

Diagnosics of forecasts for polar regions

Linus Magnusson

*ECMWF, Shinfield Park, Reading
RG2 9AX, United Kingdom
linus.magnusson@ecmwf.int*

1 Introduction

European Centre for Medium-range Weather Forecasts (ECMWF) has produced global deterministic (i.e. single realization without uncertainty estimation, HRES hereafter) forecasts since 1981. Since 1992, an ensemble of forecasts is run at a lower resolution to estimate forecast uncertainty and provide probabilistic forecasts. Because all forecasts are using a global model, the polar regions are included, but the model is not specifically tuned for these areas. In this report, with a focus on the polar regions, an overview of the improvement of the forecast quality will be given and various aspects of the ensemble (ENS) will be discussed. A study on a similar topic was presented in [Jung and Leutbecher \(2007\)](#).

The polar regions are in many aspects different to lower latitudes. From the observational system point of view, the challenging environment lead to difficulties deploying in-situ based observations. This results in fewer observations in the data assimilation system as well as little data to verify the forecast results against. Regarding the atmospheric interactions, the connection between the troposphere and the stratosphere is believed to be stronger around the poles than elsewhere and the interaction between snow-cover, sea-ice and the atmosphere is unique for these regions.

The aim of ensemble prediction is to forecast uncertainties and provide a basis for probabilistic forecasts. The uncertainties in the forecasts originate from uncertainties in the data assimilation (analysis) together with uncertainties in the model formulation and the truncated model resolution. In order to simulate the forecast uncertainty, 50 perturbed forecasts are run. Each of them is initialised from a perturbed analysis to simulate the uncertainty in the initial conditions. The initial perturbations for the ECMWF ENS are generated from an ensemble of data assimilations ([Buizza *et al.*, 2008](#); [Isaksen *et al.*, 2010](#)) together with singular vectors ([Molteni *et al.*, 1999](#)). To simulate the model uncertainty, Stochastically Perturbed Physical Tendencies (SPPT) scheme is used together with a stochastic backscatter scheme ([Palmer *et al.*, 2009](#)).

The way the differences between the perturbed forecasts evolve in time depends on the flow in the atmosphere. In situations of high predictability, the difference between the ensemble members is expected to grow slowly with time, while uncertain conditions should result in a large spread among the ensemble members. As a first order measure of the reliability of the simulated uncertainties, the standard deviation of the ensemble should match, on average, the root-mean-square-error (RMSE) of the ensemble mean. However, this assumes that the random component of the forecast error dominates over the systematic component (bias).

To simulate the uncertainties correctly, one needs to include all components that can contribute to the diversity among the ensemble members. If the model system does not include e.g a sea-ice model, the uncertainties in the atmosphere due to the sea-ice evolution will not be captured. Furthermore, if the model lacks variability in a component, the uncertainty will be underestimated. Another important aspect is that the ensemble system can only simulate uncertainties larger than the grid box scale. In other words, small scale variability inside the box will not be captured as long as the value from the model

is supposed to represent the entire grid box. The latter issue is especially a problem near the ground surface if the grid box is inhomogeneous in properties such as land-sea, orography and snow cover.

The data in this report are based on the ECMWF IFS forecasting system. The scores are mainly based on forecasts from 2012, when the HRES used spectral resolution T_L1279 (16 km) with 91 vertical levels, and the ENS T_L639 (32 km) with 62 vertical levels. We will also use the control forecast for ENS, which is an unperturbed forecast run with the same resolution as ENS.

In Section 2 the general predictability of weather in the polar regions and its development during the past 25 years. Even though the forecasts have improved in general, the system occasionally produces bad forecasts and such a case will be evaluated in Section 3. In section 4 the ability of simulating the uncertainties in the Arctic are evaluated. The effect of satellite observations on the EDA in the polar regions is presented in Section 5 and the properties of the SPPT scheme in the Arctic is discussed in Section 6. Finally the results are summarised in Section 7.

2 Predictability in the Polar regions

In this report we define the Arctic (N.Pole hereafter) as north of 65°N and the Antarctic as south of 65°S (S.Pole hereafter). The northern hemisphere (N.Hem) is defined as 20°N - 90°N and southern hemisphere (S.Hem) as 20°S - 90°S (which both exclude the tropics).

Figure 1 shows the root-mean-square-error (RMSE) for ENS control forecasts (red) and HRES forecast (blue) for the 500 hPa geopotential height (z500) from 2012. We compare results for the polar regions (solid) with the hemispheric results (dashed). First of all, the impact of model resolution is small (control and HRES have very similar RMSE). Comparing the N.Hem and N.Pole, the difference in the short-range error (around day 2) is small, while they are somewhat larger for S.Pole than S.Hem for the same lead time. For both poles, the saturation level of the errors is higher for the polar regions than for the hemispheres (N.Hem and S.Hem). The difference is largest for northern hemisphere where the atmospheric activity is higher for N.Pole compared to N.Hem (not shown). For S.Pole the major part of the difference is found to be due to systematic model errors (model drift).

Figure 2 shows the time-series of errors for short-range (2-day) forecasts (upper panels) and medium-range (6-day) forecasts. The plots show data for the HRES forecasts evaluated against ERA-Interim analyses. The reason for verifying against the ERA Interim reanalysis instead of the operational analysis is to minimise the effect of shared errors between the forecast and analysis system. The figure corresponds to Figure 3 in [Magnusson and Källén \(2013\)](#), which shows similar results for N.Hem.

For the short-range forecast errors, we see a clear decrease between 1997 and 2001. During this period important upgrades of the data-assimilation took place, by the introduction of 4D-Var in the end of 1997 and subsequent changes in the observation usage in the data assimilation ([Simmons and Hollingsworth, 2002](#)). One important change here was the upgrade of the usage of raw microwave radiances from the TOVS and ATOVS satellite-borne instruments in 2000, which especially seems to affect S.Pole-results.

To visualise the daily variability of the forecast quality, Figure 3 shows times-series of daily RMSE values for z500 from 2012 for N.Pole(a) and S.Pole(b). The data are for 6-day control forecasts. For both poles a considerable variability in the scores is present, showing that the predictability is changing from day-to-day also in the polar areas. We also see a seasonal cycle in the errors, which is most pronounced for the S.Pole where the errors are lowest during Nov-Dec-Jan (summer). Occasionally spikes are present in the errors, indicating bad forecasts. For Europe, such a case was investigated in [Rodwell *et al.* \(2012\)](#). In the next section the case with the highest errors for S.Pole during 2012 will be investigated further.

