Environnement et Changement climatique Canada
Environment and Climate Change Canada

Global Ensemble Prediction System (GEPS)

Update from version 5.0.0 to version 6.0.0

Canadian Meteorological Centre

Technical Note

3 July 2019

Revisions			
Version	Date	Authors	Remarks
1.0	2019/04/15	H. Lin X.-X. Deng J.-S. Fontecilla A. Peterson	First draft
2.0	2019/04/24	H. Lin X.-X. Deng	Inputs from R. Muncaster, J.-S. Fontecilla and P. Houtekamer
2.1	2019/04/26		More inputs from P. Houtekamer and J.-S. Fontecilla
2.2	2019/04/26		Inputs from N. Gagnon
3.0	2019/04/29	H. Lin J.-S. Fontecilla A. Peterson	Modification of 6.2 including structure change. Figure changes for 6.3
3.1	2019/04/30	X.-X. Deng H. Lin	Minor revision
3.6	2019/07/02	X.-X. Deng, A. Peterson, H. Lin	Parallel run results added.

Table of Contents

Upgrade of the Global Ensemble Prediction System (GEPS) from version 5.0.0 to version 6.0.0

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

Hai Lin[2], Xingxiu Deng[1], Andrew Peterson[2], Juan Sebastian Fontecilla[1], Greg Smith[2], Ryan Muncaster[2], Normand Gagnon[1], Pieter L. Houtekamer[2], Martin Charron[2]

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

SUMMARY

In this implementation, the GEPS is upgraded to the version 6.0.0 which applies a global atmosphere-ocean coupled model, i.e., the GEM atmospheric model is coupled with the Nucleus for European Modelling of the Ocean model (NEMO), for the forecasts. The data assimilation component still uses the uncoupled GEM model.

The GEM version is upgraded to 4.8-LTS.16 which allows for the evolution of vegetation and ozone during the forecast and reforecast integrations.

The assimilation component benefits from the ECMWF-hybrid gain configuration. The EnKF now assimilates T–Td data from Radiosondes up to 100 hPa instead of 200 hPa. The assimilation component is affected by changes in the processing of the observations made to the new GDPS (version 7.0.0) which applies GEM 5 and will be implemented at the same time. For the surface fields of GEPS 6.0.0, a surface pseudo-analysis using GEM4.8 on the GDPS 6.0 25 km resolution grid is performed to maintain the compatibility. For the GEPS prediction component, GEM 4.8 is coupled with NEMO 3.6 to have an atmosphere-ocean-sea ice fully coupled model.

The new GEPS 6.0.0 takes advantage of sources of predictability associated with the air-sea-sea ice coupling in the forecast. It also predicts the ocean and sea ice conditions. Overall, the above changes lead to improvement in the trial field quality, and better forecast performance compared to the previous GEPS 5.0.0.

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 provide 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, Deng et al. 2018). 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). 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 that are available at: http://weather.gc.ca/ensemble/index_e.html (MSC public weather web site). Since 2015, once a week (Thursday at 0000 UTC), the forecasts are extended to 32 days to produce forecasts for the next 4 weeks (Gagnon et al. 2013b, Lin et al. 2016). The official monthly temperature forecast is produced each Thursday which covers the 4 weeks starting on the following Monday.

It is 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). The REPS is producing up to 72 h operational ensemble forecasts over North America at 15 km horizontal resolution. Note that the new REPS which will be implemented at the same time is upgraded to have 10 km horizontal resolution.

This document describes the latest modifications to the Canadian Global Ensemble Prediction System, becoming GEPS version 6.0.0, hereafter referred to as the GEPS 6.0.0. A major upgrade is that the forecast of GEPS 6.0.0 is made with the GEM-NEMO coupled model, in contrast to the uncoupled GEM in the previous versions of GEPS. 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 June 2019 to replace GEPS version 5.0.0. Section 5 describes modifications to the reforecast procedure. Evaluation results of the final series test and the parallel run are detailed in Sections 6 and 7. Information is given in Section 8 on performance of dependent systems and in Section 9 on availability of products. Finally, we summarize results in Section 10.

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 4.8-LTS13 to 4.8-LTS.15. This upgrade permits the evolution of vegetation and ozone during the integration of the model that is used to produce the forecasts in the GEPS prediction component. This upgrade is also done for the model that is used to produce the trial fields in the GEPS assimilation part, but the vegetation and ozone evolution are not activated. During the implementation of the forecast component for the parallel run, the GEM model was further updated to version 4.8-LTS.16 in order to correct a problem of excessive over-forecasting for soil and screen-level temperature in case of rain falling on snow. The GEM model version remains 4.8-LTS.15 for the data assimilation component.

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 ECMWF-hybrid gain

There are two data assimilation algorithms at our center. One is the EnKF used to initialize the GEPS, and the other is the 4DEnVar used to initialize the GDPS. To maximize the benefits of having the two co-evolving algorithms, various hybrid gain configurations, including EnKF-control, EnVar-control, ECMWF-hybrid gain and the newly developed CMC-hybrid gain, were investigated and compared (Houtekamer et al. 2018). It was found that the ECMWF-hybrid gain performs better than either the EnKF-control or the EnVar-control and that the tuned CMC-hybrid gain gives the best results. With the CMC-hybrid gain algorithm, one half of the ensemble members is re-centered around the 4DEnVar analysis, while the other half of the ensemble members remain unchanged. This CMC-hybrid gain algorithm, however, will lead to changes in the ensemble spread, especially in the stratosphere (see Figure 1 of Houtekamer al. 2018). This will have an impact on our centre's 4DEnVar, which uses the EnKF flow-dependent background-error covariances. Since no comprehensive test has been done for this change in the 4DEnVar, the CMC-hybrid gain is dropped for now (but will be delivered in a future implementation). The ECMWF-hybrid gain is adopted for this implementation. With the ECMWF-hybrid gain, all ensemble members are re-centered around the mean of the EnKF analyses and the 4DEnVar analysis. Here the 4DEnVar analysis is produced on the same model grid as for the EnKF while using the ensemble mean background trajectory from the EnKF as first-guess. When compared to the EnKF, the 4DEnVAR, which is relatively less costly, takes three more data types (SSMIS, CSR and GB) and assimilates almost four times more observations (see Table 2 of Houtekamer et al. 2018).

3.2 Accept T- Td from radiosondes up to 100 hPa instead of 200 hPa

With the previous implementation of our centre's GDPS version 6.1.0 in 2018, the quality control for humidity observations was greatly improved. Following this improvement, the new EnKF is now assimilating dew-point depression (T – Td) data from Radiosondes up to 100 hPa instead of 200 hPa.

3.3 Changes coming from the upgrade of the GDPS

Same as for the presently operational version, the EnKF is two-way coupled with the 4DEnVar of the GDPS, thus any changes in the processing of the observations in the new GDPS will have impact on the EnKF as well. In our centre's GDPS 7.0.0 to be implemented concurrently (see the technical note of the GDPS 7.0.0), a major change is the upgrade of the GEM model from version 4.7 to 5, which has an impact on the bias correction of radiance observations. The GDPS also provides surface fields to the EnKF for use in integrating the GEM model. The surface fields include analyses of snow depth (SD), screen-level temperature (TS) and pseudo-analyses of ISBA fields (such as soil temperature I0 & soil moisture I1). The GEM model used in the EnKF remains at version 4.8. During the tests of final cycles, it was found that the new soil moisture is much drier and the new soil temperature is much warmer for the summer season than those in the operational run. This inconsistency affects the quality of the EnKF initial conditions, but to a lesser degree than the impacts on the quality of the GEPS forecast, especially the week-2 forecast. The main reason is that the GEM version 5 has improved precipitation forecasts and generally produces less precipitation in summer than the GEM version 4 does. As will be mentioned in section 4, in order to solve this inconsistency issue, a pseudo-analysis module had to be implemented using trial fields from the GEM model version 4.8 for the GEPS forecast component.

3.4 New SST and sea ice

As with other uncoupled systems, a new sea surface temperature (SST) analysis at 0.1 degree resolution and new ice analysis are applied in the assimilation. In the forecast component, the ocean and sea ice are initialized with the GIOPS analysis.

4 Modifications to the forecast component of GEPS

The forecast component of GEPS 6.0.0 applies the GEM-NEMO atmosphere-ocean-sea ice coupled model, in contrast to the previous versions of GEPS that used the uncoupled GEM.

4.1 New version of GEM

As described in section 2, the atmospheric model in the GEPS 6.0.0 forecast component is updated to GEM 4.8-LTS.16 from GEM 4.8.LTS.13. This version of GEM allows the evolution of vegetation and ozone during the forecast and this was activated. Both fields are now varying linearly between mid-month climatological values. This has a very small positive impact on long-range forecast.

4.2 The NEMO ocean model

The ocean model NEMO (Nucleus for European Modelling of the Ocean, http://www.nemo-ocean.eu) is NEMO 3.6 ORCA 0.25° with 50 vertical levels. The CICE 4.0 (Community of Ice CodE, Hunke and Lipscomb 2010) model is used for the sea-ice component.

GEM and NEMO are both run with a 15-minute time step. They are coupled through the GOSSIP coupler at every time step.

4.3 Ocean and sea ice initializations

The ocean and sea ice initial conditions in the forecast come from the CMC GIOPS analysis (Smith et al. 2016). For the reforecast, the ocean and sea ice initial conditions are described in section 5.2. All ensemble members have the same ocean and sea ice initial conditions.

4.4 Land Surface initial condition

For several years, GEPS predictions have been initialized with the land surfaces fields coming from the deterministic prediction system GDPS. Until recently, both the deterministic and the ensemble prediction systems use the same version of the atmospheric model, GEM 4.8, and the surface fields produced by the deterministic system were then compatible with the GEPS system. However, the GDPS is upgraded (GDPS 7.0) to a newer version of the GEM model (the version 5) while the GEPS is still using GEM 4.8. The surface fields produced by this new version of GEM are significantly different to those produced by GEM 4.8. The land surface fields generated by GEM 5.0.0 are generally warmer and drier than those from GEM 4.8 for the summer season. As a result, when using the land surface initial conditions from GDPS 7.0 the forecast skill of the GEPS is degraded, especially in week 2 during summer season, comparing to the current operational GEPS 5.0.0. The impact in the winter season is small.

In order to solve this problem of inconsistency, a surface pseudo-analysis (called PSEUDO25) is performed using the GDPS 6.0 25 km resolution grid to generate the land surface fields. The trial fields feeding this pseudo-analysis procedure are generated with the GEM 4.8 model using the previous GDPS (6.0) configuration. The land surface fields generated by PSEUDO25 are close to those of GDPS 6.0. Experiments with these PSEUDO25 land surface initial conditions confirm that the forecast skill in warm season is significantly improved with this fix.

It should be noted that this inconsistency problem arises from the different versions of GEM used in GDPS and GEPS to be implemented at the same time. The above-discussed solution is just temporary. Underway GEPS development includes migrating GEM from version 4.8 to 5 and testing the pseudo-analysis procedure on the GEPS grid. When the GEPS is updated to GEM version 5 in a future implementation, the land surface initial conditions for GEPS could again come from the concurrent operational GDPS system. It is also possible that the GEPS will have its own pseudo-analysis procedure for both data assimilation and forecast component, which will prevent inconsistency problems in the future.

5 Modifications to the reforecast procedure

5.1 Use of GEM-NEMO

As mentioned in the Introduction, the ECCC official monthly forecast is produced by extending the GEPS forecast to 32 days once a week on Thursday (00Z). With the model integration extended, model drift and systematic model errors become a serious problem that contaminates the forecast quality. A common practice in subseasonal to seasonal (S2S) prediction is to perform a historical reforecast to estimate and correct the systematic error. Therefore, the reforecast is an important component of the monthly forecast system. As in GEPS 5.0.0, the reforecast covers 20 years with 4 members of 32 day integration once per week (Thursday 00Z) to generate a historical database of 80 reforecasts for a given date as explained in Gagnon et al. (2014b). In the reforecast of GEPS 6.0.0, the GEMO-NEMO coupled model is now used, which has, as required, the same configuration as in the forecast. The period of reforecast is 1998-2017.

5.2 Initial conditions of the reforecast

The atmospheric initial conditions are generated in the same way as in GEPS 5.0.0, i.e., random isotropic perturbations are added to the ERA-interim reanalysis (Dee et al. 2011). As in GEPS 5.0.0, the land surface initial conditions are generated by running the offline Surface Prediction System (SPS; Carrera et al. 2010) forced with the ERA-interim re-analyses.

The ocean initial fields come from the ORAP5 ocean reanalysis (Zuo et al. 2015). The monthly values of temperature, salinity, zonal and meridional currents are interpolated to daily values. The sea ice concentration initial fields are the Had2CIS, which was prepared by Woo-Sung Lee of CCCma, who combined the digitized sea ice charts from the Canadian Ice Service (CIS) with the HadISST2.2. The original HadISST2.2 employs an ice chart-based bias correction of the passive microwave record (Titchner and Rayner 2014). The monthly HadISST2.2 data, the digitized CIS weekly sea ice charts over the Arctic region and the weekly CIS "Great Lakes ice charts" over Great Lakes are interpolated to daily data before the combination. The sea ice thickness is interpolated from the monthly ORAP5 data. The near surface air temperature from ERA-interim and the snow depth from SPS are used for the initialization of these variables over sea ice in CICE.

For the period after 2016 when the ORAP5 and the Had2CIS data are not available, the ocean and sea ice initial conditions come from the CMC GIOPS analysis. As the ORAP5 and the Had2CIS sea ice have a similar climatology as the GIOPS analysis, such change in data source does not cause significant discontinuity in the reforecast behavior. During the implementation of the reforecast component, ORAP5 will be replaced by ORAS5, which is a more recent version of the ocean reanalysis from ECMWF.

5.3 Evaluation of the reforecast

With 20 years of data, the reforecast can be verified against the historical observations or reanalyses. It should be stressed that this exercise is not meant to quantify the performance of the GEPS system. Instead, it is to provide an estimate of the quality of the reforecast itself, as a benchmark for future versions. As the reforecast has a much smaller ensemble size (4 members) than the real-time forecast (21 members) and the quality of initial condition is not as good as in the real-time forecast, the skill obtained here can only be considered as the lower bound of the forecast skill of the system.

The verification is performed against the ERA-interim reanalysis to compare the reforecast of GEPS 6.0.0 with GEPS 5.0.0. For the winter season (DJF), reforecasts of three dates of January 7, February 4 and December 1 are used, while for the summer season (JJA), the three dates of June 2, July 7 and August 4 are analyzed. The reforecast analyzed here is for the 20 years from 1995 to 2014, and each season has 60 forecasts of 4 members. The evaluation is done for weekly mean variables, where the week refers to the calendar week of 7 days from Monday to Sunday. As the forecast starts on Thursday, week 1 corresponds to days 5-11, week 2 is days 12-18, and so on.

Figure 1 compares the systematic error of T2m in DJF from week 1 to week 3 between GEPS 5.0.0 (left panels) and GEPS 6.0.0 (right panels). As can be seen, GEPS 5.0.0 has a strong warm bias in the North Polar Region. This is likely caused by the inaccurate sea ice specification in the reforecast of GEPS 5.0.0 and the 3% minimum fraction of sea ice leads in the uncoupled GEM.

In GEPS 6.0.0, with the interactive sea ice as well as the initial Had2CIS and corrected sea ice temperature, the T2m bias in the polar region is much reduced. During the JJA season, the difference of T2m bias between GEPS 6.0.0 and GEPS 5.0.0 is small (not shown).

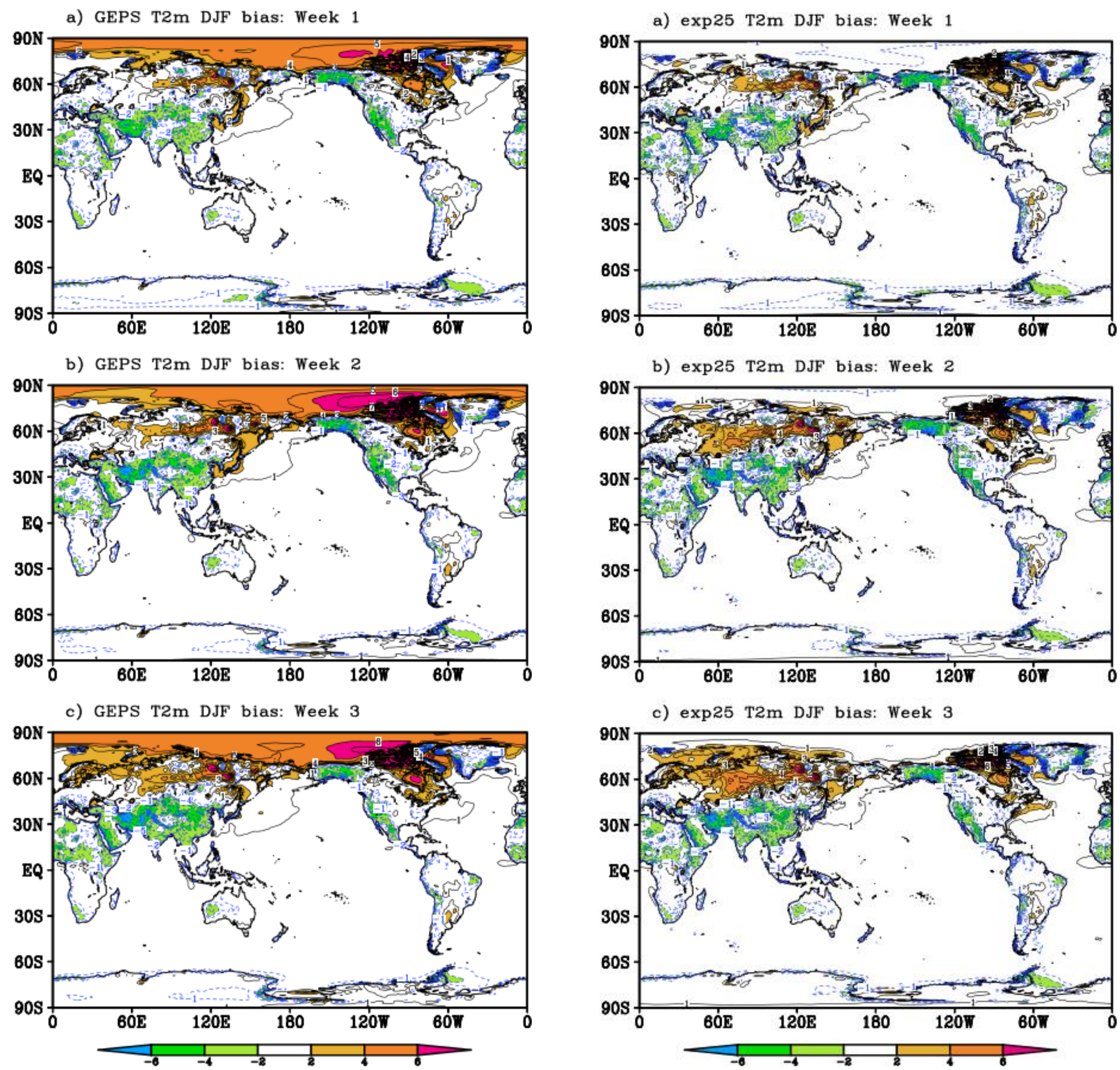


Figure 1. T2m bias for DJF reforecast against ERA-Interim reanalysis from week 1 to week 3 for GEPS 5.0.0 (left panels) and GEPS 6.0.0 (right panels). Unit:°C.

Shown in Figure 2 are the anomaly correlation skills for the 4-member ensemble mean T2m forecast in the DJF season averaged in different regions. It is clear that GEPS 6.0.0 reforecasts outperforms GEPS 5.0.0 reforecasts in all regions, especially in weeks 3 and 4.

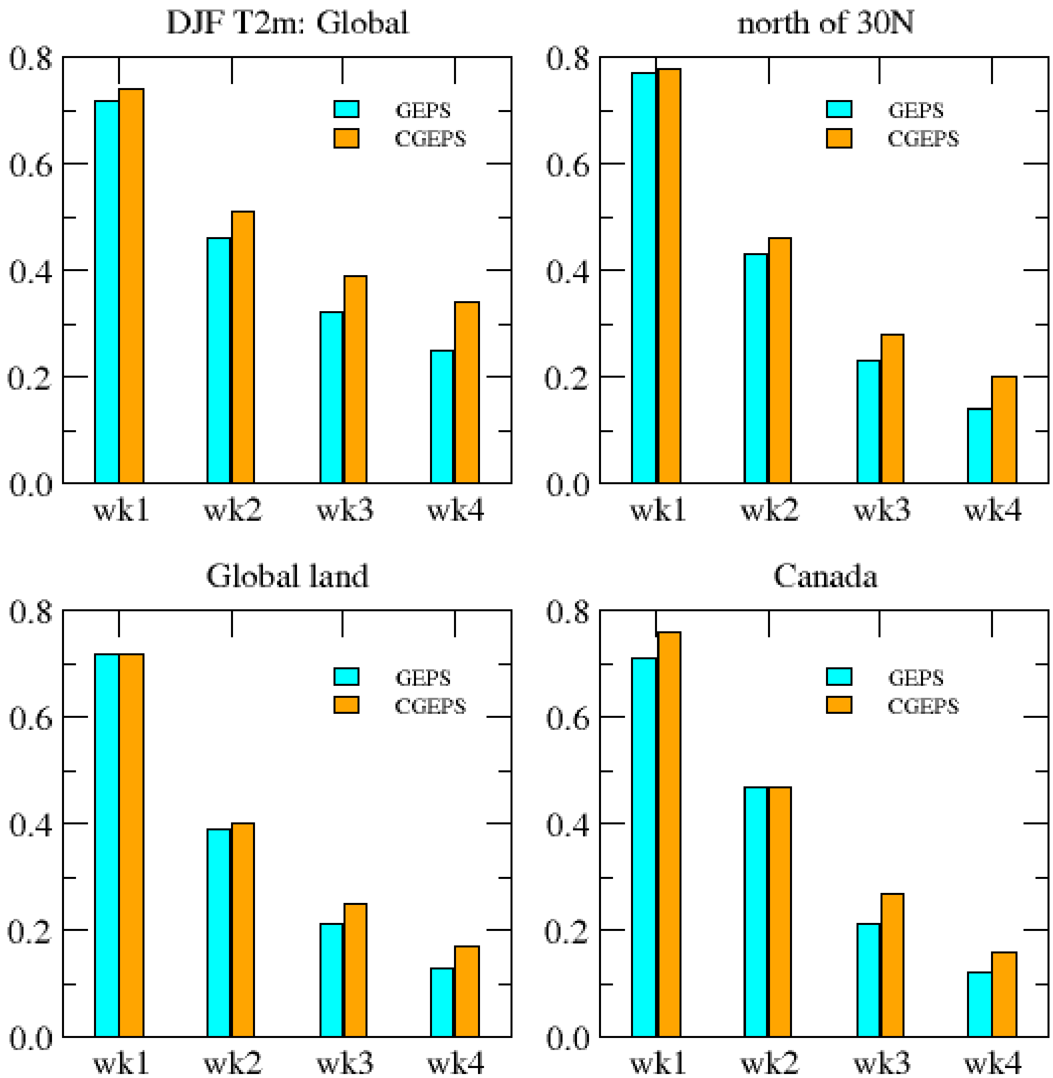


Figure 2. Anomaly correlation skill for the 4-member ensemble mean T2m forecast in the DJF season during the 20 year reforecast period averaged over the globe (upper left), north of 30N (uuper right), global land (lower left), and Canada (lower right). The light blue and orange bars represent GEPS 5.0.0 and GEPS 6.0.0, respectively.

The difference of correlation skill between GEPS 6.0.0 and GEPS 5.0.0 is presented in Figure 3. It is evident that over the globe the GEPS 6.0.0 reforecast is statistically more skillful than GEPS 5.0.0 for all lead times. In the other regions, significant differences can be found for weeks 3 and 4.

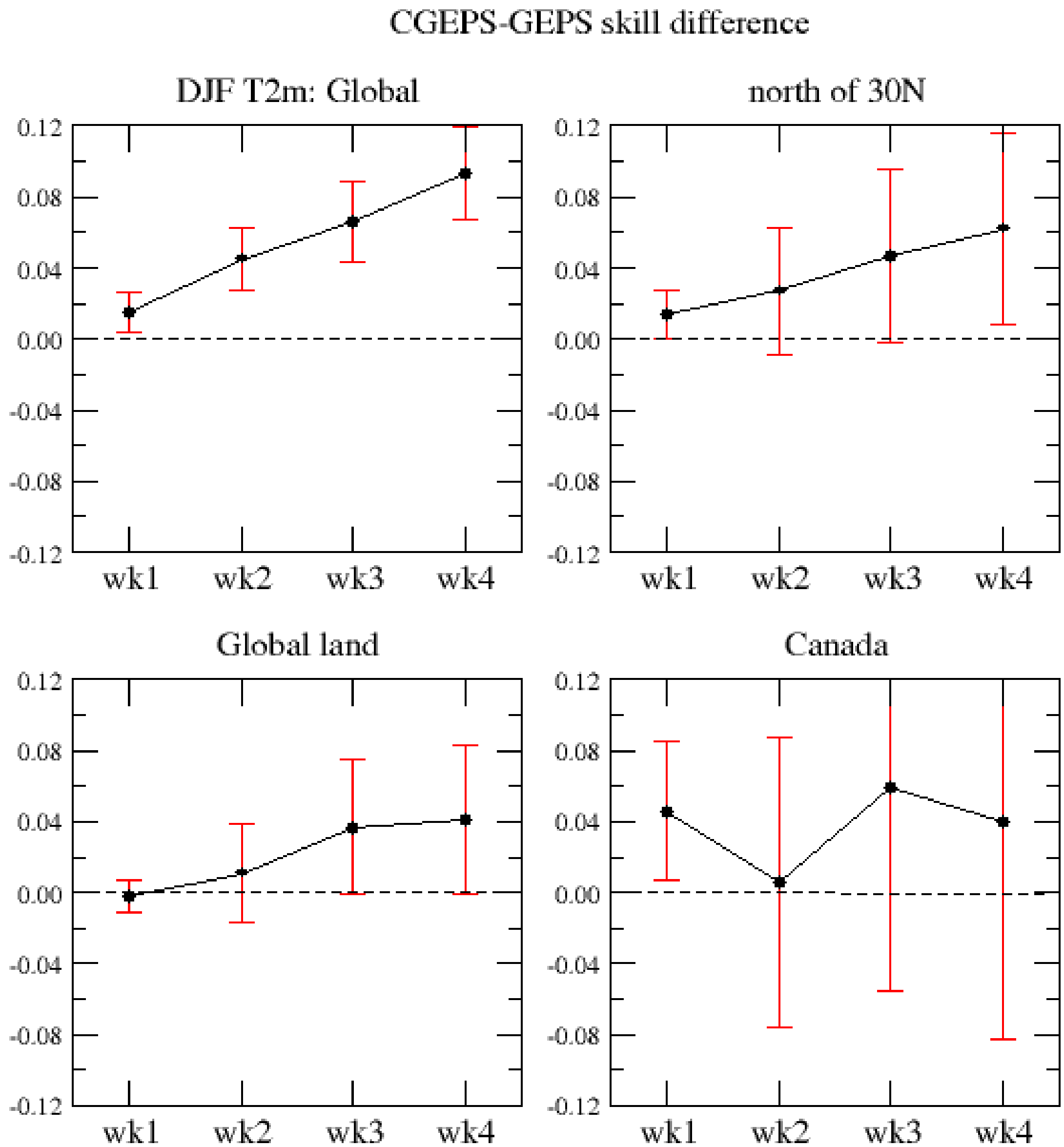


Figure 3. DJF T2m Skill difference between GEPS 6.0.0 and GEPS 5.0.0. The error bars represent the 95% statistical significance according to a bootstrap resampling test.

