TECHNICAL MEMORANDUM

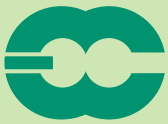
708

Spread of the ensemble of data assimilations in radiance space

Niels Bormann and Massimo Bonavita

Research Department

October 2013



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

This memorandum evaluates the spread of the EDA in radiance space against observation departures, with the aim to validate the use of the EDA spread as an estimate of the expected size of the background errors for use in the analysis quality control decisions. The investigated approach replaces similar estimates previously derived with a randomisation method.

It is found that the EDA spread shows good skill in identifying regions of larger observation departures, but it appears to be under-dispersive especially in the extra-tropics. Some inflation of the EDA spread in radiance space is hence required to account for this. After this calibration step, the EDA-based background error estimates and observation departure statistics show very consistent characteristics. The temporal evolution of the EDA spread over the subsequent assimilation window has also been investigated, with indications that the EDA spread does not grow as quickly as observation departures do.

The use of the EDA-based background error estimates in FG-check decisions has been investigated in assimilation experiments. Together with a re-tuning of the FG-check limit, the approach allows the use of more observations that were previously rejected.

1 Introduction

The prime deterministic assimilation system used at ECMWF has recently moved to a hybrid 4-dimensional variational data assimilation (4DVAR) approach (Isaksen et al. 2010, Bonavita et al. 2012). An Ensemble of Data Assimilations (EDA) is used to estimate situation-dependent aspects of the background error, and these are subsequently used in a standard higher-resolution 4DVAR. The EDA currently consists of a 10-member ensemble run at lower spatial resolution than the reference assimilation system, and for each member different perturbations are added to the observations, sea surface temperatures, and aspects of the forecast model. The variances resulting from the spread of the EDA are used in the deterministic high-resolution 4DVAR to provide "background errors of the day", adding a flow-dependent component to the background error specification. Flow-dependent correlations for the background error are planned for a future upgrade, but currently the error correlations are still static, specified on the basis of long-term statistics derived from an EDA. For the flow-dependent variances, a calibration step is included, that ensures that the estimates based on the EDA spread are consistent with typical short-term forecast errors as measured by comparing forecasts with analyses (Bonavita et al. 2012). In addition, spatial filtering is performed to address sampling noise arising from the use of relatively few members.

In this memo, the EDA is used to directly estimate background errors in radiance space, for use in quality control decisions for radiance assimilation. Knowledge of background errors in observation space is useful in the so-called First Guess (FG)-check which rejects observations that deviate further than expected from the FG, with rejection limits normally based on observation error and background error characteristics. The estimate of the background error investigated here is based directly on the EDA spread in radiance space, and replaces an estimate calculated on the basis of a randomisation approach (Fisher and Courtier 1995). The background error estimates from the randomisation field are further described in appendix 6, and they will be referred to as "EF" fields.

There are several motivations for replacing the current estimate of the background error in radiance space with a more direct estimate derived from the EDA spread: firstly, the current scheme uses a 500 hPa error growth model to model error growth for all radiances, including those primarily sensitive to water vapour, and this is considered unsatisfactory. Secondly, the randomisation method requires that more iterations are performed during the first minimisation than would be necessary to achieve convergence, adding to the computational cost of the analysis. The only use of the randomisation method is now for

the estimation of background errors in radiance space, so an alternative calculation of background errors in radiance space allows significant savings in computational cost during a time-critical component of the assimilation system. For observations of conventional geophysical variables, the estimate for the background error used in the FG-check has already been switched to a calibrated version of the EDA-spread.

A further motivation for basing a background error estimate for radiances directly on the EDA spread is that it allows further observation-based diagnostics on the reliability of the EDA spread for estimating background errors (e.g., Flowerdew and Bowler 2011). Estimates of the EDA spread in radiance space can be compared with the size of observation departures, and hence used to evaluate the performance of the EDA. This is particularly beneficial given the global coverage of polar satellite data. Currently, the calibration and validation of the EDA is done primarily against analysis-based estimates of the background error, but at the lead-times considered, analysis error can add a sizeable uncertainty to these statistics. Although diagnostics are not the primary focus of the work presented here, they are an insightful side-product.

The structure of this memo is as follows: first we introduce the calculation of the EDA spread in radiance space used in the investigated approach and provide an initial evaluation. Next we discuss spatial filtering introduced to account for sampling noise, and present the approach to calibrate the EDA spread using observation departures. We then present results from using the EDA spread in radiance space for quality control decisions, followed by an evaluation of the evolution of the EDA spread over the 12-hour assimilation window. Our conclusions are summarised in the last section. The investigated approach is applied to ATOVS and closely related data. Although the discussion primarily focuses on AMSU-A and MHS instruments, the results also apply to HIRS channels sounding in similar parts of the atmosphere.

2 EDA radiance spread

2.1 Calculation

The EDA radiance spread used in this memo is calculated in grid-point space from direct radiance calculations performed with RTTOV for all members of the EDA. It is calculated as the standard deviation of the radiance fields at the beginning of the subsequent assimilation window, and stored as “ES” type GRIB field in MARS. In theory, it would be possible to calculate the radiance spread for each observation in the subsequent assimilation window separately, based on all the EDA members and taking fully into account the viewing geometry and instrument differences between satellites. However, this would be computationally relatively costly and is therefore not chosen here. Of course, such calculations are performed as part of the EDA during the assimilation over the subsequent assimilation window. But these results are not yet available when the background error estimates in radiance space are needed for the FG-check in the deterministic 4DVAR system.

The EDA radiance spread is currently calculated for AMSU-A, MHS, and HIRS, and the scheme could be extended to other sensors. The radiance calculations in grid-point space are performed for all channels of one selected satellite only at one selected zenith angle. As outlined in appendix 6, similar choices have been made for the EF-field which is also computed in grid point space and also uses computations on single instruments of selected satellites, and nadir viewing conditions.

The influence of restricting the calculations to one satellite per instrument and to one choice of zenith angle is relatively small for the considered instruments. For AMSU-A and MHS, the differences between

the pass-band characteristics of instruments on different satellites are small and usually neglected in radiative transfer calculations. For HIRS, the differences in response function between instruments on different satellites are usually considered significant, but the influence of these differences on the spread calculations is nevertheless relatively small ($<$ 10 %, except for HIRS channel 15, where they can reach 35 %). The choice of a single zenith angle means that variations in the weighting functions arising from different viewing geometries of these cross-track scanning instruments are not taken into account. In contrast to the old scheme, we use a zenith angle of 25 °for all cross-track scanning instruments, more in line with a mean zenith angle for these instruments than the nadir choice of the old scheme. Differences between EDA spread calculations for this choice of zenith angle and for typical extreme zenith angles (0 and 45 °) are mostly below 15 %, except for surface-sensitive channels where the differences can reach 30 %. All these differences are considered small compared to the uncertainties inherent in the estimation of flow-dependent background errors.

The treatment of errors in the surface emission is conceptually significantly different in the radiance spread calculations from the approach used for the EF-fields. The old scheme uses random perturbations for the skin temperature and emissivity in addition to the perturbations drawn from the background error covariance, with the size of these perturbations determined by prescribed errors that differ over sea, land, and sea-ice. In contrast, for the EDA spread calculations, perturbations in the skin temperature are taken from the EDA (ie they are consistent with the sea surface temperature perturbations applied in the EDA) and no explicit error due to the surface emissivity specification is added. Over sea, the emissivity errors will therefore be limited to those given by the spread of the EDA for the variables that are input to the surface emissivity model, whereas over land, pre-scribed typical emissivity values are used, and no variation of emissivity is included between EDA members. This choice is motivated by the fact that we are aiming to estimate background errors, whereas errors due to the emissivity modelling are considered radiative transfer errors and should therefore not be included. This is a marked conceptual difference to the old scheme, and leads to very different estimates of “background errors” for surface-sensitive channels over land and sea-ice.

Consistent with EDA spread calculations for other geophysical variables, the radiance spread is derived on a N80 reduced Gaussian grid at T159 resolution for an EDA with a T399 model resolution. The EDA radiance spread could be calculated at any position in time over the subsequent assimilation window, to take into account the temporal evolution of the EDA spread. However, initially we will consider only calculations valid at the start of the subsequent assimilation window. This is consistent with the treatment of background error estimates already provided from the EDA for observations of other geophysical variables for quality control purposes, and also consistent with the old scheme for satellite radiances. The main reason for this restriction is to save computational cost and to limit the number of fields that need to be archived. We will consider aspects of the temporal evolution of the EDA spread over the subsequent assimilation window in the section 5.

Examples of the EDA radiance spread are shown in Fig. 1. They highlight the situation-dependence of the EDA spread, with a considerable dynamic range. For instance, an area with larger spread is apparent for the tropospheric channels in the region of the low-pressure system south of the Bering Strait, indicating larger uncertainty in the positioning of this system. Channel 5 of AMSU-A also shows larger spread over regions of high terrain (e.g., the Himalaya), reflecting the larger surface dependence and skin-temperature errors in these regions. Also notable is that there are large areas for which the EDA spread is very low, with values less than 0.04 K for the AMSU-A channels, much smaller than the instrument noise for these channels.

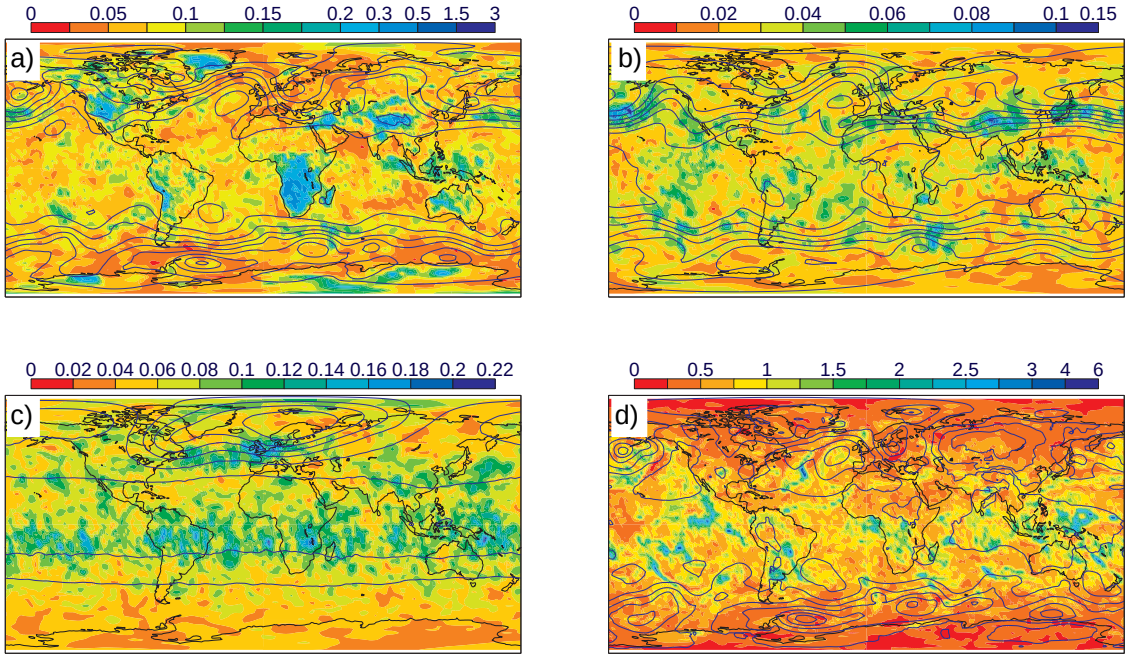


Figure 1: Examples of the EDA radiance spread [K] for selected channels valid on 15 February 2012 at 9 Z. Shading indicates the radiance spread, whereas blue contours provide another field for orientation. The selected channels are: a) AMSU-A channels 5 with 700 hPa geopotential, b) AMSU-A channel 8 with 200 hPa geopotential, c) AMSU-A channel 12 with 10 hPa geopotential, d) MHS channel 3, with the mean sea level pressure.

2.2 Initial evaluation of radiance spread

We will now evaluate the EDA radiance spread through a comparison with variances of background departures. If the EDA correctly identifies regions of larger uncertainty in the background fields through increased spread, these regions should also show increased variances of FG-departures. Similar evaluations are routinely done as part of the calibration step when flow-dependent background errors are estimated from the EDA. There, EDA spread is evaluated against forecast errors as estimated from differences between forecasts and analyses. In this step, the analyses are treated as truth, and analysis errors and their correlations with background errors are neglected. When performing the evaluation against observations in radiance space, neglecting observation errors is not possible, as their contribution dominates the standard deviation of the FG-departures.