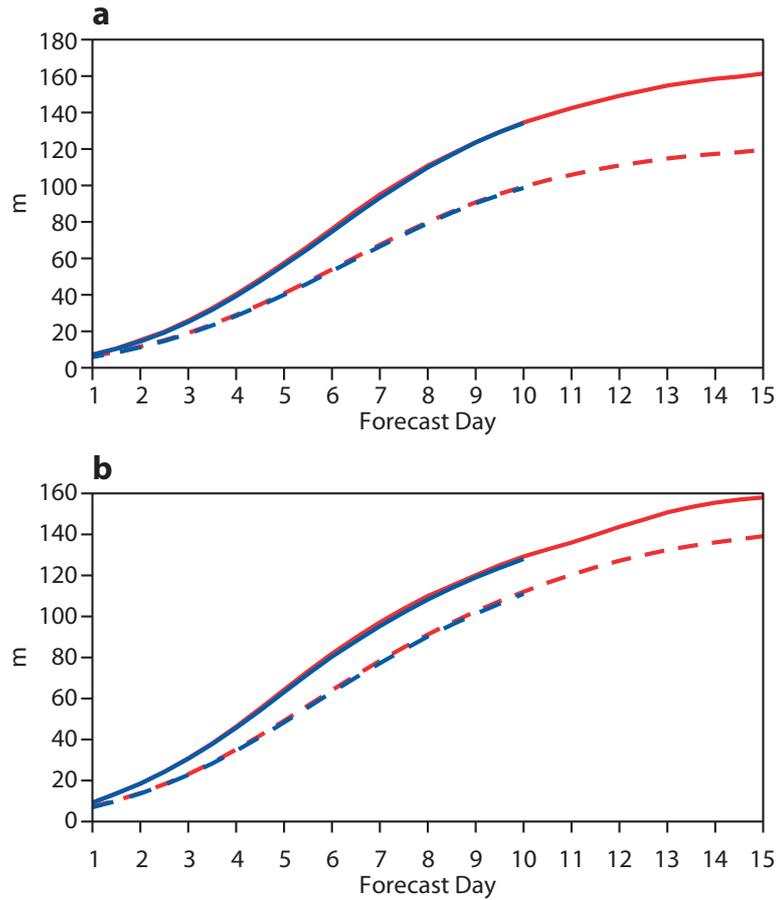


Figure 1: RMSE for z500 for ENS control forecasts (red) and HRES (blue) from 2012. In the upper panel the scores are for N.Pole (solid) and N.Hem (dashed) and in the lower panel for S.Pole (solid) and S.Hem (dashed).

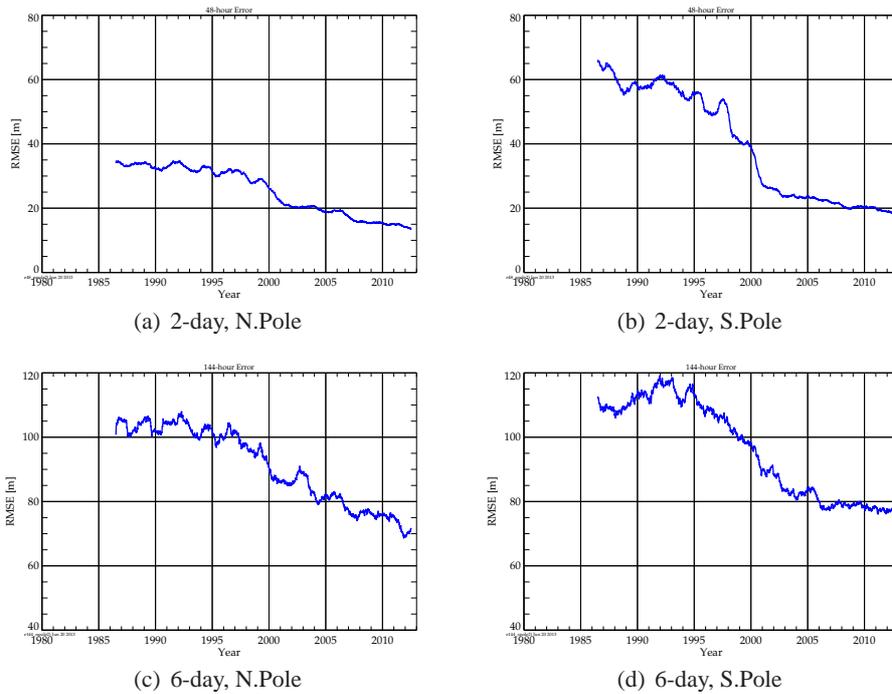
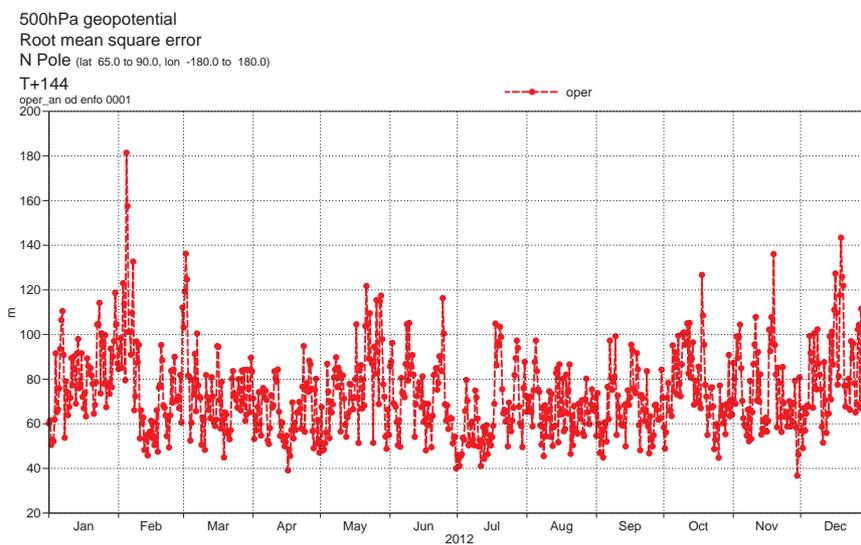
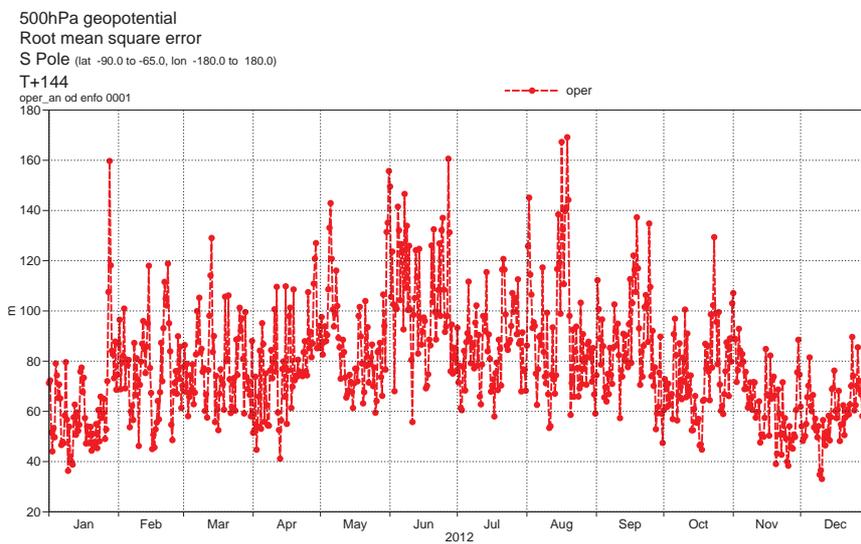


Figure 2: Evolution of RMSE for HRES forecasts from 1986 to 2013 for z500 over the polar regions.



(a) N.Pole



(b) S.Pole

Figure 3: Daily RMSE of z500 for 6-day control forecasts from 2012.

3 Example of a problematic forecast for the Antarctic

Figure 4(a) shows the same as Figure 3(b), but only for August 2012. Between the 15 August and 19 August the forecasts resulted in a period of large errors. In Figure 4(b) a map of the error (forecast-analysis) for the 6-day forecast for the 19 August 00 UTC is plotted. A large part of the Antarctic experienced a negative error, while the error equatorward of 70°S had a positive sign. This indicates that it is not a single synoptic weather system causing the major part of the error but rather a more large-scale feature or a combination of both.

Figure 5 shows the observed time-series of the mean z500 south of 70°S. Here a rapid increase in the geopotential height after 20 August appears, with a peak on 29 August. In the figure all ensemble members from 19 August 00 UTC are plotted (red, thin) together with the ensemble mean (blue, thick). Most of the members did not capture the increased geopotential height, causing a negative error. However, we find a few members that at least partially captured the development in the atmosphere. In the lower panel the 2 best and the 2 worst members are plotted and will be analysed further.

