.5-m temperature, 1.5-m dew-point depression (ES1.5m) and mean sea-level pressure (MSLP) for the North American region are shown at the Figure 20. The temperature forecasts are improved only in the first 5 days. In contrast to the summer results, the ES1.5m is now better up to day 8. The MSLP is clearly better even more than during the summer.

a) T1.5m *b)*

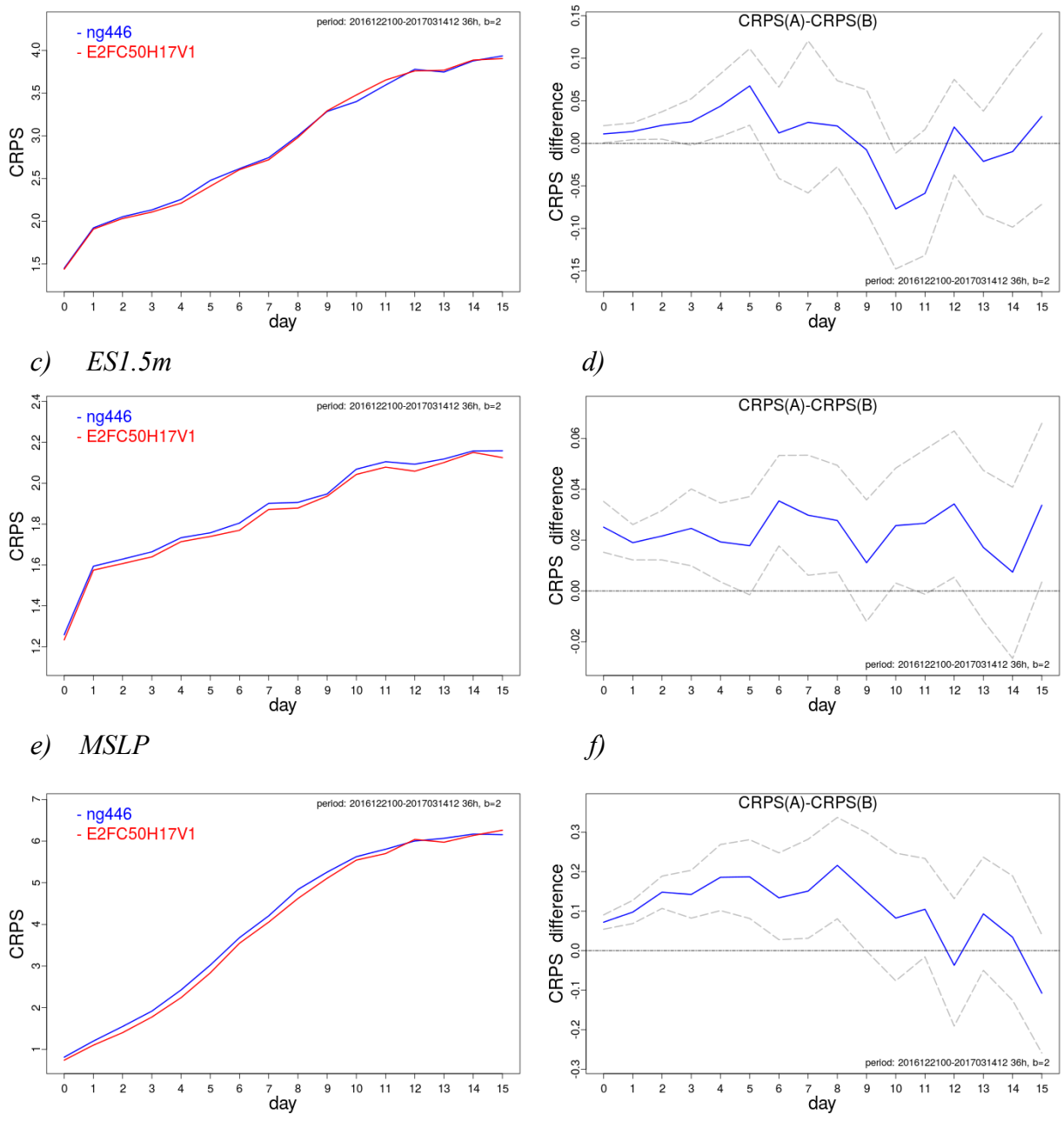


Figure 20. Left panels: CRPS in the North America of the GEPS 4.0.0 (experiment NG446 in blue) and the GEPS 5.0.0 (experiment E2FC50H17V1 in red) during Winter 2017 for the 2-m temperature (a), the 2-m dew-point depression (c) and mean sea-level pressure (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three fields.

The 10-m wind speed suffers from the same problem in winter described above for summer but only for the first 3 days after that it is neutral (not shown).

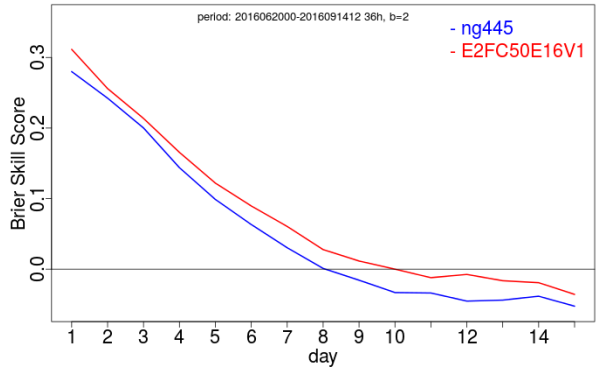
6.4 Quality of GEPS forecasts: precipitation scores

The 24 hr QPF (quantitative precipitation forecasts) were evaluated for the two systems, using the global surface synoptic stations network (approximately 7000 stations). The accumulation of all types of precipitation combined (ex.: total of rain, snow, freezing rain and ice pellets) forecast over a 24 hour period is verified against these stations observations. Brier skill score (BSS) and its decomposition in resolution and reliability are used as main measure of quality for the precipitation forecasts.

6.4.1 Results for the summer 2016 period

The BSS for the 2.5 mm and more category is significantly improved for all lead times in Northern Hemisphere as shown in Figure 21a. The reliability is the main contributor to this improvement (see Figure 21c) but the resolution (see Figure 21d) is also better up to day 7. This kind of significant improvement is not frequent for the precipitation forecasts. It is a combination of better humidity analyses and forecasts in the Troposphere as well as the enhanced horizontal resolution. Nice improvements are also noticed for higher categories up to 25 mm during the first week as seen in Figure 22. Although the skill is still modest in the medium-range, the forecasts seem clearly better for pretty much all lead categories.

a) BSS PCP > 2.5 mm



b) BSS difference

c) BSS-reliability PCP > 2.5 mm

d) BSS-resolution PCP > 2.5 mm

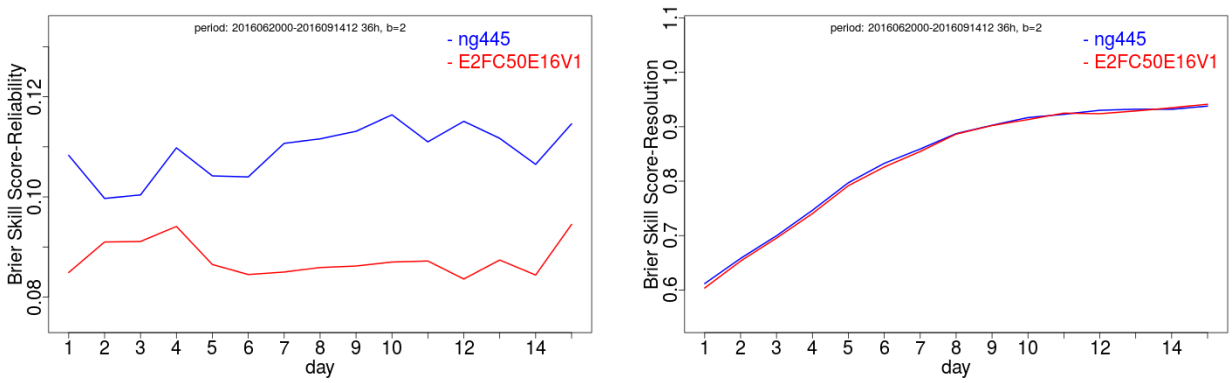


Figure 21. Brier skill score (in a) for the 2.5 mm and more category of 24 hour precipitation accumulation in Northern Hemisphere for GEPS 4.3.0 (experiment NG445 in blue) and GEPS 5.0.0 (experiments E2FC50E16V1 in red) in summer 2016. In b, the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping. The Brier skill score reliability component is shown in c while the resolution component is illustrated in d, respectively.