For the JJA season, similar conclusions as for DJF are obtained for the T2m anomaly correlation skill (Figure 4), which shows improvement of skill from GEPS 5.0.0 to GEPS 6.0.0.

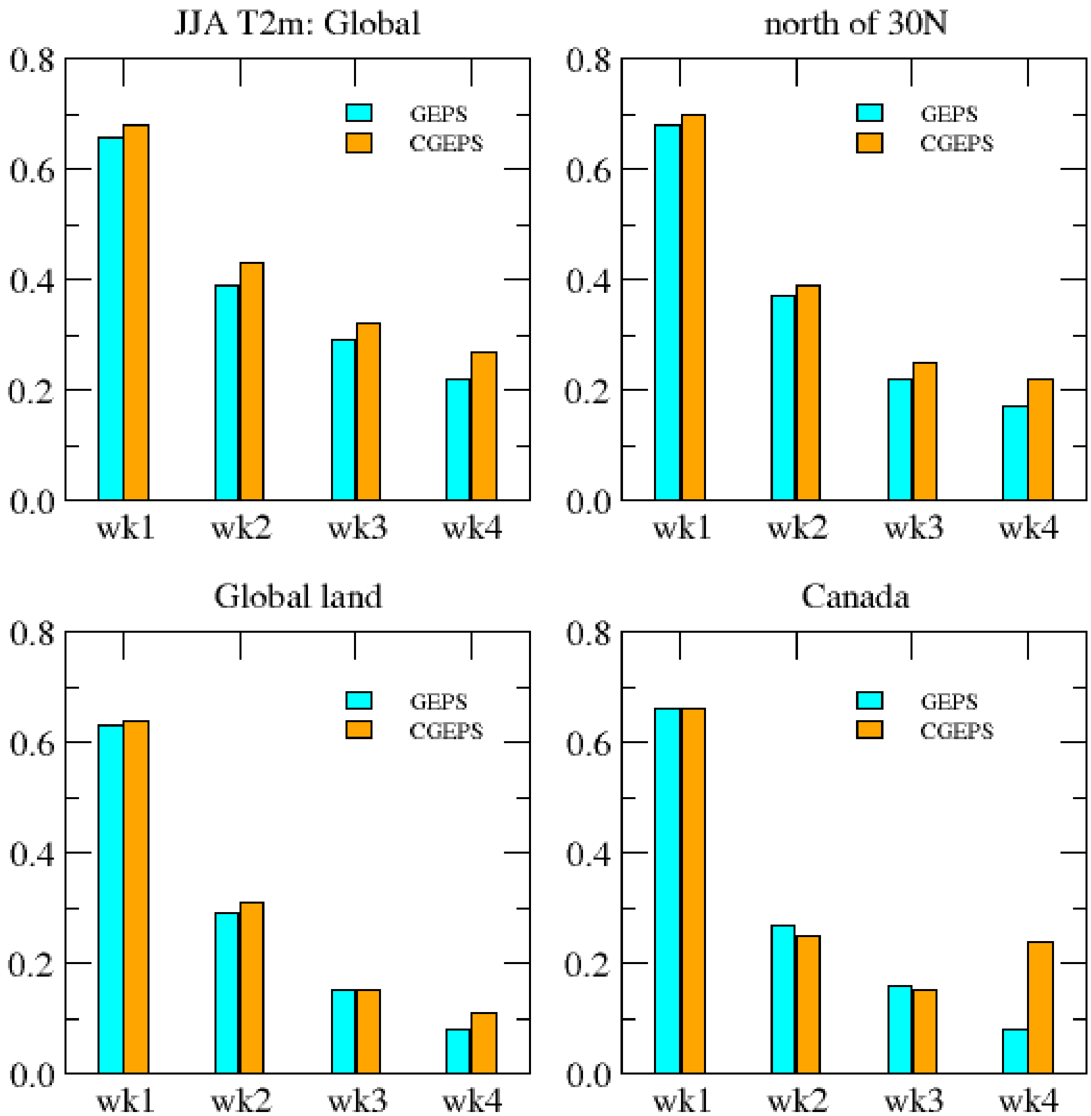


Figure 4. Same as Figure 2, but for JJA.

The skill difference between GEPS 6.0.0 and GEPS 5.0.0 in JJA is shown in Figure 5. In general, GEPS 6.0.0 is more skillful than GEPS 5.0.0, although the difference over the global land and Canadian regions is not statistically significant for weeks 1-3.

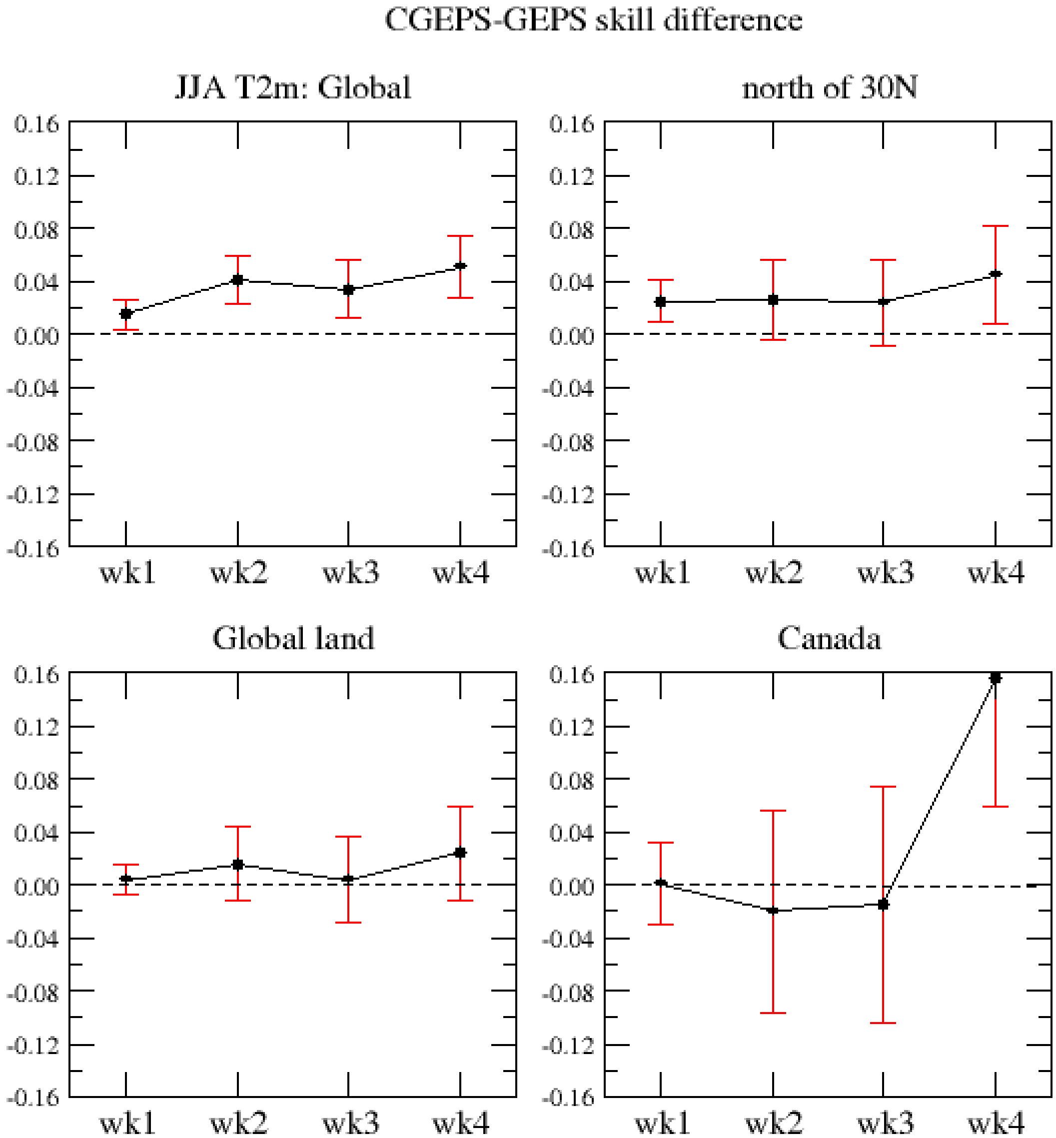


Figure 5. Same as Figure 3, but for JJA.

The Madden Julian Oscillation (MJO) is the dominant mode of variability in the tropics, which is characterized by a large-scale 30-50 day tropical wave coupled with convection propagating eastward along the equator (Madden and Julian 1971). A large number of studies have shown that the MJO significantly influences the global weather and provides an important source of predictability on the subseasonal time scale (e.g., Waliser et al. 2003, Lin and Brunet 2009, Lin et al. 2010a, b). The forecast skill of the MJO is a measure of the general quality of a subseasonal prediction system (e.g., Gottschalck et al. 2010, Vitart 2017). Shown in Figure 6 is the correlation skill of the Real-time Multivariate MJO (RMM) index (Wheeler and Hendon 2004) in the DJF season for the GEPS 5.0.0 (black) and GEPS 6.0.0 (red) reforecasts. The correlation of 0.5 is usually considered as a skillful MJO forecast. As can be seen, GEPS 5.0.0

has a skillful MJO forecast up to about 17 days. The skill in GEPS 6.0.0 is improved by about 4 days, reaching about 21 days. Similar improvement of the MJO skill is obtained for the JJA season (not shown), although the skill is lower than for DJF.

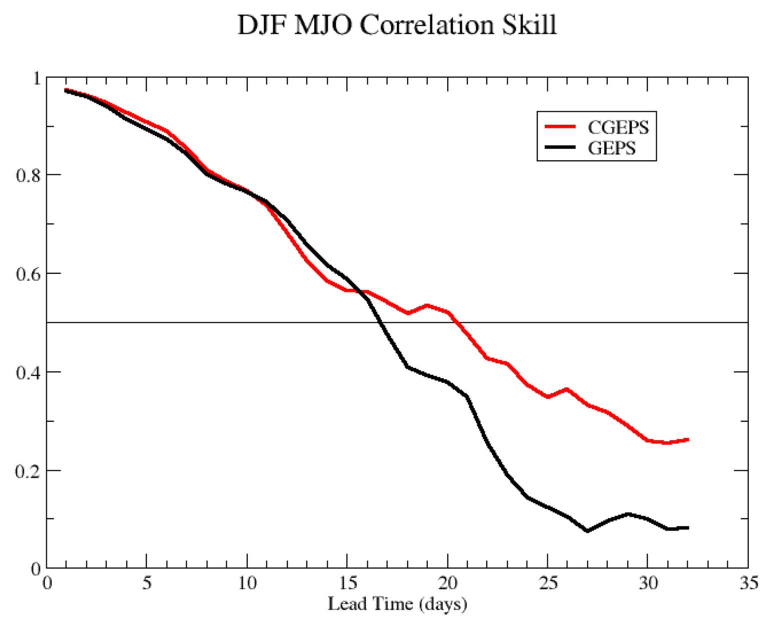


Figure 6. MJO bivariate correlation skill as a function of lead time for the 4-member ensemble mean of the reforecast for GEPS 5.0.0 (black) and GEPS 6.0.0 (red).

6 Objective evaluation of the final series of tests

In this section, we present the evaluation of the performance of the new system GEPS 6.0.0 in comparison to GEPS 5.0.0 during the final series of tests. The final series of tests, coupled with the GDPS 4DEnVar, were warm-started from the trials of the previous final series of tests with GEPS 5.0.0. To evaluate the impact of all the changes described above in the analyses and forecasts, the objective verifications were performed for summer 2016 and winter 2017.

6.1 Quality of the atmospheric initial conditions

To measure the quality of trial fields produced by the EnKF, the innovation statistics have been calculated using the ensemble mean trial field at the central analysis time and the 3D radiosonde observations (the so-called ARCAD verifications). The so-called SQLITE verifications with interpolated (to observation space) ensemble mean background trajectories against other data sets, such as GPS-RO and radiance data were also performed. The averaged results over different regions including the globe, the Northern extratropics, the Southern extratropics and the tropics are calculated for the winter (from 1 January to 28 February 2017) and summer (from 1 July to 31 August 2016) cycles respectively.

For both summer 2016 and winter 2017, the ARCAD verifications globally show significant improvement in biases and especially in standard deviations for all the variables (see Figure 7 for summer, the figure for winter with similar features is not shown). We note very slight degradation in biases for the U-wind component in the lower stratosphere, for temperature from 200 to 100 hPa and near the surface, and for geopotential heights below 700 hPa. Over the Southern extratropics we also see improvement in standard deviations for all the variables but to a lesser degree and significant reduction in biases of geopotential heights (Figure 8 for summer, the figure for winter with similar results is not shown). Apparently, the improvement in standard deviation in the global verifications mostly comes from the Northern extratropics (figures not shown). Over the tropics, we note a very big reduction in the biases of geopotential heights from 1000 hPa to 10 hPa and in the standard deviations of the U- and V-wind components (figure not shown). A slight degradation in biases is found for the lower-stratosphere U-wind and near-surface temperature. The ARCAD verifications for the final series of tests were performed against radiosonde observations from the 2016-2017 operational GDPS quality-controlled "postalt" files, which includes radiosonde humidity observations from some stations with large biases in the higher troposphere/lower stratosphere. This helps explain why we see large biases in dew-point depression for the proposed new system. The SQLITE verifications against each system's own assimilated radiosonde dew-point depression show big improvement in both biases and standard deviation from the middle troposphere to the lower stratosphere (Figure 9). Note that new observations are assimilated above 200 hPa in the new EnKF leading to a higher data count. We note slight dry biases below 925 hPa. The dry bias near the surface may result from the new drier soil moisture fields from the upgraded GDPS (see section 3.3). For the winter season, this dry bias extends from the surface to 750 hPa, which is one of the few negative impacts from using the ECMWF-hybrid gain.

Verifications against screen-level temperature show significant reduction in standard deviation, and also reduction in biases for all latitudes except near the South pole with larger warm bias (for the summer cycle, see Figure 10) and the North pole with larger cold bias (for the winter cycle, figure not shown). The larger cold bias near the North Pole may be related to the I7/I9 issue in the final cycles of the GDPS. Verifications against screen-level dew-point depression show neutral to mixed results. For both the summer (see Figure 11) and winter season (figure not shown), we note a decrease in data count near the South pole, which results in large changes to standard deviation and bias that are not statistically significant.

The summer and winter verifications against radiance observations both show improvement for AMSU-A in bias for channel 4-6 & 13 (but degradation for channel 14), and in standard deviation especially for channel 13 & 14 (figure not shown). For AMSU-B (figure not shown), we note significant improvement in standard deviation for all channels and in bias for channel 3

& 4 but degradation in bias for channel 2, which is sensitive to near surface humidity. Figure 12 shows verifications, for the summer season, against ATMS, whose channels 5 - 15 are AMSU-A like temperature sensitive and channels 17 – 22 are AMSU-B like humidity sensitive. Thus, similarly, we observe significant improvement in standard deviation for upper-level temperature sensitive channels and all humidity sensitive channels. With regard to biases, we note slight improvement except for channel 15 (upper-level temperature sensitive channel) and 17 (low-level humidity sensitive channel).

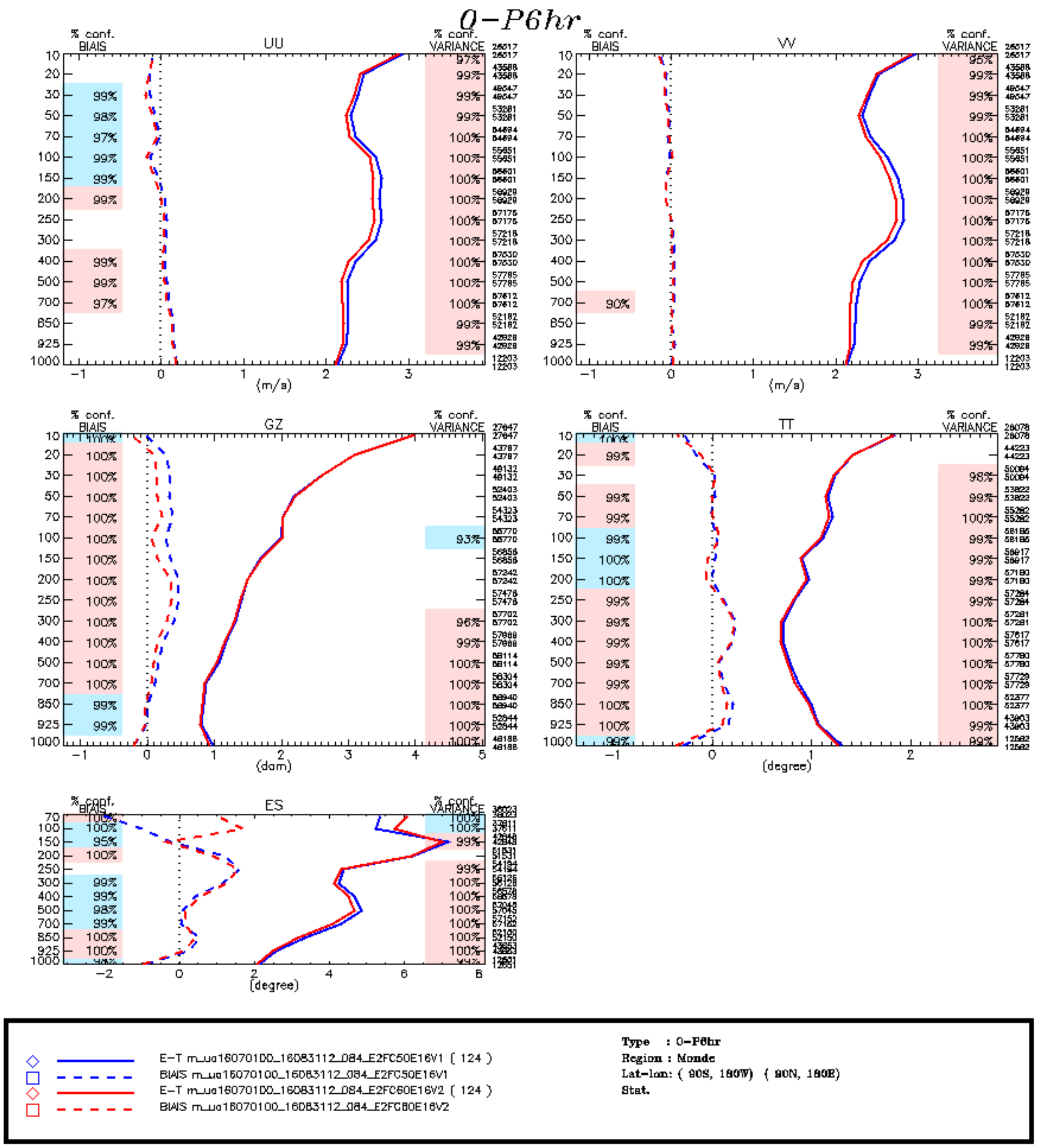


Figure 7. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the globe in the summer 2016 for the EnKF experiments with GEPS 5.0.0 (blue) and GEPS 6.0.0 (red).

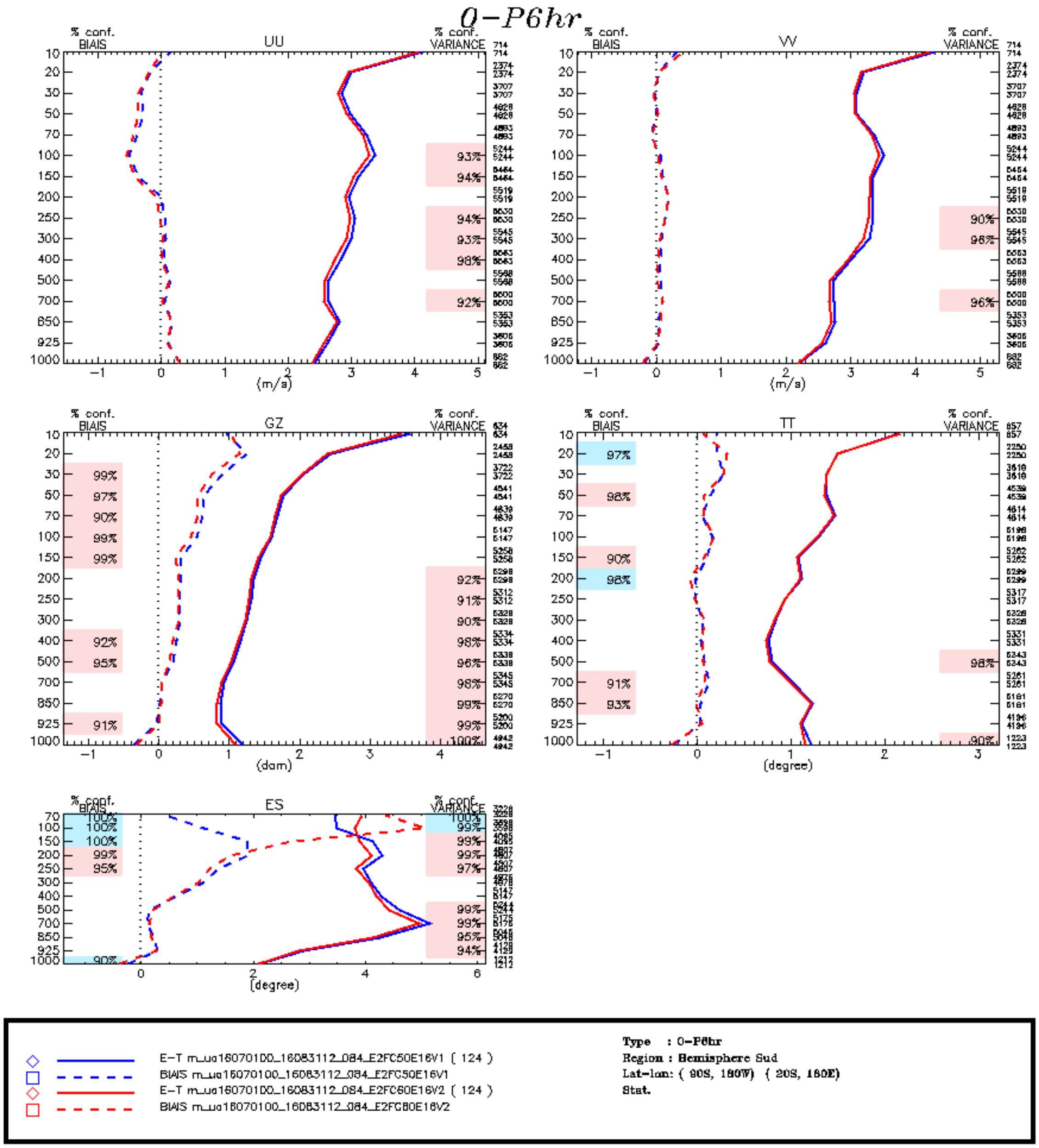


Figure 8. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the Southern Extratropics in the summer 2016 for the EnKF experiments with GEPS 5.0.0 (blue) and GEPS 6.0.0 (red).

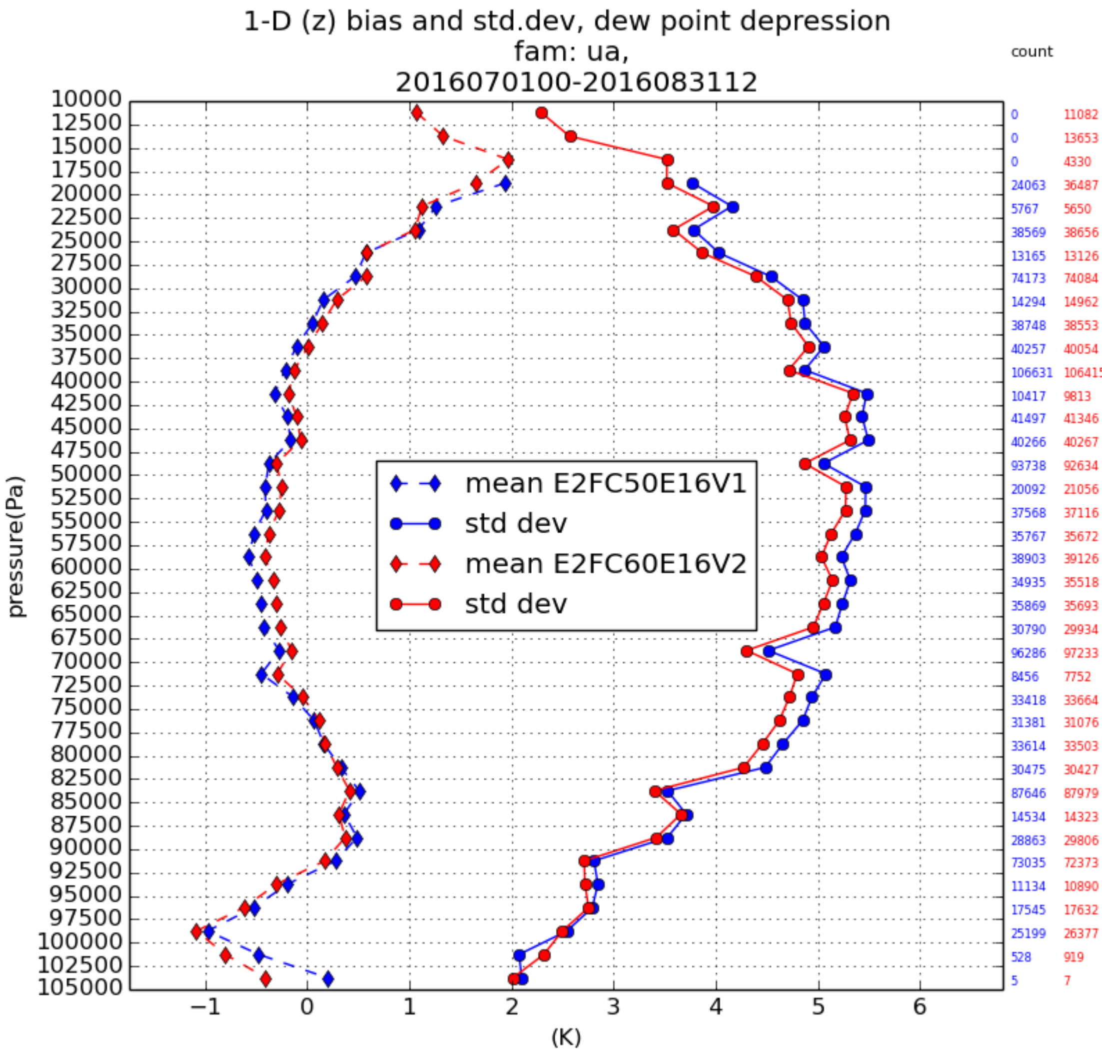


Figure 9. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosonde dew-point depression in the summer 2016 for the EnKF experiments with GEPS 5.0.0 (blue) and GEPS 6.0.0 (red).