The panels in Figure 6 show maps of errors for the best and the worst ensemble member from 19 August 00 UTC. Here we see a large difference in the large scale, where the worst member has the negative error over the Antarctic, while the best member does not. However, the best member has still large errors in synoptic scale structures.

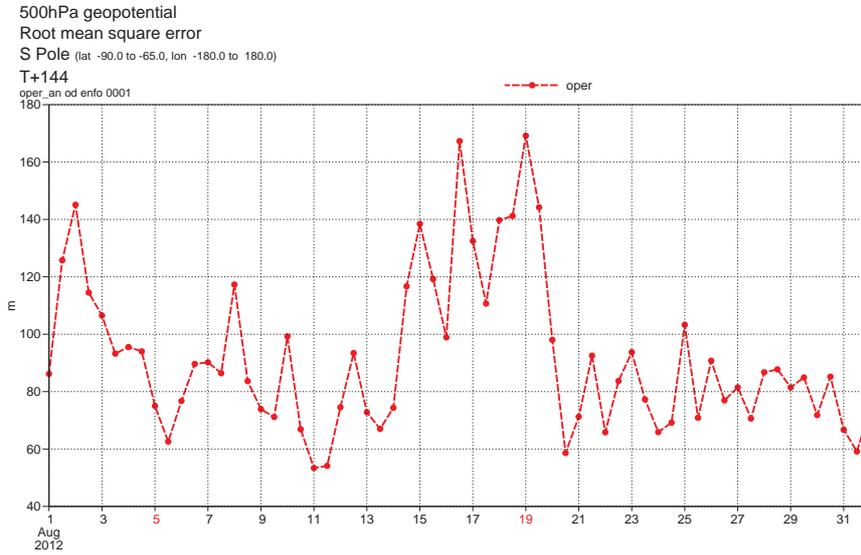
The large-scale structure of the error for this case points to an external forcing. Such a candidate is the stratosphere. Anomalous evolutions in the stratosphere over Antarctica have been investigated in the context of ECMWF forecasts in [Simmons *et al.* \(2005\)](#). In order to investigate the evolution of the stratosphere, a Hovmöller diagram of the temperature anomaly from ERA Interim is plotted in Figure 7. Here we see a warm anomaly in the stratosphere, which starts to propagate downwards after 13 August. One can suspect that this stratospheric warming influenced the troposphere. Figure 8 shows the evolution for the best and worst ensemble members for the temperature at 50 hPa. Here both the best and worst members captured the evolution reasonably and, if anything, the best members broke down the warming event too early. So one could speculate that the failing development is due to the connection between the stratosphere and the troposphere in the model.

These results suggest that the error was because the downward propagation of a stratospheric warming event was not well captured, although we cannot rule out other sources for the error. Most of the ensemble members did not capture the evolution in the troposphere, but a few ensemble members did. A deeper investigation is needed to see what aspects of the perturbations caused the event to be better captured. In general, [Vitart \(2013\)](#) showed that the current version of the ECMWF model has a too weak connection between the stratosphere and the troposphere for the Arctic. This could be the case also for the Antarctic. However, more cases of large errors needs to be investigated to see if they normally originate from the stratosphere.

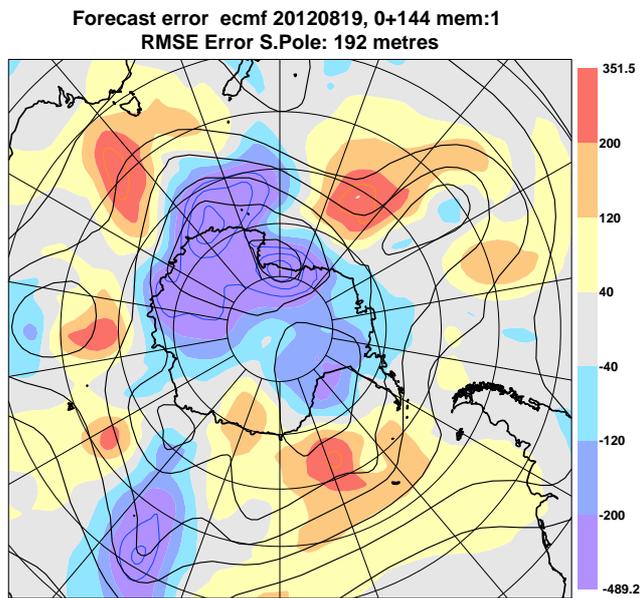
4 Simulating uncertainties in the Arctic

Figure 9 shows the ensemble mean RMSE and the ensemble standard deviation (hereafter referred to as ensemble spread) for boreal winter forecasts from 2012. For a perfect ensemble, these quantities should match each other for all lead times (see [Leutbecher and Palmer \(2008\)](#)). For the temperature at 500 hPa over the Arctic (Figure 9(a)), the relation holds well for all lead times. Further down in the troposphere (850 hPa, Figure 9(b)), the ensemble is somewhat under-dispersive. However, in this diagnostic we have not accounted for errors in the verification data set ([Saetra *et al.*, 2004](#)). By accounting for these errors, it is possible that the ensemble turns out to be somewhat over-dispersive for 500 hPa temperature.

Figure 9(d) shows the RMSE (solid) and ensemble standard deviation (dashed) for 2-metre temperature

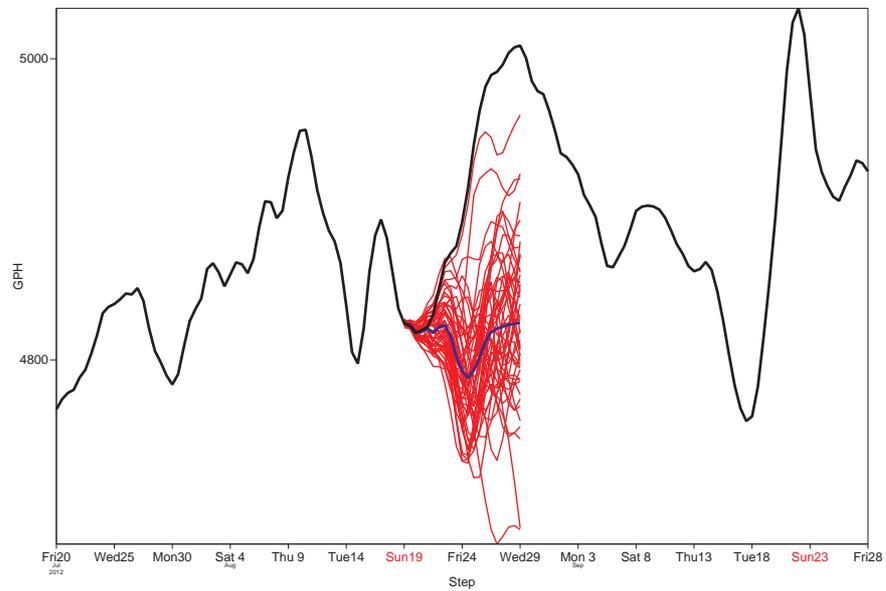


(a) Daily RMSE of z500, 6-day control forecasts of z500 from August 2012, S.Pole. Initial date on the x-axis.