a) BSS 24 hour lead time *b) BSS difference*

c) BSS 144 hour lead time *d) BSS difference*

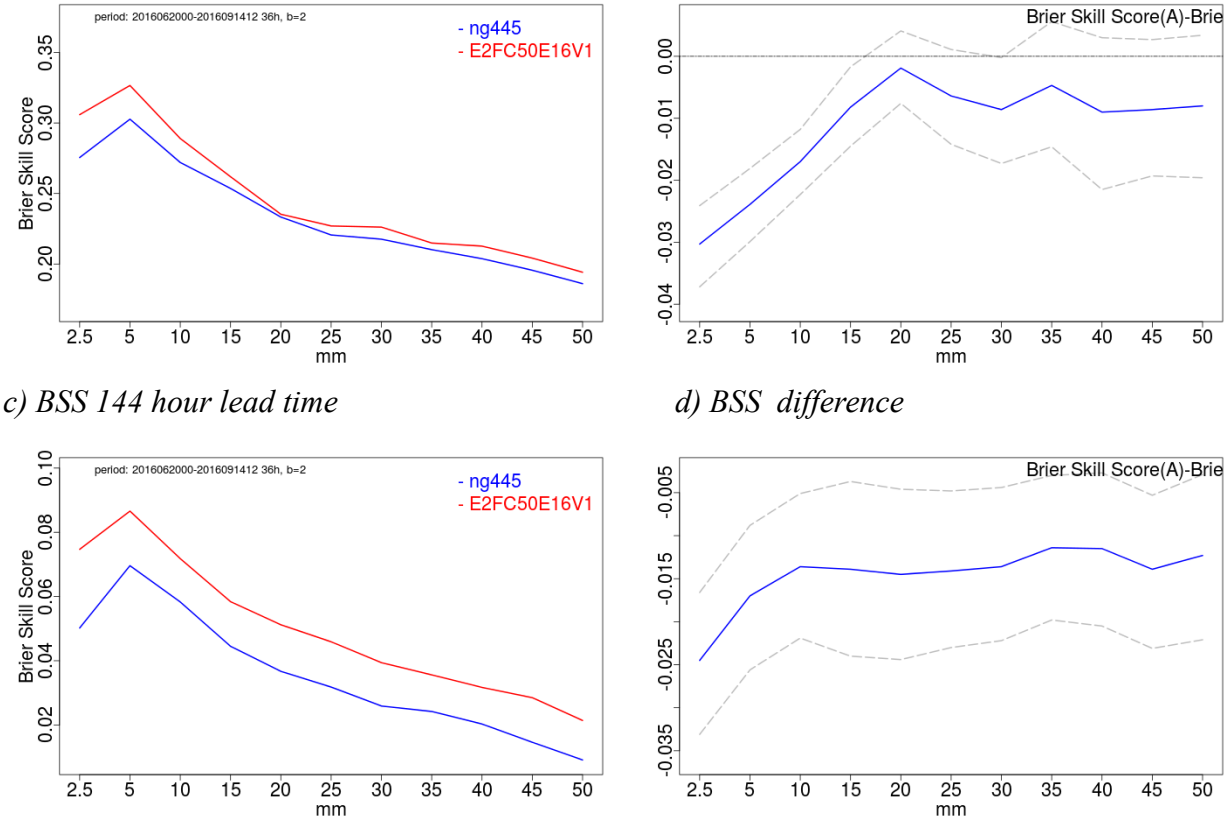


Figure 22. Brier skill score for the 24 hour (in a) and the 144 hour (in c) lead time is shown in function of the different threshold and more category of 24 hour precipitation accumulation in

Northern Hemisphere for GEPS 4.3.0 (experiment NG445 in blue) and GEPS 5.0.0 (experiments E2FC50E16V1 in red) in summer 2016. In b and d, the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping.

6.4.2 Results for the winter 2017 period

The results in winter 2017 are also positive but to a lesser degree similar to what was seen for the upper air fields. The BSS is improved at early lead times mainly for the low categories (not shown) but most of the times these differences are not statistically significant.

7 Objective evaluation of the parallel run

The parallel run, in which the new GEPS system cycled alongside the operational system, took place during the period from April to September 2018. We have performed a similar verification for the parallel run trial fields and forecasts as we did for the final series of tests. Here is a summary of the results of the objective evaluation against observations.

7.1 Quality of initial conditions

The ARCAD verifications were performed using the "postalt" files from the GDPS parallel run with bad humidity observations rejected. The mean verification scores are computed over a period from 20 April to 26 July 2018.

Over the globe (Figure 23), the ARCAD verification results from the parallel run are very similar to those from the final cycles (mentioned in section 6.1) except that we now also see improvement in the dew point depression, due to the rejection of bad humidity observations in the postalt files used for verification. The ARCAD verification results in Northern Extratropics (figure not shown) are very similar to those over the globe, namely improvement in the standard deviations for winds and temperature in the troposphere (mainly due to moving to Yin-Yang 39 km) and in the upper stratosphere (mainly due to raising the model top). Over the Southern Extratropics (Figure 24), the impacts from the proposed changes are very similar but with even bigger improvement for winds and geopotential heights. Over Tropics, similar to what was observed for the final series of tests, the biases of geopotential heights are greatly reduced mainly due to replacing DF with IAU. The verification against GPS-RO observations over the parallel-run period is shown in Figure 25. Here we see significant improvement in the standard deviation and the biases, especially in the upper troposphere/lower stratosphere, which is mainly due to raising the model top as seen in section 3.2. The verification against mean sea level pressure (MSLP) observations shows substantial reductions of standard deviation and bias (Figure 26). The substantial reductions in standard deviation are in part due to the use of a Huber norm (evident here from the reduced observation counts) that rejects outlying observations and in part due to the use of IAU. The reduction in bias of MSLP is largely due to the use of the IAU. The

performance of surface pressure is very similar to that of MSLP. Similar to what were observed for the final tests, reduction in standard deviation and data count are also found for surface temperature, dew-point depression, and also winds but with a lesser degree (figures not shown). The global SQLITE verifications also show nice improvement for AMSU-A channels 7, 11 & 12 and ATMS channels 9 -13 (Figure not shown), as seen in the final tests. But the verification against AMSU-B shows less improvement in the parallel run than in the final tests (figure not shown). Overall, the objective verification results from the parallel run confirm the positive impacts in the ARCAD verifications against Radiosonde observations and the positive impacts in the SQLITE verifications against GPS-RO, AMSU-A, ATMS, mean sea level pressure, and other surface variables, as seen in the final tests.

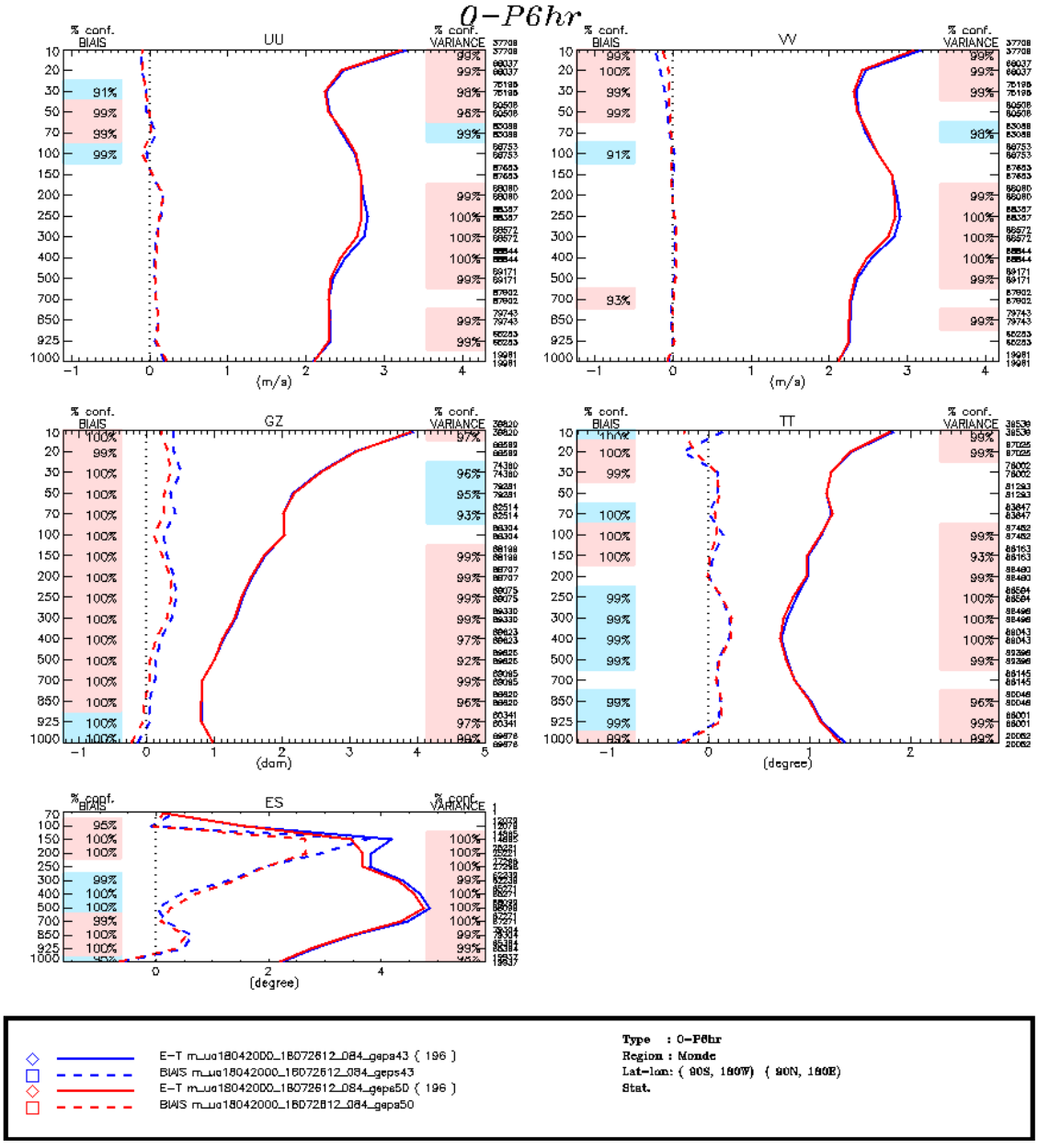


Figure 23. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the globe for a period from 20 April to 26 July 2018 for EnKF experiments with the operational GEPS 4.3.0 (blue) and the parallel GEPS 5.0.0 (red).