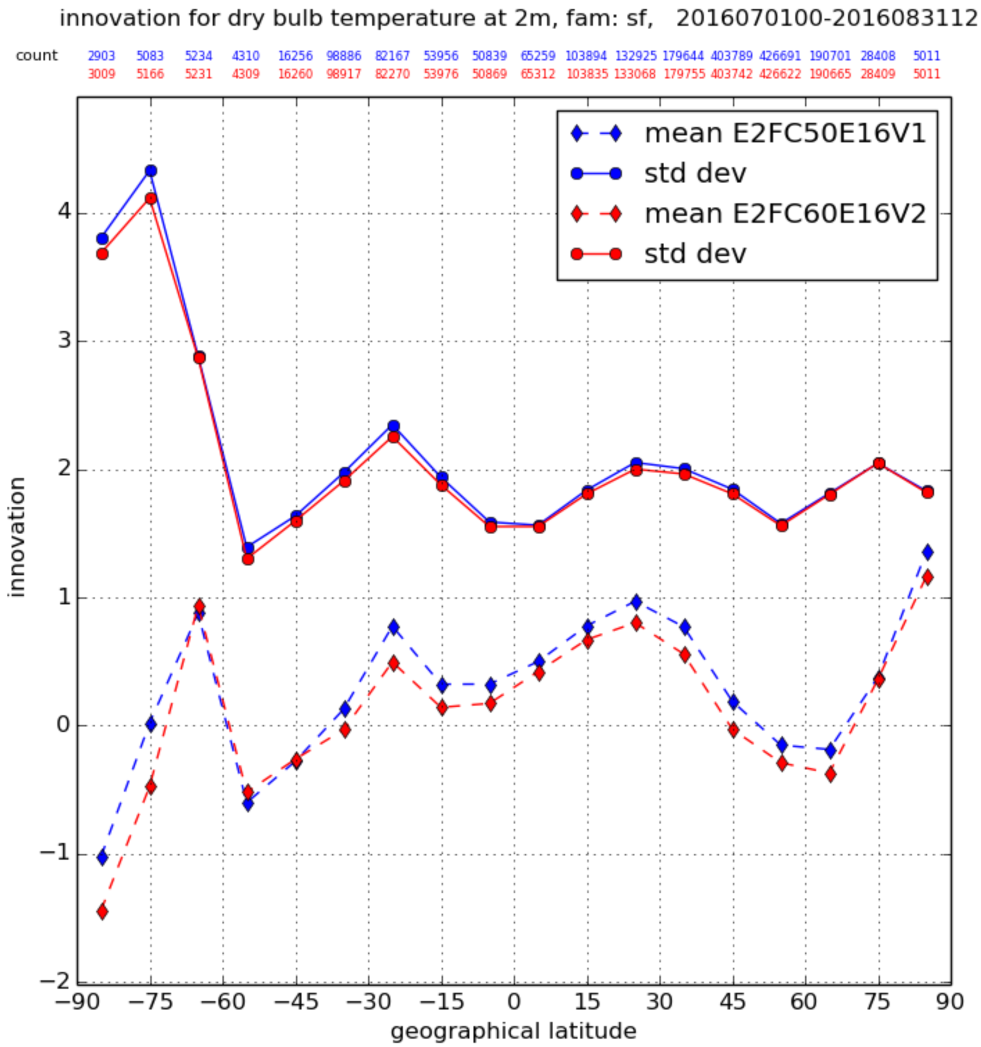


Figure 10. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against surface temperature in the summer 2016 for the EnKF experiments with GEPS 5.0.0 (blue) and GEPS 6.0.0 (red).

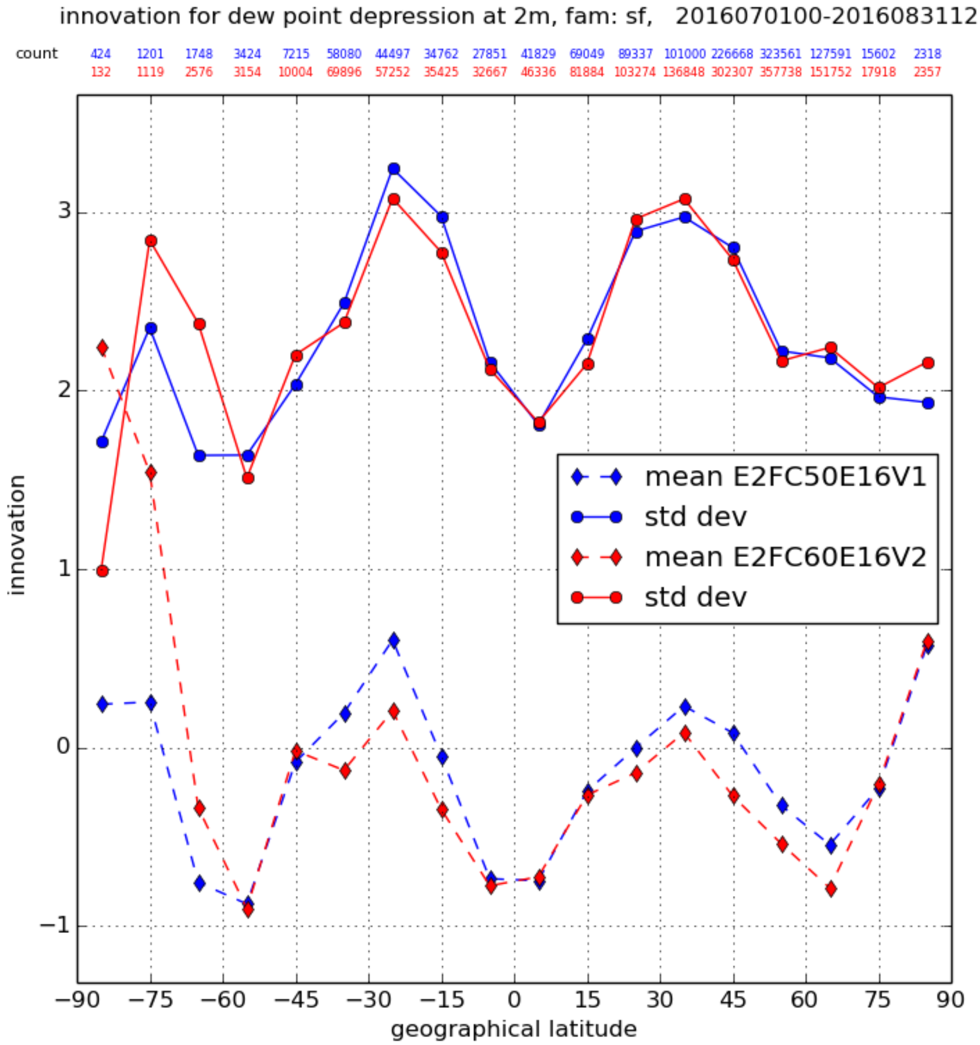


Figure 11. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against surface dew-point depression in the summer 2016 for the EnKF experiments with GEPS 5.0.0 (blue) and GEPS 6.0.0 (red).

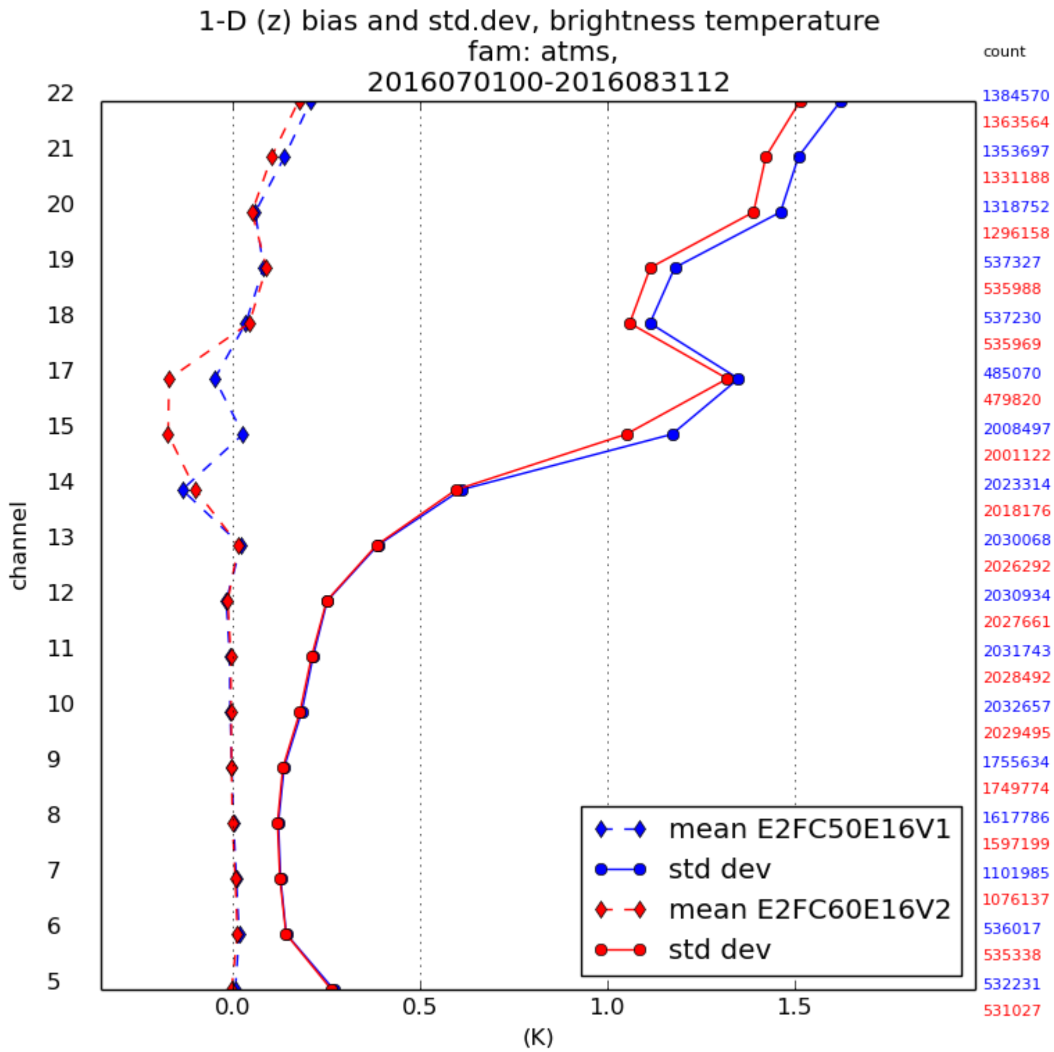


Figure 12. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) in the Summer 2016 over the globe against ATMS for the EnKF experiments with GEPS 5.0.0 (blue) and GEPS 6.0.0 (red).

6.2 Quality of the GEPS forecasts

For the summer season, 15-day forecasts were generated at every 36 hours for the period from 2016061600 to 2016083112 with a total of 52 cases. In winter, the period from 2016121512 to 2017022812 was used (52 cases). Verifications were carried out by comparing the forecasts with the observations. For the upper air, the observations used are the global radiosonde network (approximately 600 stations) at seven pressure levels: 10, 50, 100, 250, 500, 850 and 925 hPa. For the surface fields, we use the objective surface analyses at CMC of the long cut-off (named G6) that were produced based on the global network of surface synoptic station data (approximately 10000 stations). For precipitation, the 24 hr QPF (quantitative precipitation forecasts) were evaluated, using the global surface synoptic stations network (approximately 7000 stations). The accumulation of total precipitation forecast over a 24 hour period was verified against these station observations.

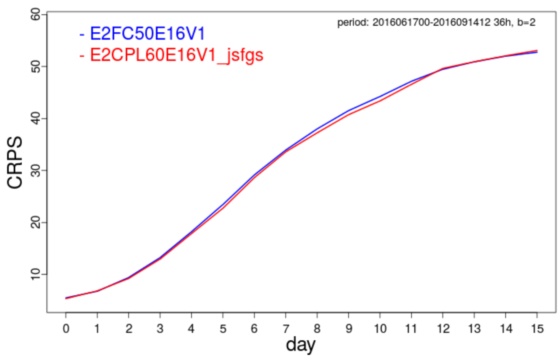
The quality of the GEPS 6.0.0 forecast is compared with that of GEPS 5.0.0 by evaluating the skill scores of two experiments in each season, namely, E2CPL60E16V1_jsfgs and E2FC50E16V1 in summer 2016, that are conducted with GEPS 6.0.0 and GEPS 5.0.0, respectively, and E2CPL60E17V1_jsfgs and E2FC50E17V1 in winter 2017. Skill scores considered include root mean square error (RMSE), bias (BIAS), continuous rank probability score (CRPS), CRPS reliability and CRPS resolution for the upper air and surface fields, and Brier skill score (BSS), BSS reliability and BSS resolution for precipitation of different accumulation criteria.

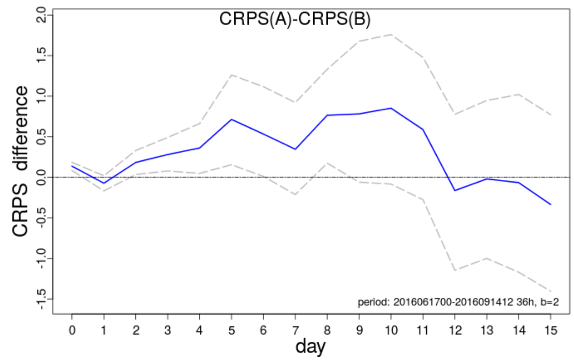
6.2.1 Results for the summer 2016 period

Many different skill score measures are calculated for different variables over different regions. It is not feasible to show all of them. Here we show an example of atmospheric evaluation for geopotential height with CRPS. After that, a summary of the skill scores for all variables will be presented as scorecards and discussed.

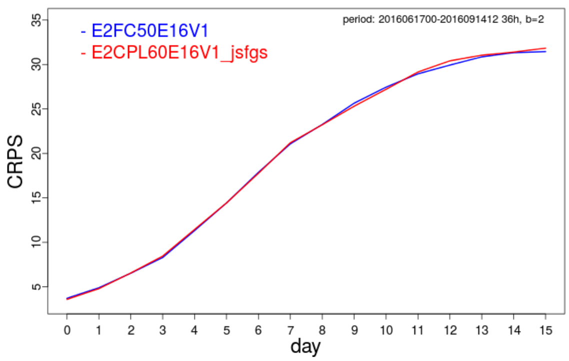
On the left panels of Figure 13, we compare the CRPS for the geopotential heights at 3 levels in the troposphere between GEPS 6.0.0 and GEPS 5.0.0 in the North America. The corresponding CRPS differences between the two systems with their 90% confidence intervals are shown on the right panels. These intervals were computed using a bootstrapping method (Candille at al. 2007). A general improvement of skill is seen in the upper troposphere at 250 hPa in the first 10 days. After that, the two versions of GEPS have a comparable skill. For the middle and lower troposphere (500 and 850 hPa), the skill differences are quite neutral.

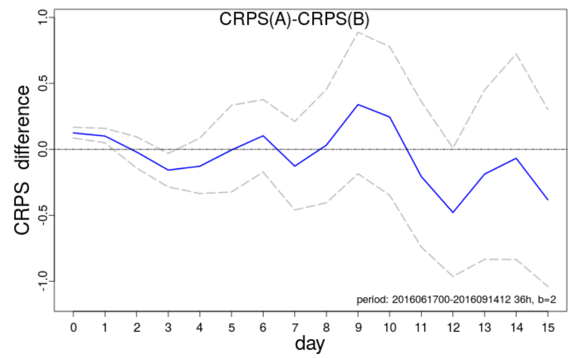
a) GZ250 *b)*



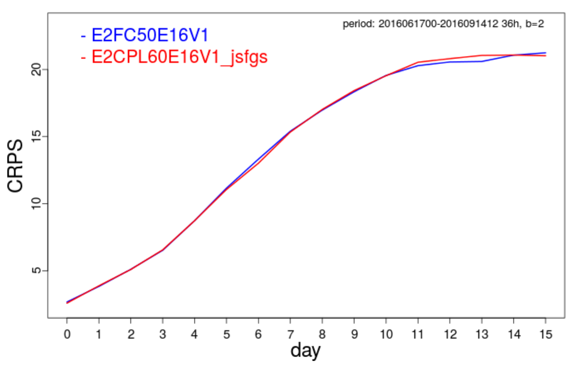


c) GZ500 *d)*





e) GZ850 *f)*



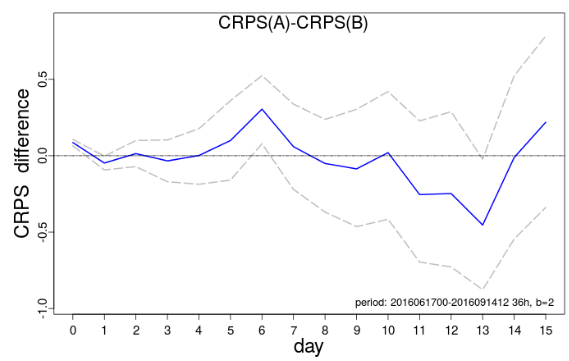


Figure 13. Left panels: CRPS of geopotential heights in the North America for GEPS 5.0.0 (experiment E2FC50E16V1 in blue) and the GEPS 6.0.0 (experiment E2CPL60E16V1_jsfgs in red) during summer 2016 for three levels 250 hPa (a), 500 hPa (c) and 850 hPa (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three vertical levels.

Figure 14 shows the CRPS scores of geopotential height over the globe. Here we see that the improvement over the globe is more evident than that over North America, especially in the lower troposphere.

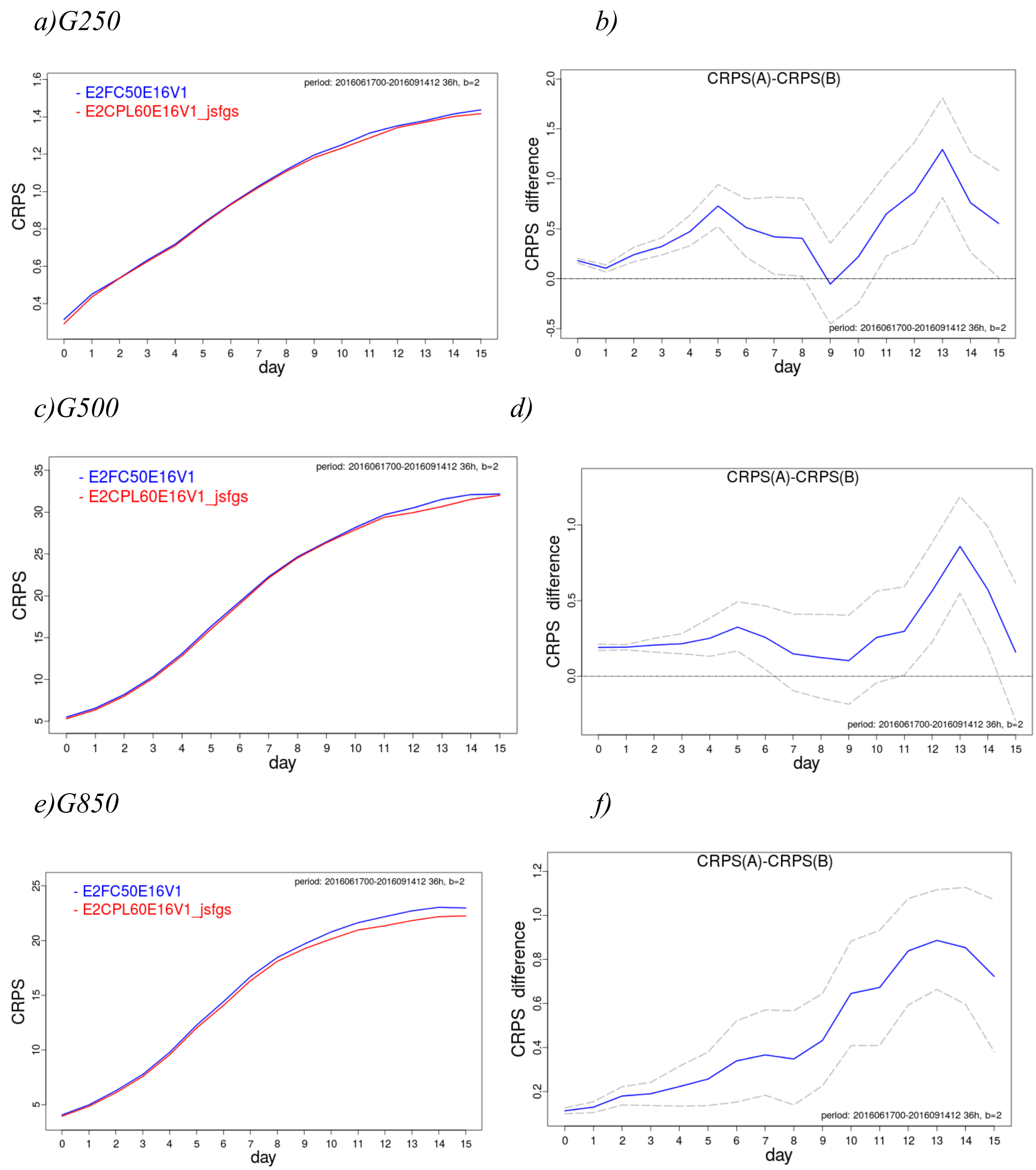


Figure 14. Same as Figure 13, but for the globe.

To summarize the performance differences for all variables with all skill measures between the two experiments, a scorecard is produced for each region and each week. In this scorecard are compiled the differences of all score measures between the two experiments for each forecast day each variable and each level. Every time the skill score of GEPS 6.0.0 in a particular day is significantly improved or degraded with respect to GEPS 5.0.0, a "+" or a "-" is registered. When there is no significant difference, a blank appears. The percentages of "+", "-" and blanks are summarized and printed at the end of the precipitation title line.

```
                                                              Printed by Juan-Sebastian Fontecilla
Apr 12, 19 14:55        RESUME_E2FC50E16V1_vs_E2CPL60E16V1_jsfgs_Noam_000-144              Page 1/1

-------------------- UU --------------------   -------------------- VV --------------------

    |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|       |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 50 | ++ +  |-++--- |       |- -----| ++++++|    50 | +    +|  -    |   -   | ---  -| +    +|
100 |++ +++ |- ++   |+   ++ |+---   |++ +++ |   100 | ++   +|       |       | ---  -| ++ + +|
250 |+++ ++ |  --   |++++++ |+ --   |++++++ |   250 |+ +++  |       |++++++ |+-     |++++++ |
500 |+++    |  -    |+++    |+--    |++++   |   500 |++     |   -   |+++ +  |+ -    |+++++  |
850 |++     |-+++ + |++     |+-    +|+++    |   850 |++     |   +   |+++    |+---   |+++    |
925 |+   -  |-+++   |+   -  |+--    |++  -  |   925 |+      |       |+      |+--   -|+++    |

-------------------- GZ --------------------   -------------------- TT --------------------

    |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|       |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 50 |+++--  |++-----|+++--- |+ -----|+++++++|    50 |++  -  |- -----|++  -- |+ -----|+++++  |
100 |+++++++|+ +++  |+ +++++|+-+++  |+++++++|   100 |+-     | -+++++|+--    |+-  +++|+ -    |
250 |+      | -    -|+ +++++| - ----|+  ++++|   250 |+      |-      |+  ++  |---    |+  ++ +|
500 |++ -   | ------|++ -   |-------|+++    |   500 |++     | +     |++ +   |-+--   |++ +   |
850 |+      | ---  +|+-    +|---    |+ +   +|   850 |++ +  -|       |++     |----   |++     |
925 |+-     | ---  +|+-    +|---    |+-    +|   925 |+  +   | - +   |+  +   |----   |++ +   |

-------------------- ES --------------------   ------------------ SURFACE ------------------

    |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|       |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
250 |+ +++  | --  --|+ +++ -|---  --|+++++  |    PN |+      |+++++++|+      |+  ++++|+      |
500 |++   + | -     |++ + + |---  + |++++ + |    TT |+   -  |+++++++|+-- -  |+--    |+   -  |
850 |++     | ----- |+    - |+--- - |++     |    UV |+ -    | ----- |+---   |+------|+++++++|
925 |++++   |+      |+      |+--    |++++   |    ES |++++  +|-- ++++|+  +   |----   |++++   |

---------------------------------------- PRECIP ----------| +  27.69%  -  19.65%   52.66%

SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
BSS    |      |      |--   -|---   |  -   |  -   |  -- -|  -- -|  -- -|  -  -|  -  -|
BSSrel |--- --|---   |---  -|---  -|---   |----  | --- -|  -- -|- -- -|  -  -|- -  -|
BSSres |++ +  |++ +  |      |  -   |    + |  -   |  -   |  --  |      |  -   |      |

Friday April 12, 2019        RESUME_E2FC50E16V1_vs_E2CPL60E16V1_jsfgs_Noam_000-144              1/1
```

Figure 15. Scorecard for the experiment E2FC50E16V1 (GEPS 5.0.0) compared to E2CPL60E16V1_jsfgs (GEPS 6.0.0), over North America for the summer, week 1.

Presented in Figure 15 is the scorecard for the week 1 forecast over North America. As can be seen, for the upper air, the scorecard shows more "+" than "-", indicating that the forecast quality is in general improved. The improvement of skill is evident in the upper troposphere, especially for the RMSE, CRPS and CRPS resolution scores. We see that GEPS 6.0.0 has a better CRPS score for all the variables than GEPS 5.0.0 at 250 hPa for most days of this week.

For the week-1 forecast of surface variables in North America, improvement is seen for BIAS of mean sea-level pressure, 1.5-m temperature and 1.5-m dew-point depression (ES), but some degradation for 10-m wind speed (Figure 15). The CRPS scores for the surface variables are quite neutral, with some improved skill in ES1.5m but degraded skill in TT1.5m near the end of the week.

For precipitation, we see degradation. Only the resolution of the Brier Skill Score shows some improvement.

As indicated by the percentage numbers of "+", "-" and blanks in the scorecard of Figure 15, there is an overall improvement for the first week over North America, where 28% of the scores show improvement, comparing to 20% for deterioration. For the second week, the forecast quality difference is quite neutral (not shown), where we have 10% of scores for improvement and 9% for deterioration.