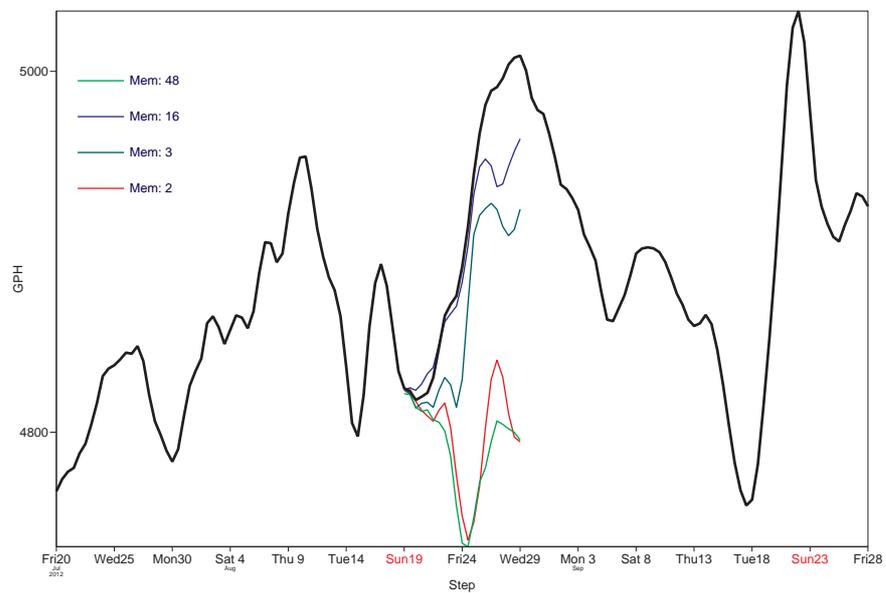


(b) Map of the error (fc-an) for the 6-day of z500 forecast from 19 August 2012 00 UTC. The distance between the circles are 10°.

Figure 4: Times-series of RMSE for August 2012 and maps of the error for day 6 forecast initialised 19 August 2012 00 UTC. Both panels for z500 and the control forecast.

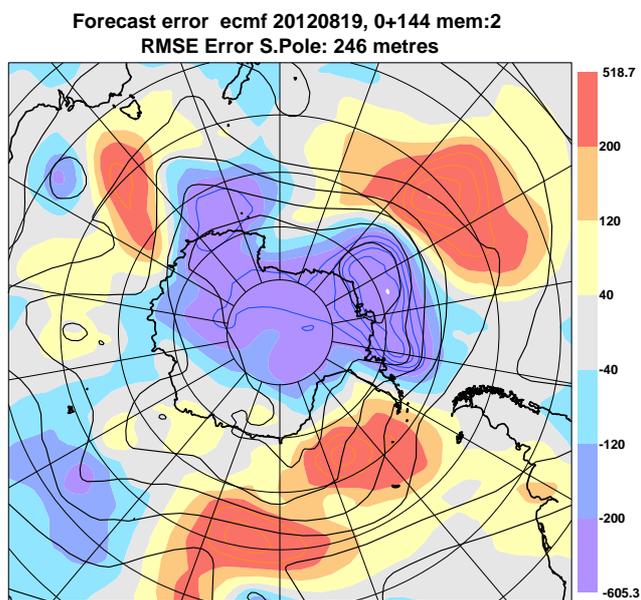


(a) All ensemble members (red, thin), ensemble mean (blue, thick) and analysis (black)

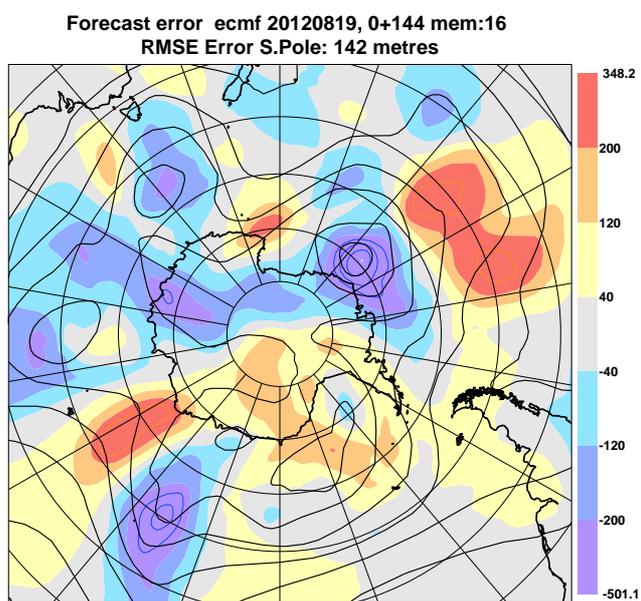


(b) The 2 best and worst members

Figure 5: Forecast of mean z_{500} as a function of date for $70^{\circ}S-90^{\circ}S$, initialised 19 August 2012 00 UTC.



(a) Worst member (number 2)



(b) Best member (number 16)

Figure 6: Map of z_{500} -error for the best and worst member after 6 days for forecasts initialised 19 August 2012 00 UTC.

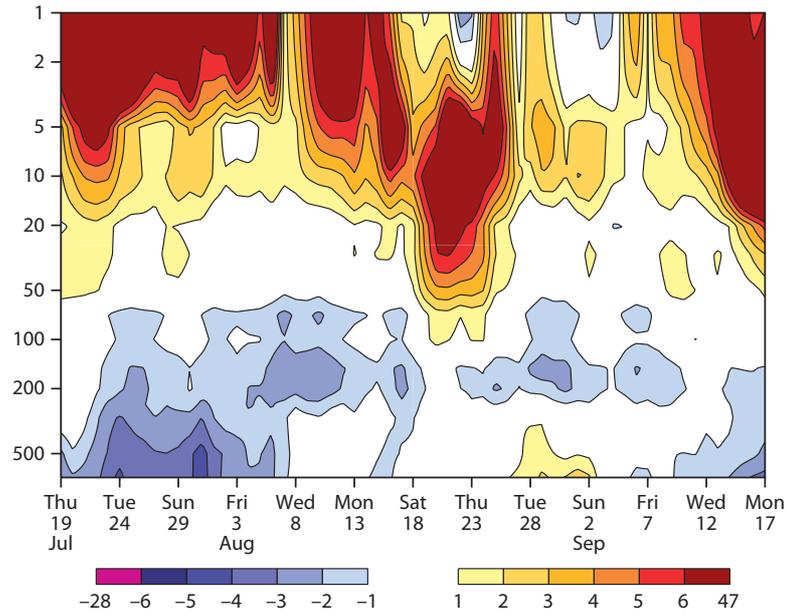


Figure 7: Hovmöller diagram of the mean temperature anomaly from ERA Interim, south of 70°S. X-axis represents time and y-axis pressure.

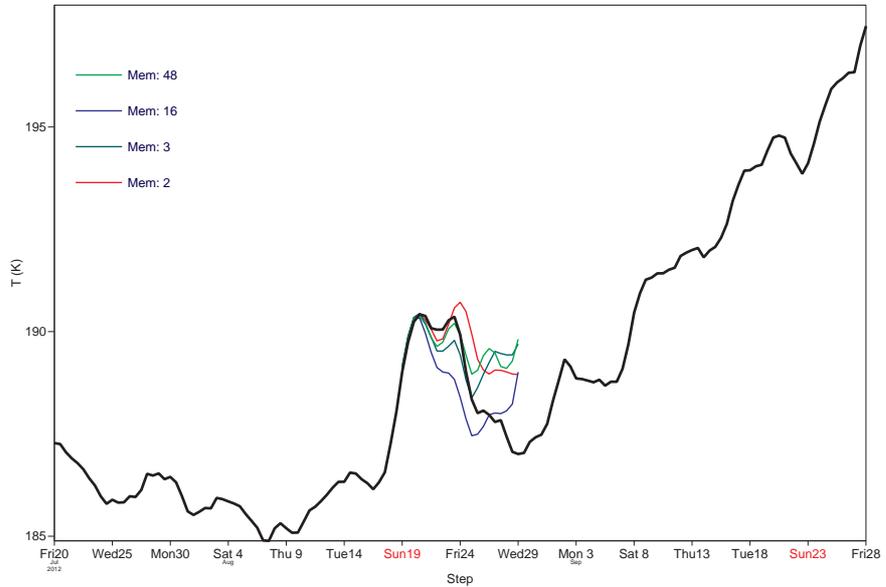


Figure 8: Mean temperature for 50 hPa south of 70°S as a function of date. Analysis (black) and same ensemble members as in Figure 6(b).