Provided the background and observation errors are uncorrelated, the variances of FG-departures for a single channel follow the usual relationship

$$< (y - \mathbf{Hx})^2 > = \mathbf{HBH}^T + \sigma_O^2 \tag{1}$$

where y and σ_O are the observation and its observation error, $\mathbf{x}$ and $\mathbf{B}$ the background and its error covariance, $\mathbf{H}$ the observation operator, and $< . >$ the expectation operator. Note that inter-channel or spatial error correlations are not considered here. If the true background error is highly correlated with the EDA variance Var_{EDA} and the EDA variance is not correlated to the observation error, we have

$$< (y - \mathbf{Hx})^2 > \approx b \; Var_{EDA} + \sigma_O^2 \tag{2}$$

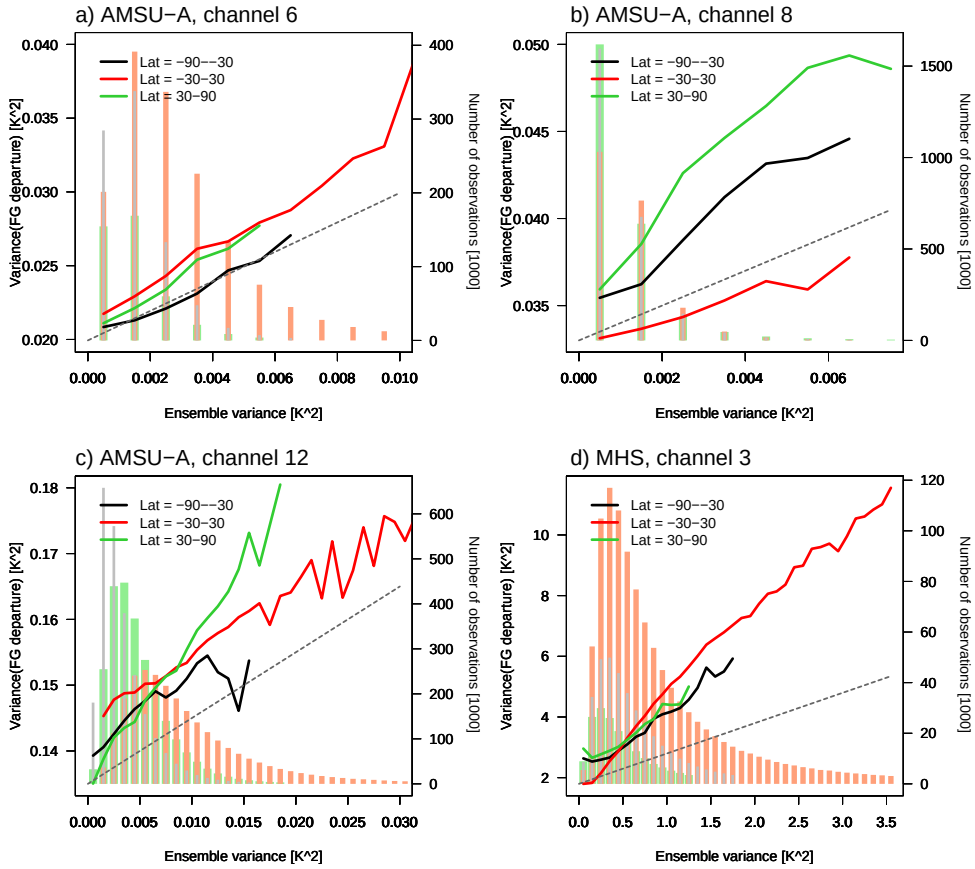


Figure 2: a) Variance of FG-departures for AMSU-A channel 6 on METOP-A as a function of the EDA variance for February 2012. The statistics are based on data over sea and after cloud screening and geographical quality control, but before the FG check. They are separated into three zonal bands: 90S-30S (black), 30S-30N (red), and 30N-90N (green). Also shown is the population of each bin as vertical bars (right y-axis; these use lighter versions of the colours for the three zonal regions); every bin contains at least 3000 observations. In addition, for orientation, a line with the slope 1 is also displayed (grey dashed). b) As a), but for channel 8 over land and sea, c) As b), but for channel 12, and d) As b), but for MHS channel 3.

Here, b represents a simple scaling factor that ensures that $\mathbf{HBH}^T \approx b\ Var_{EDA}$, ie we assume that the true background error is proportional to the EDA variance. More complex relationships between the true background error and the EDA variance are possible, but here we consider only this simple relationship.

We examine the above relationship using FG-departures taken from a deterministic 4DVAR experiment that uses the background error variances derived from an EDA for the specification of the background error. For each observation, we extract the corresponding EDA spread as calculated above. We then bin the observations by EDA spread, and calculate standard deviations of FG-departures, after bias correction. Bias correction is done using variational bias correction (VarBC), as described in Dee (2004). We consider only observations after cloud screening and geographical quality control, but before the FG-check.

The ensemble spread in radiance space shows good skill in identifying areas with larger standard deviations of FG-departures for almost all channels considered here (see, for instance, Fig. 2). The relationship is fairly linear, as suggested by equation 2. However, the slope of the relationship is larger than one for

many channels and regions, ie the increase in the size of the background departures is larger than would be suggested by the EDA spread. As mentioned earlier, we assume that the EDA spread and the observation error are uncorrelated, and the statistics therefore suggests that the ensemble variance appears to underestimate the true background error variance in many areas. Calibration of the EDA spread is therefore considered necessary to provide estimates of background errors in radiances space, analogous to what is done for the control vector background errors used in ECMWF's hybrid 4DVAR system.

It is worth mentioning here that the assumption that the EDA spread and the true observation error are uncorrelated may not always be strictly valid. The observation error includes the error of representativeness and quality control error. It is possible that both of these may be correlated with the EDA spread, as the EDA spread tends to be larger in areas of active wether systems, where quality control and representativeness errors may also be larger. While it is less likely that this is an issue for the temperature-sounding provided by AMSU-A, it is more difficult to exclude this possibility for humidity-sounding channels of MHS. If such error correlations are present, they would lead to an over-estimation of the required scaling.

3 Spatial filtering and calibration of radiance spread

In the hybrid 4DVAR system currently used operationally at ECMWF, the EDA spread for the control variables is spatially filtered and calibrated before it is passed to the deterministic 4DVAR to provide flow-dependent estimates of the background error variances (Bonavita et al. 2012). In the following, we describe the analogous steps for the radiance calculations.

3.1 Filtering of sampling noise

The spatial filtering applied to the ensemble spread in radiance space aims to reduce sampling noise which can be significant given the small number of ensemble members used in the EDA. The method used is the same as that used for all other geophysical variables. It is based on the wavenumber-dependent signal correlation, as described in detail in Bonavita et al. (2012). The idea is to retain the signal from those wavenumbers for which results from different sub-ensembles of a large ensemble show strong correlation as the signal dominates, whereas to filter out the contribution from wavenumbers with low correlation, as here the sampling noise dominates. The filter retains broader spatial features and dampens small-scale detail associated with high wavenumbers. Fig. 3 shows the filtered versions of the radiance spread examples shown in Fig. 1. The filter tends to retain more spatial detail for humidity-sensitive radiances (e.g., Fig. 3d showing MHS channel 3), whereas smaller scales are filtered out most heavily for stratospheric temperature sounding channels (e.g., Fig. 3c showing AMSU-A channel 12).

3.2 Method for the spread calibration

The calibration of the radiance spread is done directly in observation space on the basis of FG-departures, using the spread-skill relationships illustrated in the previous section. The calibration consists of a simple scaling of the EDA-spread, with the scaling factor derived by fitting a linear model to the FG-departure variance/EDA variance relationship. This is done for each AMSU-A, MHS, and HIRS channel assimilated in the system. We use a fit weighted by the number of observations in each bin and allow different intercepts for different satellites, but a single channel-dependent scaling factor that applies to all satellites. As can be seen from equation 2, the intercept plays the role of σ_O^2, and this can differ from satellite

to satellite due to differences in the instrument noise. In contrast, the scaling required for the EDA spread is expected to be independent of the satellite.

Different fits are calculated and applied for three zonal bands (90S-30S, 30S-30N, 30N-90N), similar to what is done in the current calibration of the EDA spread for the control vector background errors (e.g., Bonavita et al. 2012). Such geographical separation was motivated by the finding that different scaling appears justified for different regions, as seen in Fig. 2. For instance, statistics for AMSU-A channel 8 suggest that more scaling is required for the extra-tropics than the tropics, whereas for channel 12 stronger scaling seems to be required for the Northern Hemisphere over February 2012. The geographical dependence of the scaling indicates that the EDA spread may represent background errors differently in different areas, reflecting the observational coverage or the realism of the perturbations added to represent model error. During the experimentation, other geographical separations were considered (e.g., land/sea separation), but the chosen simple zonal separation showed overall the most distinct signatures in the required scaling over all channels.

The temporal evolution of the calibration is taken into account by calculating the scaling factors for each analysis cycle from observations used over the last 5 days. This is similar to what is done for the calibration of the assumed background errors.

The scaled and spatially filtered ensemble spread in radiance space is calculated for each analysis cycle of the EDA and stored as field with type SES in MARS. Note that in the calibration step only an inflation of the ensemble spread is allowed (not a deflation), and a further inflation of 1.3 is applied globally for all channels. Both of these aspects are designed to avoid an underestimation of the background error in radiance space. In addition, we limit the scaling factor to be below 3 to avoid excessive scaling of the EDA spread.

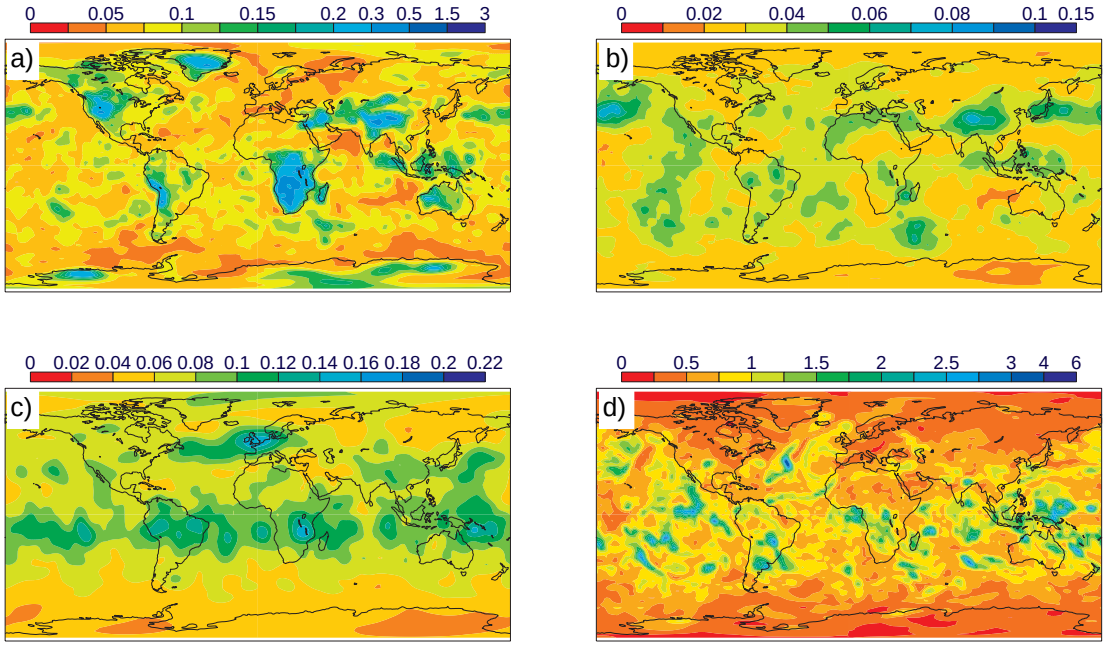


Figure 3: As Fig. 1, but for the spatially filtered versions of the EDA radiance spread [K], showing AMSU-A channels 5 (a), 8 (b), and 12 (c), and MHS channel 3 (d).

Table 1: Fitting statistics for NOAA-18, NOAA-19, and METOP-A AMSU-A and MHS for the Southern Hemisphere (30S-90S), derived over February 2012. We show the square root of the intercept [K] (playing the role of σ_O in equation 2). For comparison, we also list a σ_O estimate based on the Desroziers diagnostic [K] and estimates of the instrument noise (NedT, K) from the direct instrument monitoring.

	NOAA-18			NOAA-19			METOP-A		
Channel	Inter-cept$^{0.5}$	Des-roziers	NedT	Inter-cept$^{0.5}$	Des-roziers	NedT	Inter-cept$^{0.5}$	Des-roziers	NedT
AMSU-A									
5	0.21	0.18	0.19	0.19	0.16	0.17	0.21	0.18	0.16
6	0.15	0.14	0.14	0.15	0.13	0.14	0.15	0.14	0.12
7	0.18	0.17	0.17	0.24	0.24	0.24	n/a	n/a	n/a
8	0.21	0.20	0.23	n/a	n/a	0.78	0.19	0.18	0.16
9	0.18	0.16	0.16	0.18	0.16	0.16	0.18	0.17	0.16
10	0.22	0.20	0.19	0.22	0.20	0.20	0.22	0.21	0.19
11	0.24	0.23	0.21	0.24	0.22	0.20	0.27	0.25	0.23
12	0.34	0.32	0.30	0.35	0.34	0.30	0.38	0.36	0.30
13	0.48	0.47	0.39	0.49	0.47	0.39	0.53	0.51	0.40
14	0.80	0.88	0.62	0.82	0.80	0.62	0.87	0.86	0.66
MHS									
3	1.57	1.38	0.65	n/a	n/a	n/a	1.44	1.24	0.55
4	1.33	1.23	0.50	1.32	1.24	0.55	1.25	1,12	0.45
5	1.20	1.13	0.45	1.16	1.11	0.30	1.16	1.08	0.45

3.3 Analysis of the calibration fit

The above calibration method has been applied to analysis experiments, performed for the period 1 January - 31 March 2012 and 1 June - 31 August 2012. The experiments use a spatial resolution of T511 and a 12-hour window 4DVAR with three inner loop incremental analysis updates at T95/T159/T255 resolution. They include all observations assimilated operationally at the time. Ten-day forecasts were run for each 00 UTC assimilation cycle.