Shown in Figure 16 and Figure 17 are the scorecards over the globe for the week 1 and 2, respectively. Bigger improvement is seen over the globe than over North America. The upper air and surface scores are dominantly improved.

The BSS of the precipitation of GEPS 6.0.0 has also been improved compared to the GEPS 5.0.0, especially for quantities greater than 20 mm the second week. The resolution has been improved for weeks 1 and 2, and for all the quantities.

An overall significant improvement is obtained over the globe, especially for the second week, with 55% for improvement versus 22% for deterioration in the first week and 47% versus 14% for the second week.

```
                                                                        Printed by Juan-Sebastian Fontecilla
Apr 12, 19 14:55        RESUME_E2FC50E16V1_vs_E2CPL60E16V1_jsfgs_Global_000-144                 Page 1/1

 -------------------- UU --------------------      -------------------- VV --------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|          |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 | -- ---| ------| ------| ------| + ++++|       50 |++++  +| +-+   |+      |+--- --|+++++++|
 100 |+++++++|   ----|+++++++|+------|+++++++|      100 |+++++++| -+ +  |+++++++|+--- --|+++++++|
 250 |+++++++|+  ++++|+++++++|-      |+++++++|      250 |+++++++| -+-  -|+++++++|+-     |+++++++|
 500 |++++++ |+++++++|++++++ |+      |+++++++|      500 |+++++++|      +|+++++++|+ -    |+++++++|
 850 |++++++ |++++++ |++++++ |+---   |+++++++|      850 |++++ ++|     + |+++++++|+------|+++++++|
 925 |++++++ |+++++  |++++++ |+----  |+++++++|      925 |++++ ++| ++    |+++++++|+------|+++++++|

 -------------------- GZ --------------------      -------------------- TT --------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|          |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 | ----  |- -----| ----- | ------| - ++++|       50 |++ ----|+------|++ ----|+------|+++++++|
 100 | --  + |+  ----|+- ++++|+------|  +++++|      100 |+ +++++|+ +++++|+ +++++|+-+++++|+++ +  |
 250 |+++++++|+- ----|+++++++|+------|+++++++|      250 |+++++++| +-----|+++++++|-+-----|+++++++|
 500 |++++++ |+ ---  |+++++++|+ -----|+++++++|      500 |+++++++|++-----|+++++++|-+-----|+++++++|
 850 |+++++++|-++++++|+++++++|- +++++|+++++++|      850 |++++++ |++++++ |++++++ | --    |++++++ |
 925 |+++++++| ++++++|+++++++|- +++++|+++++++|      925 |++++++ | +++   |++++++ |-------|++++++ |

 -------------------- ES --------------------      ------------------ SURFACE ------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|          |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 250 |+++++  |  +++  |++++++ |--+++  |++++++ |       PN |+++++++|--+++  |+++++++|-- ++  |+++++++|
 500 |+++++++|+++    |+++++++|----   |+++++++|       TT |++++++ |+++++++|+++++++|+++++++|+    - |
 850 |+++    |+------|+   -  |+------|++     |       UV |+----  |+------|+----- |+------|+++++++|
 925 |++++  +|+------|+      |+------|++++   |       ES |+++++++|- -----|+-  +++|------ |+++++++|

 --------------------------------------- PRECIP ----------| +  54.93%  -  22.46%   22.61%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    |---   |--    | - +  |  +   |      |  -   |      |+     | +    |++    | +    |
 BSSrel |------|------|----- |--- - |---   |----  |----  |----  | --   |  -   |  -   |
 BSSres |++++++|++++++|+++++ |+++++ |++ ++ |++ +  |++ +  |++   +|++++  |++++  |+++   |
Friday April 12, 2019          RESUME_E2FC50E16V1_vs_E2CPL60E16V1_jsfgs_Global_000-144                    1/1
```

Figure 16. Scorecard for the experience E2FC50E16V1 compared to E2CPL60E16V1_jsfgs , for the summer, over the globe, week 1.

Printed by Juan–Sebastian Fontecilla

Apr 12, 19 14:56	RESUME_E2FC50E16V1_vs_E2CPL60E16V1_jsfgs_Global_168–312	Page 1/1

```
 -------------------- UU --------------------    -------------------- VV --------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|        |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |--  +++|-------| -  +++|-------|+ +++++|     50 |++   ++| +    +|     + |-------|++++ ++|
 100 |+++++++|-      |+++++++|--     |+++++++|    100 |+++++++|+++ +++|+++++++|-----  |+++++++|
 250 |   ++++|       |   ++++| -     |   ++++|    250 |      +|    +  |     ++|-  -- -|     ++|
 500 |    +++|+++++++|    +++|     + |    +++|    500 |       |++     |   +  +|---  - |   +  +|
 850 |+ +++++|+++++++|+ +++++|   +++ |+++++++|    850 |++ ++++|++ ++++|+++++++|   +  +|+++++++|
 925 |+ +++++| +++++ |+ +++++|       |+ +++++|    925 |++ ++++| ++++++|+++++++|   +   |+++++++|

 -------------------- GZ --------------------    -------------------- TT --------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|        |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 | ++++++|--  +++|  +++++|-- ++++|+++++++|     50 |-------|-------|-------|-------|+++++++|
 100 | ++++++|-------|+++++++|-----  |+++++++|    100 |+++++++|+++++++|++ ++++|+++++++|+    ++|
 250 |    +++|-----  |++  +++|-----  |++  +++|    250 |+++++++|----  +|+++++++|----  +|+++++++|
 500 |     ++|       |    +++|------ |   ++++|    500 |+   +++|----  +|++ ++++|---  ++|++ ++++|
 850 |+++++++|+++++  |+++++++|+++++++|+++++++|    850 |  -    |      +|       | --    |       |
 925 |+++++++|++++   |+++++++|+++++++|+++++++|    925 |+ -    |    +++|  -    |- - -  |  -    |

 -------------------- ES --------------------    ------------------ SURFACE ------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|        |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 250 |       |-------|       |-------| ++++ +|     PN |   ++++|+      |+++++++|       |+++++++|
 500 |+++++++|+++++++|+++++++| + ++  |+++++++|     TT |   -   |+++++++| +    +|+++++++|   --  |
 850 |+    ++|-------|+    ++|---  ++|+    ++|     UV |   ++++|----   |     ++|----   |+++++++|
 925 |+  + ++|------ |      +|--    +|+     +|     ES |+++  ++|   ++++|+++++++| +    +|+++++++|

 --------------------------------------- PRECIP ----------| +  46.74%  -  14.11%   39.15%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    |- ----|-     |      |      | +   +| +   +| +   +| ++ + | ++ + |  +++ |  +++ |
 BSSrel |------|---- -|-     |     +| + ++ | + ++ | ++++ | ++++ | ++++ | ++++ | ++++ |
 BSSres | +    | ++   | +    |      | +    |      |      |      |      | +    |      |
```

Friday April 12, 2019	RESUME_E2FC50E16V1_vs_E2CPL60E16V1_jsfgs_Global_168–312	1/1

Figure 17. Scorecard for the experience E2FC50E16V1 compared to E2CPL60E16V1_jsfgs, for the summer, over the globe, week 2.

6.2.2 Results for the winter 2017 period

As in the previous subsection, here we show an example of evaluation with the CRPS, but on surface variables this time. Then, we will show the scorecards for North America and the globe, for weeks 1 and 2.

The CRPS of 1.5-m temperature, 1.5-m dew-point depression (ES1.5m) and mean sea-level pressure over the North American region as a function of lead time are shown in the Figure 18. The temperature and MSLP forecasts are clearly improved for almost the whole period of the integration, while that of ES1.5m is quite neutral.

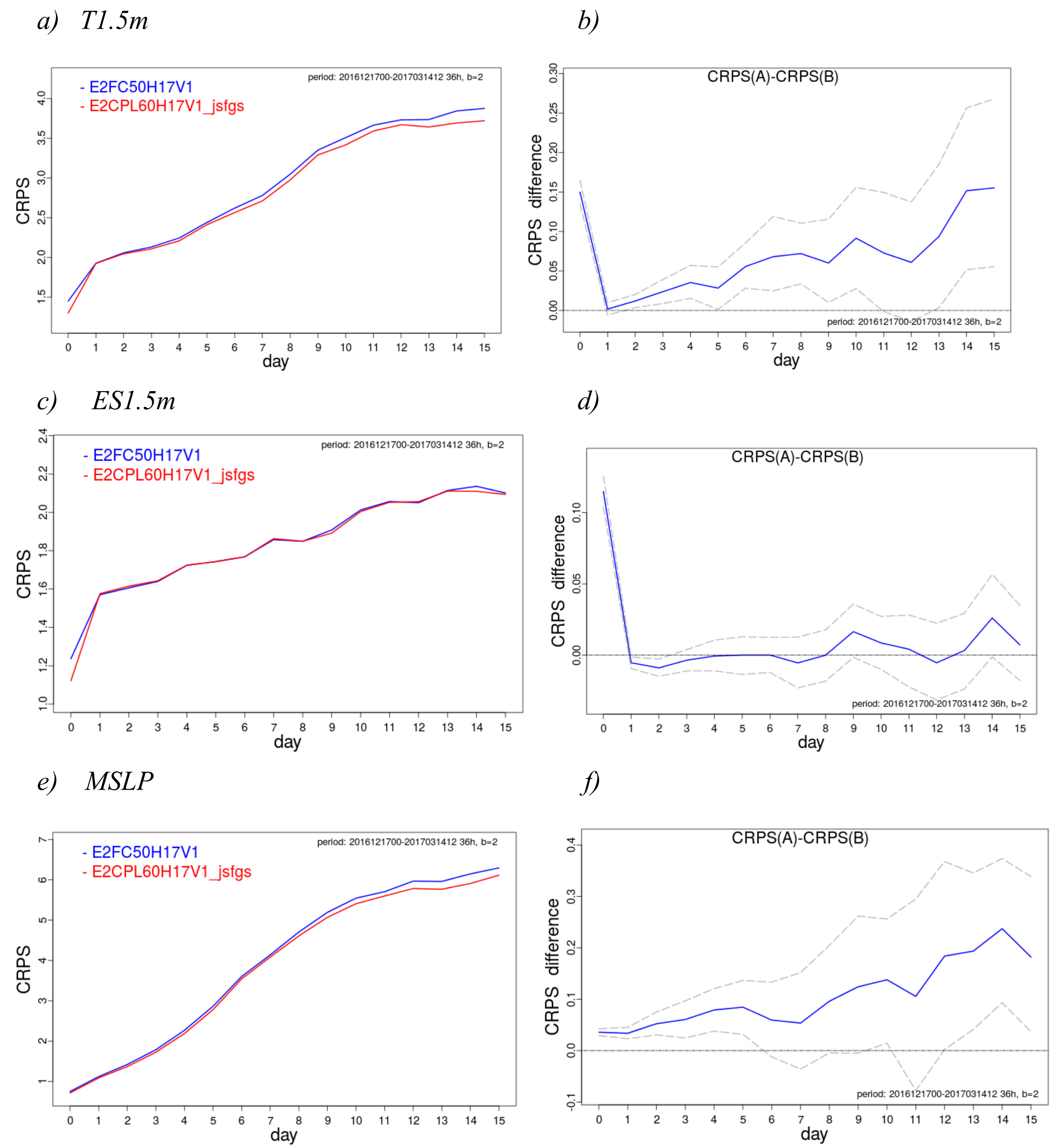
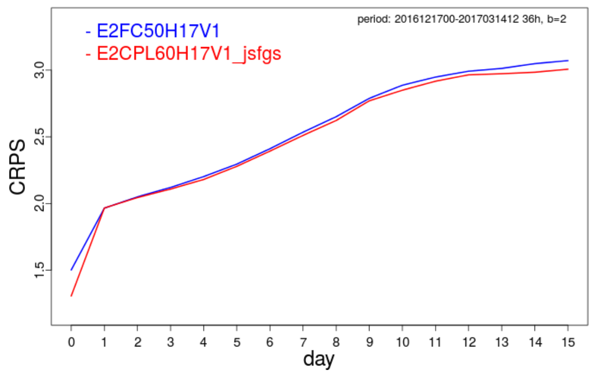


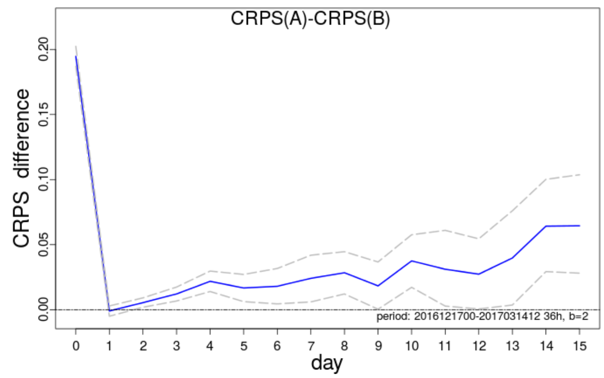
Figure 18. Left panels: CRPS in North America of the GEPS 5.0.0 (experiment E2FC50H17V1 in blue) and the GEPS 6.0.0 (experiment E2CPL60H17V1_jsfgs in red) during Winter 2017 for the 2-m temperature (a), the 2-m dew-point depression (c) and mean sea-level pressure (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three fields.

Figure 19 shows the CRPS scores for the same surface variables as in the Figure 18, but over the globe. The results are quite similar to those for North America. We observe improvement for the temperature and mean sea level pressure, and a neutral impact for the dew point depression.

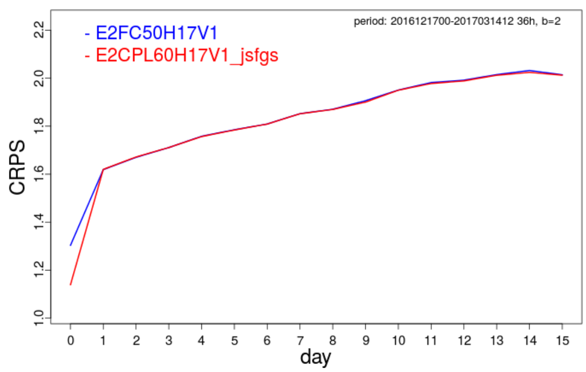
a) T1.5m



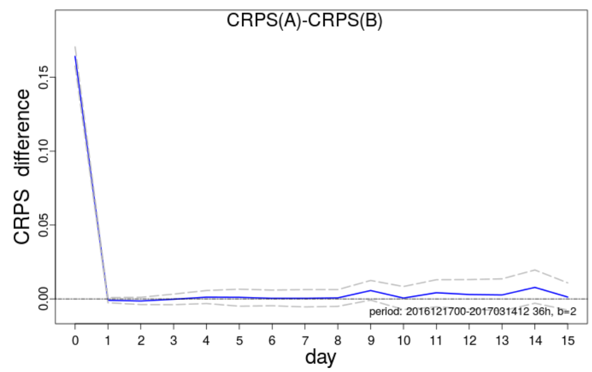
b)



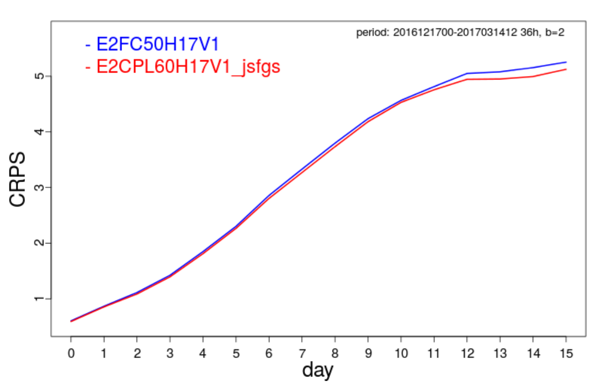
c) ES1.5m



d)



e) MSLP



f)

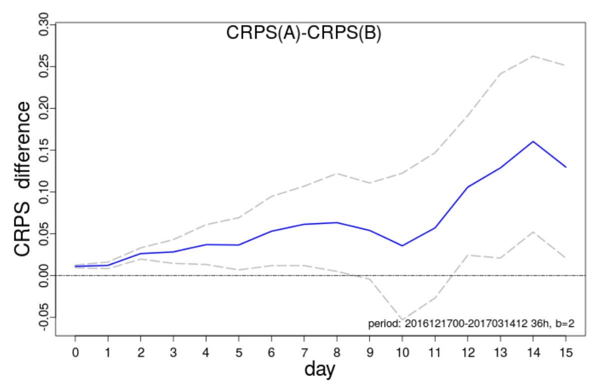


Figure 19. Left panels: CRPS over the globe for GEPS 5.0.0 (experiment E2FC50H17V1 in blue) and GEPS 6.0.0 (experiment E2CPL60H17V1_jsfgs in red) during Winter 2017 for the 2-m temperature (a), the 2-m dew-point depression (c) and mean sea-level pressure (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three fields.

Figure 20 shows the scorecard over North America for week 1. The improvement of forecast performance in winter 2017 is even larger from GEPS 5.0.0 to GEPS 6.0.0 than that in summer 2016. The overall improvement is 51% for the first week versus 7% of deterioration. We can also see improvement in mean sea-level pressure and temperature at 1.5 m, and relatively small improvement in dew point depression and winds at 10 m. The precipitation is also improved over North America, during the winter.

Printed by Juan–Sebastian Fontecilla

Mar 17, 19 11:28 **RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Noam_000–144** Page 1/1

```
 -------------------- UU --------------------     -------------------- VV --------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 | ++    |   ----|  +    |  +----| +++++ |      50 |     ++|       |       |       |  +  ++|
 100 |+++++++|       |+++++++|+      |+++++++|     100 |+  + ++|   -   |+  + + |+      |++ + ++|
 250 |+++++ +| -     |+++++++|+     +|+++++++|     250 |+++++++|-      |+++++ +|+      |+++++ +|
 500 |++++++ |       |++++++ |+      |++++++ |     500 |+++++++| +     |+++++++|+++    |+++++++|
 850 |++++++ |-++    |+++ +++|+      |+++++++|     850 |+++ +++|       |+++++++|+      |+++++++|
 925 |+++ ++ |-      |+++ ++ |+ +    |+++ ++ |     925 |+++ +++|       |+++++++|+      |+++++++|

 -------------------- GZ --------------------     -------------------- TT --------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |+++++++|+ +++++|+++++++|+ +++++|++ +++ |      50 |++ ++  |-+    -|++ +  -|+    --|++ ++  |
 100 |+++++++|+++++++|+++++++|+++++++|+++++++|     100 |+++++++|-++++++|+++++ +| ++++++|++++   |
 250 |+++++++| ++++  | ++++++| ++    | ++++++|     250 |+++++++| ---   |+++++++|+--    |+++++++|
 500 | ++++++| -     | ++++++| -     | ++++++|     500 |+++++++|   -   |+++++++|-      |+++++++|
 850 |++++++ | - --  |++++++ |-  -   |++++++ |     850 |+++++++|       |+++++++|   ++  |+++++++|
 925 |++++++ |+  --  |++++++ |-  --  |++++++ |     925 |+ +++++| +  ---|+++++++|+      |+++++++|

 -------------------- ES --------------------     ------------------ SURFACE -----------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 250 | ++++++|-++++++| ++++++|-++++++|+++++++|      PN |+++++++|+  --- |++++++ |+  --  |++++++ |
 500 |+++++  | --    |+ +++  |--  +  |+++++  |      TT |+++++++|+ -----|+ +++++|+--  --|+++++++|
 850 |+++    |+     -|++++   |---  --|++++ + |      UV |+   +  |+--   -|+--    |+--  - |+++ +  |
 925 |++ +  +|+  + ++|++ + ++| --    |++++ ++|      ES |++     | ++++++|+--    |---    |++     |

 --------------------------------------- PRECIP ----------| +  50.68%  -   7.21%   42.11%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    | ++   |  + + |    ++|    ++| - ++ |    ++|  +   |      |      |  -   |  - - |
 BSSrel |-++   |-    +|-    +|--    | -    | -    |      |      |      |  -   |    - |
 BSSres |+++   |+ +  +|+   ++|+   ++|+   ++|      |      | ++   |      |   +  |      |
```

Sunday March 17, 2019 RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Noam_000–144 1/1

Figure 20. Scorecard for the experience E2FC50H17V1 compared to E2CPL60H17V1_jsfgs, for the winter, over North America, week 1.

The second week forecasts also shows significant improvement (Figure 21), where the overall improvement is 21% versus 3% of deterioration. We observe bigger improvement in precipitation than the first week, with no degradation for all thresholds and score measures (BBS, BSS resolution or BSS reliability).

Printed by Juan–Sebastian Fontecilla

```
Mar 17, 19 11:28     RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Noam_168-312     Page 1/1

 -------------------- UU ---------------------     -------------------- VV ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |       |--     | +     |       |++     |      50 |       |       | +     |       | +     |
 100 |+++    |       |+++    |+++    |+++    |     100 |+++    |       |+++    |       |+++    |
 250 |+      |      -|+      |       |+      |     250 |+      |       |++     |   +   |++     |
 500 |       |       |       |       |       |     500 |+      |       |       |       |       |
 850 |+      |+      |       |+      |       |     850 | +     |       | +     |       | +     |
 925 |       |++     | +     | +    -|       |     925 | +     |       | +     |       | +     |

 -------------------- GZ ---------------------     -------------------- TT ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |+++    |+      |+++    |+++    |++ +   |      50 |    ---|-------| ------|-------|       |
 100 |++++   |++++   |+++++  |++++   |+++++  |     100 |+++    |+++++++|+++    |+++    |  +    |
 250 |+++    |       |+++   +|  ++   |+++   +|     250 |++     |       |+++    | +     |+++    |
 500 |++    +|       |++     |  ++  +|++     |     500 |+++    |       |+++    |+      |+++    |
 850 |  ++  +|       |   + ++|       |   +  +|     850 |+     +|       |+      |       |+      |
 925 | + +  +|       |   + ++|       |   + ++|     925 |++ +  +|- --   |+  +  +|       |+++++ +|

 -------------------- ES ---------------------     ------------------ SURFACE ------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 250 |++++   |+++++++|++++   |+++++++|++     |      PN |   +  +|       |   + ++|       |   +  +|
 500 |       |    -- |       | +     |       |      TT |++ +   |- -----|++++  +|       |+++++++|
 850 |   +   |       |       |       |       |      UV |     - |--     |     - |       |  + -  |
 925 |+      |++     |       |       |       |      ES |       |+++    |       | +++   |-      |

 -------------------------------------- PRECIP ----------| +  21.32%  -   3.34%   75.34%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    |  +   |+ +   |+ ++  |  ++  |+ ++  |+ ++  |+ +   |+     |+     |+     |+     |
 BSSrel |  +   |  ++  |  ++  |  ++  |+ ++  |+ ++  |+ +   |   +  |      |      |      |
 BSSres |  +   | ++   |  +   |      |+     |+     |+     |      |   +  |+     |   +  |

Sunday March 17, 2019        RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Noam_168-312        1/1
```

Figure 21. Scorecard for the experience E2FC50H17V1 compared to E2CPL60H17V1_jsfgs, for the winter, over North America, week 2.

Over the globe, we obtain dominant improvement for both weeks, where the overall improvement is 60% versus 11% of deterioration for the first week (Figure 22), and 42% versus 5% for the second week (Figure 23).

Again, we see significant improvement for precipitation, especially for quantities greater than 10 mm.

Printed by Juan–Sebastian Fontecilla

```
Mar 17, 19 11:29      RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Global_000–144      Page 1/1

 -------------------- UU ---------------------   -------------------- VV ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|       |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |++++ + | ------|+++   +|+   ---|+++++++|    50 |+++++++|++     |++++ + |+ -    |+++++++|
 100 |+++++++| ++   +|+++++++|+      |+++++++|   100 |+++++++|-  -   |+++++++|+-     |+++++++|
 250 |++++++ |+      |++++++ |++++   |+++++++|   250 |+++++ +|- -    |+++++ +|+      |+++++ +|
 500 |++++++ |       |++++++ |++++ + |++++++ |   500 |+++++++|       |+++++++|+++++++|+++++++|
 850 |+++++++|++ -   |+++++++|+      |+++++++|   850 |++++++ |+      |+++++++|+-     |+++++++|
 925 |+++++++|+++    |+++++++|+   +  |+++++++|   925 |++++++ |+   -  |+++++++|+-     |+++++++|

 -------------------- GZ ---------------------   -------------------- TT ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|       |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |++++++ |+++++++|++++++ |+++++++|++++   |    50 |++  ---|+ -----|+   ---|+------|+++++++|
 100 |  +++++|+++++++|+ +++++|+++++++|  +++++|   100 |+++++++|+++++++|+++++++|+++++++|++++   |
 250 |++++++ |+++++  |+++++++|++++   |+++++++|   250 |+++++  |++-----|+++++  |++-----|+++++++|
 500 |+++++++|   +   |+++++++| ++    |+++++++|   500 |++++++ | -++++ |++++++ |-+++++ |++++++ |
 850 |+++++++|-+ - - |+++++++| ++    |+++++++|   850 |+++++++|- ---- |+++++++|---  ++|+++++++|
 925 |+++++++| + - - |+++++++| ++   +|+++++++|   925 |+++++++|++++   |+++++++|++ ++++|+++++++|

 -------------------- ES ---------------------   ----------------- SURFACE -----------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|       |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 250 | ---   |--  +++|---    |--- ++ |       |    PN |+++++++|       |+++++++|- + +++|+++++++|
 500 |+++++++|++--   |+++++++|--  ++ |+++++++|    TT |+++++++|+    --|+ +++++|+--  --|+++++++|
 850 |++++   |-      |++++   |----- -|++++   |    UV |+---   |+------|+--- - |+------|+++++ +|
 925 |+++   +| ++++ +|+++   +|+---   |++++  +|    ES |+++    |+++++++|+      |----   |+++++  |

 ------------------------------------ PRECIP ----------| +  60.32%  -  10.93%   28.76%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    |      | + +  |  ++++|  ++++| - +++|    ++|    ++| -  ++|     + | -  ++| -  ++|
 BSSrel |---   |----  |---   |--    |--- + |--  ++|--  ++| -  + | -     | --   | -   +|
 BSSres |++++++|++++++|++++++|++++++|+ ++++|+  +++|+ + ++|+ + ++|  +++ |  +++ |    + |

Sunday March 17, 2019      RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Global_000–144      1/1
```

Figure 22. Scorecard for the experience E2FC50H17V1 compared to E2CPL60H17V1_jsfgs, for the winter, over the globe, week 1.