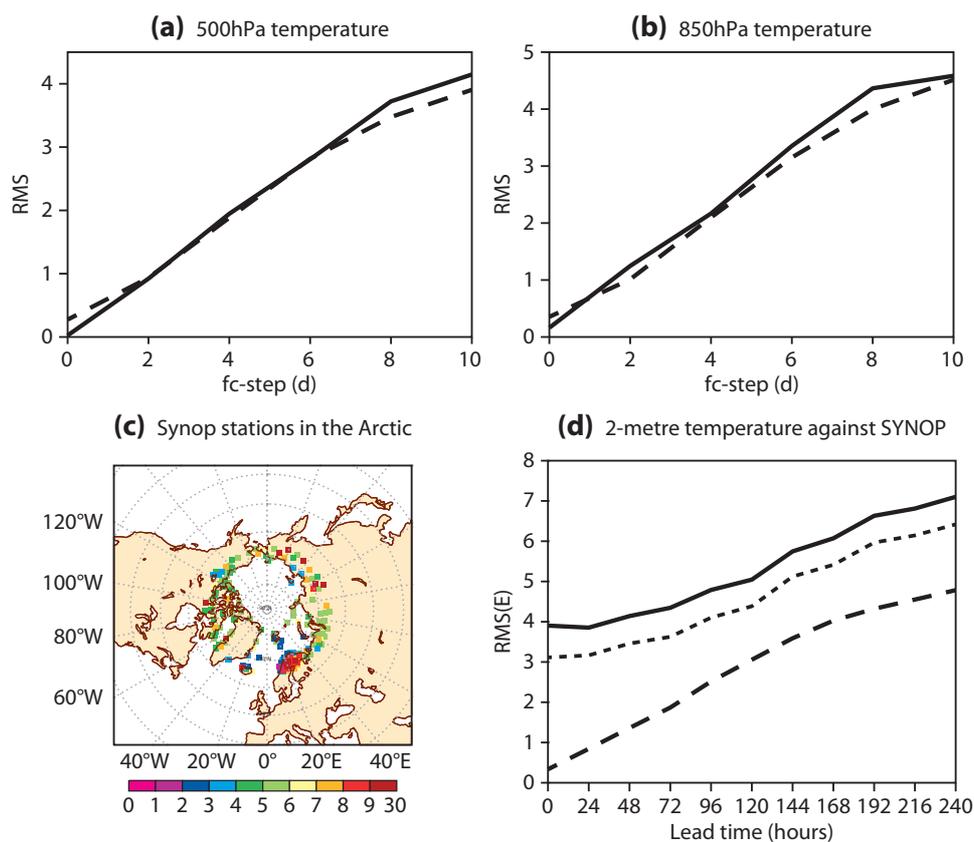


Figure 9: RMSE (solid) and ensemble spread (dashed) for the Arctic (65-90N). For the SYNOP verification, the standard deviation of the error is included (dotted).

verified against SYNOP observations. The observation network used here is shown in Figure 9(c). For the 2-metre temperature there is a large difference between the error and the spread. While the amplitude of the spread has the same magnitude as for 850 hPa, the error is much larger. Several factors can play a role here. Firstly, systematic model errors will increase the error level. For example the cloud modelling is a well-known source of uncertainty (see e.g. Svensson and Karlsson (2011); Karlsson and Svensson (2011)). In order to remove the effect of model bias, the standard deviation of the error is included in Figure 9(c).

Another source of differences between observations and model data is sub-grid variability. The sub-grid variability could be due to small-scale weather features such as convective cells, but also to sub-grid variability in the boundary conditions. Examples of such variabilities are orography, land-sea-lake contrasts and snow conditions. For the 2-metre temperature in the polar areas these three items constitute a large source of variability, especially considering that many of the observation stations are located either close to the sea or in low level terrain, affected by strong inversions in winter. To illustrate the problem, Figure 10 shows a 3-day HRES forecast from 7 February 2013 00 UTC and corresponding observations of 2-metre temperature for the stations Tarfala and Nikkalouta in northern Sweden. The stations are separated by 17 kilometres, but while Tarfala is located far up a steep, down-sloping valley, Nikkaluota is located in the bottom of a gentle valley. For Tarfala the temperature forecast is in good agreement with the observations, while for Nikkalouta the forecast is about 20°C too warm. The forecasts for both stations are very similar. The large difference in the observed values for Nikkalouta is due to a strong inversion. The inversion temporary broke up at midday on the 8 February and the temperatures became higher than the forecast. To address the issue with strong local inversions, one needs either a much higher model resolution or a parametrisation of the sub-grid uncertainty. For an ensemble system, the latter is essential in order to catch the true forecast uncertainty.

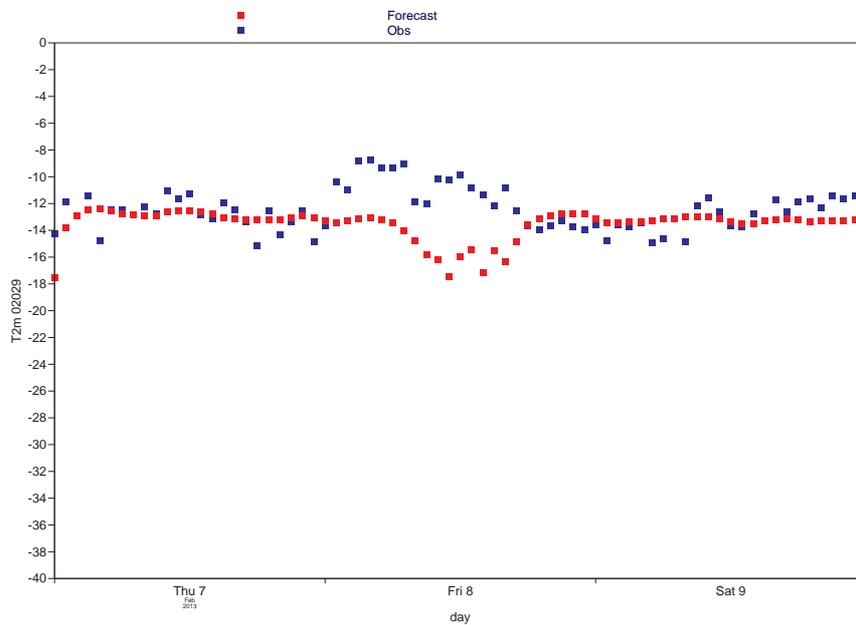
In this section we have pointed out some difficulties in diagnosing the spread in the ensemble system. It is hard to disentangle the the lack of spread in the ensemble system from systematic model errors and representativeness errors between the model grid and the observations. If one derives the uncertainties in the verification data set, that component could be accounted for following Saetra *et al.* (2004).