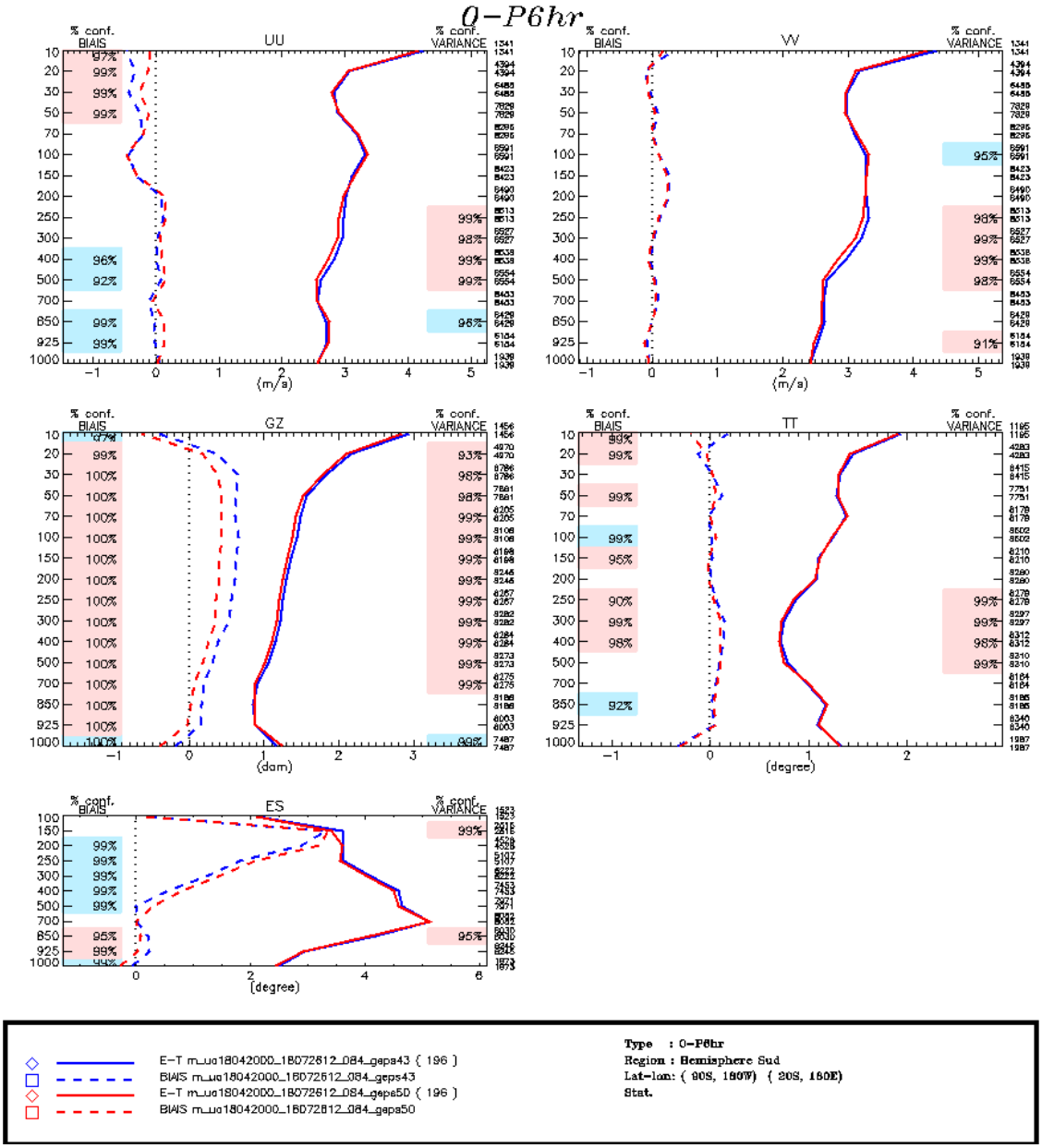


Figure 24. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the Southern Extratropics for a period from 20 April to 26 July 2018 for EnKF experiments with the operational GEPS 4.3.0 (blue) and the parallel GEPS 5.0.0 (red).

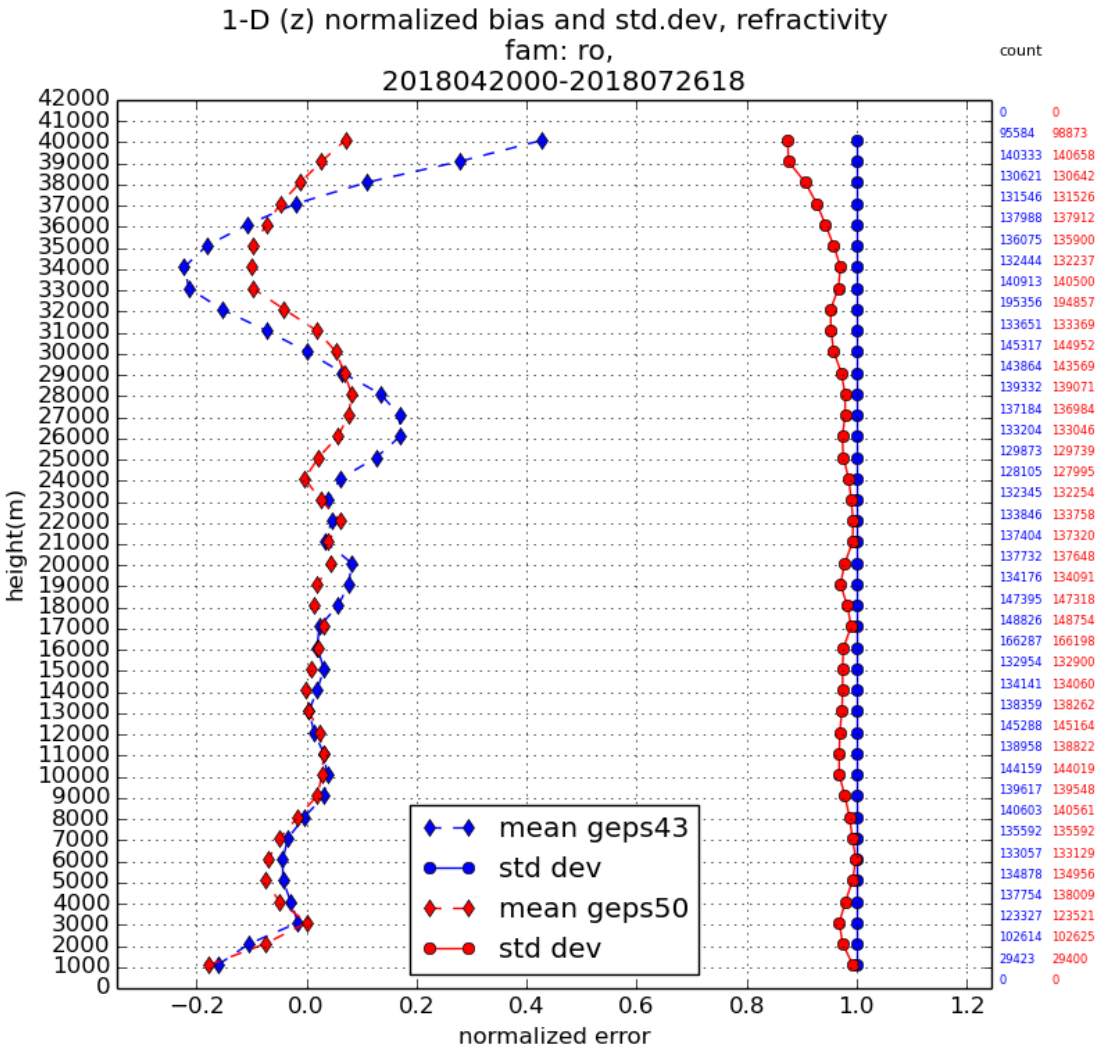


Figure 25. Normalized standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against GPS-RO globally averaged over a period from 20 April to 26 July 2018 for EnKF experiments with the GEPS 4.3.0 (blue) and the GEPS 5.0.0.

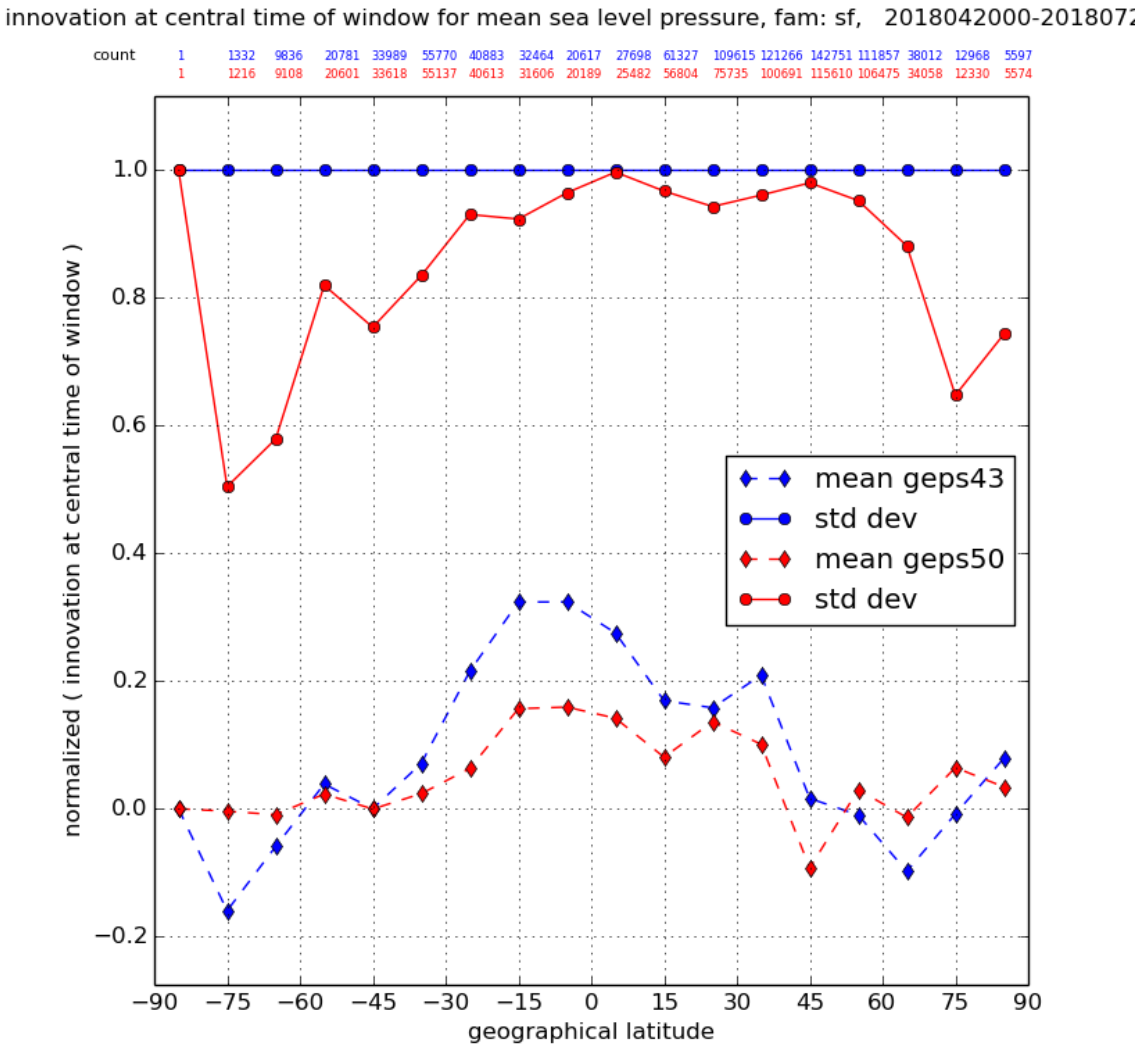
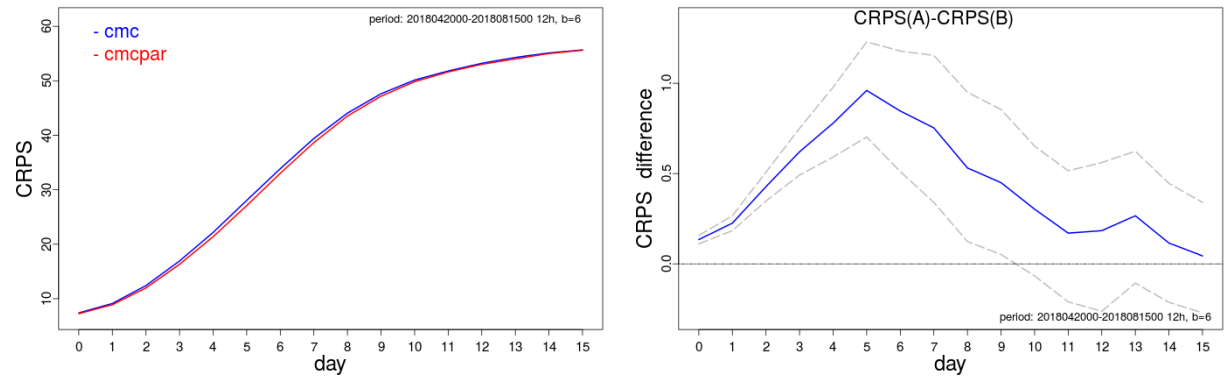