Printed by Juan–Sebastian Fontecilla

Mar 17, 19 11:29 **RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Global_168–312** Page 1/1

```
 -------------------- UU --------------------    -------------------- VV --------------------

     |RMSE    |BIAS    |CRPS    |CRPSrel|CRPSres|       |RMSE    |BIAS    |CRPS    |CRPSrel|CRPSres|
  50 |  +    +|------  | ++     |-       | ++ +   |    50 |       +|       +|       +|        |       +|
 100 |++++++++|++      |++++++++|+++    +|++++++++|   100 |+++   ++|        |++++++++|        |++++++++|
 250 |++    ++|        |+++    +|+       |+++    +|   250 |   +++  |        |+  +++  |   +    |+  +++  |
 500 | +    + |        | +      | +      |++      |   500 |   ++   |       +|   +++  | +++    |   +++  |
 850 |+     ++|        |++    ++|        |++    ++|   850 |   ++   |        |   +++++|    +   |   +++++|
 925 |+     + |        |+     ++|        |+     ++|   925 |   +++ +|        |   +++  |        |   +++  |

 -------------------- GZ --------------------    -------------------- TT --------------------

     |RMSE    |BIAS    |CRPS    |CRPSrel|CRPSres|       |RMSE    |BIAS    |CRPS    |CRPSrel|CRPSres|
  50 | ++     |++      | ++     |++++    |        |    50 |----    |--------|--------|--------| +      |
 100 |++++++++|++++++++|++++++++|++++++++|+++++   |   100 |++++++++|++++++++|++++++++|++++++++|     + +|
 250 | +++++++|-       |++++++++|        |++++++++|   250 |        |------  |        |--------| ++++   |
 500 | +++++++|        |++++++++|   +    |++++++++|   500 |  ++   +|    --  | +++   +|+++     | +++   +|
 850 |+++   ++|        |++++  ++|        |++++  ++|   850 |++  ++++|        |+  +++++|++++++++|+  +++++|
 925 |+++   ++|        | +++  ++|+++     | +++  ++|   925 |+  +++++|        |++++++++|++++++++|++++++++|

 -------------------- ES --------------------    ------------------ SURFACE ------------------

     |RMSE    |BIAS    |CRPS    |CRPSrel|CRPSres|       |RMSE    |BIAS    |CRPS    |CRPSrel|CRPSres|
 250 |        |++++++++|+       |++ +++  |        |    PN |      ++|        |++    ++|+       |++    ++|
 500 | +++++++|        |  ++++++| ++    +|  ++++++|    TT |++ +    |--------|++++++++|  -     |++++++++|
 850 |    +   |        |     +  |--      |    +   |    UV |    -   |---     |    -   |---     |+  -    |
 925 |+     + |+++     |        |       -|        |    ES |        |+++     |        |  +     |        |

 -------------------------------------- PRECIP ----------| +  41.58%  -   5.08%   53.34%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    |  + ++|  +   |+ +++ |  +++ |  ++++|  +++ |+ +++ |+ ++++|+++++ |  + ++| ++ + |
 BSSrel |      |  +   |  +++ |  +++ |  +++ |  +++ |+ +++ |+ +++ |+++++ | ++ + | ++ + |
 BSSres |  + ++|  +++ |++++++| +++++| +++++|  ++++|+ ++++|+ + ++|  +++ |  + + |   +  |
```

Sunday March 17, 2019 RESUME_E2FC50H17V1_vs_E2CPL60H17V1_jsfgs_Global_168–312 1/1

Figure 23. Scorecard for the experience E2FC50H17V1 compared to E2CPL60H17V1_jsfgs, for the winter, over the globe, week 2.

6.3 Quality of GEPS forecasts: ocean

GEPS 6.0.0 has introduced an interactive ocean and sea ice component to the ensemble forecast, and so this represents the first opportunity to investigate the deterministic and probabilistic skill in these components over the 15 day forecast here and later over the full monthly period as well. As the previous system did not have an interactive ocean and sea ice component, we will not be able to show any improvements in skill over GEPS 5.0.0 which used persisted anomalies for sea surface temperature (SST) and dynamically adjusted sea ice concentration. Both of these quantities were purely deterministic in nature. Nevertheless continuous ranked probability scores (CRPS) will be scored for SST along with spatial probability scores for sea ice concentration. Furthermore, ensemble mean error and CRPS scores will also be provided for some sub-surface ocean variables for future reference.

6.3.1 Results for the summer 2016 period

We have investigated the spatial averages and spatial fields for ensemble mean temporal mean (mean) and root mean squared (rmse) error for the summer 2016 forecasts. The summer 2016 version of these forecasts constituted forecasts every 36h from 2016-06-22 00Z to 2016-08-31 12Z. Unless otherwise specified, all the quantities investigated were 24 hour means, and so restricting this to a full calendar day mean for validation required sub-setting this set of forecasts to the 00Z forecasts performed every 72h (26 in all). For ocean variables, the fields were validated against the first 24 hour forecast of a similar period forecast of the Global Deterministic Prediction System version 7.0 (GDPS 7.0). Since both of these products include a diurnal cycle this is preferable to validating SST against an analysis of foundation (pre-dawn) SST, such as CMCSST. For subsurface variables, however, this is simply the most practical source available, as it will be the most applicable estimate of forecast error as opposed to initial state error. However, for validation of the persisted anomalies used in GEPS 5.0.0, which would have no diurnal cycle, validation was performed against the CMC SST analysis, used as initial conditions. As instantaneous values were used, we were also able to use both the 00Z and 12Z forecasts, but were only able to validate at the 36h intervals the system was initiated at.

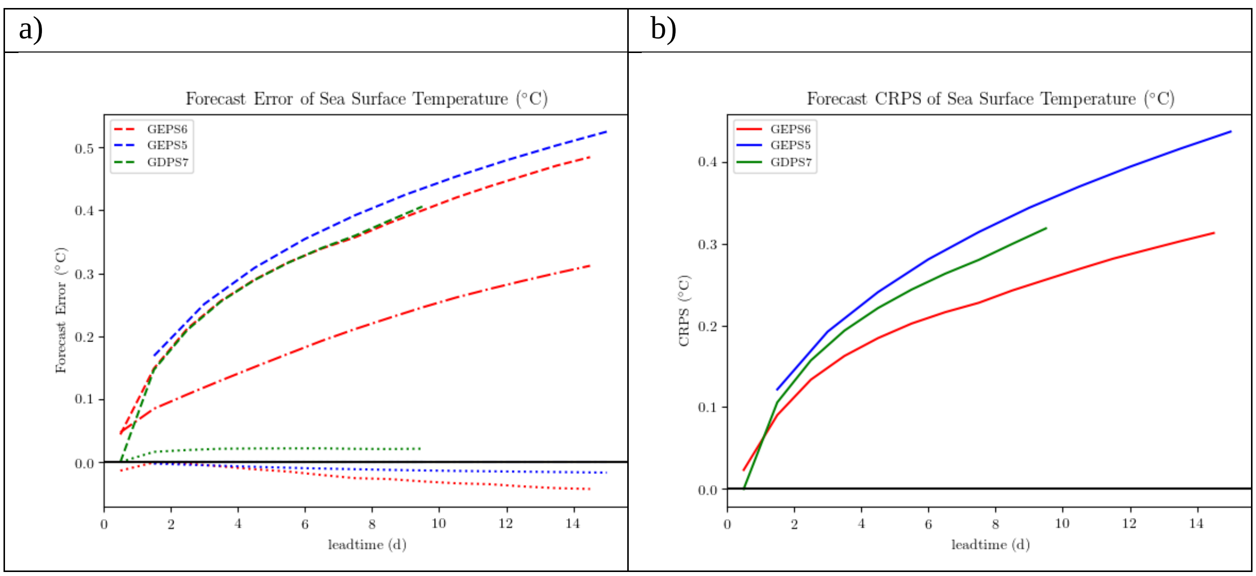


Figure 24. a) Root Mean Square Error (RMSE; dashed line), mean error (dotted line), and ensemble variance (dot-dashed line) for SST in the summer forecast set relative to the validating analysis. Plotted are the errors for the ensemble mean GEPS 6.0.0 forecast (red), the persisted anomaly SST used in GEPS 5.0.0 (blue) and in GDPS 7.0 (green). GEPS 6.0.0 and GDPS 7.0 were validated against the first 24h GDPS forecast (and therefore GDPS 7.0 has zero error at 12h), while GEPS 5.0.0 was validated against CMC SST. b) Continuous Ranked Probability Score (CRPS) for SST in GEPS 6.0.0 (red), GEPS 5.0.0 (blue), and GDPS 7.0 (green) using the same validating analysis as for errors. GEPS 5.0.0 and GDPS 7.0 were treated as single ensemble members.

As shown in Figure 24 above, the coupled and dynamic ocean RMSE easily outscores the persisted SST of GEPS 5.0.0, and over the first 10 days performs equally well to the deterministic coupled system GDPS 7.0, which although it has an identical resolution ocean, has a higher resolution atmospheric component. However, both coupled systems have larger mean error (bias) than the persisted GEPS 5.0.0 SST. Also note that the ensemble GEPS 6.0.0 system is under dispersive, with the ensemble variance (Figure 24a, red dot-dashed line) being about a half of the ensemble mean RMSE (Figure 24a, red dashed line). Figure 24b further shows that the ensemble system easily surpasses the 1-ensemble member GDPS 7.0 and persisted SST values of GEPS 5.0.0 for CRPS. This information is unlikely of major importance, but it, along with the RMSE and ensemble spread relationship shown in Figure 24a is offered up as a benchmark for future versions. Spatial maps (Figure 25) and regional lead time plots (Figure 26) show that GEPS 6.0.0 does particularly well in the NINO3.4 region, as well as at the extra-tropical and high northern latitudes (Figure 26b), but underperforms compared to GEPS 5.0.0 persisted SST in the southern ocean (Figure 26c; note however**,** the higher resolution of the ORCA025 grid compared to the atmospheric grid in GEPS 5.0.0 would allow better resolution of the southern ocean eddies, creating larger "double penalty" errors).

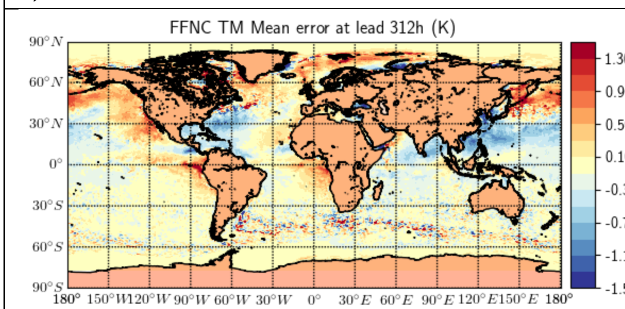
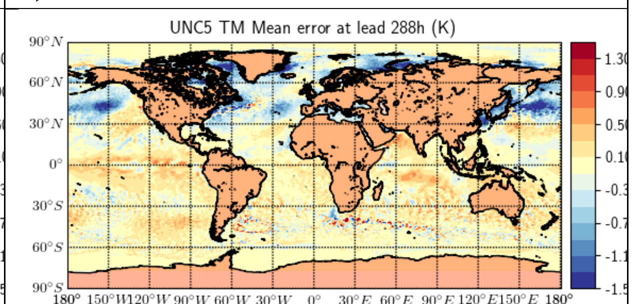
a)	b)
	

Figure 25: *a) Mean error in GEPS 6.0.0 at 12.5d lead time. b) Mean error in GEPS 5.0.0 at 12.0d lead time.* ***Note:*** *We are penalizing GEPS 6.0.0 an additional 0.5d here to account for the fact that it is being validated against a 12h GDPS forecast*.

a)	b)	c)

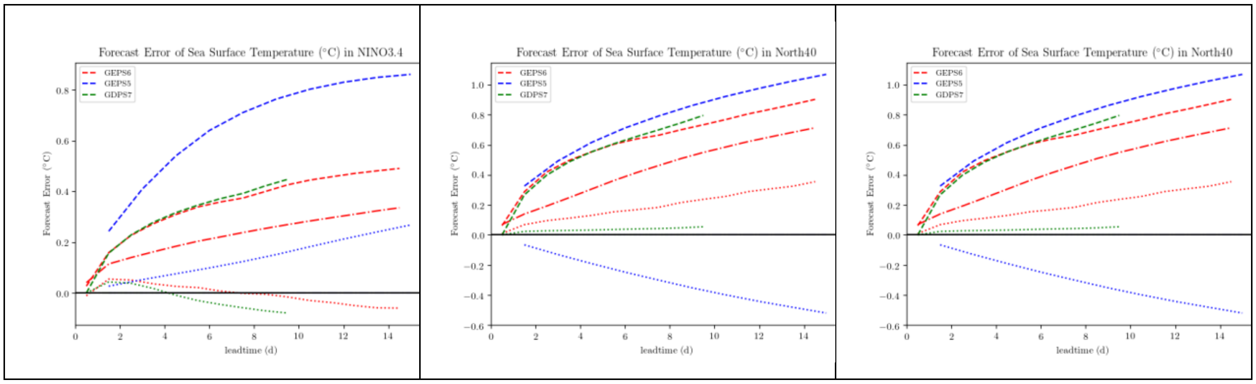


Figure 26: *Same as Fig. 24a, except for regions a) NINO3.4 [170 -120W, 5S- 5N], b) north of 40N, c) south of 40S.*

Finally, having a full ocean now offers the opportunity to investigate errors in the ocean sub-surface. As this is a new capability, we can only offer references for future benchmarks. Figure 27 below shows 9.5 day lead errors for GEPS 6.0.0 (left column), compared to GDPS 7.0 (right column), along with the globally averaged errors at a function of lead time (4^{th} row) for tropical cyclone heat potential (1st row; heat content of column of water above the 28C tropical cyclone threshold), mixed layer depth (2nd row; depth at which density change relative to 10m depth is 0.01kg/m3), and the 30 m depth averaged x-coordinate velocity (3rd row; approximately the east/west velocity for much of the domain except north polar regions. Note that the spatial average in Figure 27i is only over the 20S-20N tropical region). Comparing the errors between the two systems, tropical cyclone heat potential is underestimated relative to GDPS errors with potential implications for tropical cyclone generation and sustenance. Similarly, the wintertime mid-latitude mix layer depth is overestimated in GEPS 6.0.0 relative to GDPS 7.0 errors, with possible implications for winter storms. Equatorial velocities, however, may be better estimated by the ensemble mean of the GEP 6.0.0 system relative to GDPS 7.0 (Note: Equatorial velocities are primarily westward/negative, so red shading would be indicative of underestimations of the westward equatorial currents.) and may influence the smaller NINO3.4 SST biases seen in GEPS 6.0.0.

a)	b)

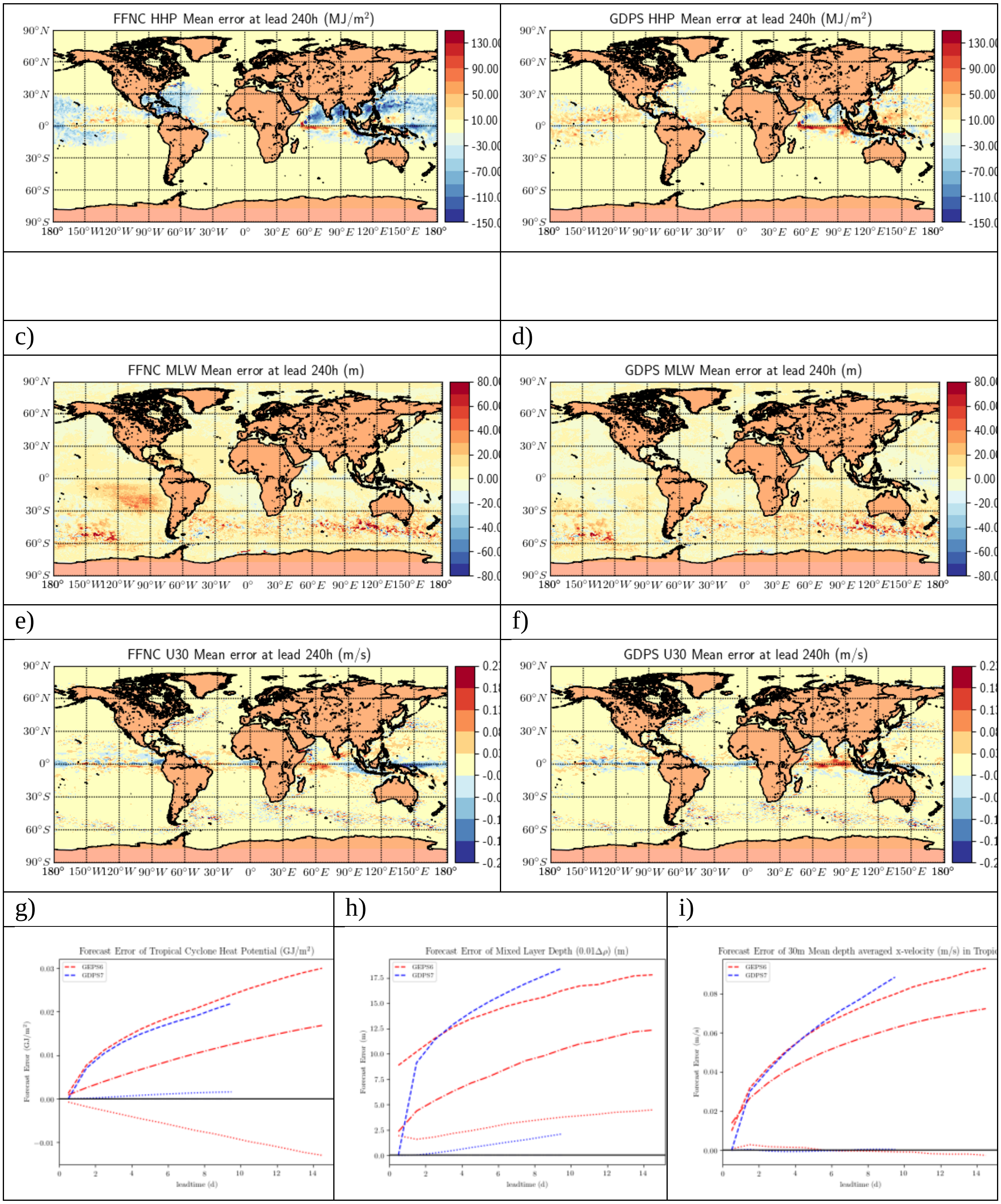


Figure 27: First Row: Estimates of Error in Tropical Cyclone Heat Potential (MJ/m3) in a) GEPS 6.0.0, b) GDPS 7. Second Row: Same as above except for Mixed Layer Depth (m). Third

Row: As above, but for 30m mean depth averaged x-component of velocity. g-i) Global averaged values for the above quantities, respectively, as a function of lead time. Spatial average in i) is over the 20S-20N tropical region instead of globally.

6.3.2 Results for the winter 2016/2017 period

We can now repeat this exercise for the winter 2016/17 forecast that spanned the period 2016-12-15 12Z to 2017-02-28 12Z, but again only using the 00Z forecasts (26 start dates in total).

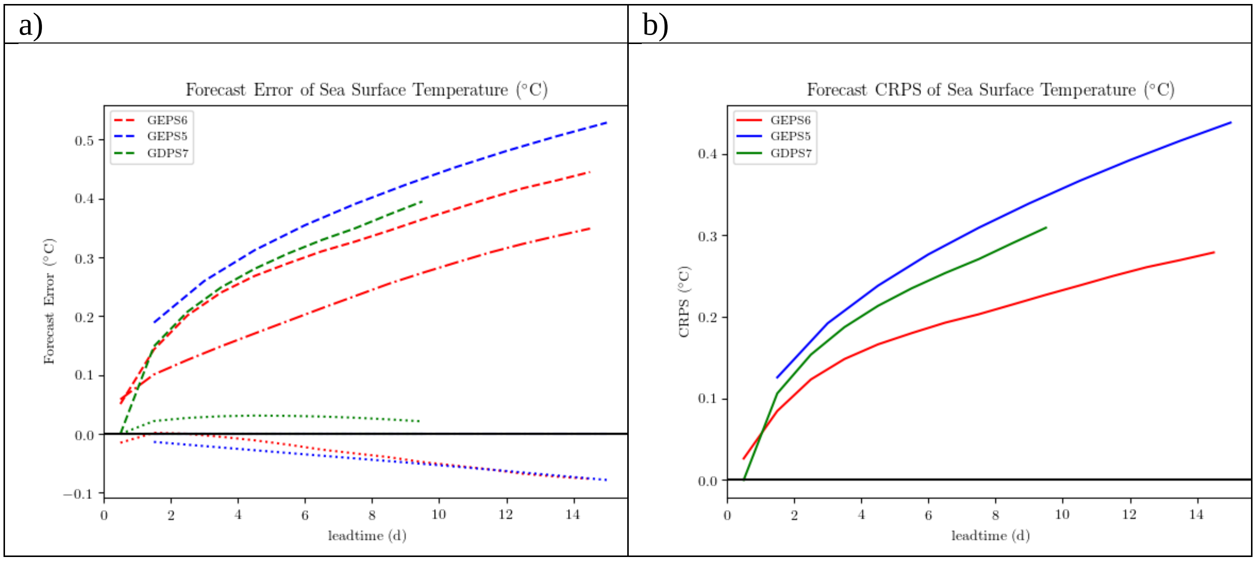


Figure 28: Same as Figure 24 except for winter set of forecasts. a) RMSE (dashed-line), mean error (dotted line) and ensemble spread (dot-dashed line). GEPS 6.0.0 (red), GEP 5.0.0 (blue), and GDPS 7.0 (green).

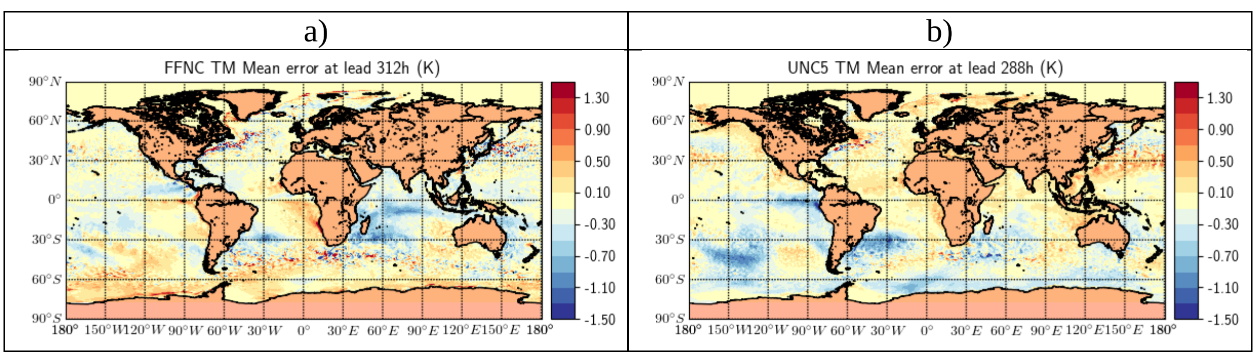


Figure 29: a) Mean error in GEPS 6.0.0 winter forecasts at 12.5d lead time. b) Mean error in GEPS 5.0.0 winter forecasts at 12.0d lead time. ***Note:*** *We are penalizing GEPS 6.0.0 an additional 0.5d here to account for the fact that it is being validated against a 12h GDPS forecast.*

a)	b)	c)

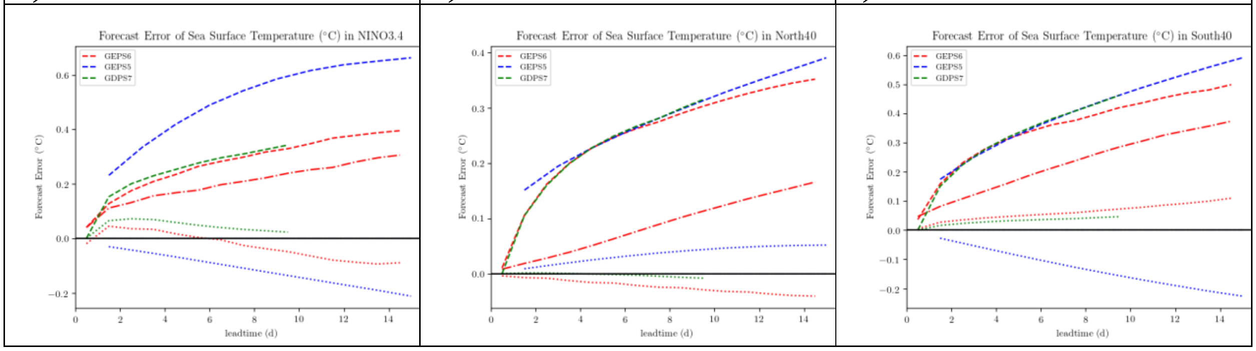


Figure 30: *Same as Figure 28a, except for regions a) NINO3.4 [170 -120W, 5S- 5N], b) north of 40N, c) south of 40S.*

As with the summer case, the coupled ocean easily outperforms the persisted SST's used in GEPS 5.0.0 and gives comparable performance to the GDPS system. Looking regionally, GEPS 6.0.0 again gives the best improvement of skill over the NINO.3.4 region, but with the transfer of GEPS 5.0.0 polar summer time deficiencies to the southern hemisphere, the regional performance at mid to high latitudes is now approximately the same in both hemispheres, with improvements in SST due to coupling in GEPS 6.0.0 arriving around day 6 and 8 in the southern and northern mid to high latitudes respectively. As before, CRPS shows the superiority of the ensemble system for this probabilistic score and once again, it is offered up as a benchmark for future versions.