3.3.1 Intercept

Following equation 2, the square root of the intercept of the fitted linear model can be interpreted as an estimate of σ_O. Note, however, that this interpretation hinges of course on the assumptions made in the derivation, especially that the true background error variance is highly correlated with the EDA variance and that the relationship is approximately linear. If this is not the case, structures in the standard deviations of FG-departures that are due to background error variations, but do not map onto the EDA-spread will lead to an intercept that is larger than that given by true observation errors.

The square root of the intercept compares relatively well with other observation error estimates (Table 1). One estimate provided here is the Desroziers diagnostic (e.g., Desroziers et al 2005, Bormann and Bauer 2010) derived from FG and analysis departures from an active assimilation of the data, and another estimate is the instrument noise from the routine monitoring of the instrument's telemetry. For AMSU-A, the three estimates are very consistent. For MHS, the σ_O estimate based on the intercept is larger than the Desroziers estimate which again is much larger than the instrument noise. The latter is likely a result of other errors such as errors of representativeness or forward model errors which contribute to the obser-

vation error estimates for the humidity sounding channels. The larger intercepts in the EDA calibration may be a result of background error contributions that do not map onto the structures represented by the EDA spread fields. The neglected temporal evolution of the background error over the assimilation window is partly responsible for this.

The finding that the fitted intercepts are nevertheless overall consistent with alternative observation error estimates gives further confidence in the chosen approach.

3.3.2 Scaling factor

The scaling factors for all AMSU-A channels can be seen in Fig. 4a. For the month considered, they show that the EDA appears to be under-dispersive for upper tropospheric temperature over the extra-tropics, whereas the spread appears more consistent with departure statistics over the tropics.

The scaling factors derived for AMSU-A have also been compared to scaling factors derived from radiosonde temperature observations, or from an analysis-based estimate of short-term forecast error. The radiosonde statistics have been derived in the same way as those for AMSU-A, except that the ensemble spread has been calculated directly, based on the temperature spread on model levels. For the analysis-based estimation of short-term forecast error, we computed differences between short-term forecasts and the operational analysis on a N80 reduced Gaussian grid, and again compared these to the ensemble spread. This comparison is similar to the one chosen in the operational calibration of the EDA. A selection of the resulting spread-skill statistics is shown in Fig, 5.

The slopes of the spread-skill relationships and hence the scaling factors suggested from the AMSU-A statistics are rather different from the radiosonde ones. The scaling suggested by the radiosondes is typically 2-3 times larger than that suggested by the AMSU-A statistics, even though some qualitative agreements exist, such as stronger scaling for the Northern Hemisphere extra-tropics for upper tropospheric and stratospheric channels/levels for both observation types (cf Fig. 4a and b) for this period.

The reasons for the different scaling is likely due to the different representativeness in the vertical and the horizontal of the two types of observations. AMSU-A measures temperature over broad weighting functions, whereas the radiosondes measure single-level values. One interpretation is that the results suggest that the EDA represents differently perturbations associated with such different vertical scales. The EDA spread appears to be better able to capture errors on broad vertical layers (and the associated broad horizontal lengthscales) as seen by radiances, whereas more scaling is required for finer-scale structures. This may be a result of the small number of ensemble members used in the EDA, meaning that structures with fine vertical scales are poorly represented and dominated by sampling noise, similar to what is observed spatially for higher wavenumbers in spectral space (e.g., Bonavita et al. 2012). An alternative interpretation is that the true observation error for radiosondes is not constant over the considered ensemble spread range. For instance, the spatial representativeness errors for radiosonde data may be larger in active areas for which the ensemble spread also tends to be larger. This would mean some of the increase in the variances of FG-departures with ensemble variance is due to an increase of the representativeness error, and the relationships shown in Fig, 5 would hence overestimate the required scaling. Lastly, there is also a sampling difference: while the AMSU-A data provides global sampling (excluding strongly-cloud affected areas for the tropospheric channels), the radiosonde data is mainly concentrated over land.

The comparison to the analysis-based evaluation and calibration is even more difficult, as it crucially depends on the assumptions made about analysis errors. In the operational calibration of the EDA spread, the analysis is treated as the truth and analysis errors are neglected. Scaling factors are calculated by

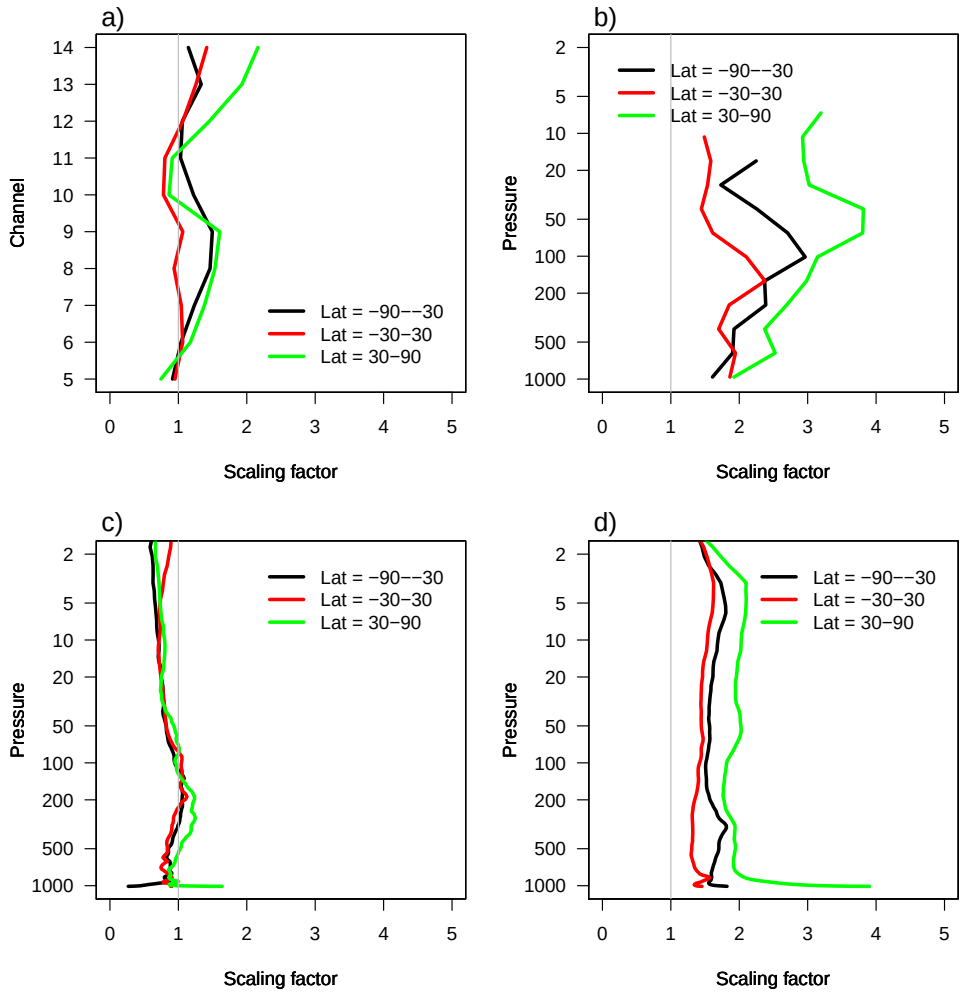


Figure 4: Intercomparison of scaling factors for the EDA spread derived in different ways for February 2012. a) Scaling factors derived from AMSU-A data as described in the main text. b) As a), but using radiosonde temperature observations. c) Scaling factors derived from analysis-based spread-skill relationships as shown in Fig. 2, adopting the same approach as applied to the observation-based scaling. d) Mean scaling factors calculated in the way adopted for the calibration of the EDA spread for assumed background errors.

splitting the ensemble variance into 10 equally sized bins, and 10 different scaling factors are calculated from the ratio between the perceived FG error and the ensemble spread for each bin (Bonavita et al. 2012). As can be seen from Fig. 5d-f this will result in large scaling factors for small variances, and progressively smaller scaling factors for larger variances. This spread-dependence of the scaling factors makes it impossible to directly compare the operationally-derived scaling factors with the scaling factors presented here for AMSU-A or radiosondes. Nevertheless, Fig. 4d shows the mean scaling factors actually applied in the EDA calibration of the temperature field, and these are mostly between the scaling factors derived for AMSU-A and the radiosondes.[1] However, bearing in mind the dependence of the

[1]Note that this comparison is based on the so-called "SES" fields only, whose calibration aims to harmonise the geographical distribution of the assumed background errors and to approximately match their statistical distribution with that of the perceived analysis errors. It should be borne in mind, however, that the globally averaged magnitude of the assumed background errors is not determined by the "SES" fields, but by the full backround error covariance matrix used (JB) in the 4DVar assimilation together with the "REDNMC" factor.

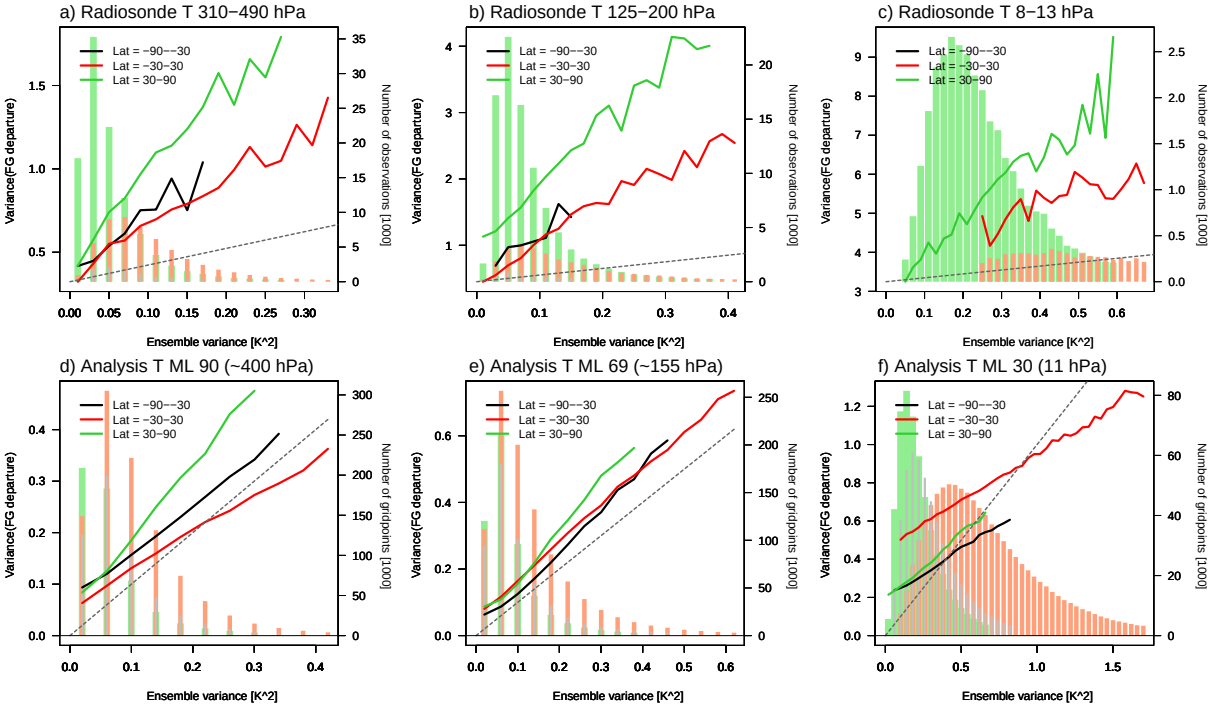


Figure 5: Spread-skill relationships as in Fig. 2, but using radiosonde temperatures (a-c) and analyses (d-f) for evaluation. For the latter, the FG error variances have been calculated from differences between the 12-hour forecasts and a high-resolution analysis on a reduced Gaussian grid at T159 resolution. The three columns evaluate the EDA spread in different parts of the atmosphere, chosen to coincide with the AMSU-A channels selected in Fig. 2a-c for comparison purposes. Panels a-c are based on single-level radiosonde data over the layer indicated in the titles, whereas panels d-f use forecast minus analysis differences on model levels 90, 69, and 30, roughly at the centre of the radiosonde layers.

scaling on the ensemble spread itself, we stress that this comparison is not very meaningful. Alternatively, we can also calculate scaling factors in exactly the same way as for the observations, that is we allow for an analysis error (which replaces σ_O in equation 2) that is constant with the ensemble variance. The assumption that the analysis error is constant with the ensemble variance is likely to be untrue, just as the earlier assumption that the analysis error is zero. Nevertheless, using this approach gives scaling factors shown in Fig. 4c. These factors are much smaller than the mean applied scaling factors, and even smaller than the AMSU-A ones. The finding that these scaling factors are so small likely reflects the presence of correlations between errors in the short-term forecast and the verifying analysis used here. If neglected, as done in our approach, such correlations would act to underestimate the required scaling.[2]

The above analysis shows that there are further insights to gain when evaluating the ensemble spread with observations. The scale-dependence of the scaling appears worth investigating further, as are methods that further reduce the reliance on neglecting analysis errors in the calibration step for the assumed background error variances.