Figure 26. Normalized standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against mean sea level pressure globally averaged over a period from 20 April to 26 July 2018 for EnKF experiments with with the GEPS 4.3.0 (blue) and the GEPS 5.0.0. (red)

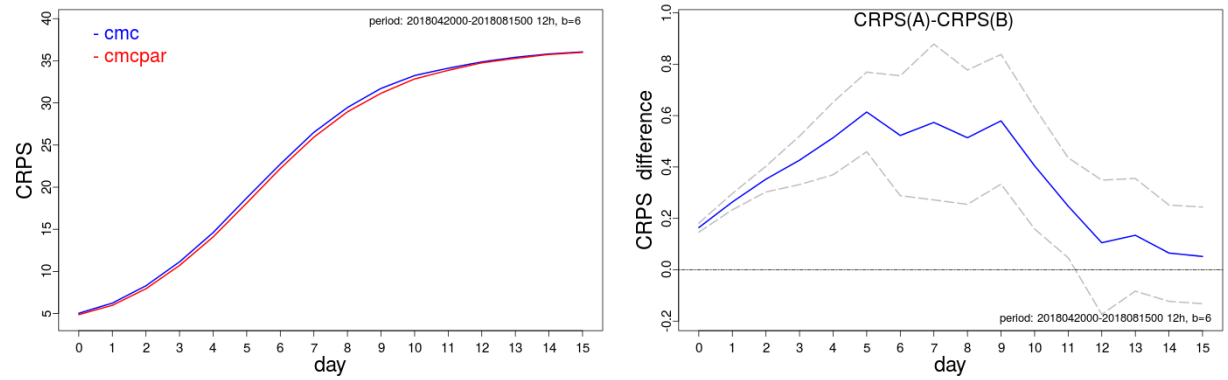
7.2 Quality of GEPS forecasts: upper air scores

The improvements noticed in the summer 2016 final series of tests are generally found in the parallel run for the medium-range forecasts. In fact, the improvements are slightly greater as can be seen in Figure 27 for the geopotential heights at 250, 500, and 850 hPa against Northern Hemisphere radiosondes (for the period of April 20th to August 15th). The geopotential heights forecasts of the GEPS 5.0.0 parallel run are even better than what could be expected from the summer 2016 test.

a) GZ250 *b)*



c) GZ500 *d)*



e) GZ850 *f)*

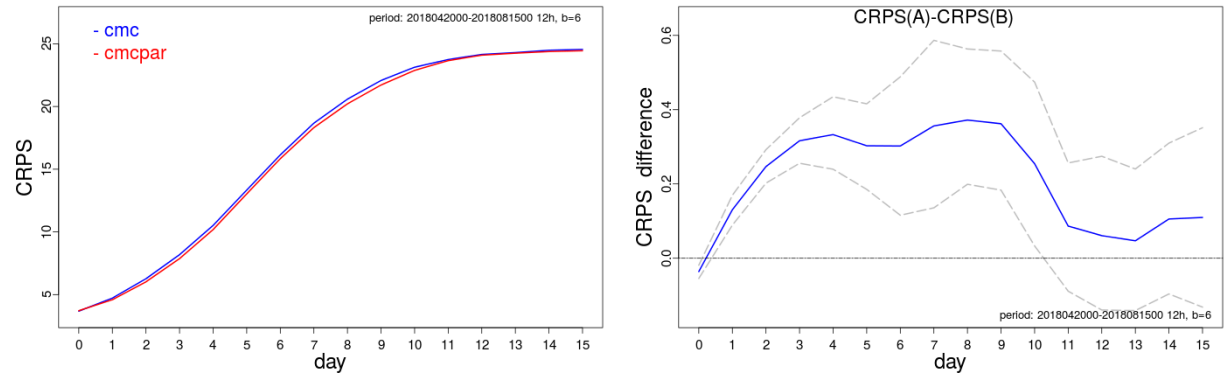


Figure 27. Left panels: CRPS of geopotential heights in Northern Hemisphere of the GEPS 4.3.0 (experiment 'cmc' in blue) and the GEPS 5.0.0 (experiment 'cmcpar' in red) during summer 2018 for three levels 250 hPa (a), 500 hPa (c) and 850 hPa (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three vertical levels.

However, the improvement of the temperature and the winds (not shown) are as estimated by the pre-parallel tests (as shown for example in Figure 14 and Figure 15). As for the geopotential

heights, the dew-point depression forecasts seems to be more accurate than expected as displayed in Figure 28.

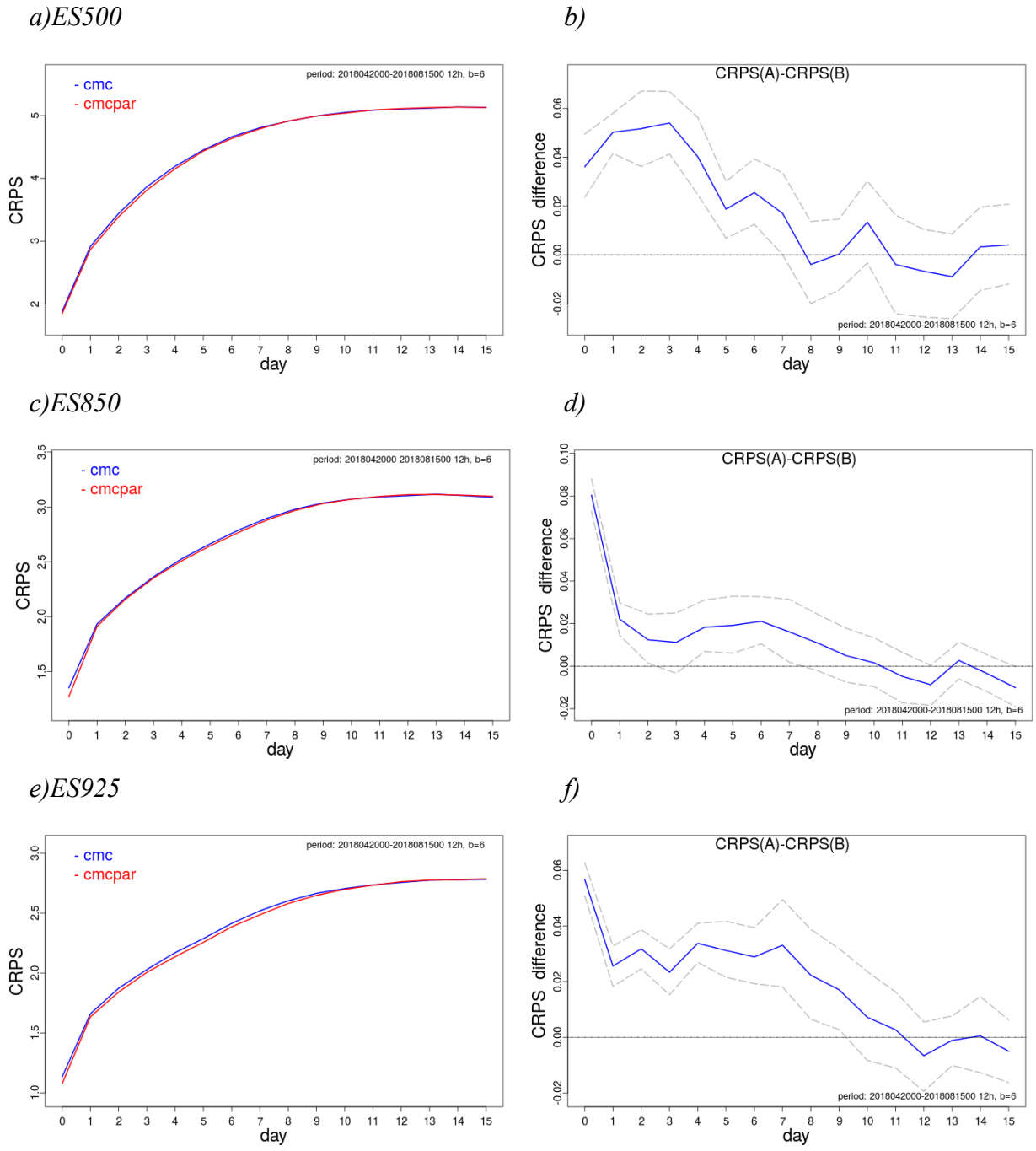
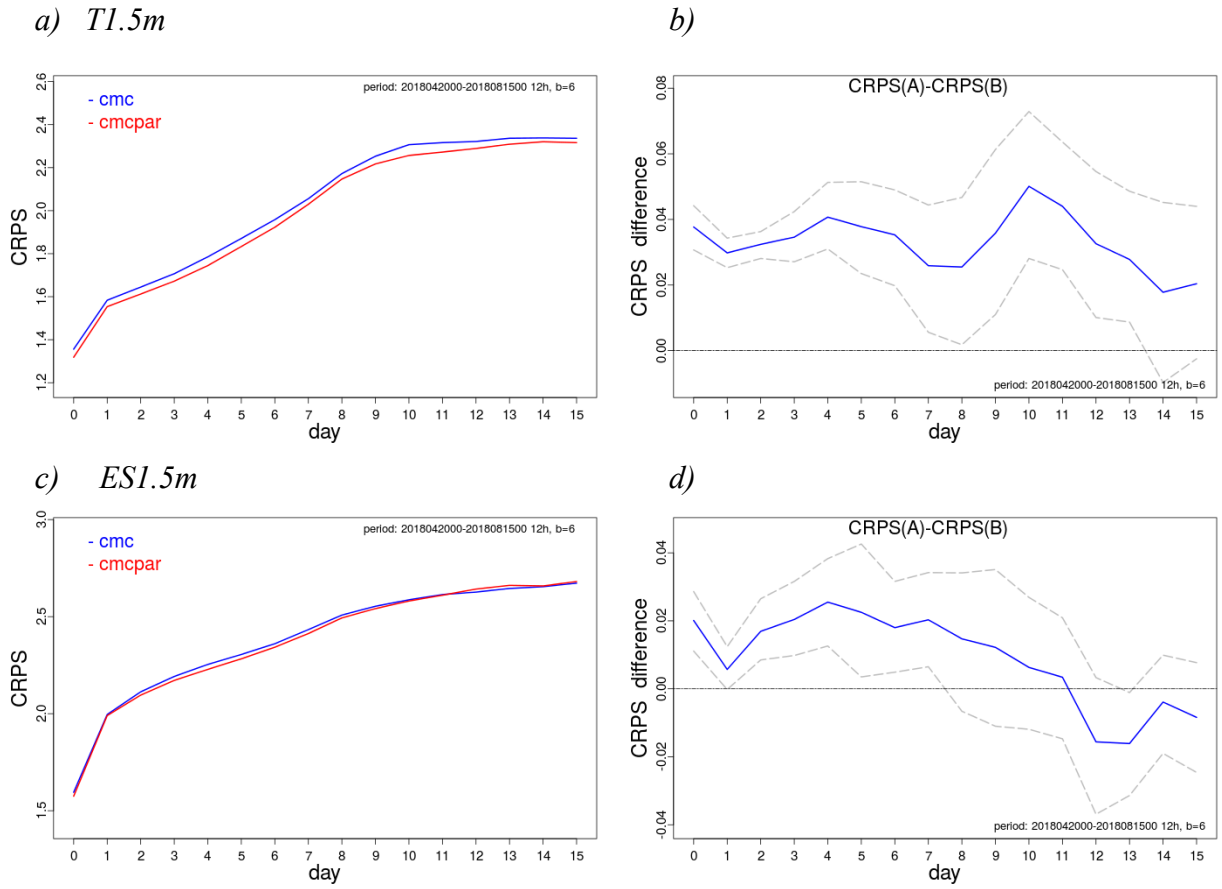