5 Impact of polar orbiting satellites on EDA standard deviation

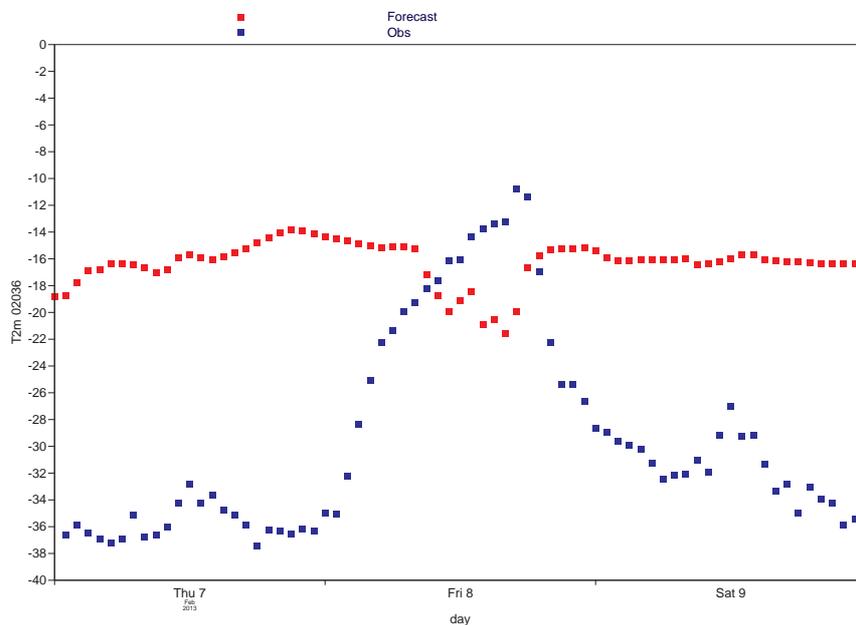
Since June 2010, ECMWF is operationally using an ensemble of data assimilations (EDA) to generate a part of the initial perturbations for the ensemble forecasts and since May 2011 to scale the background error variances for the data assimilation system (Buizza *et al.*, 2008; Isaksen *et al.*, 2010). The concept of the EDA is illustrated in Figure 11. The EDA consists of 10 independent 4DVAR data-assimilation cycles, which uses observations perturbed according to observation uncertainty. For the forecasts between the assimilations (first guess forecast), the model uncertainty is simulated by the SPPT scheme (Palmer *et al.*, 2009).

Although the observations are perturbed, their presence should reduce the dispersion between the EDA members to realistic levels. The trivial example is when we do not have any observations to assimilate, which should lead to a dispersion as large as the climatological variability after a number of data assimilation cycles.

As discussed in Section 2, the introduction of assimilation of polar orbiting satellites clearly reduced the forecast error in the polar regions. The effects of such data are documented in McNally (2006); Andersson (2006). In this section we investigate the impact on the EDA standard deviation (spread). For this purpose, an EDA experiment without data from polar orbiting satellites (but still using MODIS AMV) was run between 10 October 2012 and 11 November 2012 (hereafter referred to as NoPol). The experiment has been evaluated for the impact on hurricane Sandy in McNally *et al.* (2013); Magnusson *et al.*



(a) Tarfala



(b) Nikkaluota

Figure 10: Forecasts (from HRES) and observations of 2-metre temperature.

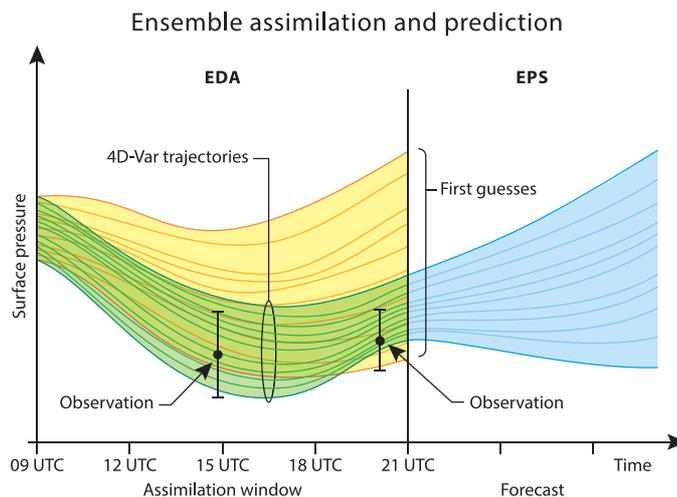


Figure 11: Concept of Ensemble of data assimilations.

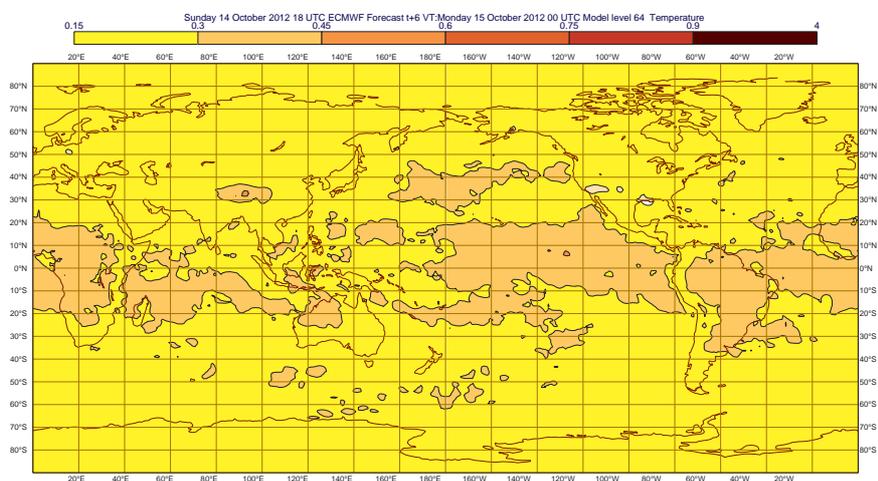
(2013). Optimally, as the analysis error and forecast error increase with less data, the ensemble standard deviation should do as well.

Figure 12 shows the EDA spread for the temperature at 500 hPa, for the operational EDA (top panel) and the NoPol experiment (bottom panel). Without the data from the polar orbiting satellites the spread increases over the oceans outside the tropics (over the tropics the data from geostationary satellites are dominating). The largest difference is present in the southern hemisphere where the spread increases by more than 3 times over some areas. The difference is much less pronounced over the Arctic. The less difference in the Arctic could be due to more conventional observations in that region, compared to Antarctica.

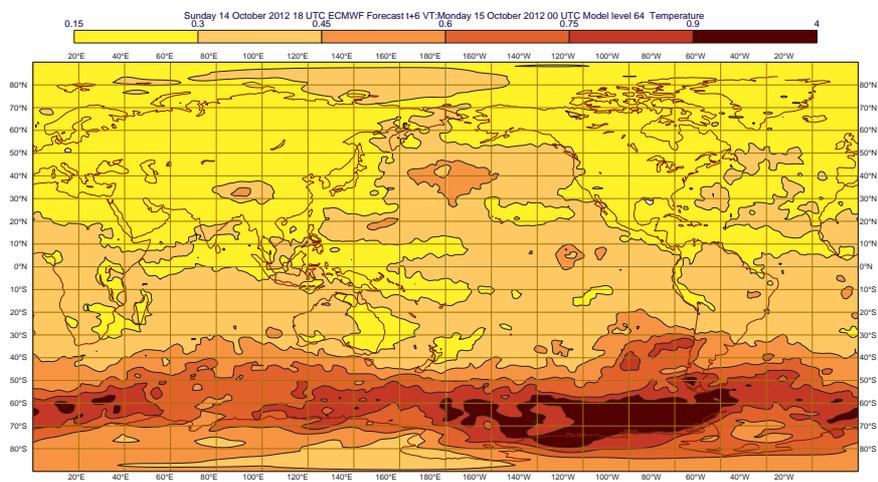
Figure 13 shows the verification of an ensemble experiment using EDA perturbations from the experiment discussed above. The experiment is initialized from HRES analysis that also had the polar orbiting satellite data omitted. 8 ensemble forecasts have been run with start dates from 2012-10-21 to 2012-10-28. One caveat with these results is the limited sample. The figure shows the RMSE of the ensemble mean (left panels) and the ensemble spread (right panels) for N.Pole (upper panels) and S.Pole (lower panels). Together with the NoPol ensemble experiment (blue) a control experiment initialised from a HRES analysis and EDA perturbations using all observations (red) is plotted. The forecasts are verified against the operational HRES analysis. For both polar regions the forecast error increased without the satellite data. The largest change is seen for S.Pole where the 2-day error is more than doubled. For the ensemble spread for S.Pole, we see a similar increase, which is a clear sign that the simulation of the forecast uncertainty is capturing the increased forecast error caused by the loss of data. This is not apparent for N.Pole, where the ensemble spread is similar for the two experiments, although the forecast error increased. One reason for this could be the more complex observation system over the Arctic (more observations), which makes the EDA spread more sensitive to the tuning of the errors from different types of observations.