The above analysis has also highlighted an important further aspect: the calibration of the EDA radiance spread as presented here is completely decoupled from the calibration of the EDA spread used to specify the background error in the assimilation: the derived scaling factors can be quite different. On the one

[2]We found this aspect especially severe when the experiments' own analysis is used as "truth" rather than the operational analysis; this choice leads to even smaller estimates for the scaling factors.

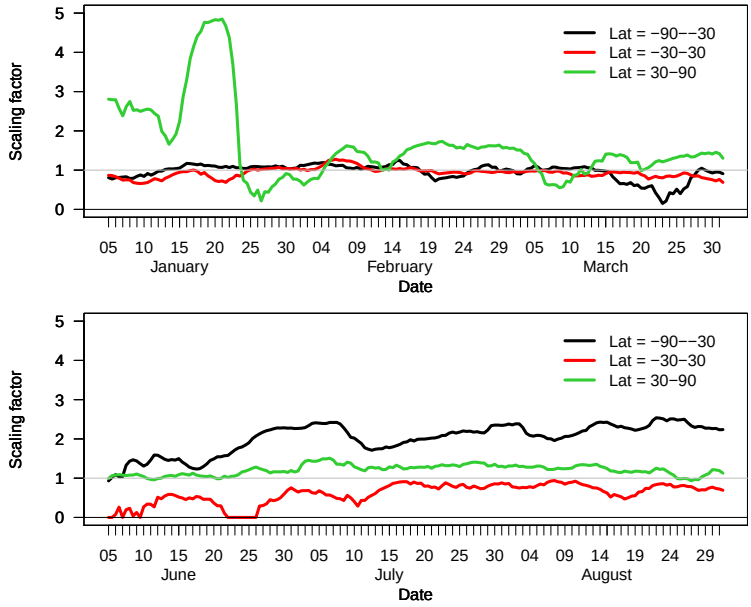


Figure 6: Time-series of the zonal scaling factors ($\sqrt{b}$) for AMSU-A channel 12 for January-March and June-August 2012. The scaling factors have been calculated from observations covering 5 days, and they are plotted here at the end of each 5-day interval.

hand, this has the advantage that the calibration provides an independent assessment of the performance of the EDA, which is useful for diagnostic purposes, as already highlighted above. On the other hand, it means that the resulting estimates of the background errors in radiance space are not necessarily consistent with the assumed background errors used in the assimilation. In particular, the calibration in radiance space may give quite different results if the correlation between the true background errors and the EDA variance is poor, or if the EDA spread is correlated with the observation errors. The fact that the calibration is decoupled from that used for the calibration of the assumed background errors is an important caveat to bear in mind with the reformulation of the background error estimate in radiance space. As our primary interest here is to use the background error estimate for quality control or diagnostic purposes this caveat is not considered a fundamental drawback.

3.3.3 Temporal evolution

There is considerable temporal evolution of the calibration. While the intercepts tend to be fairly constant (consistent with the interpretation as estimates of observation error, not shown), the scaling factors show variations on seasonal time-scales as well as on time-scales of a few days. This is particularly evident for the stratospheric temperature sounding channels. For instance, Figure 6 shows how AMSU-A channel 12 requires larger scaling factors for the Southern Hemisphere in July/August compared to January-March 2012. This relates to larger uncertainty around the polar vortex associated with gravity wave activity; this uncertainty is represented in the EDA spread, but the observations suggest that it is underestimated. Also, there is a striking episode with particularly large scaling factors for the Northern Hemisphere extratropics in January 2012, lasting a few days. This is associated with a Sudden Stratospheric Warming (SSW) event, again under-represented in the EDA-spread. The latter event will be described in more detail in section 4.3. Also worth pointing out are periods for which the scaling factors drop to zero

(e.g., around 25 June 2012, Fig. 6). This tends to occur when the spread-skill relationships are fairly flat, which is most often a result of low background error contributions that are overall small compared to the instrument noise of the channel in question. It can also indicate a poor correlation between the background errors and the ensemble spread.

3.4 Comparison between SES and EF fields

A comparison between the new SES and the EF field from the randomisation calculations shows that the two estimates for background errors in radiance space are rather different. An example is provided in Fig. 7 for selected AMSU-A and MHS channels, and a number of points are immediately apparent:

- For surface-sensitive channels, the most notable feature is the large difference between the two background error estimates over land (e.g, for AMSU-A channel 5, see Fig. 7a and b). This is a result of the different approaches to treating surface emission errors, as discussed in section 2.1 The two fields should therefore not be compared over land and sea-ice.
- The SES field tends to show a larger dynamic range of the background error estimates, with areas of smaller as well as larger values than in the EF fields.
- Although there are some common features between the two fields (e.g., for AMSU-A channel 8 the larger uncertainty south of the Bering Strait), the geographical variations are mostly very different, particularly for the MHS channels (cf Figures 7g and h).

The very clear differences may seem surprising, given that the EF field is based on the assumed background error which in turn has a flow-dependent component derived from the EDA spread. However, a number of aspects contribute to this. Firstly, the SES fields take the temporal evolution of the background error better into account, as the ensemble uses the atmospheric model to produce the spread field valid at the beginning of the next assimilation window. In contrast, the error evolution in the EF field is based on a static error growth model which will, for instance, be unable to handle the propagation of weather systems and the associated error evolution. Also, although the assumed background errors are based on the EDA, their representation is subject to the modelling of the background error, such as the treatment of correlations, the formulation of control variables and their treatment of flow-dependency. The latter aspect has changed considerably in recent cycles, with the introduction of flow-dependent components for variables other than vorticity.

The EF and SES fields have been evaluated by examining spread-skill relationships for observation departures. For our purposes of using the field in FG-check decisions, the main aspect is that the background error estimate correctly identifies areas for which the departures are larger, as a result of larger errors in the background.

The ability of the SES field to identify areas of larger background error can be compared to a similar analysis for the EF field (compare left and right columns in Figures 8 and 9). For tropospheric temperature sounding channels, both schemes perform relatively well, although the calibration is better for the SES field, leading to slopes in the spread-skill relationships that are closer to 1. This is to be expected, since the calibration scheme has been designed to achieve exactly this. For the stratospheric temperature sounding channels, the spread-skill relationship again suggests a good calibration for the SES field, whereas the EF field tends to under-estimate areas with larger background departures (e.g., Fig. 8e, f). For humidity channels, the spread-skill relationship suggests rather little skill for the EF field to identify areas of larger departures, whereas the SES field shows a very good calibration, and in particular good

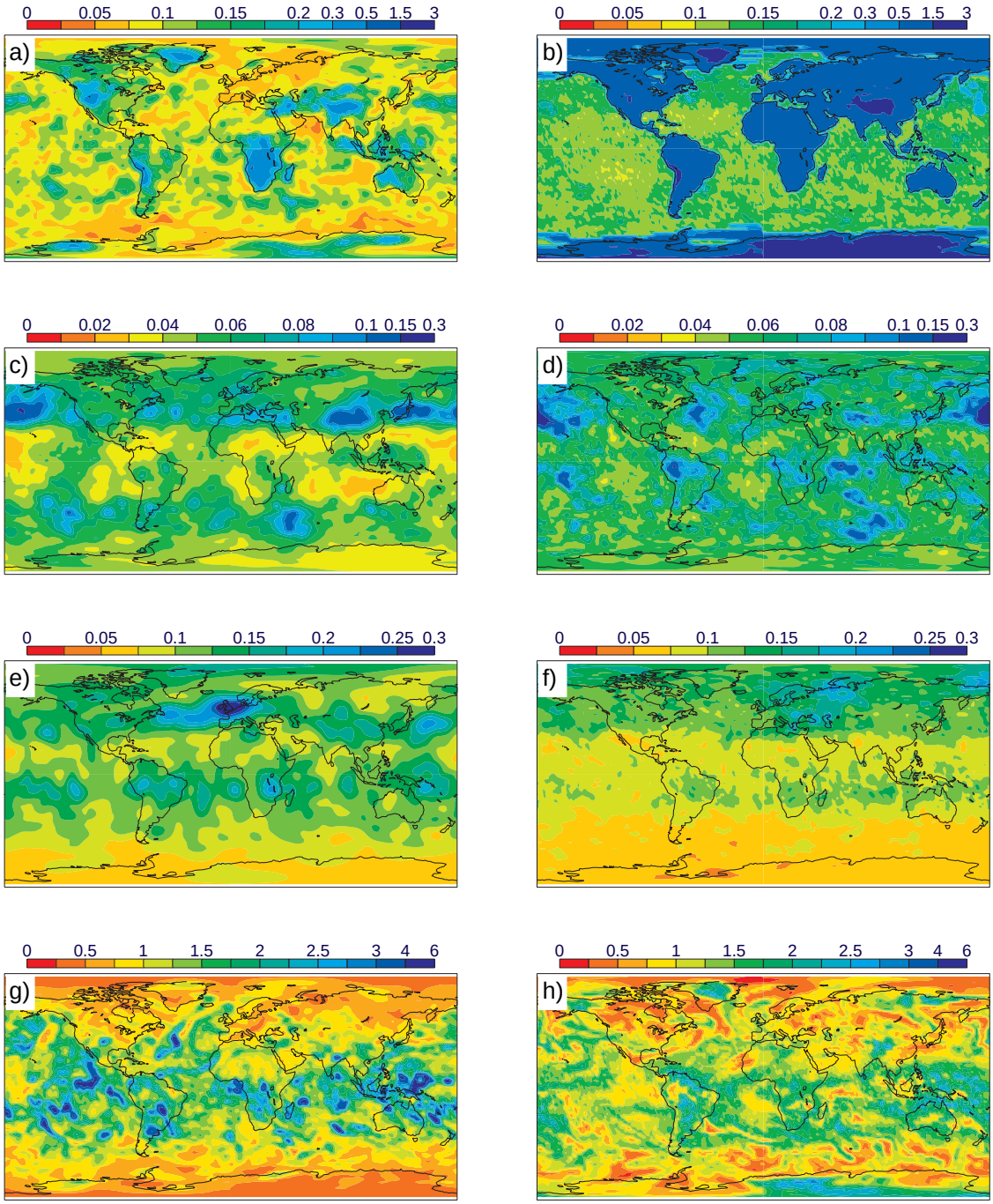


Figure 7: Examples of calibrated and filtered EDA spread fields (left), in comparison to the equivalent EF fields (right) for 15 February 2012, 9 UTC, for AMSU-A channels 5 (a,b), 8 (c,d), and 12 (e,f), and MHS channel 3 (g,h).

skill in identifying the regions with the largest FG-departures as a result of larger background errors (Fig. 9). The better results for the SES field for the humidity sounding channels are likely to be related to the larger temporal variability of humidity which is better handled in the SES fields

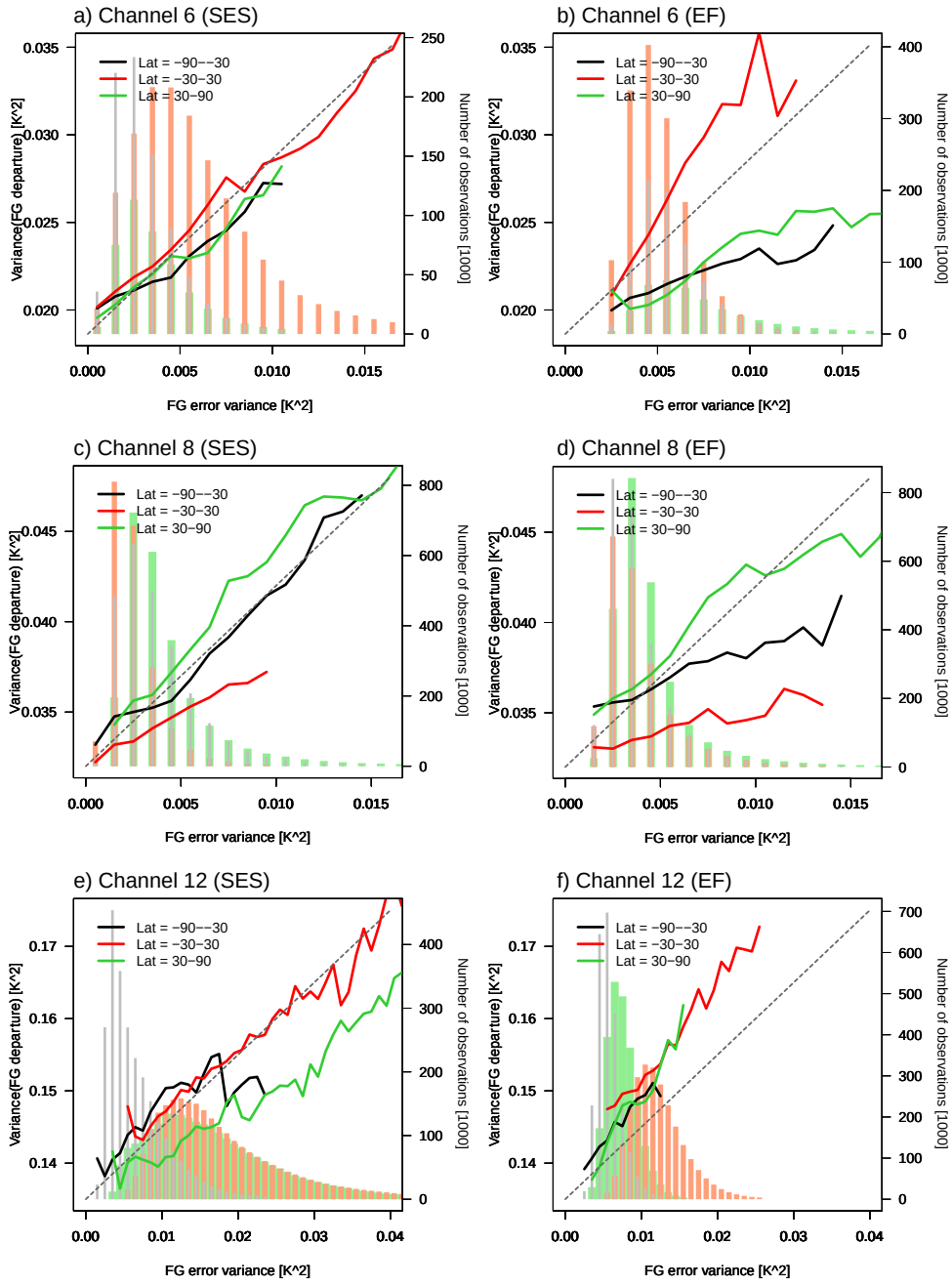


Figure 8: Spread-skill relationships for AMSU-A as in Fig. 2, but for the SES estimates of background error (left column) and the old EF estimates (right column), based on data for February 2012. The panels show channel 6 (a, b, over sea only), 8 (c, d), and 12 (e, f).