Figure 28. Left panels: CRPS of dew-point depression in Northern Hemisphere of the GEPS 4.3.0 (experiment 'cmc' in blue) and the GEPS 5.0.0 (experiment 'cmcpar' in red) during summer 2018 for three levels 500 hPa (a), 850 hPa (c) and 925 hPa (e). Right panels: the difference

between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three vertical levels.

7.3 Quality of GEPS forecasts: surface scores

The forecast performance of the surface fields over North America during the parallel run is illustrated in Figure 29 (temperature and dew-point depression at 1.5m and mean sea-level pressure) and Figure 30 (10-m wind speed). Generally, the CRPS are better than the ones obtained for the summer 2016 tests (see Figure 18 and Figure 19). Improvements are seen for all fields in first forecast week except for the 10-m wind speed, which is degraded as expected.



e) MSLP *f)*

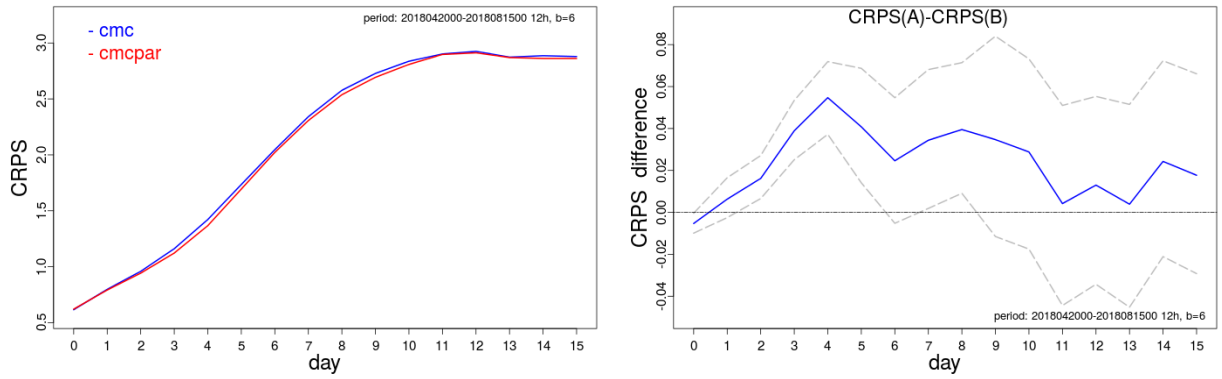


Figure 29. Left panels: CRPS in the North America of the GEPS 4.0.0 (experiment 'cmc' in blue) and the GEPS 5.0.0 (experiment 'cmcpar' in red) during summer 2018 for the 2-m temperature (a), the 2-m dew-point depression (c) and mean sea-level pressure (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three fields.

a) CRPS UV10m *b)*

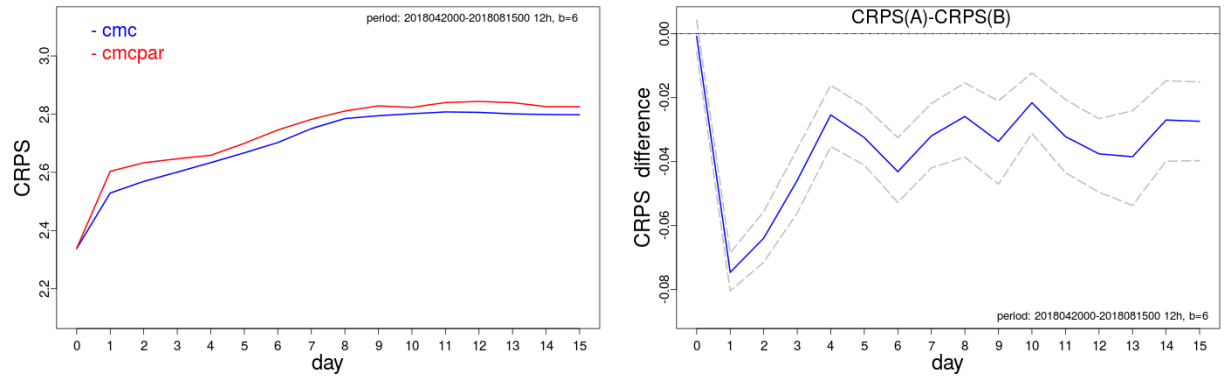
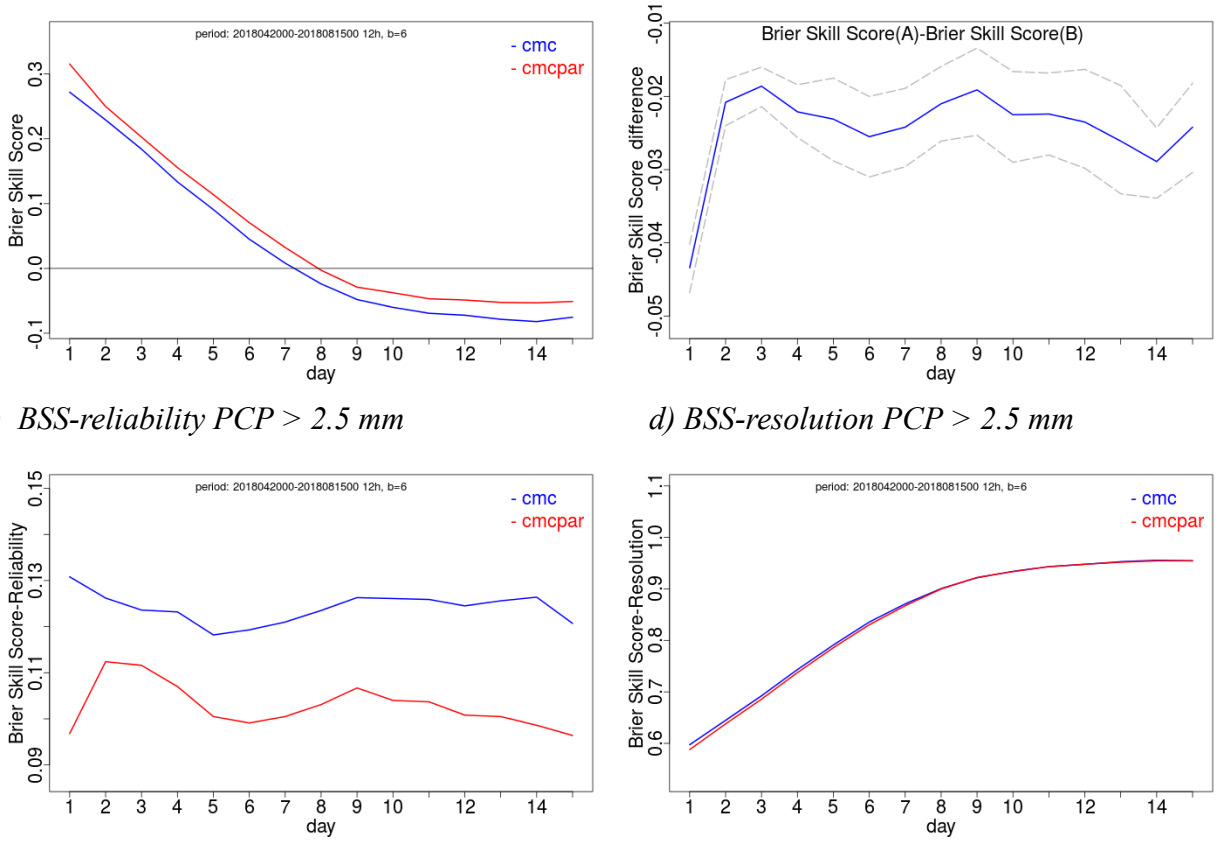


Figure 30. Idem as Figure 29 but as 10-m wind speed.