As with the summer case, we also investigate the deterministic skill for the ensemble mean value of the sub-surface variables of tropical cyclone heat content, mixed layer depth, and 30m depth averaged x-component of velocity. The too little in summer tropical cyclone heat potential and too deep mid-latitude winter time mixed layer depth have been transferred to the seasonally appropriate hemisphere and will presumably effect the tropical and extra-tropical storm characteristics in those hemispheres. Once again, there is some indications in the RMSE as function of lead time plots for velocity (Figure 31i) that the ensemble system can outperform the deterministic system after approximately 5 days. However, where that improvement is most prevalent is no longer particularly obvious, although once again GEPS 6.0.0 does show improvements over GDPS 7.0 in the equatorial Indian Ocean. In this season, the two systems have completely opposite bias in the Atlantic. As was the case with the summer forecasts, GEPS 6.0.0 seems to have significant errors in the western Pacific, however. This may be due to the variety of convection schemes used in the atmospheric ensemble generation.

a)	b)

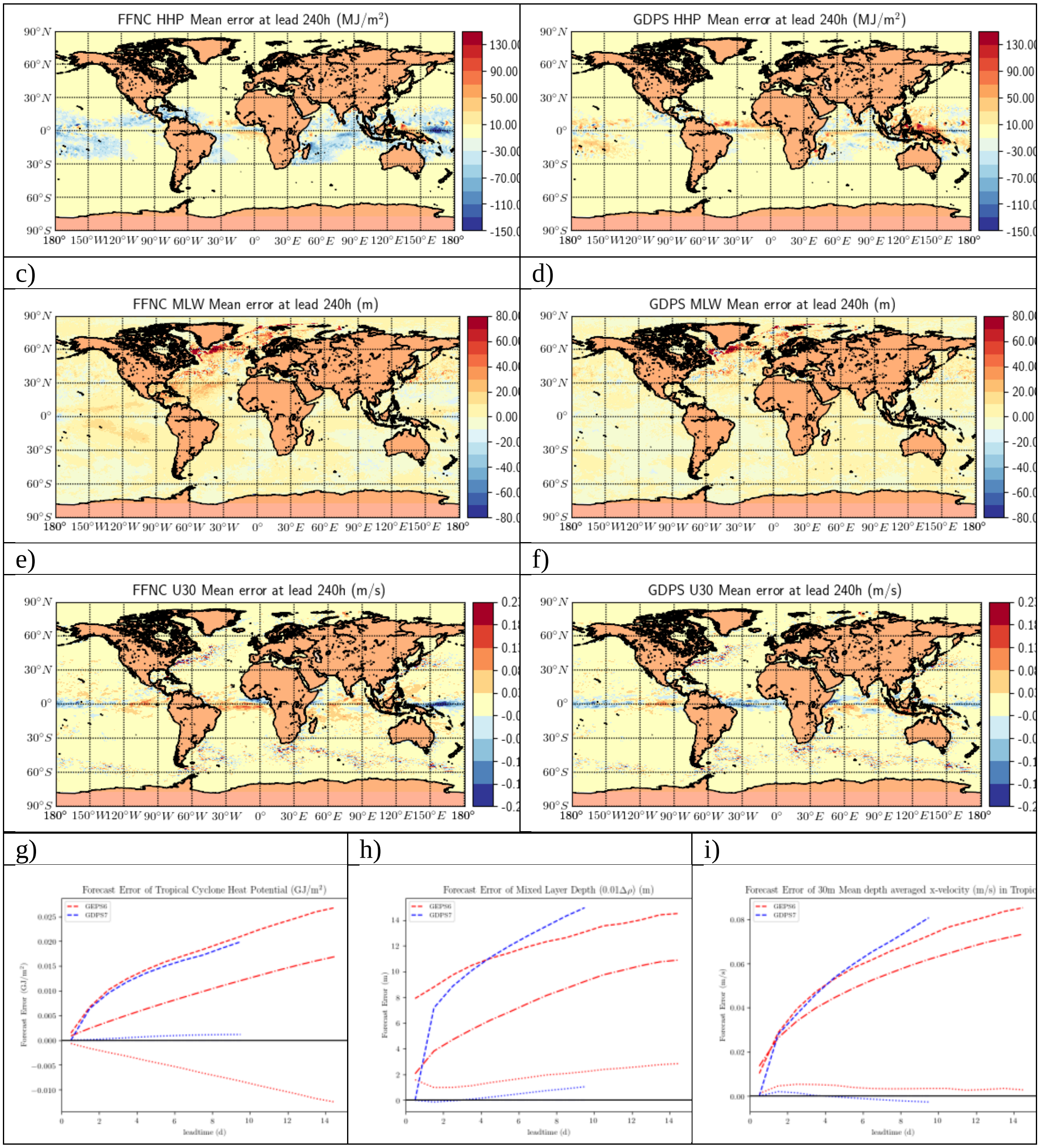


Figure 31: *First Row: Estimates of Error in Tropical Cyclone Heat Potential (MJ/m3) in a) GEPS 6.0.0, b) GDPS 7.0 and Second Row: Same as above except for Mixed Layer Depth (m). Third Row: As above, but for 30m mean depth averaged x-component of velocity. g-i) Global averaged values for the above quantities , respectively, as a function of lead time. Spatial average in i) is over the 20S-20N tropical region instead of globally.*

6.4 Quality of GEPS forecasts: sea ice

Sea Ice concentration is a particularly non-Gaussian variable, being constrained to an interval of [0, 1]. Furthermore, only within the marginal ice zone is this variable widely varying. For this reason, the sea ice diagnostics will be based on correctly forecasting whether ice is present (ice concentration > 0.15) or absent. To investigate this, we will quantify this skill with two metrics: the ensemble mean of Integrated Ice Edge Error, and the Spatial Probability Score.

$$IIEE = \frac{1}{n}\sum_{i=1}^{n}\int dA \mid F^i(x) - O(x) \mid = \int da \mid P(x) - O(x) \mid = \int_{ice} da \cdot P_0 \ (x) + \int_{no\ ice} da \cdot P(x)$$

$$SPS = \int da\,(P(x) - O(x))^2 = \int_{ice} da \cdot (P_0(x))^2 + \int_{no\ ice} da \cdot (P(x))^2,$$

where $F^i(x)$ is the forecast of ice present (=1) or absent (=0) in each ensemble member i, $O(x)$ is the observed value of ice present (=1) or absent (=0), $P(x) = \frac{1}{n}\sum_{i=1}^{n} F^i\,(x)$ is the probability of ice being present, and $P_0 = (1 - P(x))$ the probability of ice begin absent. We are able to bring the ensemble summation within the absolute value as the forecast minus the observed value will have the same sign for all ensemble members (value 1 or 0 if O=0, 0 or -1 if O=1). Note both IIEE and SPS are sums of the missed forecast areas, but IIEE is weighted by the probability of the incorrect forecast, while SPS is weighted by the square (energy) of the probability. Therefore, $SPS \leq IIEE$, with the equality applying for a deterministic forecast (P=1/0), with the difference between the two values indicative of the usefulness of probabilistic information, as

$$IIEE - SPS = \int da\, P \cdot (1 - P) = \int da\,(P_0 \cdot P)$$

represents the area of uncertain forecast ice coverage. In the plots below, besides the ensemble averaging as described above, an additional simple arithmetic mean over all forecast start dates has been applied.

Spatial figures will simply show the mean error averaged over start date and ensemble member

$$\sum_{t=1}^{n_t}\sum_{i=1}^{n_e}(F^i(x) - O(x)).$$

6.4.1 Results for the winter 2016/2017 period

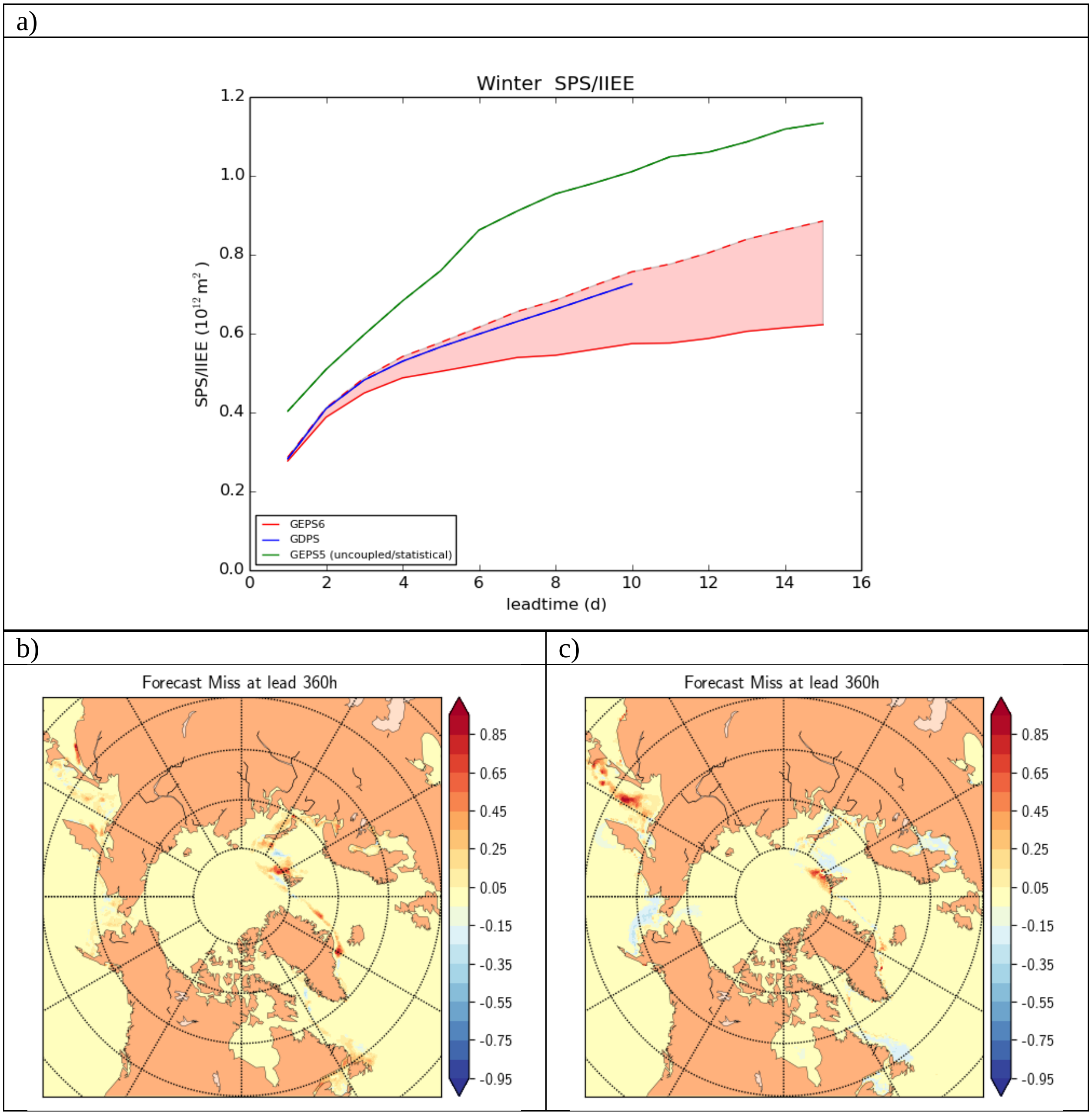
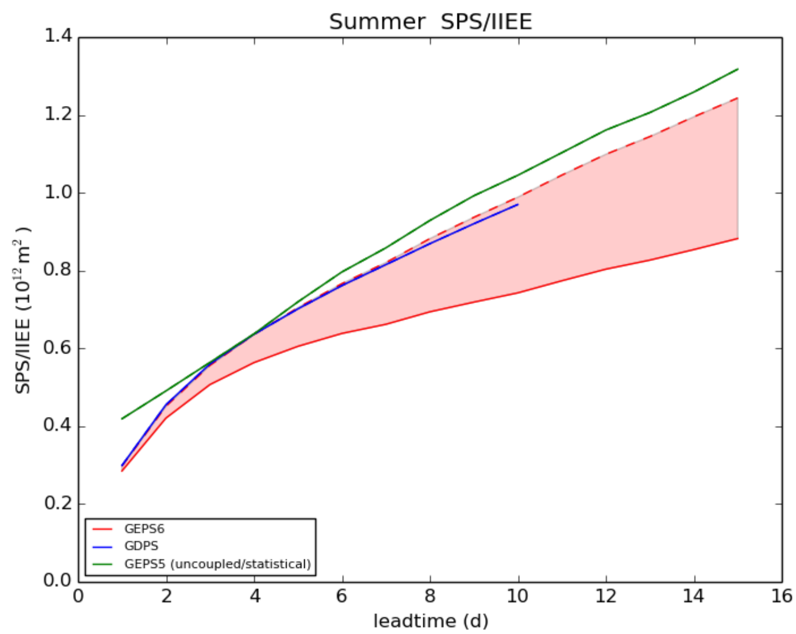


Figure 32 a) Values of SPS and IIEE (defined above) as a function of leadtime for GEPS 6.0.0 (red), GEPS 5.0.0 (statistical ice; green) and GDPS 7.0 (blue). Note: For deterministic systems SPS and IIEE are identical. For probabilistic systems, the difference (red shading for GEPS 6.0.0) quantifies the value of the ensemble. 15 day mean forecast probability minus observed value of sea ice (sea ice concentration threshold value of 0.15) for b) GEPS 6.0.0 and c) GEPS 5.0.0 statistical sea ice.

As can be seen by the IIEE value, GEPS 6.0.0 offers good improvement over the statistical sea ice used in GEPS 5.0.0, and has similar skill to the deterministic GDPS system. However, the significant reduction of the value of SPS with respect to the IIEE value shows the significant benefits offered up by the ensemble system – also suggesting that the ensemble variance is adequate to significantly capture the uncertainty in the forecast. The spatial maps suggest much of the improvement in the GEPS 6.0.0 system is in regions (Eurasian sector of Arctic and Pacific) of reduced interest for us, however, previous winter time deficits of sea ice in the Labrador Sea and Gulf of St. Lawrence have been replaced by milder surpluses. Regional indices will be added when available.

6.4.2 Results for the summer 2016 period

a)



b)

c)

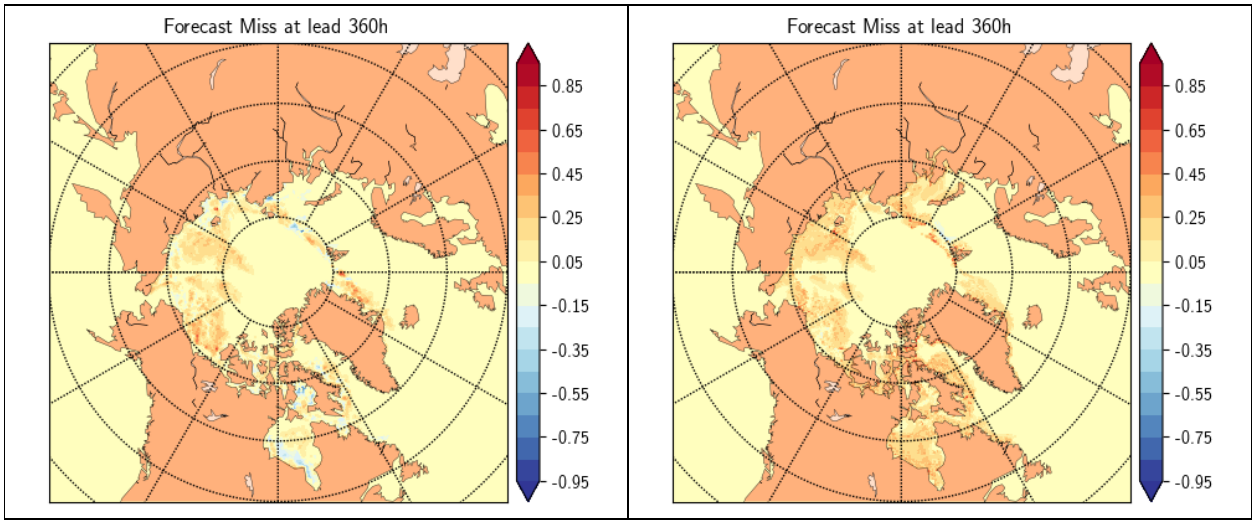


Figure 33. a) Values of SPS and IIEE (defined above) as a function of lead time for GEPS 6.0.0 (red), GEPS 5.0.0 (statistical ice; green) and GDPS 7.0 (blue). 15 day mean forecast probability minus observed value of sea ice (sea ice concentration threshold value of 0.15) for b) GEPS 6.0.0 and c) GEPS 5.0.0 statistical sea ice.

The summer forecasts show smaller improvements in the IIEE values of GEPS 6.0.0 with respect to GEPS 5.0.0, with a portion of that (to approximately day 5; not shown) being down to re-gridding aspects required for validation. The GEPS 6.0.0 system shows considerable improvements over the GEPS 5.0.0 statistical sea ice in the eastern Arctic and Hudson Bay regions with a large surplus of ice being replaced with mild deficits. In the western Arctic, sea ice surpluses may be marginally worse.

Compared to the winter forecasts, the summer forecasts show smaller improvements in the IIEE values of GEPS 6.0.0 with respect to GEPS 5.0.0 (Figure 33a). Regions in which the GEPS 6.0.0 system shows considerable improvements over GEPS 5.0.0 are in the Eastern Canadian Arctic and Hudson Bay, where large over forecasts of ice in GEPS 5.0.0 have been replaced with small under forecasts in GEPS 6.0.0 (Figure 33b/c). There are no significant improvements in the sea ice forecasts in the western Canadian Arctic or the Beaufort Sea.

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 February to June 2019 for the GEPS data assimilation component (from April to June 2019 for the GEPS forecast component). 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 atmospheric initial conditions

The ARCAD verifications were performed using the “postalt” files from the GDPS operational data assimilation cycle. The mean verification scores are computed over a period from 13 February to 30 May 2019.

Over the globe (Figure 34), the ARCAD verification results from the parallel run are very similar to those from the final cycles (discussed in section 6.1). The ARCAD verification results in the Northern extratropics (figure not shown) are very similar to those over the globe, namely significant improvement in the standard deviations for winds, dew-point temperature and temperature, but slight degradation in biases of upper-level U-wind component, lower-level geopotential heights and near-surface temperature. Over the Southern extratropics (Figure 35), the impacts from the proposed changes are very similar but with slightly less significant improvement for standard deviation and no degradation in the biases of geopotential heights. Over the tropics, similar to what was observed for the final series of tests, the biases of geopotential heights are greatly reduced and the standard deviations are significantly improved for both wind components and dew-point depression (figure not shown).We also note slight degradation in the biases for the lower-stratosphere U-wind and near-surface temperature. The verification against radiosonde dew-point depression observations over the parallel-run period is shown in Figure 36. Here we see significant improvement in standard deviation except below 1000 hPa where data count is very small, and also in biases except below 700 hPa where slightly larger dry biases are observed, which validates negatively between 700 and 825 hPa, positively between 825 and 925 hPa and negatively below 925 hPa. The verification against screen-level temperature observations shows substantial reductions of standard deviation and bias from 60 degree south to 60 degree north (Figure 37). In both the North- and South-Pole regions, slightly larger warm biases are noted. The verification scores against the screen level dew point depression are very similar to what was observed during the final cycles (Figure 38). Verifications against ATMS radiance observations are very similar to those of the final series of tests (Figure 39). Significant reduction in standard deviation is also found for aircraft data, Scatterometer winds, and Satellite winds. Probably because the 4DEnVar assimilates much more IASI and CrIS observations than the EnKF, the ECMWF-hybrid gain brings the largest positive impacts on the verifications against those observations.

Overall, the objective verification results from the parallel run confirm the general significantly positive impacts seen with the ARCAD verifications against independent Radiosonde observations and with the SQLITE verifications against assimilated observations, like AMSU-A, AMSU-B, ATMS, Radionsondes and surface observations. However, as seen in the final series of tests, there are again some slight negative impacts such as for zonal wind bias in the Stratosphere and for dew-point depression bias in the lower Troposphere. .

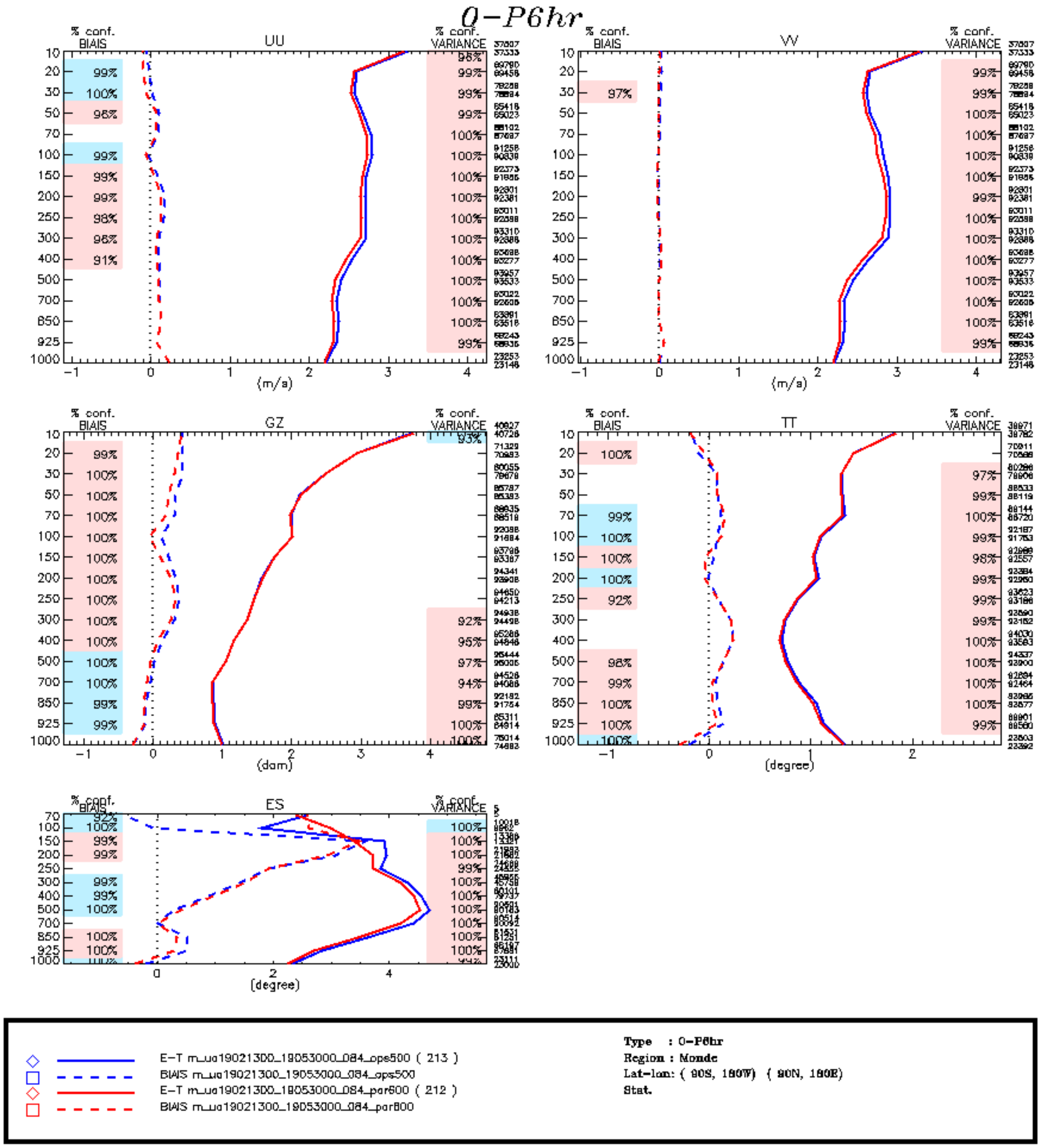


Figure 34. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the globe for a period from 13 February to 30 May 2019 for EnKF experiments with the operational GEPS 5.0.0 (blue) and the parallel GEPS 6.0.0 (red).

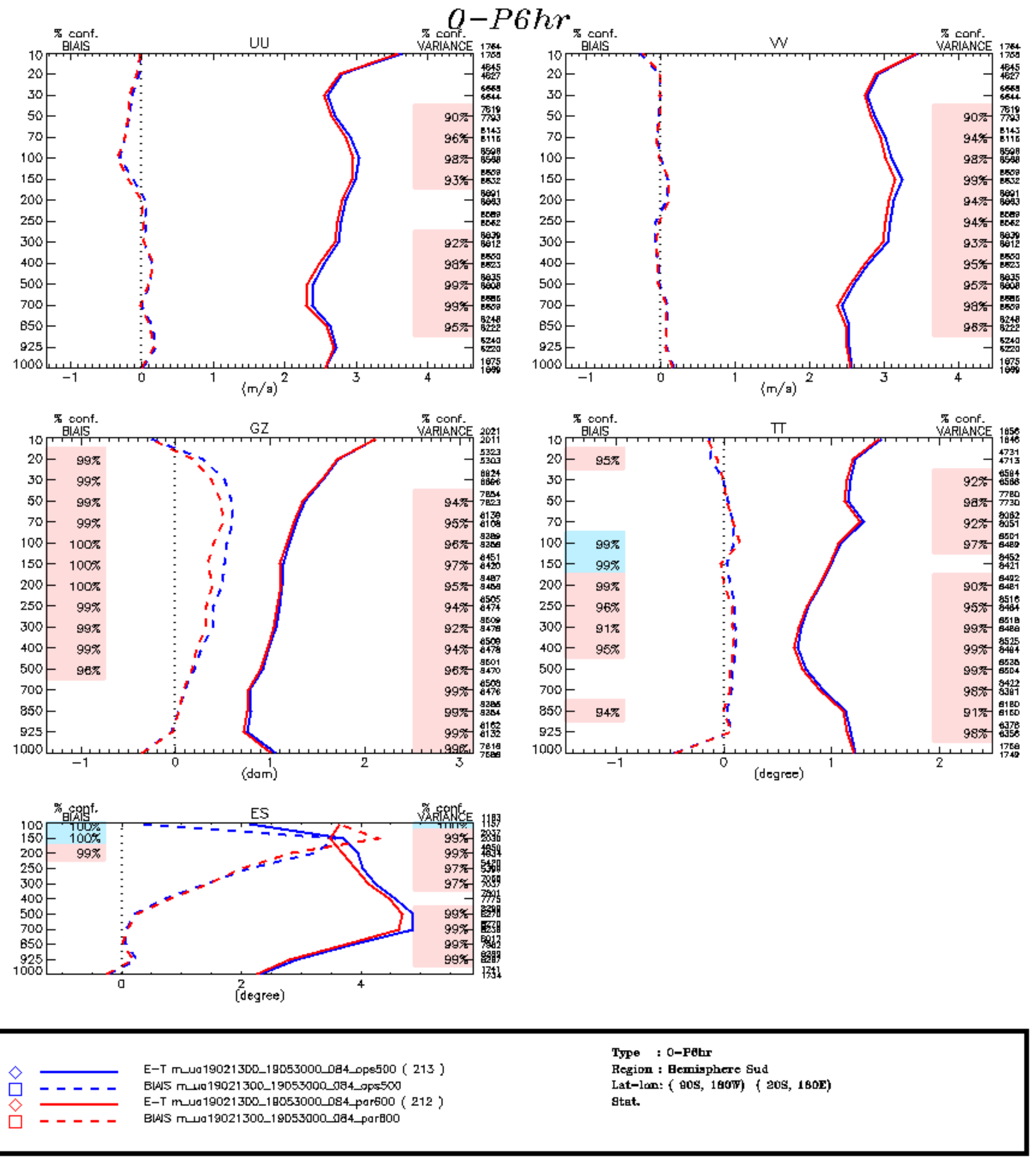


Figure 35. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosondes over the Southern Extratropics for a period from 13 February to 30 May 2019 for EnKF experiments with the operational GEPS 5.0.0 (blue) and the parallel GEPS 6.0.0 (red).