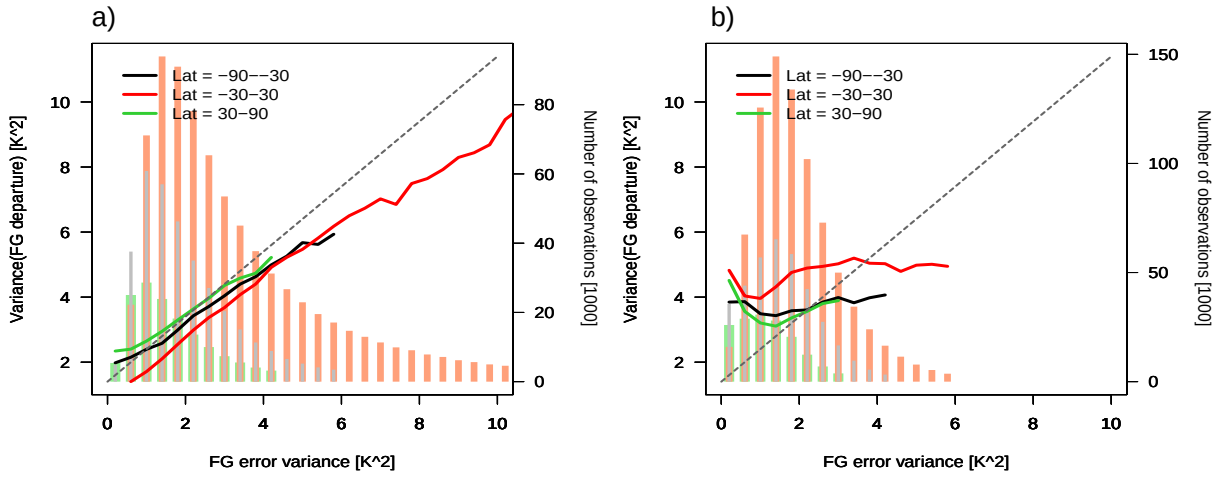


Figure 9: As in Fig. 8, but for MHS channel 3, with data over sea only.

4 Quality control with the scaled EDA radiance spread

We will now use the estimates of the background errors from the EDA spread (SES fields) together with observation errors to identify observations that deviate unexpectedly far from the background. This is primarily done to eliminate gross outliers in the observations, and it is usually referred to as the FG-check. In the FG-check, the expected standard deviation of FG-departures is estimated from assigned observation and background errors, and if the absolute value of the FG-departure exceeds a certain multiple of this the observation is rejected. That is, observations are rejected if

$$|y - \mathbf{H}\mathbf{x}| \; > \; l\sqrt{(\sigma_B^2 + \sigma_O^2)} \tag{3}$$

with the rejection limit l. These rejection limits are usually chosen on the basis of departure distributions, with the aim to reject observations whose departures are outliers from a Gaussian distribution. As part of the present revision, the rejection limits l have also been reviewed.

4.1 Revision of rejection limits

The FG-check limits have been revised on the basis of departure histograms as displayed in Fig. 10. For AMSU-A, the distributions are fairly Gaussian, with relatively few clear outliers, especially for channels 8-14. For the tropospheric channels, the distribution shows a standard deviation that is relatively close to that suggested by the assigned observation errors and estimated background errors, albeit somewhat smaller (as can be seen by comparing the thick grey and black lines in Fig. 10). Based on these statistics, the FG-check limit l has been relaxed to 3.5 for all AMSU-A channels, compared to the previous choice of 2.5. Note that this translates to values of approximately 4-5 times the actual standard deviations of the FG-departure distributions. The humidity channels of MHS and HIRS show a less Gaussian distribution, possibly a result of remaining shortcomings in the estimates of the geographical variation of the background error. Nevertheless, the limits were relaxed to $l = 3$ for MHS channels 3 and 4 and HIRS channel 11 and 12. For other HIRS channels, the assumed (diagonal) observation errors are rather high,

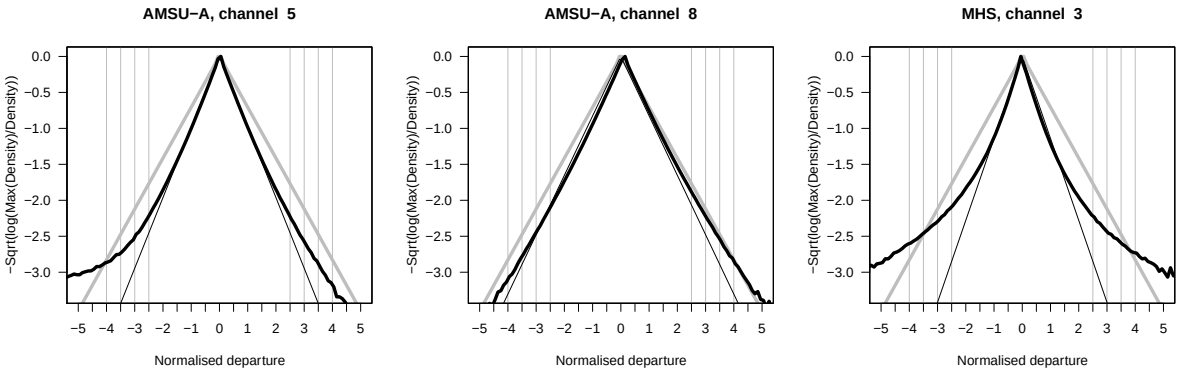


Figure 10: Examples of transformed departure histograms for AMSU-A channel 5 (left), AMSU-A channel 8 (middle) and MHS channel 3 (right). The histograms are displayed in terms of departures normalised by the expected departure standard deviation (derived from observation errors and the new estimates for the background error). Black thick lines show the histograms, thin black lines indicate a Gaussian distribution, fitted at the centre of the actual distribution, and thick grey lines show a Gaussian with a standard deviation of one. The histograms are based on all data after geophysical quality control for February 2012, with data from all assimilated instruments combined and departures after bias correction.

to counteract error correlations, and the rejection limit is kept at $l = 2.5$. Parameters for the variational quality control were also updated (Andersson and Järvinen 1999), and relatively few observations are now substantially downweighted by the variational quality control.

4.2 Assimilation results

The use of the SES fields in the FG-check together with the updated rejection limits has been evaluated in analysis experiments covering the period 1 January - 31 March 2012 and 1 June - 31 August 2012. The experiments are based on cycle 38r2 and use 12-hour 4DVAR with a spatial model resolution of T511 (≈40 km), a final incremental analysis resolution of T255 (≈80 km) and 137 levels in the vertical. The experiments include all observations assimilated operationally with cycle 38r2. Ten-day forecasts were run for each 00 UTC assimilation cycle. The following experiments were run for each period:

EF: Control experiment which uses fields obtained from the randomisation method (EF fields) in the FG-check for ATOVS and geostationary satellite radiances, and in which the FG-check limits are as in the operational configuration[3]

SES: Experiment in which the SES fields described above are used instead, and the rejection limits have been updated as described above.

The re-tuning of the FG-check limits together with the changes to the estimates for the background errors in radiance space lead to an increase in the number of assimilated AMSU-A and MHS observations (Fig. 11). The increase reaches up to about 2 % for some channels, and is accompanied with an increase

[3]For technical reasons, the EF fields in this experiment are fetched from an equivalent high-resolution T1279 experiment that uses the same EDA for the background error specification as the EF and SES experiments, rather than calculated during the minimisation.

in the standard deviation of the FG-departures. The latter is to be expected, as the change specifically allows observations to deviate more from the FG. Similar changes can also be reported for the humidity channels on HIRS and the geostationary satellite radiances which use the background error estimates from the HIRS water vapour channels.

There are no consistent changes to the FG or analysis fit for other observations, suggesting a similar quality of the analysis or the FG. The only noteworthy exception is a small reduction in the standard deviation of FG-departures for upper tropospheric humidity channels of AIRS and IASI of around 0.5-1 %. Even though the effect is small it may be a reflection of improvements to the humidity field as a result of allowing more observations from the humidity channels of MHS and HIRS to influence the analysis.

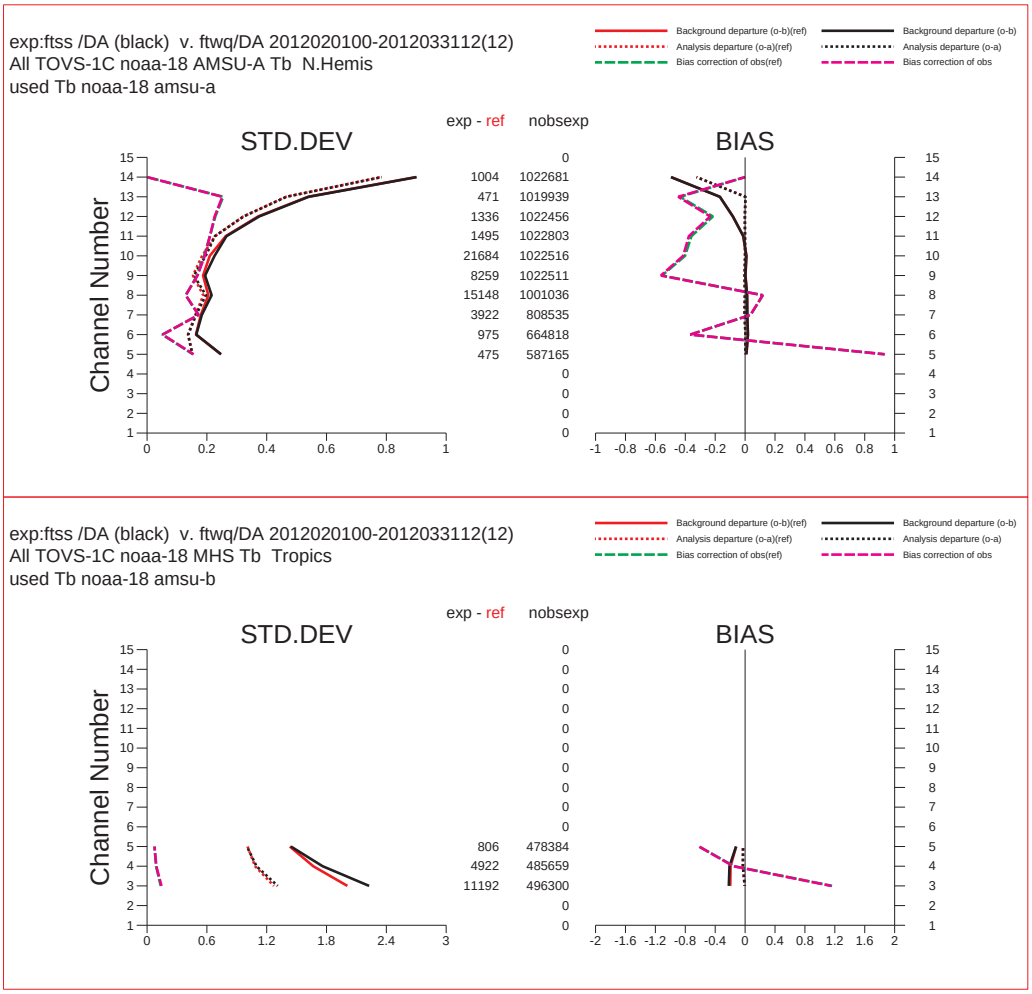


Figure 11: Departure statistics for used NOAA-18 AMSU-A radiances over the Northern Hemisphere (top) and for used NOAA-18 MHS radiances over the tropics (bottom) for February/March 2012. Black lines indicate statistics for the SES experiment, red lines for the EF experiment, with FG-departure statistics displayed in solid lines and analysis-departure statistics in dotted lines. Also shown are statistics for the bias correction for the SES experiment in magenta and the EF experiment in green. Standard deviations are on the left, mean values on the right, and the numbers between the two columns show the number of assimilated observations and the differences between the SES and the EF experiment.