7.4 Quality of GEPS forecasts: precipitation scores

The precipitation forecasts of the new system were found to be better than the operational system by a large margin during the parallel run (see Figure 31). In fact, for the lowest category (2.5 mm+ shown in Figure 31) the gain in predictability is as big as 12 hours up to day 7. Both the reliability and the resolution components are contributing to the enhanced forecast quality.

a) BSS PCP > 2.5 mm *b) BSS difference*



c) BSS-reliability PCP > 2.5 mm

d) BSS-resolution PCP > 2.5 mm

Figure 31. Brier skill score (in a) for the 2.5 mm and more category of 24 hour precipitation accumulation in Northern Hemisphere for GEPS 4.3.0 (experiment 'cmc' in blue) and GEPS 5.0.0 (experiments 'cmcpar' in red) in summer 2018. In b, the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping. The Brier skill score reliability component is shown in c while the resolution component is illustrated in d, respectively.

For the higher categories, the improvements are smaller and more limited to the first days of the forecasts (not shown). One behavior that was observed is the fact that the higher bins (30 mm +) are not as good as they were in the summer 2016 test (see *Figure 32* a and b) but this change is not necessarily statistically significant because the number of cases for this type of event is of course less than for the lower bins. However, the bootstrapping technique is tagging these differences to be significant at the 144h lead time (see *Figure 32*c and d). Since at these lead times the BSS is very low (<0.1), we should be careful in drawing conclusions from these scores .

a) BSS 24 hour lead time

b) BSS difference

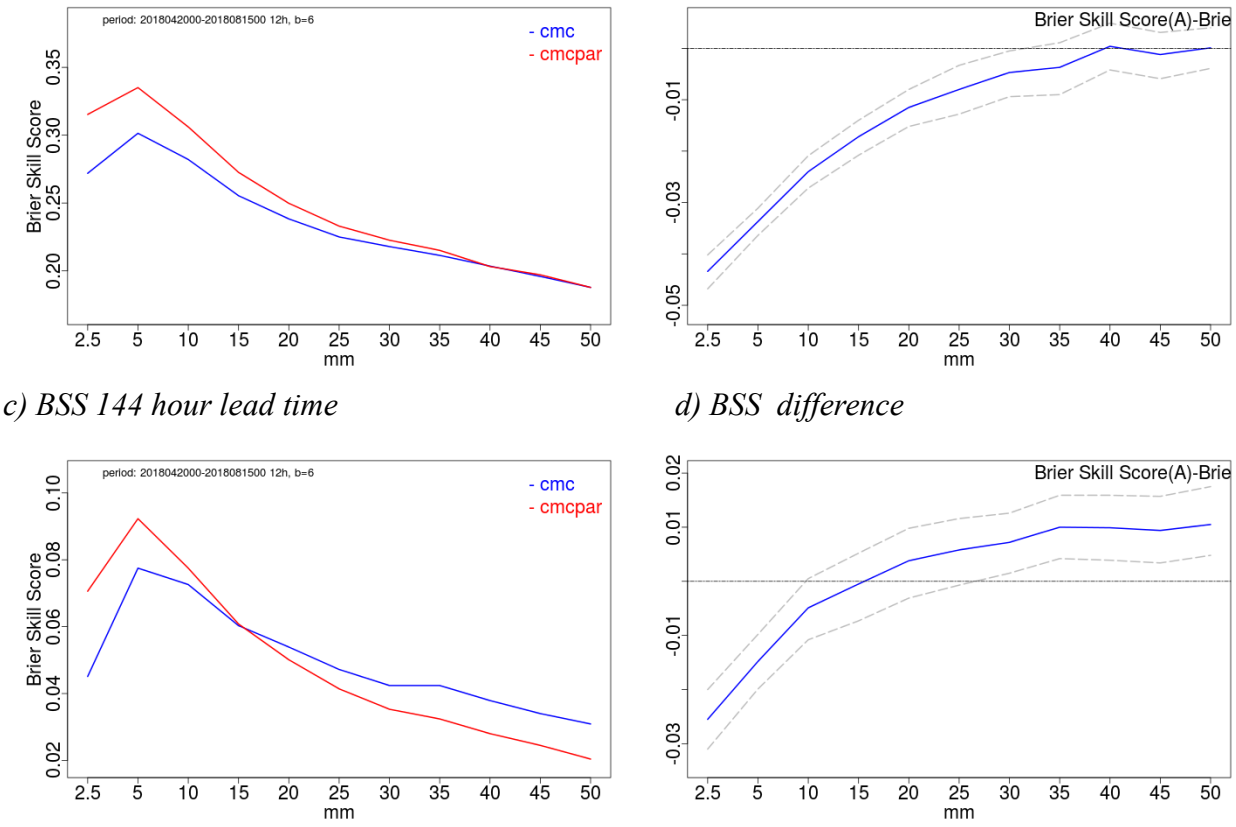


Figure 32. Brier skill score for the 24 hour (in a) and the 144 hour (in c) lead time is shown in function of the different threshold and more category of 24 hour precipitation accumulation in Northern Hemisphere for GEPS 4.3.0 (experiment 'cmc' in blue) and GEPS 5.0.0 (experiments 'cmcpar' in red) in summer 2018. In b and d, the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping.

8 Subjective evaluation

A comparative evaluation between the GEPS-4.3.0 (operational system, to be replaced) and the GEPS-5.0.0 (new system, in parallel) was carried out by the team of operational meteorologists in the "Analysis and Prognostics" section (A and P) of the CMC. The evaluation was conducted over a 71-day period from June 5 to August 15, 2018. Only the 00Z runs were evaluated as part of this evaluation. The evaluation consisted of visually comparing the depressions trajectories of the operational GEPS with the trajectories of the GEPS running in parallel. The trajectories of each system were compared to an analysis of the position of the depression in order to judge which of the two forecasts was the better. The evaluation area covered North America and adjacent oceans. In total, 236 trajectories of meteorological systems were evaluated. Figure 33 shows an example of the visualization tool used for this evaluation as well as the area covered.

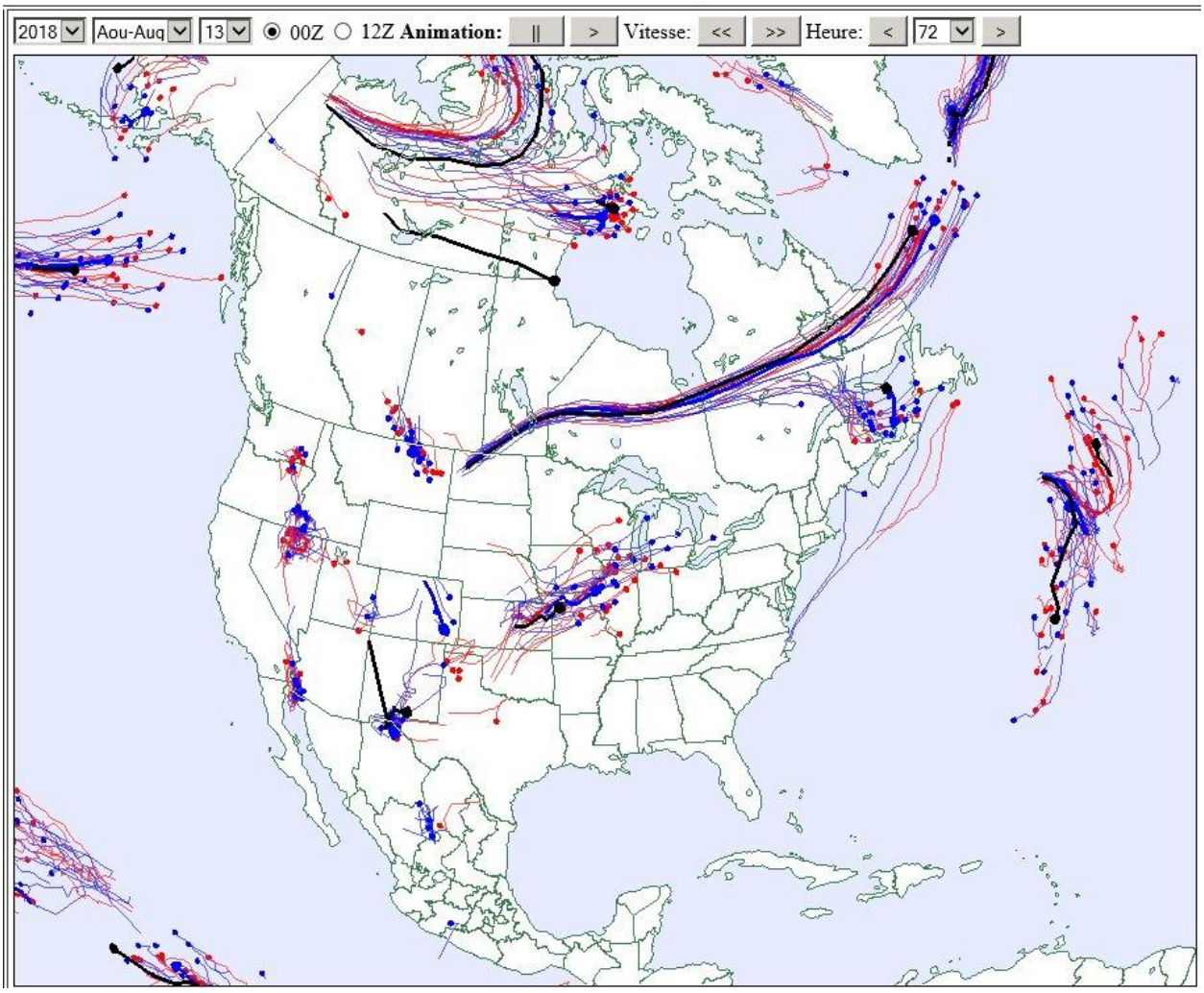


Figure 33. The analysis of the position is shown by the black line highlighted in bold. The red lines represent the trajectories of each individual member of the operational GEPS while the blue lines represent those of the GEPS in parallel.