6 SSPT scheme in the polar areas

To simulate model errors, the ensemble prediction system at ECMWF uses the Stochastically Perturbed Physics Tendency (SPPT) scheme together with the stochastic backscatter scheme (Palmer *et al.*, 2009). The SPPT scheme perturbs the tendencies from the physics schemes in the model, which includes the convection, cloud, radiation, vertical diffusion and dissipation. Together with the dynamics scheme (mainly advection), these tendencies give the evolution of the forecast during the integration of the



(a) Operational



(b) Without polar orbiting sat.

Figure 12: Ensemble spread for 500 hPa temperature for 6-hour EDA forecasts with and without polar orbiting satellites, based on forecasts with start dates from 15-30 October 2012, twice a day.

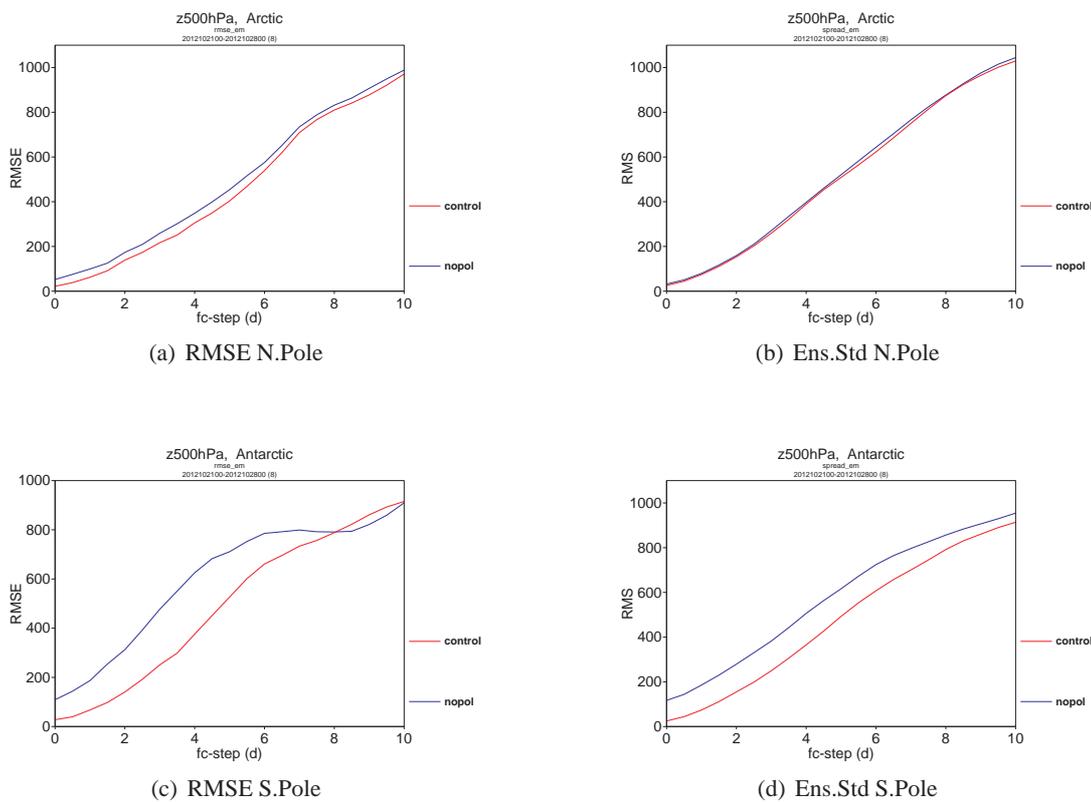


Figure 13: Average RMSE of ensemble mean and ensemble spread for 500 hPa geopotential height for experiments with and without polar orbiting satellites. Forecasts initialised 21-28 October 2012, 0 UTC.

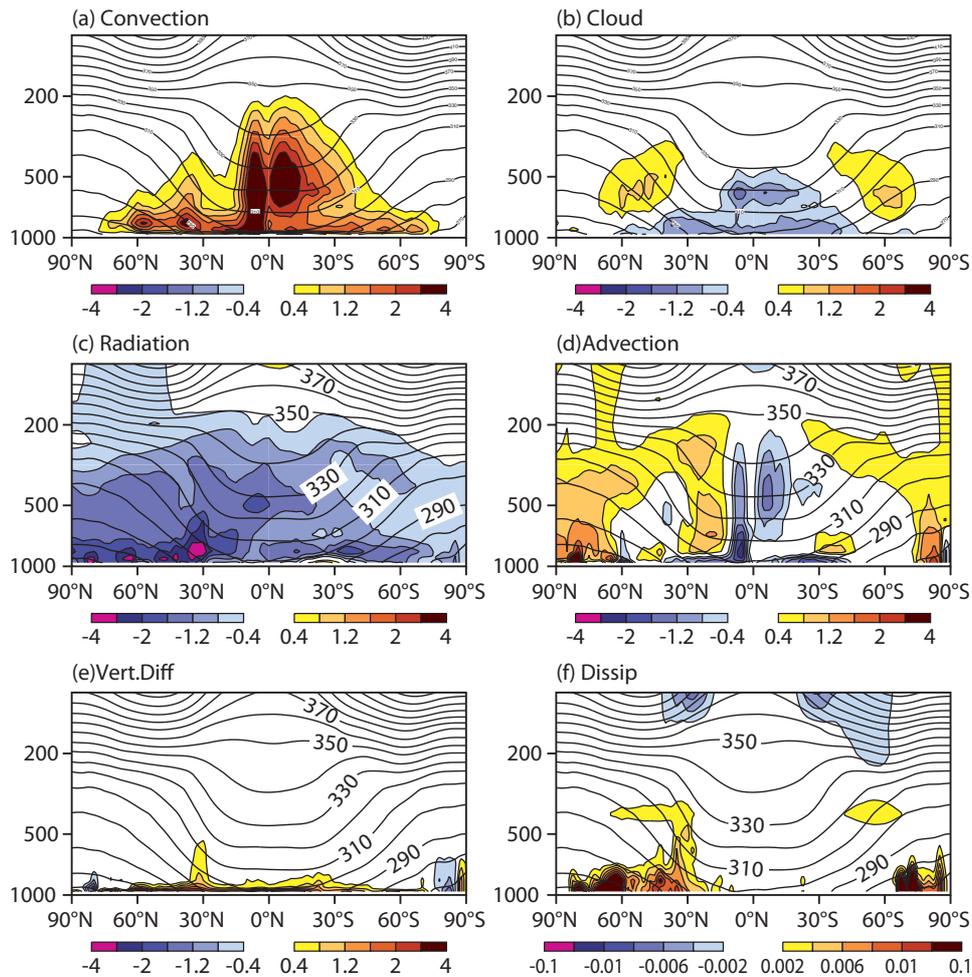


Figure 14: Zonal average of mean temperature tendencies from forecast day 0 to 5 (mean from 7 forecasts).

model. By scaling the physical tendencies by a random number the forward integration of the model is slightly changed. The random numbers are generated by a pattern simulator in order to include a spatial scale to the perturbations. The random numbers are also correlated in time. For the boundary layer and the stratosphere no perturbations from the SPPT scheme are used.