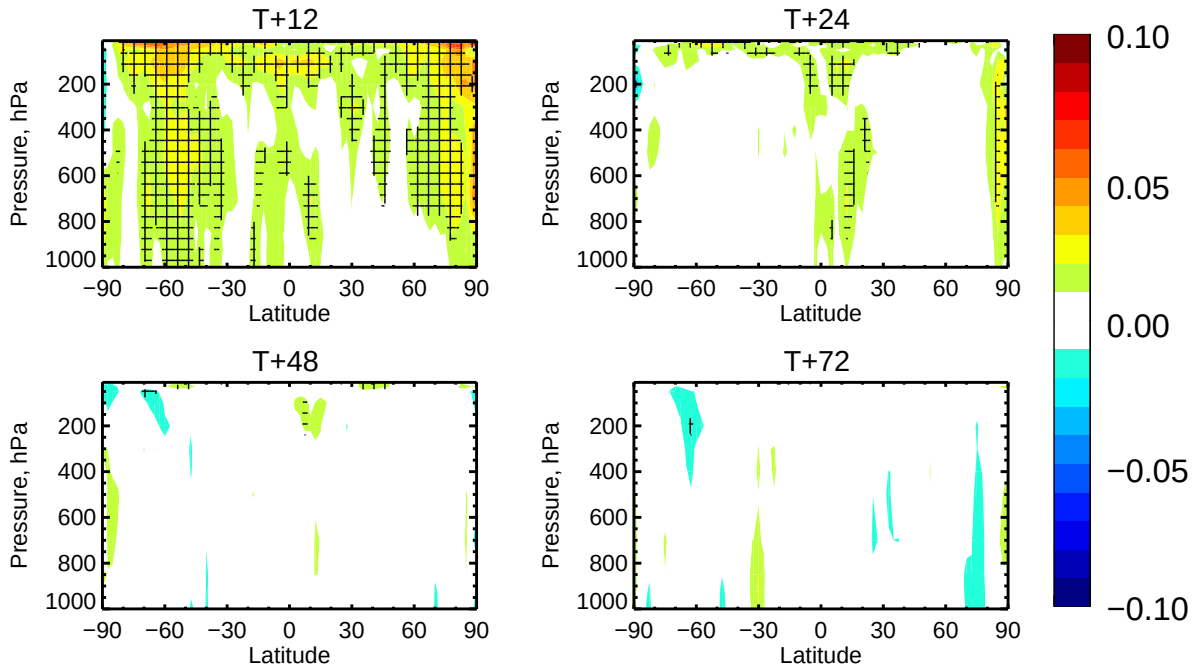


Figure 12: Zonal means of the normalised difference in the root mean square error for the geopotential, for the two periods combined (i.e., 1 January - 31 March 2012 and 1 June - 31 August 2012, 180-183 cases). Each experiment has been verified against its own analysis.

Increments are generally increased through the modifications, as can be seen in the upper left panel in Fig. 12. This is a result of relaxing the FG-check which specifically allows more observations with larger departures to pass our quality control. Given the lack of change in the FG-fit for other observations, the increased increments are not considered problematic. Forecast scores for day two and beyond are overall neutral, when verified against the own analysis (e.g., Fig. 12), the operational analysis, or observations (not shown).

4.3 Influence on stratospheric warming event

In the following, we will take a closer look at the performance of the new background error estimation scheme during the sudden stratospheric warming (SSW) event of January 2012. As highlighted earlier, during this event the scaling coefficients shown in Fig. 6 indicate the need for particularly strong inflation of the ensemble spread. SSWs are associated with rather large FG-departures for stratospheric AMSU-A channels in the ECMWF system (e.g., Fig. 13). At times, the FG-departures can be so large that the observations are rejected through the FG-check. It is therefore an illustrative example of the performance of the new setup under very specific conditions. Note that the episode of larger FG-departures coincides with the enhanced scaling of the EDA-spread noted earlier, showing that the calibration reacts quickly to this situation.

Figure 14 shows an example of the coverage of used AMSU-A channel 13 (peaking around 7 hPa) during the peak of the large departures. For the control, a gap in coverage can be seen around Novaya Zemlya and Franz Josef Land, caused by rejections of the data through the FG-check. Over this area, FG-departures reach up to 6 K, indicating that the FG is considerably too cold. In comparison, standard

deviations of FG-departures for this channel are typically around 0.5 K, and the observation error is assigned to be 0.85 K. The EDA spread suggests considerable enhanced uncertainty in the FG over the region, but only up to a modest 0.6 K (Fig. 15). Following the filtering and calibration, the estimate for the background error reaches 1.8 K over the affected area. This is sufficient for most of the observations to pass the FG-check (Figures 13 and 14b). However, the impact of this additional data on the analysis and subsequent forecasts is relatively limited.

It is worth mentioning here that the problems during the SSWs are more severe at higher model resolutions, and although the new estimates for the background errors in radiance space help to reduce the number of rejections, they do not solve the problem at the operational resolution of T1279. This is illustrated in Fig. 16 which shows the same coverage plot as in Fig. 14, but for experiments run at a resolution of T1279. The area with rejected observations is much larger in this case, and the FG departures reach a stunning 15 K over the affected areas (compared to the 6 K encountered for the T511 experiments). The use of the SES field in the FG-check achieves only a fairly modest reduction of the hole in the coverage. Any statistical FG-check method is likely to struggle under such extreme conditions, and instead a more fundamental change to the quality control such as using the Huber norm approach (Tavolato and Isaksen 2010) could be considered. The resolution-dependence of the underlying problem is linked to the generation of spurious gravity waves in the forecast model, which leads to substantial large-scale

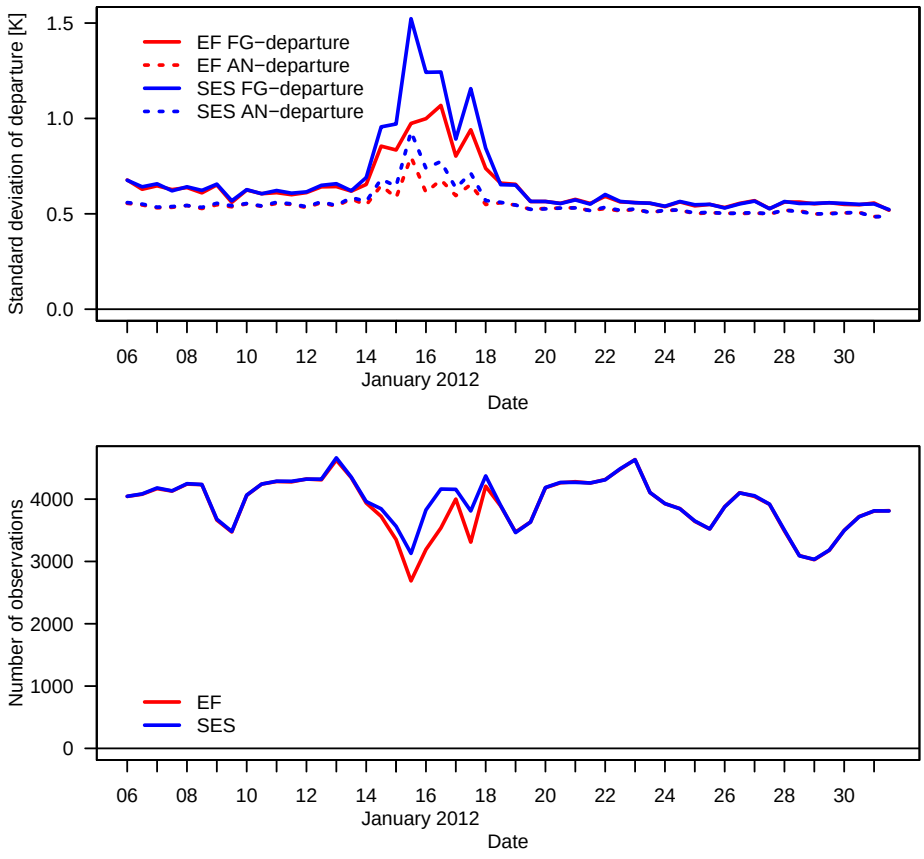


Figure 13: Time-series of standard deviations of FG-departures (top, solid lines) and analysis departures (top, dashed lines) and the number of used data (bottom) for Metop-A AMSUA-A channel 13 over the North Polar region (north of 60N), taken from the EF (red) and the SES (blue) experiment over January 2012.

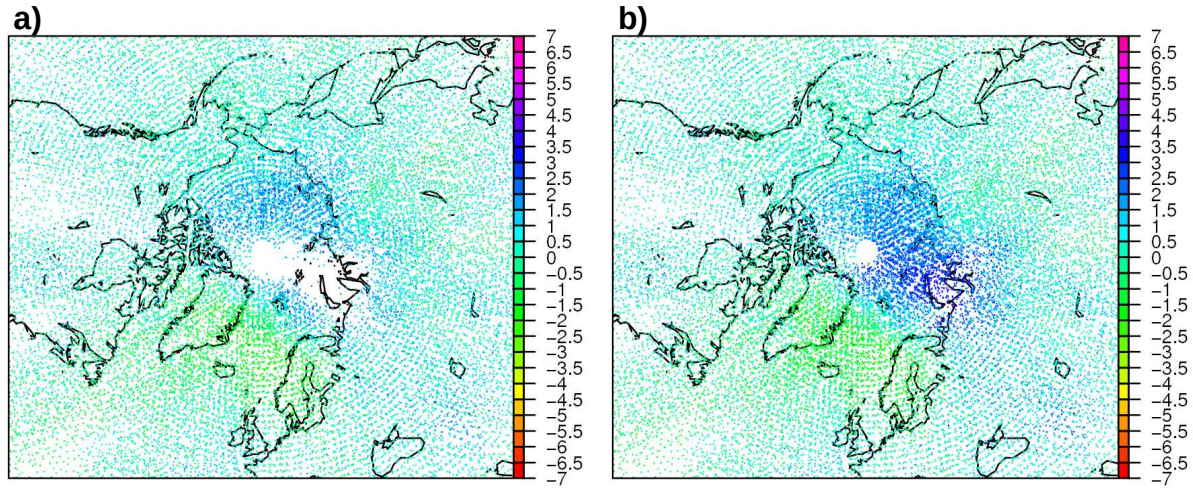


Figure 14: FG departures for all used AMSU-A channel 13 data for the assimilation cycle covering 15 January 21 UTC to 16 January 9 UTC 2012, taken from the T511 experiments. a) EF experiment, b) SES experiment.

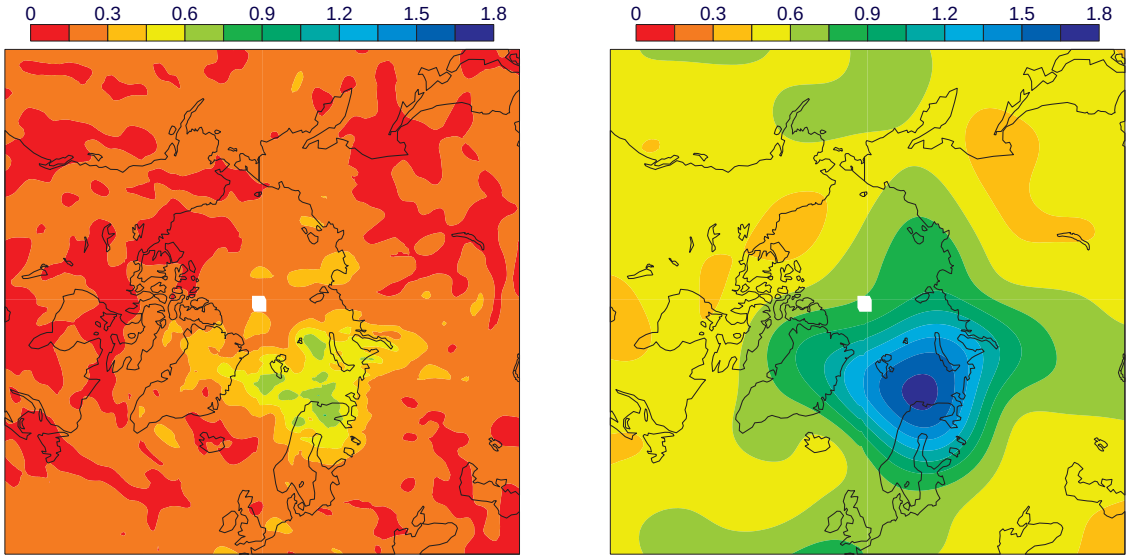


Figure 15: Un-calibrated and unfiltered ensemble spread for AMSU-A channel 13 (ES, left) and corresponding SES field (right) for 15 January 21 UTC.

changes of the temperature field in the upper stratosphere. The generation of spurious gravity waves is more severe at higher spatial resolution. An alternative approximation in the semi-Lagrangian scheme is currently being investigated to solve this problem, showing very promising results (Diamantakis 2013, pers. communication).