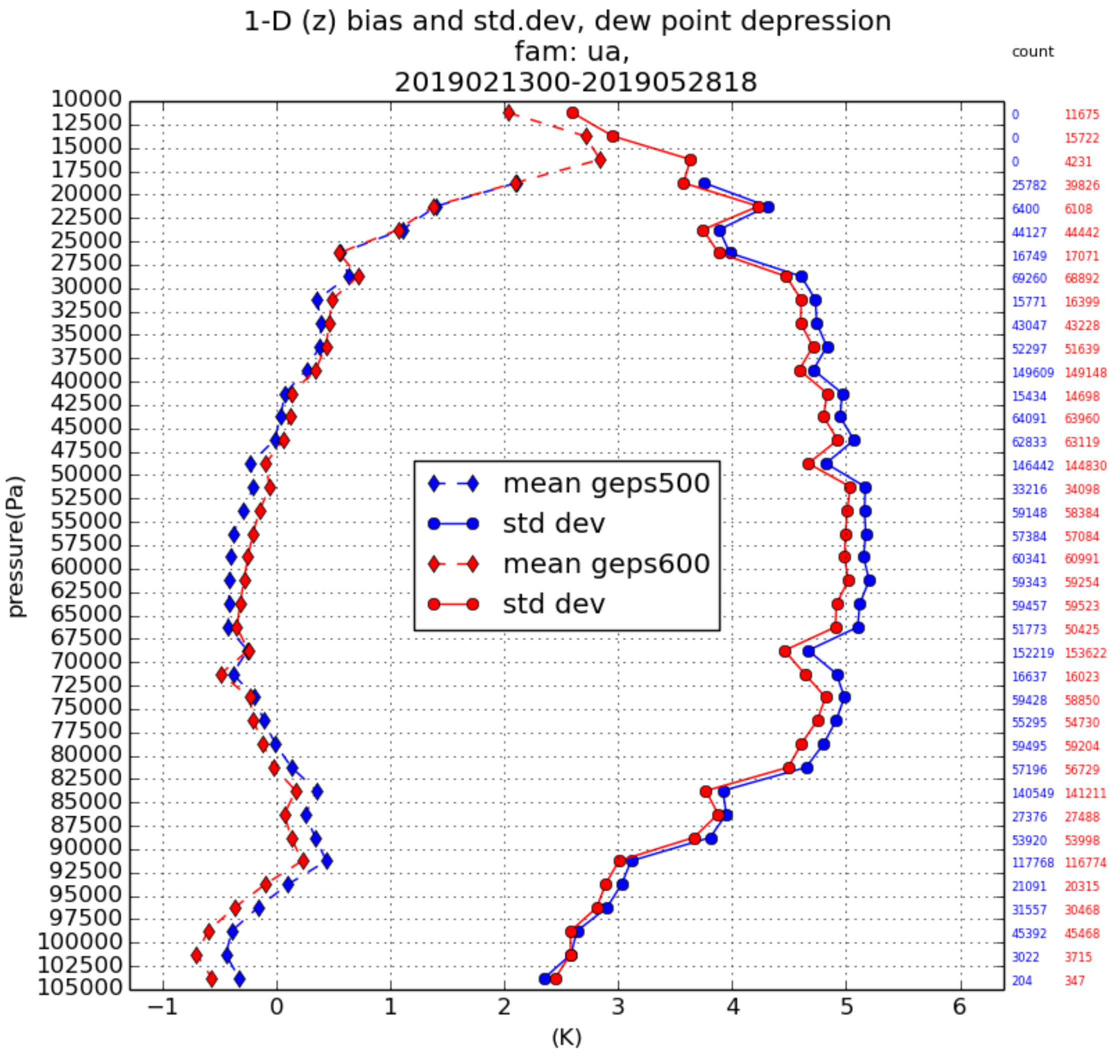


Figure 36. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosonde dew-point depression globally averaged over a period from 13 February to 28 May 2019 for EnKF experiments with the operational GEPS 5.0.0 (blue) and the parallel GEPS 6.0.0 (red).

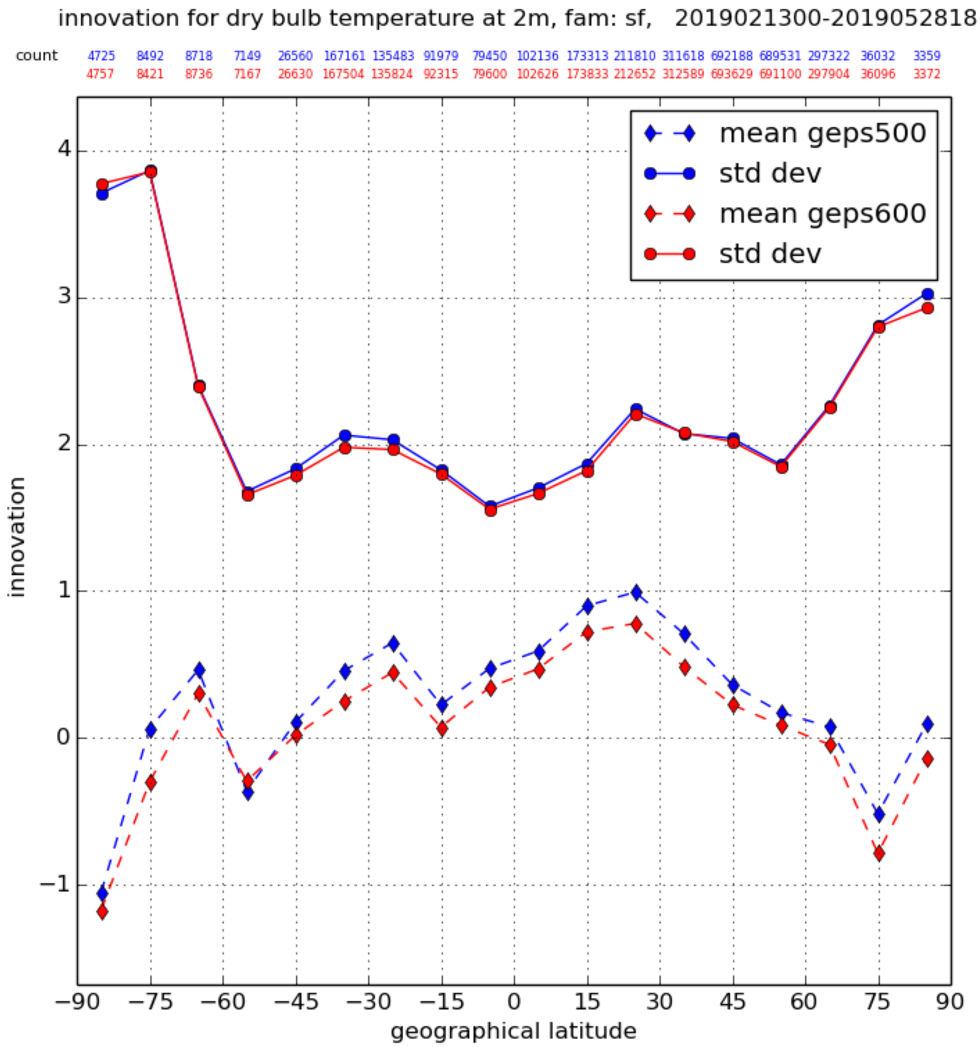


Figure 37. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosonde dew-point depression globally averaged over a period from 13 February to 28 May 2019 for EnKF experiments with the operational GEPS 5.0.0 (blue) and the parallel GEPS 6.0.0 (red).

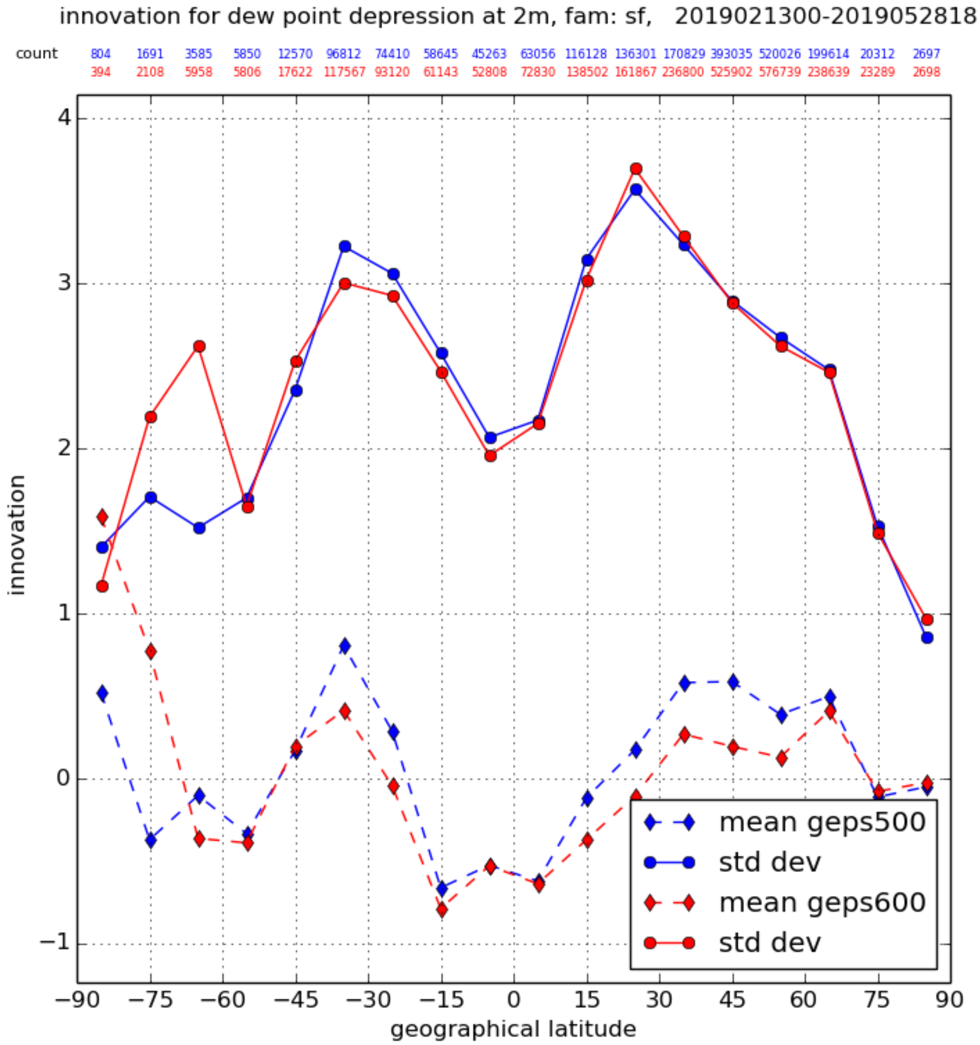


Figure 38. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosonde dew-point depression globally averaged over a period from 13 February to 28 May 2019 for EnKF experiments with the operational GEPS 5.0.0 (blue) and the parallel GEPS 6.0.0 (red).

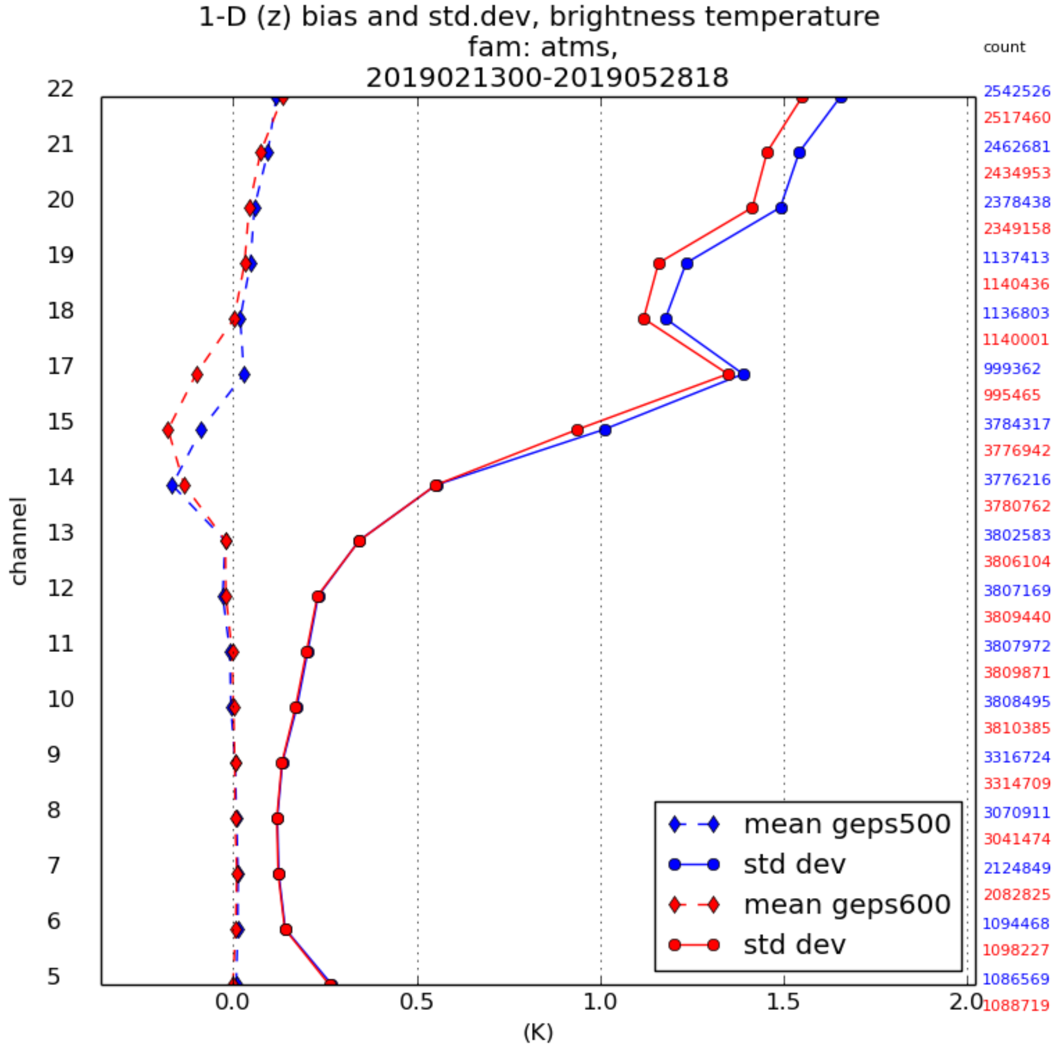


Figure 39. Standard deviation (solid line) and bias (dashed line) of the trial error (observation minus trial fields) against radiosonde dew-point depression globally averaged over a period from 13 February to 28 May 2019 for EnKF experiments with the operational GEPS 5.0.0 (blue) and the parallel GEPS 6.0.0 (red).

7.2 Quality of GEPS forecasts

Here we show an example of evaluation with the CRPS for the parallel run period of 7 April to 16 June 2019, in comparison to the operational GEPS 5.0.0. Then, we will show the scorecards for the globe, for weeks 1 and 2.

Figure 40 shows the CRPS scores of geopotential height at the 250, 500 and 850 hPa levels over the globe. We observe improvement of skill for geopotential height in the first week at all levels.

In the second week, the improvement at the 250 hPa level is more obvious than the lower levels. A positive trend of improvement toward the end of the second week can also be found.

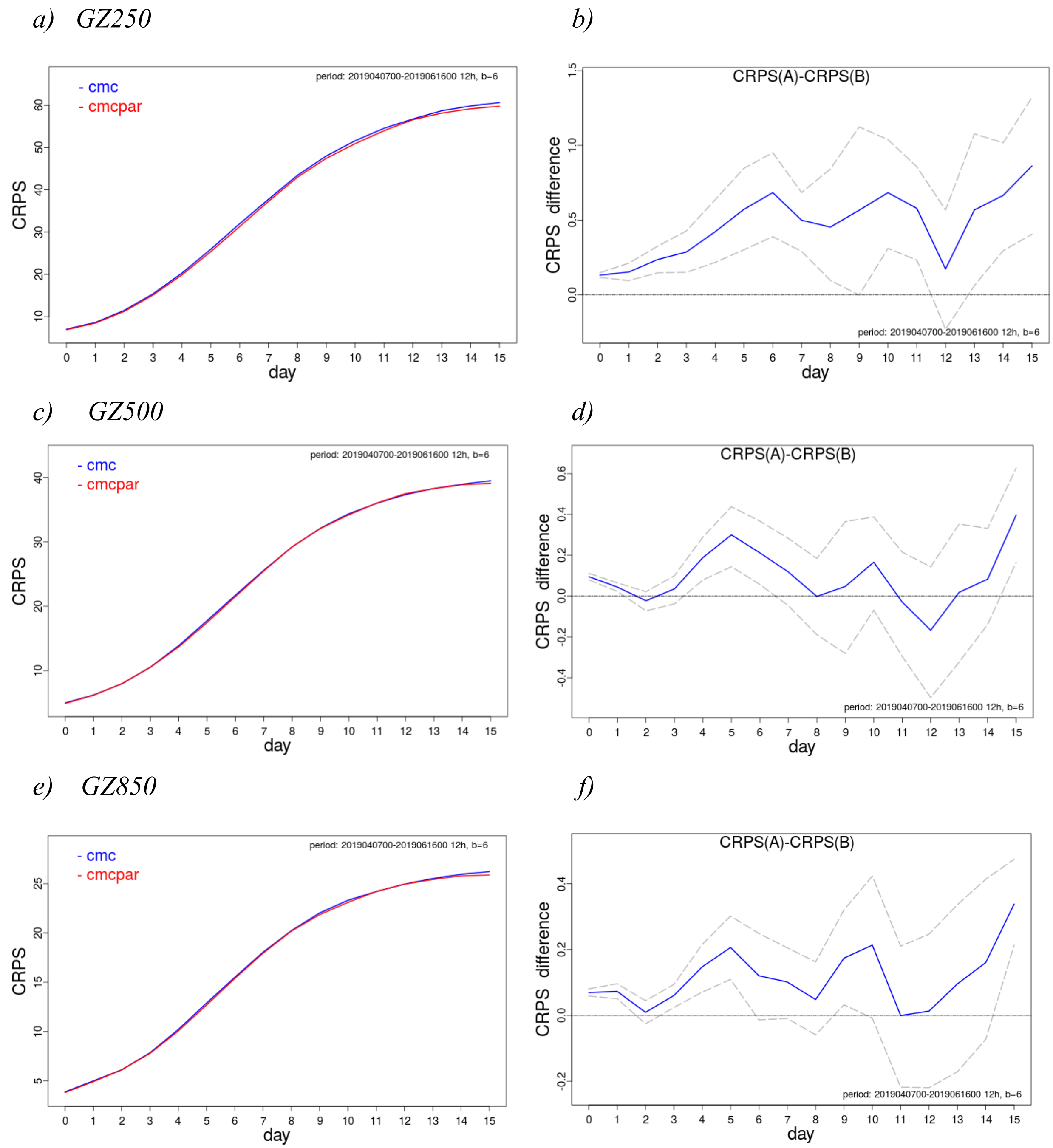
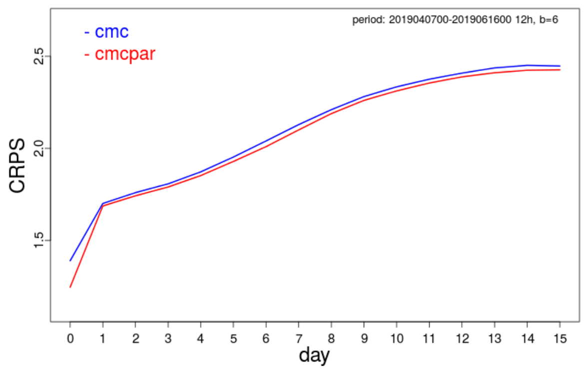


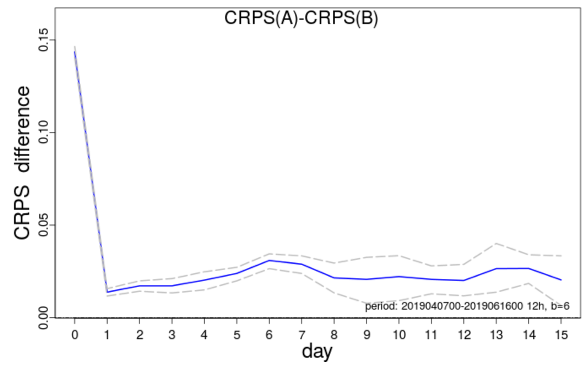
Figure 40. Left panels: CRPS over the globe for the operational GEPS 5.0.0 (in blue) and the parallel run of GEPS 6.0.0 (in red) during 7 April to 16 June 2019 for geopotential height at 250 hPa (a), 500 hPa (c) 850 hPa (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three fields.

The CRPS of 1.5-m temperature, 1.5-m dew-point depression (ES1.5m) and mean sea-level pressure over the globe as a function of lead time are shown in *Figure 41*. In general, the forecast skill of these variables is improved, especially for the temperature.

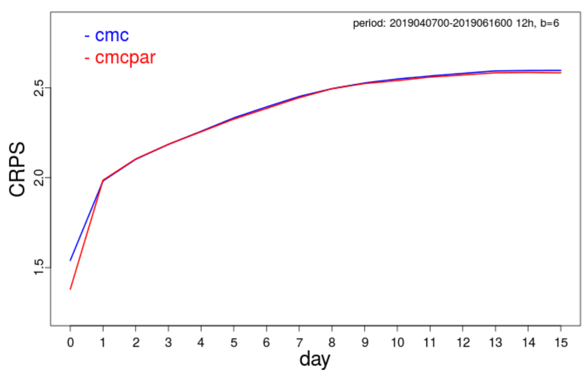
a) T1.5m



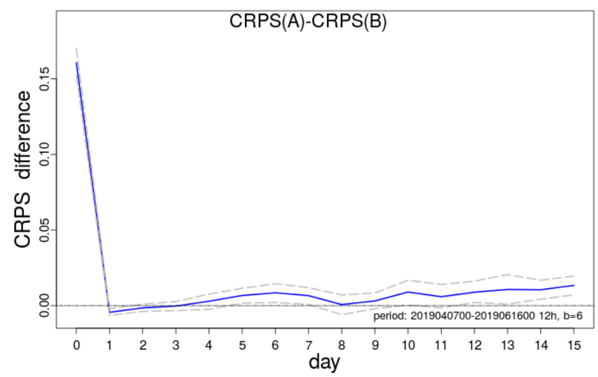
b)



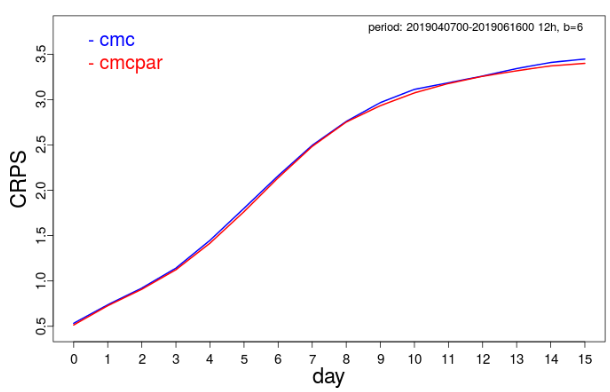
c) ES1.5m



d)



e) MSLP



f)

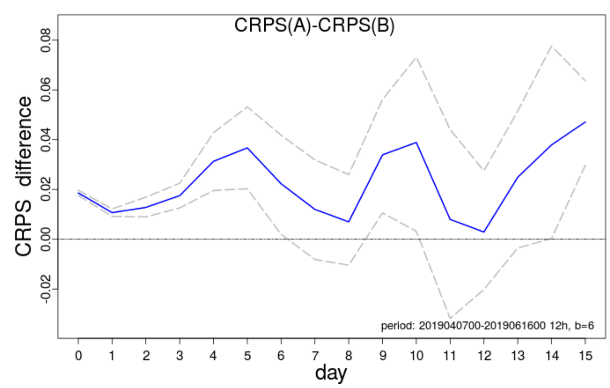


Figure 41. Left panels: CRPS over the globe of the operational GEPS 5.0.0 (in blue) and the parallel run of GEPS 6.0.0 (in red) during 7 April to 16 June 2019 for the 2-m temperature (a), the 2-m dew-point depression (c) and mean sea-level pressure (e). Right panels: the difference between the scores of the two systems is shown as well as the 90% confidence intervals calculated with block bootstrapping for the three fields.

Figure 42 shows the scorecard over the globe for week 1. The improvement of forecast performance is clear in the parallel run. The overall improvement is 54% for the first week versus 22% of deterioration.

```
                                                                              Printed by Xingxiu Deng
Jun 17, 19 14:55             RESUME_cmc_vs_cmcpar_Global_000-144                         Page 1/1

 -------------------- UU ---------------------   -------------------- VV ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|        |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |++ ----|-------|++ ----|+------|++++  -|     50 |+++++++|  --   |++     |+------|+++++++|
 100 |++++   | +-----|++++   |+------|+++++++|    100 |+++++++| ---   |+++++  |+------|+++++++|
 250 |+++++++|+-+++++|+++++++|+-   + |+++++++|    250 |+++++++|-      |+++++++|+ ---  |+++++++|
 500 |+++++++| ++++++|+++++++|+   +++|+++++++|    500 |+++++++| + -   |+++++++|+  - + |+++++++|
 850 |+++++++|++     |+++++++|+----  |+++++++|    850 |+++++++|+      |+++++++|+---  -|+++++++|
 925 |+++++++| +     |+++++++|+---   |+++++++|    925 |+++++++|+      |+++++++|+---- -|+++++++|

 -------------------- GZ ---------------------   -------------------- TT ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|        |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |+--    |- +++++|+-  +  | -+++++|+------|     50 |++ ----| ------|++ ----|+------|++++++ |
 100 |+++++++|+++++++|+++++++|+++++++|++- +++|    100 |++++ --|+++++++|++ ----|+ ++   |++-----|
 250 |+++++++|++     |+++++++|+      |+++++++|    250 |+++++++|- ---  |+++++++|-------|+++++++|
 500 |++  +++|-------|++  +++|-------|+++++++|    500 |+++++++|    ---|+++++++|--     |+++++++|
 850 |++++++ |-------|++ +++ |------ |+++++++|    850 |+++++++|-+++++ |+++++++|--  +++|+++++++|
 925 |++++++ |-------|+++++++|----   |+++++++|    925 |+++++++|++++++ |+++++++|-   +++|+++++++|

 -------------------- ES ---------------------   ------------------ SURFACE ------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|        |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 250 |       |       |       |       |       |     PN |+++++++| ------|+++++++|--     |+++++++|
 500 |+++++++|++     |+++++++| ---   |+++++++|     TT |+++++++|+++++++|+++++++|+++++++|+++++++|
 850 |++++++ |--- - -|++     |+---- -|+++    |     UV |+---   |+------|+---   |+------|+++++++|
 925 |+++++++|-------|++   + |+------|+++++++|     ES |+++++++| ------|+-   ++|-------|+++++++|

 -------------------------------------- PRECIP ----------| +  54.25%  -  21.85%   23.90%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    |----  |---  +|    ++| + ++ |      |  +   |      |      |+-    |+     |+    +|
 BSSrel |------|------|----  |----  |---   |---   |--    |---   |---   |--    |---   |
 BSSres |++++++|++++++|++++++|++++++|++++++|++++++|+++ + |+++  +|+    +|+     |+    +|

Monday June 17, 2019            RESUME_cmc_vs_cmcpar_Global_000-144                        1/1
```

Figure 42. Scorecard for the the parallel run of GEPS 6.0.0 compared to the operational GEPS 5.0.0, for 7 April to 16 June 2019, over the globe, week 1.