For each meteorological system, the meteorologist had to determine where the observed trajectory (the analysis) was compared with the trajectories proposed by the forecasting systems. He had to determine whether the analysis coincided with the average of the trajectories, whether it was in the outer quarter of the trajectory envelope, whether it was at the edge of the envelope or whether it was outside the envelope of the planned trajectories. He then had to determine which of the operational or parallel forecasting system provided the most useful forecast, or whether the two systems offered solutions of equal value.

Error! Reference source not found. shows the results of the evaluation conducted to determine which of the two systems performed better during the evaluation period. For 36% of the depressions evaluated, the parallel system was considered to offer the better ensemble solution while the operational system solution was considered better in 30% of the evaluations. In 34% of the cases, the solutions proposed by the two systems were considered equal.

Table 2. Evaluation results used determine which system performed better during the evaluation period.

	# de systèmes	Pourcentage
GEPS Parallel	85	36%
GEPS Operational	70	30%
Solutions of equal values	81	34%

Table 3 shows the number of evaluations for each of the position categories listed above. The parallel system shows a slight gain with respect to the number of cases where the analysis is positioned on the average of the ensemble solutions and also with respect to the number of solutions at the edge of the envelope of the trajectories. These results show that, in general, the difference between the two forecasting systems is limited since the number of evaluations grouping the categories where the envelope of the trajectories includes the analysis (average, 25% and border) is equal for both forecasting systems. Despite this, the parallel system was considered by the meteorologists as slightly better performing as shown in **Error! Reference source not found.**.

Table 3. Number of cases for each category of the evaluation. These criteria aim to evaluate where the analysis is in relation to the envelope of the set solution.

	GEPS Parallel	GEPS Operational
Avergae	90	84
25%	52	52
The edge	57	63
Outisde	36	36

9 Performance of dependent systems

This section describes the adjustments made to systems that depend on the GEPS 5.0.0. Specifically, we describe the adaptation of the REPS and the NAEFS.

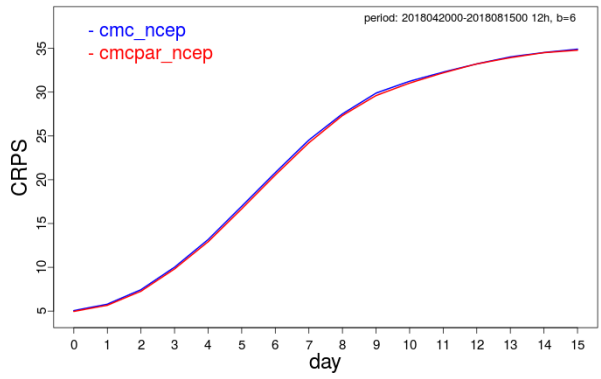
9.1 Regional Ensemble Prediction System (REPS)

The impact of the arrival of GEPS 5.0.0 on the REPS was evaluating during the same final tests and parallel run period. The results are usually positive and they are discussed in another technical document (Patoine et al. 2018).

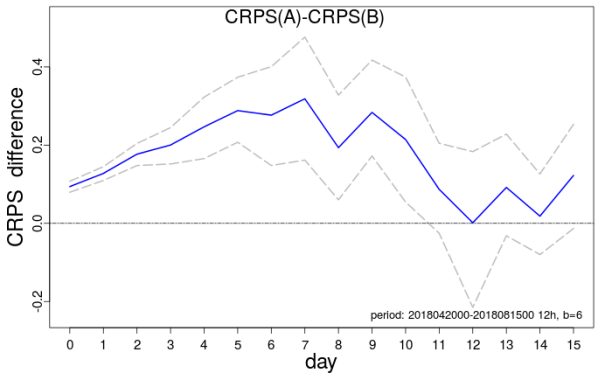
9.2 North American Ensemble Forecast System (NAEFS)

The impact on the super ensemble of the North American Ensemble Forecast System (NAEFS) was also evaluated. Verifications against observations were done during the parallel run for the operational NAEFS which is the sum of the GEPS 4.3.0 members and the NCEP GEFS members. After, the GEPS 5.0.0 (instead of the GEPS 4.3.0) members were put in the mix to form the new NAEFS. Figure 34 shows CRPS for the 500 hPa geopotential heights and MSLP as well as the BSS for the 2.5 mm and more category. Significant improvements are noticed for the new NAEFS with the arrival of the GEPS 5.0.0. As usual, the improvements are smaller because half of the super ensemble is the same (NCEP members). The degradation of the 10-m wind speed is also reduced (not shown). The NAEFS users should expect better forecasts in week 1 with the replacement of the GEPS 4.3.0 by the GEPS 5.0.0.

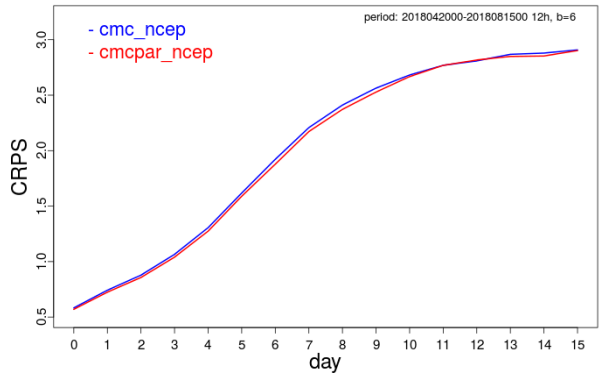
a) GZ500



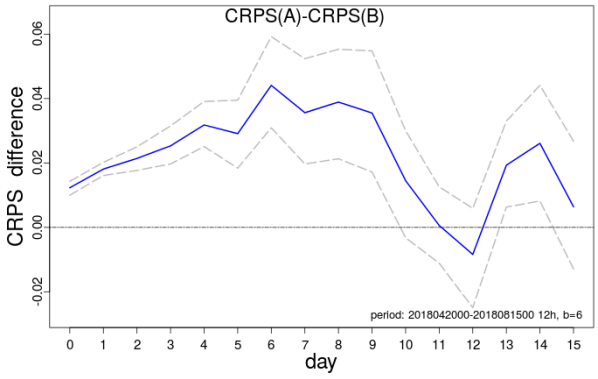
b)



c) MSLP



d)



e) PCP at 24h lead time

f)

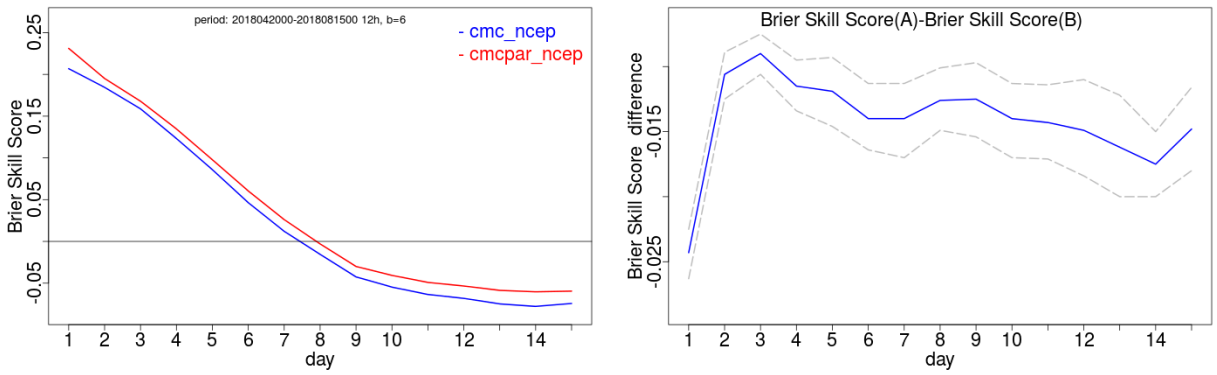


Figure 34. Left panels: CRPS in Northern Hemisphere of the GEPS 4.3.0 (experiment 'cmc' in blue) and the GEPS 5.0.0 (experiment 'cmcpar' in red) during summer 2018 for 500 hPa geopotential heights (a) and mean sea level pressure (c), Brier skill score for the 2.5 mm and more in 24 hour precipitation category (in e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping.

10 Availability of products

The enhanced horizontal resolution and addition of vertical levels in the forecast model lead to an increase in the computing time to do the numerical forecasts. We are expecting a delay of 10 minutes in the production of the forecasts. Therefore, the products will be available slightly later than usual following the implementation of the GEPS 5.0.0.

11 Summary of the results

- The quality of the new EnKF trial fields is improved especially in the Troposphere and upper Stratosphere, as well as at surface (especially surface pressure, surface temperature and mean sea level pressure).
- The forecast performance of the new system GEPS 5.0.0 is generally higher in the first 7-10 days in Northern Hemisphere for all upper air fields.
- The forecast surface fields (MSLP, 1.5-m temperature and dew-point depression and precipitation) are also improved significantly during the first week except the 10-m wind speed which is degraded.
- The forecast spread is usually greater during days 8 to 15 for most of fields. The wind spread is now smaller during the first 5 days in lower troposphere.
- The NAEFS users should expect better forecasts in week 1 with the replacement of the GEPS 4.3.0 by the GEPS 5.0.0.
- The enhanced horizontal resolution and addition of vertical levels in the forecast model lead to an increase of 10 minutes in the computing time to complete the numerical forecasts.

12 Acknowledgements

We would like to thank the following colleagues for their various contributions to this project:

Seung-Jong Baek, Bin He, Mark Buehner, Cécilien Charette, Stéphane Laroche, Pierre Koclas, Ervig Lapalme, Sylvain Heilliette, Josée Morneau, Judy St-James, A. Erfani, Hai Lin, Ryan Muncaster, Ron McTaggart-Cowan, André Plante, Vivian Lee, Michel Desgagné and Michel Valin.

13 References

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, 124, 1256–1271.