Apart from the size of the problem in high-resolution experiments, a number of reasons contribute to the less successful performance of the quality control with the new approach for high-resolution experiments for this particular case. Firstly, the scaling of the EDA spread is limited to a maximum of 3. While this is a reasonable choice for normal conditions and the value is only rarely exceeded, the SSW conditions

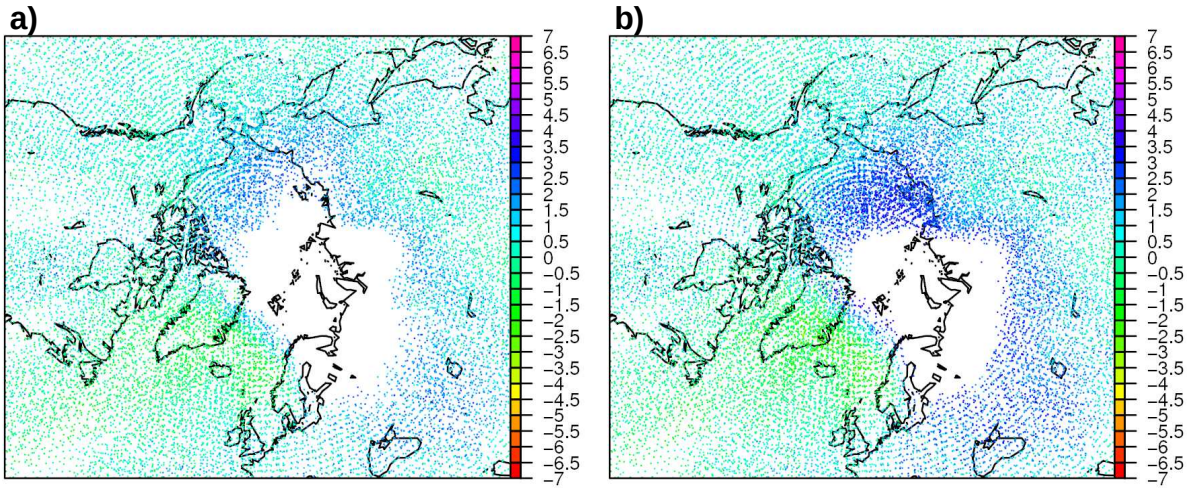


Figure 16: As Fig. 14, but for the experiments run at a resolution of T1279.

suggest scaling factors in excess of 5 (e.g., Fig. 6), so the limit of 3 is too restrictive for this extreme event, and it could be considered to relax it in the future. Secondly, given the resolution-dependence of the underlying model issues, an EDA run at a resolution of T399 cannot be expected to fully represent this uncertainty in the higher-resolution system, even if it adequately represents the underlying model uncertainties. Thirdly, the problem leading to the rejections is a large-scale bias in the forecast model for these very specific and extreme conditions, and this bias is consistent over several days. It is not clear whether this should indeed be addressed through a locally enhanced background error, or, instead, through a local model bias. Ideally, it would of course be best to address the model deficiencies themselves, and adjustments to the model numerics indeed show very promising results. The inflation of EDA errors in these circumstances should therefore be seen as a temporary solution.

5 Evolution of the EDA spread over the assimilation window

So far, our analysis has neglected the evolution of the background error over the assimilation window. We will now investigate this aspect on the basis of FG-departure statistics as well as the evolution of the EDA spread over the time window.

Variances of FG-departures for AMSU-A and MHS show a notable increase as a function of the position of the observation in the assimilation window (e.g., Figures 17 and 18): on average, observations at the beginning of the assimilation window have smaller departures, whereas observations towards the end of the assimilation window show the largest departures. This is related to the evolution of the background error over the assimilation window, together with the addition of model error. The observation error contribution is instead expected to be constant over the assimilation window. The increase is relatively small for the temperature-sounding channels of AMSU-A, but more significant for the humidity-sounding channels of MHS. A similar increase of the size of FG-departures has previously been noted for the assimilation of clear-sky radiances from geostationary satellites which are particularly suited to highlighting this thanks to their high temporal resolution and the geographically fixed viewing geometry (e.g., Munro et al 2004).

Neglecting the temporal evolution of the size of the FG-error means that the FG-check rejects more

observations towards the end of the assimilation window. This is most noticeable for MHS channel 3, where the number of observations rejected by the FG-check rises from 0.1 % at the beginning of the assimilation window to just over 0.6 % at the end (Fig. 19). The effect is less of an issue for AMSU-A, where only 0.1 % more observations are rejected at the end of the assimilation window compared to the beginning (not shown). Even though this aspect is suboptimal, the percentage of observations that are rejected by the FG-check is generally small for the ATOVS sounders (less than 1 %).

The departure characteristics shown in Figures 17 and 18 can be compared with the evolution of the EDA spread in radiance space over the assimilation window. To do so, the EDA spread was calculated in 3-hourly intervals over the assimilation window, instead of just at the beginning as done in previous sections. As can be seen in Figures 17 and 18, the EDA spread also increases over the assimilation interval, but the slope of the increase tends to be not as steep as the increase seen from the FG-departures (with the exception of AMSU-A channel 5, the lowest of the AMSU-A temperature sounding channels

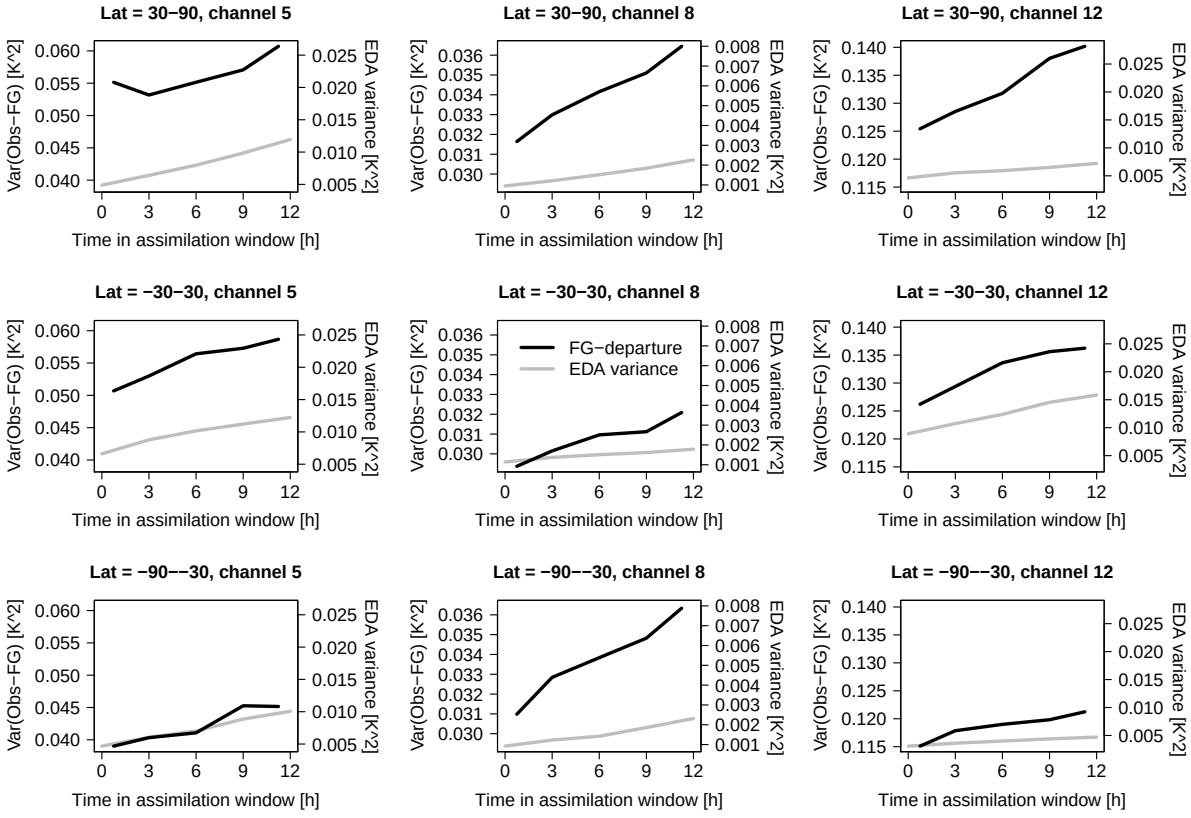


Figure 17: Variances of FG-departures (black, left x-axis) and EDA variance (grey, right x-axis) as a function of the position in the 12-hour assimilation interval for selected AMSU-A channels. The FG-departure statistics are based on data from all AMSU-A instruments combined, for February 2012, after geophysical quality control, but before the FG-check. They have been binned into 3-hourly intervals based on their observing time (the first and last bin are 1 1/2 hours wide only, to allow better comparison to the EDA spread which was only available at certain times in the assimilation window.). The panels show statistics for the Northern Hemisphere extra-tropics (top row), tropics (middle), and Souther Hemisphere extra-tropics (bottom), for channel 5 (left column), channel 8 (middle), and channel 12 (right). Although the range of the x-axes for the departure variance and the EDA variance are different, the scale of the two is consistent.

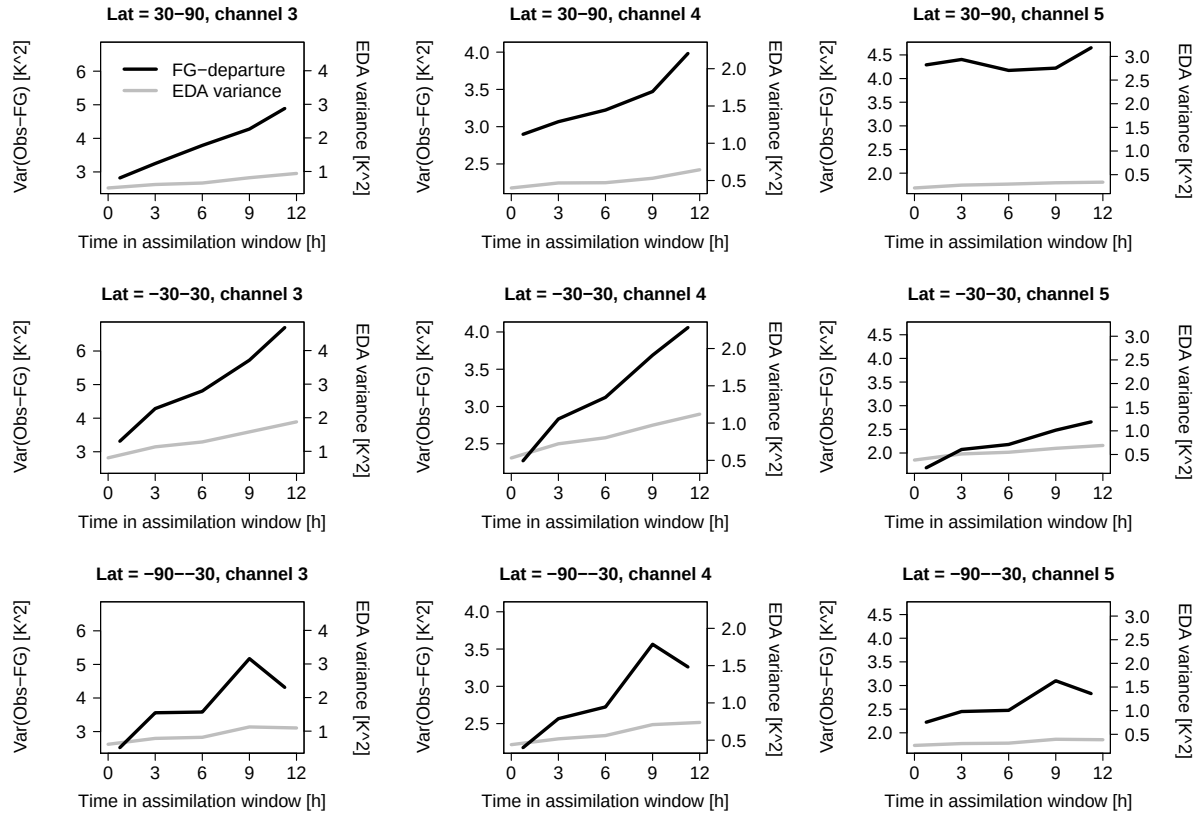


Figure 18: As Fig. 17, but for the MHS channels 3 (left), 4 (middle), and 5 (right).