Figure 14 shows vertical cross-sections of the mean tendencies from forecast day 0 to 5 for each component, contributing to the temperature evolution. For the tropics, the dominating tendencies are a warming by release of latent heat in convection balanced by vertical advection of air and the evaporation of clouds.

For the polar regions the dominant process in the free troposphere is the radiative cooling, which is compensated by advection (both horizontal and vertical). By the design of the SPPT scheme, the radiation tendencies will be the dominant contributor for the SPPT perturbations over the Arctic. One could argue that this process is well understood, and the uncertainty is low in the process in the free atmosphere during the Arctic night (by the same argument the SPPT scheme is switched off in the stratosphere). For the Arctic the largest model uncertainties are constrained to the boundary layer, which is not yet perturbed by the SPPT scheme. It is plausible that the resulting lack of model perturbations lead to an under-dispersive ensemble close to the surface (together with other sources of uncertainty such as sea-ice), as seen in Figure 9.

7 Discussion

In this report various aspects of ECMWF forecasts in the polar regions have been discussed. Since the operational forecasts started in 1986, the 6-day root-mean-square error (RMSE) has been reduced by about 30 % for the polar regions. For short-range (2-day) forecast errors the reduction is much larger for the southern hemisphere (about 60 %). The major part of the reduction took place between 1997-2001 and is believed to be due to improvements in the data assimilation leading to a more efficient use of satellite observations.

Even though the forecasts have been improved over the years, large errors appear occasionally. In this report the event with the highest errors over Antarctica during 2012 was investigated. The error seems to coincide with a sudden stratospheric warming event. To see if this is a general source of large errors, more cases need to be investigated. However, Vitart (2013) points to a too weak link between sudden stratospheric warmings and the tropospheric development (in the Arctic). This is a process that needs more diagnostic work to understand and to monitor future model development.

The ensemble is designed for the purpose of simulating the uncertainties in the forecasts. The ability to simulate the uncertainties has been evaluated for the polar regions by comparing the RMSE of the ensemble mean and the standard deviation of the ensemble (ensemble spread). Optimally these two quantities should match. However, this assumes that the model bias is negligible and that the uncertainty in the validation data set is small. This may be reasonably true in the free atmosphere over well observed areas. For the Arctic, we see a good match between the quantities for the free atmosphere (t500 and t850 verified against the analysis), but a large difference when we verify 2-metre temperature against SYNOP observations. In the Arctic during winter-time, the local variability in temperature due to strong inversions can be large. Hence, more work is needed to quantify the sub-grid variability. Without such an estimate it is hard to decide whether the ensemble is under-dispersive or not.

For the EDA, the impact of assimilating observations from polar orbiting satellites on the ensemble spread has been investigated. By reducing the number of observations, the error in the analysis is increased as expected. This seems to be well simulated by the EDA for the Antarctic, while for Arctic the ensemble spread did not increase to the same degree as the error. The presence of more conventional observations in the Arctic makes the data assimilation system more complex, and more sensitive to the tuning of the observation uncertainty.

Regarding the SPPT scheme, we recognise the problem with tapering of the perturbations in the boundary layer, where we have the largest uncertainties in the polar regions. Instead the perturbations originate mainly from the radiation tendency in the free atmosphere, a process that is less uncertain. Therefore more development should be undertaken aiming to perturb the model in the boundary layer and also include the surface modelling.

In this report some key areas of future diagnostics regarding the polar areas have been highlighted, such as the ability to obtain the correct strength in the teleconnection from the stratosphere to the troposphere and the need for a good estimate for the representativeness error. We also highlighted the impact of satellite observations in the EDA and the possibility to run data denial experiments for the EDA. Regarding the ensemble system design, the SPPT scheme is going to be revised to better target the uncertainties in the polar regions in terms of radiation and boundary layer processes, and in the autumn of 2013, initial perturbations of the surface variables will be introduced that will potentially affect the polar areas (at least on the edge of the snow cover). Finally, the uncertainty caused by sea-ice cover is not represented in the ensemble today, hence the plans for the future also includes introduction of a dynamic sea-ice model in the prediction system.

Acknowledgements

We would like to acknowledge Trond Iversen, Peter Bauer, Roberto Buizza, Massimo Bonavita, Simon Lang, Sarah Keeley, David Richardson, Thomas Haiden and many other for valuable discussions and providing material for this report and Anabel Bowen for help with the preparation of the figures.

References

- Andersson E. 2006. Data assimilation in the Polar Regions. In: *Seminar on Polar Meteorology, 4-8 September 2006, ECMWF*. pp. 89–102.
- Buizza R, Leutbecher M, Isaksen L. 2008. Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Q. J. R. Met. Soc.* **134**: 2051–2066.
- Isaksen L, Bonavita M, Buizza R, Fisher M, Haseler J, Leutbecher M, Raynaud L. 2010. Ensemble of Data Assimilations at ECMWF. Technical Memorandum 636, ECMWF.
- Jung T, Leutbecher M. 2007. Performance of the ECMWF forecasting system in the Arctic during winter. *Q. J. R. Met. Soc.* **133**: 1327–1340.
- Karlsson J, Svensson G. 2011. The simulation of Arctic clouds and their influence on the winter present-day climate in the CMIP3 multi-model dataset. *Clim. Dyn.* **36**: 623–635.
- Leutbecher M, Palmer TN. 2008. Ensemble forecasting. *J. Computational Physics* **227**: 3515–3539.
- Magnusson L, Källén E. 2013. Factors influencing skill improvements in the ECMWF forecasting system. *Mon. Wea. Rev.* **141**: 3142–3153.
- Magnusson L, Thorpe A, Bonavita M, Lang S, McNally T, Wedi N. 2013. Evaluation of forecasts for hurricane Sandy. Technical Memorandum 699, ECMWF.
- McNally T. 2006. The use of satellite data in Polar Regions. In: *Seminar on Polar Meteorology, 4-8 September 2006, ECMWF*. pp. 103–114.
- McNally T, Bonavita M, Thépaut JN. 2013. The Role of Satellite Data in the Forecasting of Hurricane Sandy. Technical Memorandum 696, ECMWF.
- Molteni F, Buizza R, Palmer T, Petroliagis T. 1999. The ECMWF Ensemble Prediction System: Methodology and validation. *Q. J. R. Met. Soc.* **122**: 73–119.
- Palmer T, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts G J, Steinheimer M, Weisheimer A. 2009. Stochastic parametrization and model uncertainty. Technical Memorandum 598, ECMWF.
- Rodwell MJ, Magnusson L, Bauer P, Bechtold P, Bonavita M, Cardinali C, Diamantakis M, Earnshaw P, Garcia-Mendez A, Isaksen L, Källén E, Klocke D, Lopez P, McNally T, Persson A, Prates F, Wedi N. 2012. Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer. Meteor. Soc.* **Accepted**.
- Saetra O, Hersbach H, Bidlot JR, Richardson DS. 2004. Effects of Observation Errors on the Statistics for Ensemble Spread and Reliability. *Mon. Wea. Rev.* **132**: 1487–1501.
- Simmons A, Hollingsworth A. 2002. Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Met. Soc.* **128**: 647–677.

- Simmons A, Hortal M, Kelly G, McNally A, Untch A, Uppala S. 2005. ECMWF Analyses and Forecasts of Stratospheric Winter Polar Vortex Breakup: September 2002 in the Southern Hemisphere and Related Events. *J. Atm. Sci.* **62**: 668–689.
- Svensson G, Karlsson J. 2011. On the Arctic wintertime climate in global climate model. *J. Climate* **24**: 5757–5771.
- Vitart F. 2013. Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years. Technical Memorandum 694, ECMWF.