Buehner, M. and co-authors, 2014: Changes to the Global Deterministic Prediction System (GDPS) from version 3.1.1 to version 4.0.0 Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/technote_gdps-400_20141118_e.pdf

Buehner, M., R. McTaggart-Cowan, A. Beaulne, C. Charette, L. Garand, S. Heilliette, E. Lapalme, S. Laroche, S. R. Macpherson, J. Morneau and A. Zadra, 2015: Implementation of Deterministic Weather Forecasting Systems based on Ensemble-Variational Data Assimilation at Environment Canada. Part I: The Global System. *Mon. Wea. Rev.* **143**, 2532-2559.

Candille, G., S. Beauregard and N. Gagnon, 2010: Bias Correction and Multiensemble in the NAEFS Context or How to Get a "Free Calibration" through a Multiensemble Approach. *Mon. Wea. Rev.*, **138**, 4268–4281.

Candille G., 2009: The Multi-ensemble Approach, The NAEFS Example, *Mon. Wea. Rev.*, **137**, 1655-1665.

Candille, G., C. Côté, P.L. Houtekamer and G. Pellerin, 2007: Verification of an Ensemble Prediction System against Observations. *Mon. Wea. Rev.*, **135**, 2688-2699.

Carrera, M. L., S. Bélair, V. Fortin, B. Bilodeau, D. Charpentier, and I. Doré (2010), Evaluation of snowpack simulations over the Canadian Rockies with an experimental hydrometeorological modeling system, *J. Hydrometeorol.*, 11, 1123–1140.Charron, M., R. Frenette and N. Gagnon, 2011: First Operational Implementation of the Regional Ensemble Prediction System at CMC (REPS 1.0.0). Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/op_systems/doc_opchanges/technote_reps_20111004_e.pdf

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell and L. Michelin, 2010: Toward Random Sampling of Model Error in the Canadian Ensemble Prediction System, *Mon. Wea. Rev*., **138**, 1877-1901.

Charron M., 2016: Generating space-time auto-correlated random fields on the sphere. Recherche en Prévision Numérique Technical Document. [Available on request from Recherche en Prévision Numérique, Environnement and Climate Change Canada, 2121 route Transcanadienne, Dorval, Québec, H9P 1J3, Canada.]

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev*., **126**, 1373–1395.

Côté, J., J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev*., **126**, 1397–1418.

Cui, B., Z. Toth, Y. Zhu, D. Hou, 2012: Bias Correction for Global Ensemble Forecast. *Wea. Forecasting*, **27**, 396–410.

Dee, D. P., and Co-authors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc., 137, 553-597.

Erfani, A., and Co-authors, 2013: The New Regional Ensemble prediction System at 15 km horizontal grid spacing (REPS 2.0.1) Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/technote_reps201_20131204_e.pdf

Fillion, L., H. Mitchell, H. Ritchie and A. Staniforth, 1995: The impact of a digital filter finalization technique in a global data assimilation system. Tellus, 47A, 304-323.

Gagnon, N., and Co-authors, 2015: Improvements to the Global Ensemble Prediction System (GEPS) from version 4.0.1 to version 4.1.1. Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/technote_geps-411_20151215_e.pdf

Gagnon, N., and Co-authors, 2014a: Improvements to the Global Ensemble Prediction System from version 3.1.0 to version 4.0.0. Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/technote_geps-400_20141118_e.pdf

Gagnon, N., and Co-authors, 2014b: Improvements to the Global Ensemble Prediction System reforecast system from version 3.1.0 to version 4.0.0. Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/Tech_Note_GEPS400_reforecast_v1.1_E.pdf

Gagnon, N., and Co-authors, 2013a: Improvements to the Global Ensemble Prediction System from version 2.0.3 to version 3.0.0. Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/op_systems/doc_opchanges/technote_geps300_20130213_e.pdf

Gagnon, N., and Co-authors, 2013b: Improvements to the Global Ensemble Prediction System from version 3.0.0 to version 3.1.0. Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121 route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/technote_geps310_20131204_e.pdf

Gagnon, N., and Co-authors, 2011: Improvements to the Global Ensemble Prediction System. Canadian Meteorological Centre Technical Note. [Available on request from Environment Canada, Centre Météorologique Canadien, division du développement, 2121, route Transcanadienne, 4e étage, Dorval, Québec, H9P1J3 or via the following web site :

http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/lib/op_systems/doc_opchanges/technote_geps_20110906_e.pdf

Gagnon, N., and Co-authors, 2007: An update on the CMC ensemble medium-range forecast system. Proc. ECMWF 11th Workshop on Meteorological Operational Systems, Reading, United Kingdom, ECMWF, 55–59.

Girard, C., 2015: Histoire d'une Inconsistance, RPN seminar: http://collaboration.cmc.ec.gc.ca/science/rpn/SEM/dossiers/2015/seminaires/2015-04-10/Seminar_2015-04-10_Claude-Girard.pdf.

Houtekamer, P. L., X. Deng, H. L. Mitchell, S.-J. Baek and N. Gagnon, 2014a: Higher resolution in an operational ensemble Kalman filter, *Mon. Wea. Rev*, **142**, 1143-1162.

Houtekamer, P.L., Mitchell H. L. and Deng X. 2009: Model Error Representation in an Operational Ensemble Kalman Filter, *Mon. Wea. Rev*., **137**, 2126-2143.

Qaddouri, A., and V. Lee, 2011: The Canadian Global Environmental Multiscale model on the Yin-Yang grid system. Quart. J. Roy. Meteor. Soc., 137, 1913-1926. http://web-

mrb.cmc.ec.gc.ca/mrb/rpn/SEM/dossiers/2015/seminaires/2015-04-10/Seminar_2015-04-10_Abdessamad-Qaddouri.pdf.

Qaddouri, A., 2015: The Future operational GDPS-YY25, RPN seminar: http://web-mrb.cmc.ec.gc.ca/mrb/rpn/SEM/dossiers/2015/seminaires/2015-04-10/Seminar_2015-04-10_Abdessamad-Qaddouri.pdf.

Qaddouri, A., and Co-authors, 2015: Changes to the Global Deterministic Prediction System (GDPS) from version 4.0.1 to version 5.0.0 - Yin-Yang grid configuration. Available at http://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/tech_notes/technote_gdps-500_20151215_e.pdf.

Tavolato, C., and L. Isaksen, 2010: Huber norm quality control in the IFS. ECMWF Newsletter (Meteorology section), No. 122, 27-31, available at https://www.ecmwf.int/node/17474.

Appendix A: Nomenclature

4DEnVar	Four-Dimensional Ensemble-Variational data assimilation
AMSU	Advanced Microwave Sounding Unit
AMV	Atmospheric Motion Vector
AIRS	Atmospheric Infrared Sounder
ARCAD	Objective verification system in use at CMC
ATMS	Advanced Technology Microwave Sounder
BSS	Brier skill score
CMC	Canadian Meteorological Center
CrIS	Cross-track Infrared Sounder
CRPS	Continuous Rank probability Score
EC	Environment Canada
ECCC	Environment and Climate Change Canada
EnKF	Ensemble Kalman Filter
ES1.5m	"Surface" 1.5-m dew-point depression
GDPS	Global Deterministic Prediction System
GEM	Global Environmental Multiscale model
GEPS	Global Ensemble Prediction System
GPS-RO	Global Positioning System - Radio Occultation data
IASI	Infrared Atmospheric Sounding Interferometer
LAM	Limited area model
MSLP	Mean sea-level pressure
NH	Northern hemisphere
NPP	Suomi National Polar-orbiting Partnership
MSC	Meteorological Service of Canada (of Environment Canada)
NAEFS	North American Ensemble Forecasting System
NCEP	(United States) National Centers for Environmental Prediction
NWP	Numerical Weather Prediction
O-P	Observations minus prediction
QPF	Quantitative precipitation forecast
RDPS	Regional Deterministic Prediction System
REPS	Regional Ensemble Prediction System
RMSE	Root-mean-squared error
SHEF	Standard Hydrological Exchange Format
T1.5m	"Surface (screen level)" 1.5-m temperature
UV10m	"Surface" 10-m wind speed
VIIRS	Visible Infrared Imaging Radiometer Suite
YY	Yin-Yang

Appendix B: Table of the GEPS 5.0.0 model configurations

#	Convection (* K&F: Kain & Fritsch)	Gravity wave drag	Mixing length	Vertical Diffusion	Orographic blocking	Deacu Z0T	Salty QSAT	SKEB	PTP
0	K&F	Standard	Bougeault	1.0	1.0	Yes	Yes	No	No
1	K&F	Strong	Blackadar	1.0	0.5	Yes	No	Yes	Yes
2	OldKuo	Strong	Blackadar	1.0	0.5	No	No	Yes	Yes
3	K&F	Weak	Bougeault	0.85	1.5	Yes	Yes	Yes	Yes
4	OldKuo	Weak	Bougeault	0.85	1.5	No	No	Yes	Yes
5	K&F	Weak	Blackadar	1.0	0.5	No	No	Yes	Yes
6	OldKuo	Weak	Blackadar	1.0	1.5	Yes	Yes	Yes	Yes
7	K&F	Weak	Bougeault	1.0	0.5	No	Yes	Yes	Yes
8	OldKuo	Weak	Bougeault	1.0	1.5	No	Yes	Yes	Yes
9	K&F	Strong	Bougeault	1.0	1.5	Yes	Yes	Yes	Yes
10	OldKuo	Strong	Bougeault	1.0	1.5	No	Yes	Yes	Yes
11	K&F	Strong	Bougeault	0.85	1.5	No	No	Yes	Yes
12	OldKuo	Strong	Bougeault	0.85	0.5	No	No	Yes	Yes
13	K&F	Weak	Blackadar	0.85	1.5	Yes	No	Yes	Yes
14	OldKuo	Weak	Blackadar	0.85	0.5	Yes	Yes	Yes	Yes
15	K&F	Strong	Blackadar	0.85	1.5	Yes	Yes	Yes	Yes
16	OldKuo	Strong	Blackadar	0.85	1.5	No	Yes	Yes	Yes
17	K&F	Strong	Blackadar	1.0	0.5	No	No	Yes	Yes
18	OldKuo	Strong	Blackadar	1.0	1.5	No	Yes	Yes	Yes
19	K&F	Weak	Bougeault	0.85	0.5	No	No	Yes	Yes
20	OldKuo	Weak	Bougeault	0.85	1.5	No	Yes	Yes	Yes