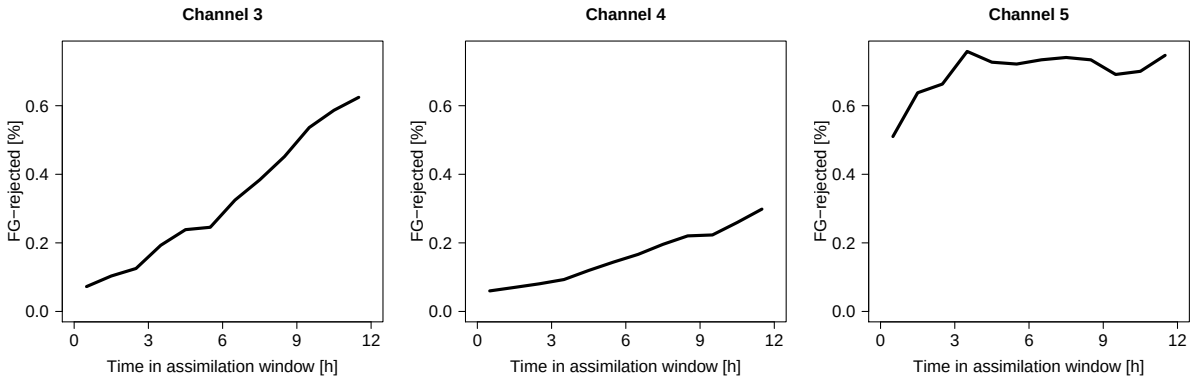


Figure 19: Percentage of observations rejected by the FG-check as a function of the position of the observation in the assimilation window for the 3 MHS channels assimilated. Statistics are based on the SES experiment for January-March 2012, using all MHS instruments considered.

considered). This is especially the case for the stratospheric channels over the Northern Hemisphere extra-tropics (see, for instance, channel 12 in Fig. 17).

Figure 20 quantifies to what extent the EDA spread underestimates the temporal evolution of the errors in

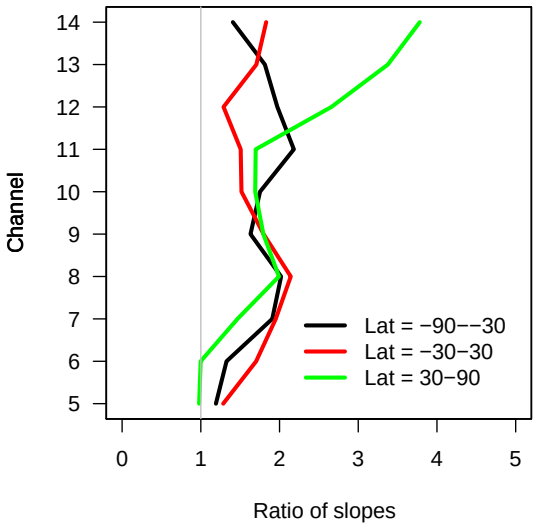


Figure 20: Square root of the ratio of the slope of the increase in the variance of FG-departures over the assimilation window and the slope of the increase in the EDA variance over the assimilation window for the assimilated AMSU-A channels. The results are based on the data shown in Fig. 17.

the FG-trajectory. To do so, we calculated the slopes of the FG-departure/time and EDA-variance/time relationships, based on linear fits to the relationships shown in Fig. 17. Figure 20 shows the square root of the ratio of these two slopes. In other words, this is the scaling of the EDA spread that would be required to achieve the temporal growth suggested by the FG-departures. For most channels, the standard deviation of FG-departures increases about twice as fast as the increase in the spread in the EDA. It is interesting to compare this to the scaling derived from the spread-skill relationships (Fig. 4a). While the latter suggests that the EDA spread requires hardly any inflation over the tropics, the temporal evolution of the EDA spread suggests that the EDA nevertheless underestimates the error growth over the 12-hour assimilation window by a factor of 1.5-2 for the tropical troposphere. In contrast, for the mid and upper stratospheric AMSU-A channels (12-14), both statistics point to a considerable underestimation of the errors in the northern hemisphere extra-tropics for this time of the year. The growth in the EDA-spread over the considered time-intervals is largely driven through the stochastic physics, and some tuning of these contributions may be possible in the future.

Given the above findings, the evolution of the EDA spread over the assimilation window could be used to better describe the evolution of the FG error over the assimilation window. However, given the relatively small number of observations affected, the added complexity is likely to be of relatively little benefit, and it has therefore not been considered for implementation.

6 Conclusions

In this memorandum we have investigated the EDA spread in radiance space for ATOVS sounders, and we developed an approach to use filtered and calibrated versions of the EDA spread in quality control decisions. The approach replaces the earlier scheme of estimating background errors through a randomisation approach. The main findings are:

- Regions of larger EDA spread in radiance space are associated with larger standard deviations of FG-departures, as would be expected if the EDA spread diagnoses areas of larger background errors.
- The radiance departures suggest that the EDA is under-dispersive, particularly over the extra-tropics, and scaling factors greater than unity are required to obtain estimates of background error in radiance space that are consistent with departure statistics.
- Comparisons between the radiance inflation suggested by radiances with that suggested by radiosonde data suggests some dependence of the required inflation on the vertical and horizontal error scales, and this could be investigated further.
- The calibrated EDA spread can be used for situation-dependent quality control decisions to filter out observations that deviate further than expected from the FG, with, on average, neutral impact on forecast scores.

Our findings confirm that there currently is a clear need to calibrate the EDA spread to provide quantitatively meaningful estimates of the background error. It is hoped, that over time the need for such calibration will reduce, with improvements to the overall performance of the EDA. It should be stressed here that the calibration of the EDA spread introduced for ATOVS data is technically decoupled from the calibration performed for the background errors that are assumed for the control variables during the 4DVAR assimilation. The calibration used here is based on observation departures, whereas the calibration for the 4DVAR-background errors is based on analyses. This is an important caveat when using the new SES fields for diagnostic purposes.

The present memorandum gives a first glimpse at using observation-based evaluations of the EDA spread. The approach taken here has been tailored to providing background error estimates for quality control decisions for observations, but the memorandum has also highlighted the potential for further observation-based diagnostics of the EDA. The main advantage of such diagnostics is that they avoid the correlations between analysis error and short-term forecast error that affects analysis-based diagnostics particularly at short forecast ranges. The following points are worth highlighting or deserve further evaluation:

- While there is qualitative agreement between aspects of the scaling performed in the analysis-based and the observation-based calibration of the EDA-spread (e.g., larger scaling required in the extra-tropics), there are also clear quantitative differences.
- There are suggestions of a possible dependence of the required calibration on the vertical scales represented in the observations.
- The growth of the EDA-spread over the 12-hour assimilation window provides an additional diagnostic and comparisons to departure-statistics may indicate how to fine-tune aspects that influence the error growth.
- For radiance observations from satellites it is clear the observation error needs to be taken into account when evaluating the EDA spread with observation departures. The statistics suggest that the size of the background error for the lower AMSU-A channels is frequently well below the instrument noise.

This memorandum also prompts the question whether observations have more of a role to play in the calibration of the EDA for the provision of background errors used in the assimilation. In practice,

it is difficult to achieve a consistent calibration for all control variables used in the assimilation that is only based on observations. This is mostly because not all control variables are directly observed equally and consistently. A direct calibration of the EDA through observations is hence not considered feasible. However, it is clear from the results presented here that observation-based diagnostics can provide additional evaluation of the calibration, without relying on the assumption that analysis errors are zero. At the same time, it is also clear that such diagnostics require careful treatment of observation errors. These diagnostic aspects need to be developed further, and may shed some light on how to relax the assumption of zero analysis errors in the analysis-based calibration step in the future.

The scheme described here opens a few other avenues, both scientifically and technically. The information provided in the new SES fields could be explored further, for instance in flow-dependent thinning decisions, to optimise the use of observations in areas where the background uncertainty is larger. Also, the estimates from the randomisation approach are not needed after the introduction of this new approach, and this allows a substantial reduction in the number of iterations in the first minimisation, with associated savings in computational cost during the analysis. The new scheme is included in cycle 40r1.

Appendix I: Randomisation-based background error estimate

In current operations, the FG-check uses an estimate of the background error in radiance space that is calculated during the first minimisation of the previous assimilation cycle. The calculation uses a randomisation approach and proceeds in three steps: 1) The background error is mapped to radiance space using a sample of 51 perturbations drawn from the background error covariance assumed in the assimilation. The calculations are done in in grid-point space using the tangent linear of the radiative transfer model. Surface emissivity and skin temperature perturbations are also included. The calculations are done for all channels of selected instruments (HIRS, MSU, SSU, AMSU-A, AMSU-B, SSMI), each using radiative transfer specifications for just one satellite and assuming nadir conditions. 2) Analysis error are estimated from the above fields, using the method described in Fisher and Courtier 1995. 3) The background error for the subsequent assimilation cycle is estimated for all channels by applying a model for the error growth in the 12 forecast, applicable to the 500 hPa geopotential. The resulting field is stored in MARS as “EF” fields. Steps 2 and 3 are necessary as the field calculated in 1) is only available during the first minimisation of the current analysis, and therefore cannot be used in the FG-check during the screening.

Figure 21 shows an example of the background error field in radiance space after step 1 and before step 2 and 3, valid for the same time as the fields shown in Fig. 7. This allows a comparison of the SES fields and the background errors applied in the assimilation mapped to radiance space through the randomisation method, without the analysis error and error propagation step. Comparing the fields in Fig. 21 with the SES fields in the left column of Fig. 7 shows that there is better agreement in the geographical structures of the errors, especially for the MHS humidity channel shown. This is because both take the temporal evolution of the error into account in the same way, whereas the EF field shown in the right column of Fig. 7 mimics the error evolution through a simple, static error growth model. Note, however, that the magnitudes of the errors shown in Fig. 21 and the left column of Fig. 7 differ considerably, due to the different calibration applied in both cases.

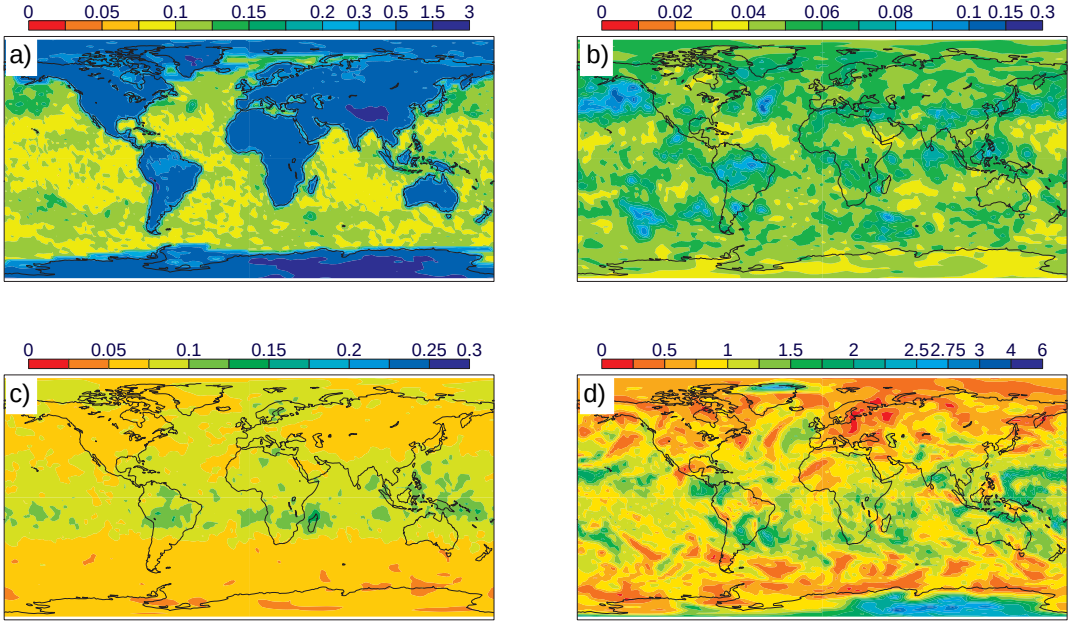


Figure 21: Examples of the background errors used for the control variables in the assimilation, mapped into radiance space with the randomisation method. The fields are valid for for 15 February 2012, 9 UTC, and the panels show AMSU-A channels 5 (a), 8 (b), and 12 (c), and MHS channel 3 (d).

References

Andersson, E., and H. Järvinen, 1999: Variational quality control. *Q. J. R. Meteorol. Soc.*, **125**, 697–722.

Bonavita, M., L. Isaksen, and E. Holm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. Technical memorandum 664, ECMWF, 31pp.

Bormann, N., and P. Bauer, 2010: Estimates of spatial and inter-channel observation error characteristics for current sounder radiances for NWP, part I: Methods and application to ATOVS data. *Q. J. R. Meteorol. Soc.*, **136**, 1036–1050.

Dee, D., 2004: Variational bias correction of radiance data in the ECMWF system. In ECMWF Workshop on Assimilation of High Spectral Resolution Sounders in NWP, ECMWF, Reading, UK, 97–112.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, **131**, 3385–3396.

Fisher, M., and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational dataassimilation. Technical memorandum 220, ECMWF, 28pp.

Flowerdew, J., and N. Bowler, 2011: Improving the use of observations to calibrate ensemble spread. *Q. J. R. Meteorol. Soc.*, **137**, 467–482.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. Technical memorandum 636, ECMWF, 45pp.

Munro, R., C. Köpken, G. Kelly, J.-N. Thépaut, and R. Saunders, 2004: Assimilation of Meteosat radiance data within the 4d-var system at ECMWF: Data monitoring, bias correction and single-cycle experiments. *Q. J. R. Meteorol. Soc.*, **130**, 2293–2313.

Tavolato, C., and L. Isaksen, 2010: Huber norm quality control in the ifs. *ECMWF Newsletter*, **122**, 27–31.