The second week forecasts also shows general improvement (Figure 43), where the overall improvement is 29% versus 15% of deterioration.

In summary, the performance of GEPS 6.0.0 during the parallel run period appears consistent with that during the final series of tests as discussed in Section 6.2.

Printed by Xingxiu Deng

Jun 17, 19 14:56 **RESUME_cmc_vs_cmcpar_Global_168–312** Page 1/1

```
 -------------------- UU ---------------------    -------------------- VV ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |------ |-----  |------ |-------|       |      50 |++     |       |+      | ---   |++     |
 100 | +++   |--     | +++  +|-----  |+++++ +|     100 |   ++++|       |   +++ |-- -   |   ++++|
 250 |++ ++  |+++    |+++++  |       |+++++  |     250 | +     |       |++    +|----   |++    +|
 500 |   +   |+++++++|+  +   |       |+  +   |     500 |+      |       |+     +|       |+     +|
 850 |+  ++  |       |+  +   |       |+  ++  |     850 |+      |       |+ +++  |   -   |+ +++  |
 925 |+  ++  |       |++++++ |       |++++++ |     925 |+      |    ++ |+ +++  |       |+ +++  |

 -------------------- GZ ---------------------    -------------------- TT ---------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
  50 |    ---|+++    |    -- |++   --|-      |      50 |------ |-------|-------|------ |       |
 100 |+++++++|+++++++|+++++++|+++++++|+++++  |     100 |-------|+++++++|-------|   ++++|-------|
 250 |       |       |++ ++ +|--     |+++++ +|     250 |    +++|   ++++|+++++++|-      |+++++++|
 500 |       |-------|       |-------|+  +   |     500 |+      |--     |+      |       |++     |
 850 |+ +    |---    |  +    |- -  - |+ ++   |     850 |+   +  |       |++ ++  |+  ++  |++ ++  |
 925 |+ ++   |--     |+ ++   |       |++++   |     925 |+   +  |       |+  ++  |+  ++  |+  ++  |

 -------------------- ES ---------------------    ------------------ SURFACE ------------------

     |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|         |RMSE   |BIAS   |CRPS   |CRPSrel|CRPSres|
 250 |       |       |       |       |       |      PN |+ ++   |--     |  ++   |       |  ++   |
 500 |+ +++++|       |+++++++|    +  |+++++++|      TT |+++ +++|+++++++|+++++++|+++++++|++  +++|
 850 |       | -- ---| -    -|--- - -| -    -|      UV |+ -    |-------|  -    |-------|+  ++ +|
 925 |       |-------|      -|-------|       |      ES |+      |-------|+  + ++|-------|+++++++|

 -------------------------------------- PRECIP ----------| +  28.98%  -  14.95%   56.07%

 SEUIL  | 2.5  |  5   |  10  |  15  |  20  |  25  |  30  |  35  |  40  |  45  |  50  |
 BSS    |+ - - |+     |+    +|+++  +| +   +|     +|     +|     +|     +|   + +|     +|
 BSSrel |------| ---- |+     | +   +| +   +| +   +| ++  +| +   +| +   +| +   +| ++  +|
 BSSres |+  +  |++++  |++++ +|++++++|++   +|   + +|   + +|     +|     +|     +|   + +|
```

Monday June 17, 2019 RESUME_cmc_vs_cmcpar_Global_168–312 1/1

Figure 43. Same as Figure 42, but for week 2.

7.3 Quality of GEPS forecasts: ocean and sea ice

An evaluation of GEPS 6.0 has been carried out for the parallel suite run starting with 16 April, 2019 for the sea ice and 25 April, 2019 for the ocean variables (evaluation of the ocean variables also required the parallel run of GDPS 7.0). It continues with all cases having validation through 27 June.

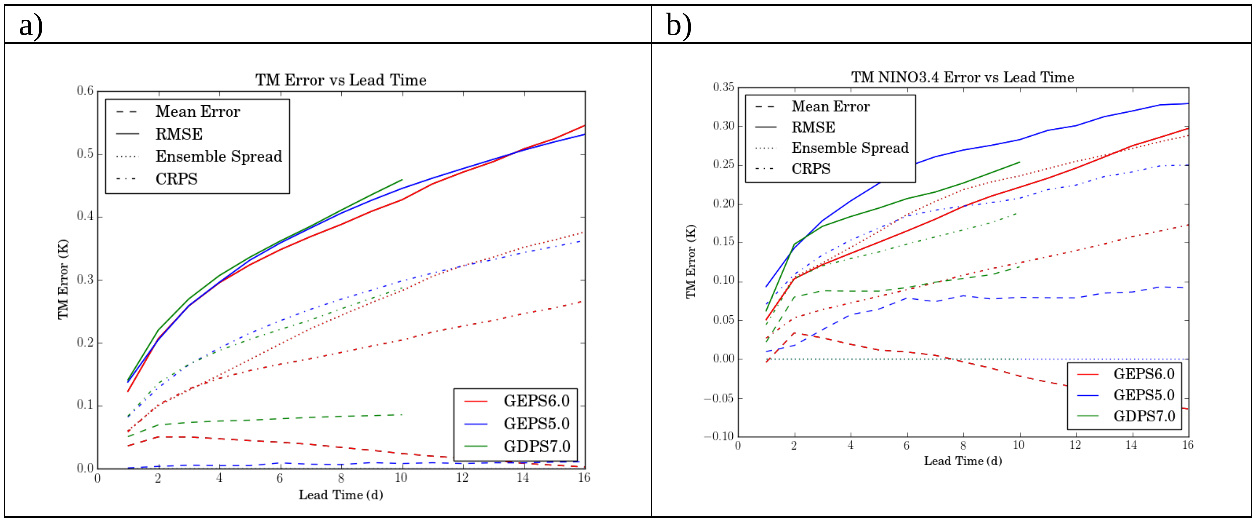


Figure 44 RMSE (Solid), Mean Error (Dashed), CRPS (dashed-dot) and Ensemble Spread (dotted) for GEPS6.0.0 (red), GEPS5.0.0 (blue) and GDPS7.0.0 (green). a) Global Area weighted average. b) NINO3.4 [170-120W, 5S-5N] area weighted average. Note: Due to technical difficulties in data retrieval from tape archive, GEPS5.0 errors are based on a shorter evaluation period.

Figure 44a shows the global area averages of Root Mean Squared Error of the Ensemble Mean (RMSE), Mean Error, Continuous Ranked Probability Score (CRPS) and Ensemble Spread for the GEPS 6.0 parallel run, the GEPS 5.0 operational run (persisted SST anomalies) and the GDPS 7.0 parallel run. Figure 44b shows similar values for the NINO3.4 region. The operational run is evaluated relative to the operational SST analysis, whereas the parallel runs are evaluated against the parallel SST analysis. All three systems perform equivalently for RMSE, although the two prognostic ocean models (GEPS 6.0 and GDPS 7.0) have larger mean errors. However, both these systems will have a residual diurnal SST cycle (shown are errors in the daily mean values) not present in the SST analysis, or the persisted SST of GEPS 5.0. In the NINO3.4 region, the GEPS 6.0 system outperforms both the persisted SST of GEPS 5.0 and GDPS 7.0. The lower scores for CRPS (dashed-dotted lines) in the ensemble GEPS 6.0 system shows the ensemble system is outperforming the single realization of the GDPS system, and the single realization persisted SST of GEPS 5.0, however, the ensemble spread of the GEPS 6.0 system (dotted red line) is about half of the RMSE, indicating under-dispersiveness in the ensemble.

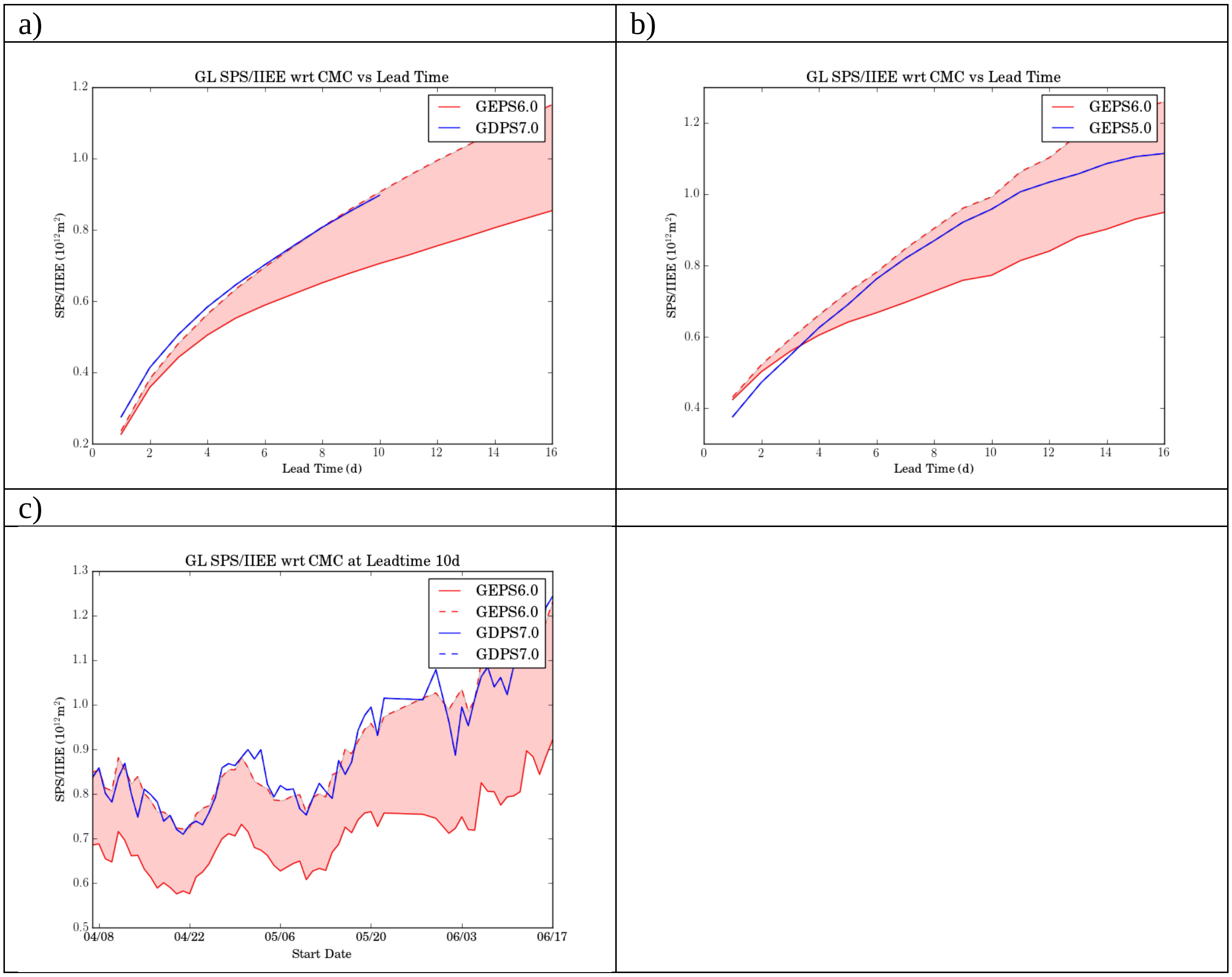


Figure 45 Spatial Probability Score (SPS: Solid) and Integrated Ice Edge Error (IIEE: Dashed) scores for a) GEPS 6.0 (red) and GDPS 7.0(blue), b)GEPS 6.0 (red) and GEPS 5.0 (blue) as a function of lead time. c) SPS and IIEE for GEPS 6.0 (red) and GDPS (blue) as a function of start date for 10 day lead time. See section 6.4 for a definition of SPS and IIEE. For the two deterministic systems, GDPS 7.0 (a/c; blue line) and GEPS 5.0 (b; blue line), SPS and IIEE are equivalent (solid lines are overlying the dashed lines). Scores in a) are evaluated from the sea ice grid output, whereas scores in b) are evaluated from the atmospheric grid output subsequently interpolated onto the sea ice grid (for binary land/sea masking). Additional error is introduced by this interpolation. Furthermore, due to technical difficulties in retrieving atmospheric data from tape, these scores are evaluated primarily using late May and early June start dates when the error was higher (c).

Figure 45 shows the Spatial Probability Score (SPS) and the Integrated Ice Edge Error (IIEE) for (a) the GEPS 6.0 parallel run, and GDPS 7.0 parallel run; (b) the GEPS 6.0 parallel run and operational GEPS 5.0 over overlapping periods. Since GEPS 5.0 sea ice is only available on the atmospheric grid, and interpolation of that onto the sea ice grid for evaluation introduces additional error, SPS and IIEE values in (b) are taken from the atmospheric grid for both GEPS 6.0 and GEPS 5.0 for comparison purposes. Figure 45a is evaluated over the whole 16 Apr – 27 June period, whereas Figure 45b is evaluated over a truncated period primarily covering late May and June due to technical issues with data retrieval. GEPS 6.0 shows comparable deterministic errors with GDPS 7.0 (IIEE; dashed line, hidden below identical solid SPS curve for single realization GDPS). However, the lower SPS (solid line) score for GEPS 5.0 indicates the added value of having a probabilistic forecast of sea ice. Over the period evaluated for the GEPS 6.0 and GEPS 5.0 comparison (Figure 45b), the dynamically adjusted sea ice of GEPS 5.0 is outscoring the GEPS 6.0 coupled sea ice for IIEE. However, this is over the late May and early June start dates in which the coupled models (both GEPS 6.0 and GDPS 7.0) are showing a large increase in ice edge error (dashed line Figure 45c), and a larger uncertainty (shaded region of Figure 45c) in the probabilistic outcomes of the ensemble system, both likely due to being in a transition to a seasonal melt period. A full parallel suite period evaluation of GEPS 6.0 vs GEPS 5.0 results is ongoing. The probabilistic SPS score (solid line in Figure 45b) still demonstrates the additional forecast information coming from the ensemble system.

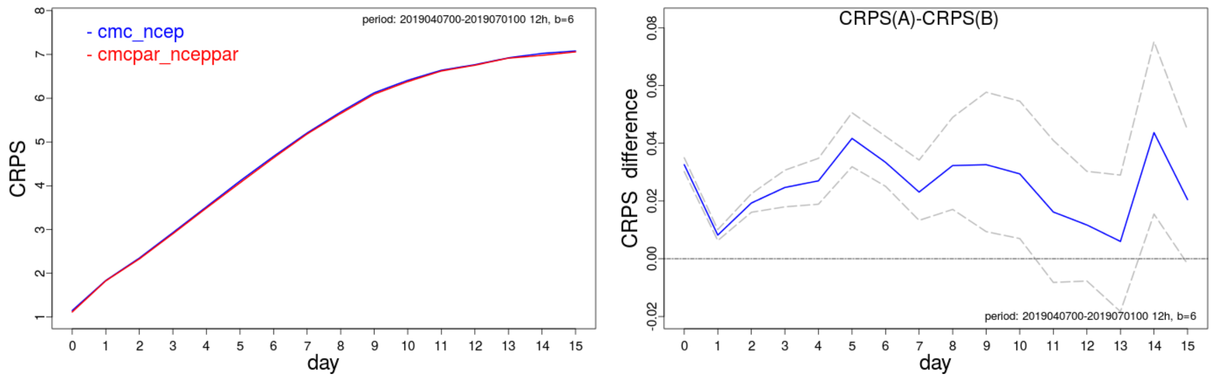
8 Performance of dependent systems

This section describes the adjustments made to the systems that are dependent on GEPS 6.0.0. In particular, we describe the adaptation of the North American Ensemble Forecast System (NAEFS). Results from the Regional Ensemble Prediction System (REPS) are discussed in another technical note (Patoine et al., 2019).

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 5.0.0 members and the NCEP GEFS members. After, the GEPS 6.0.0 (instead of the GEPS 5.0.0) members were put in the mix to form the new NAEFS. Figure 46 shows CRPS over globe for the 250 hPa zonal wind and 500 hPa geopotential heights as well as the BSS for the 15 mm and more category. Significant improvements are noticed for the new NAEFS with the arrival of the GEPS 6.0.0. As usual, the improvements are smaller because half of the super-ensemble is the same (NCEP members). The NAEFS users should expect better forecasts in week 1 over North America and in week 1 and 2 over the globe with the replacement of GEPS 5.0.0 by GEPS 6.0.0.

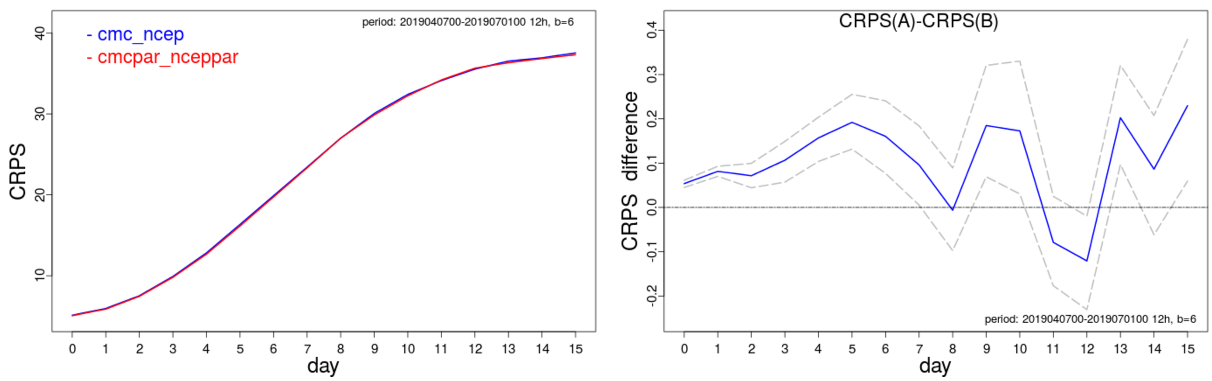
a) UU 250 hPa

b)



c) GZ 500 hPa

d)



e) PCP at 24h lead time

f)

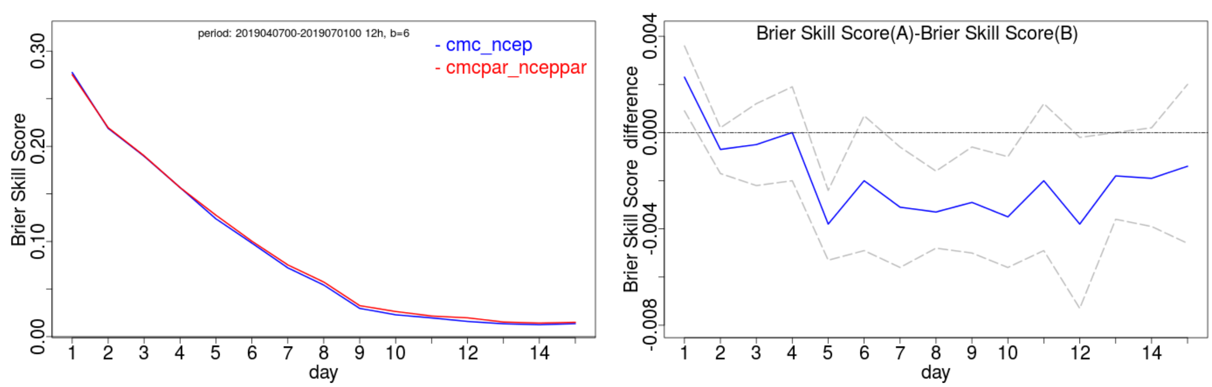


Figure 46 Left panels: CRPS over globe of the superensemble with GEPS 5.0.0 and NCEP GEFS (experiment 'cmc_ncep' in blue) and the one with GEPS 6.0.0 and NCEP GEFS (experiment 'cmcpar_nceppar' in red) during the parallel run period from 7 April 2019 to 1 July for 250 hPa zonal wind (a), 500 hPa geopotential heights (c) Brier skill score for the 15 mm and more in 24 hour precipitation category (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.

9 Availability of products

The addition of the NEMO and the sea ice model to the forecast component results in an increase in the computational time for numerical forecasts. We expect a delay of approximately 25 minutes maximum in the production of forecasts. As a result, the products will be available a little later than usual after the implementation of GEPS 6.0.0.

10 Summary of the results

- The quality of the new EnKF trial fields is improved especially in standard deviations for all the variables.
- The forecast performance of the new system GEPS 6.0.0 is improved in week 1 for North America, and in both weeks 1 and 2 for the globe for all upper air fields.
- The forecast surface fields (MSLP, 1.5-m temperature and dew-point depression and precipitation) are also improved significantly except the 10-m wind speed which is degraded.
- The GEPS 6.0.0 SST forecast over the first 10 days performs equally well to the deterministic coupled system GDPS. It does particularly well in the NINO3.4 region, as well as at the extra-tropical and high northern latitudes, but underperforms compared to GEPS 5.0.0 persisted SST in the southern ocean.
- For the sea ice, GEPS 6.0.0 offers good improvement over the statistical sea ice used in GEPS 5.0.0, and has similar skill to the deterministic GDPS system.
- The quality of the reforecast is improved. This is reflected by reduced T2m bias over the North Polar region in DJF, significantly improved T2m skill in weeks 3 and 4, and improved MJO forecast skill.

11 Acknowledgements

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

- Frederic Dupont, Jean-Marc Bélanger, Yukie Hata, and François Roy, Seung-Jong Baek, Bin He, Mark Buehner, Cécilien Charette, Stéphane Laroche, Pierre Koclas, Ervig Lapalme, Sylvain Heilliette, Josée Morneau, Judy St-James, A. Erfani, Ron McTaggart-Cowan, André Plante, Antoine Macia, Sua Lim, Vivian Lee, Michel Desgagné, Maria Abrahamowicz, Nicola Gasset, Nicolas, Katja Winger, and Michel Valin.

12 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.

Deng X.-X., and Co-authors, 2018: Global Ensemble Prediction System reforecast system (GEPS) update from version 4.3.0 to version 5.0.0. 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/tech_notes/technote_geps-500_20180918_e.pdf

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.

Gottschalck J., M.Wheeler, K.Weickmann, F.Vitart, N.Savage, H.Lin, H.Hendon, D.Waliser, K.Sperber, M.Nakagawa, C.Prestrelo, M.Flatau, W.Higgins, 2010: A Framework for Assessing Operational Model MJO Forecasts: A Project of the CLIVAR Madden-Julian Oscillation Working Group. *Bull. Amer. Meteor. Soc*., 91, 1247-1258.

Houtekamer, P. L., M. Buehner and M. De La Chevrotière, 2018: Using the hybrid gain algorithm to sample data assimilation uncertainty. *Q. J. R. Meteorol. Soc*., 1-22.

Houtekamer, P. L., X. Deng, H. L. Mitchell, S.-J. Baek and N. Gagnon, 2016: 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.

Hunke, E.C. and Lipscomb, W.H.. 2010. CICE: the Los Alamos sea ice model, documentation and software user's manual, Version 4.1. Los Alamos, NM, Los Alamos National Laboratory. (Tech. Rep. LA-CC-06-012.)

Lin, H., and G. Brunet, 2009: The influence of the Madden-Julian Oscillation on Canadian wintertime surface air temperature. *Mon. Wea. Rev*., 137, 2250-2262.

Lin, H., G. Brunet, and R. Mo, 2010a: Impact of the Madden-Julian Oscillation on wintertime precipitation in Canada. *Mon. Wea. Rev*., 138, 3822-3839.

Lin, H., G. Brunet and J. S. Fontecilla. 2010b. Impact of the Madden-Julian Oscillation on the intraseasonal forecast skill of the North Atlantic Oscillation. *Geophys. Res. Lett*., 37, DOI: 10.1029/2010gl044315.

Lin, H., N. Gagnon, S. Beauregard, R. Muncaster, M. Markovic, B. Denis, and M. Charron, 2016: GEPS based Monthly Prediction at the Canadian Meteorological Centre. *Mon. Wea. Rev*., DOI: http://dx.doi.org/10.1175/MWR-D-16-0138.1.

Madden, R.A. and P.R. Julian, *Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific*. Journal of the atmospheric sciences, 1971. **28**(5): p. 702-708.

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.

Smith, G.C., F. Roy, M. Reszka, D. Surcel Colan, Z. He, D. Deacu, J.-M. Belanger, S. Skachko, Y. Liu, F. Dupont, J.-F. Lemieux, C. Beaudoin, B. Tranchant, M. Drévillon, G. Garric, C.-E. Testut, J.-M. Lellouche, P. Pellerin, H. Ritchie, Y. Lu, F. Davidson, M. Buehner, M. Lajoie and A. Caya, 2016: Sea ice Forecast Verification in the Canadian Global Ice Ocean Prediction System. *Quart. J. Roy. Met. Soc.*, 142, 659–671, doi: 10.1002/qj.2555.

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.

Titchner, H. A., and N. A. Rayner (2014), The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, *J. Geophys. Res. Atmos., 119,* 2864-2889, doi: 10.1002/2013JD020316.

Vitart, F, 2017: Madden-Julian Oscillation prediction and teleconnections in the S2S database, *Quart. J. Roy. Met. Soc.*, 143, 2210-2220, doi: 10.1002/qj.3079.

Waliser, D. E., K.-M. Lau, W. Stern, and C. Jones, 2003: Potential predictability of the Madden-Julian Oscillation. *Bull. Amer. Meteor. Soc*., 84, 33-50.

Wheeler, M., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Mon. Wea. Rev*., 132, 1917-1932.

Zuo, H., Balmaseda, M. A. and Mogensen, K., 2015. The new eddy-permitting ORAP5 ocean reanalysis: description, evaluation and uncertainties in climate signals. Climate Dynamics, 10.1007/s00382-015-2675-1.

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
CIS	Canadian Ice Service
CICE	Community of Ice CodE
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
GIOPS	Global Ice Ocean Prediction System
GOSSIP	Globally Organized System for Simulation Information Passing
GPS-RO	Global Positioning System - Radio Occultation data
Had2CIS	HadISST2 combined with CIS digitized sea ice charts
HadISST2	Hadley Centre Sea Ice and Sea Surface Temperature Version 2
IASI	Infrared Atmospheric Sounding Interferometer
IIEE	Integrated Ice Edge Error
LAM	Limited area model
MJO	The Madden-Julian Oscillation
MSLP	Mean sea-level pressure
NEMO	Nucleus for European Modelling of the Ocean
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
ORAP5	Ocean Reanalyses Prototype 5
ORAS5	Ocean Reanalysis System 5
QPF	Quantitative precipitation forecast
RDPS	Regional Deterministic Prediction System

REPS	Regional Ensemble Prediction System
RMSE	Root-mean-squared error
SHEF	Standard Hydrological Exchange Format
SPS	Spatial Probability Score
SST	Sea Surface Temperature
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 6.0.0 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*