

# Enhancing the impact of IASI observations through an updated observation error covariance matrix

Niels Bormann, Massimo Bonavita, Rossana Dragani, Reima Eresmaa, Marco Matricardi, and Tony McNally

Research Department

July 2015

*This paper has not been published and should be regarded as an Internal Report from ECMWF.  
Permission to quote from it should be obtained from the ECMWF.*



Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/en/research/publications>

Contact: [library@ecmwf.int](mailto:library@ecmwf.int)

©Copyright 2015

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director-General. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

## Abstract

The present memorandum investigates the use of an updated observation error covariance matrix for IASI in the ECMWF system. The new observation error covariance matrix is based on observation-space diagnostics and includes inter-channel error correlations, but also assigns significantly altered error standard deviations. The use of the new observation error is investigated in detail in assimilation experiments, including an assessment of the role of error inflation and taking inter-channel error correlations into account. The influence of the observation error update on the Ensemble of Data Assimilations (EDA) used for background error specification is also examined.

The updated observation error covariance leads to a significant improvement in the use of IASI data, especially in the tropics, the stratosphere, and for humidity. The benefits are particularly strong for the short-range forecasts, whereas the impact in the medium range is less pronounced. The update also has a particularly large positive impact on the ozone analysis, related to especially large modifications in the observation error for ozone-sensitive channels. The observation error update leads to a modified spread in the EDA, with some reductions in spread in areas where improved short-range forecast impact is diagnosed.

The study highlights the benefits of taking inter-channel error correlations into account, which allows the use of an observation error covariance for IASI that is overall more consistent with departure statistics. At the same time, the study also demonstrates that error inflation can be used to partially, though not fully, compensate for neglected error correlations. Adjustments such as scaling of the originally diagnosed observation error estimates are found beneficial also when inter-channel error correlations are taken into account.

## 1 Introduction

This Technical Memorandum reports on experimentation with a new observation error covariance matrix for the hyperspectral Infrared Atmospheric Sounding Interferometer (IASI) in the ECMWF system. The aim is to make the assumed observation errors more consistent with departure statistics. A new feature is that estimates for inter-channel error correlations are explicitly taken into account, but substantial changes are also made to the diagonals of the matrix. The developments are a contribution to the more effective use of observations from hyperspectral infrared instruments, and also add to a refined treatment of observational uncertainties in data assimilation in general.

The IASI instrument has been successfully assimilated operationally in the ECMWF system for many years, with a positive influence on the assimilation system (e.g., Collard and McNally 2009). Observations from up to 191 channels are assimilated, making IASI the observing system with the largest number of assimilated observations. The majority of assimilated channels are temperature-sounding and window channels in the long-wave CO<sub>2</sub> band, but some humidity and ozone channels have also been added (e.g., Han and McNally 2010, Dragani and McNally 2013).

The ECMWF system currently uses a relatively simple specification of observation errors for IASI, in line with the initial use of IASI data at other Numerical Weather Prediction centres (e.g., Hilton et al. 2009, Guidard et al. 2011): the error is assumed to be diagonal, and the assigned observation error is a constant over three wavenumber bands (e.g., Collard and McNally 2009). The setting of this observation error is loosely based on standard deviations of background departures, but with ad-hoc, and in parts substantial inflation over certain spectral regions, as considered necessary.

The specification of observation errors is an essential step for the successful assimilation of any observation. The assigned observation errors together with the specified background errors determine the weighting of the observation in the assimilation system. Apart from measurement errors such as instru-

ment noise, the observation error should also include other sources of random error, such as forward model error, representativeness error, quality control error, etc (e.g., Daley 1993). These latter errors will, most likely, exhibit more complex structures, with situation-dependence and error correlations between different observations. While the refinement of background errors has received steady attention since the inception of data assimilation and has reached considerable sophistication (e.g., Bonavita et al. 2012), a more advanced treatment of observation errors has only recently been emerging. Accounting for situation-dependence or error correlations are avenues that are being pursued with considerable success (e.g., Forsythe and Saunders 2008, Geer and Bauer 2011, Salonen and Bormann 2012, Weston et al. 2014).

Assumed observation error characteristics should, as far as practical, reflect the statistical properties of the true total observation error, requiring a reliable estimate of this total observation error. One way to obtain this is through an error inventory, that is, by estimating the uncertainties in all contributions to the total error. This is not always straightforward, but some work in this direction has been reported by Ventress and Dudhia (2014) in the context of IASI, and effort has also commenced at ECMWF (Chun 2015, pers. communication).

Alternatively, an estimate of the total observation error can be obtained based on departure statistics from assimilation systems, using methods such as those developed by Hollingsworth and Lönnberg (1986) or Desroziers et al. (2005). These methods make use of the fact that background departures reflect the combination of true observation and background errors. The methods then aim to separate the contributions from background and observation error by making assumptions on the structure of the background errors or the weights given to observations in an assimilation system. Such methods have been applied to hyperspectral infrared data in a range of assimilation systems by numerous authors in recent years (e.g., Garand et al. 2007, Bormann et al. 2010, Stewart et al. 2013) with similar results: all authors find that the estimates for the error standard deviations are significantly lower than the values typically assigned to IASI observations, but many channels exhibit also significant inter-channel or spatial error correlations, especially in the water vapour band. These error correlations are commonly neglected when specifying observation errors in assimilation systems, and as a pragmatic counter-measure, observation errors are instead inflated.

The assimilation diagnostics have the advantage that they estimate the total observation error, consistent with departure statistics. However, the diagnostics are dependent on the assumptions made to derive them, such as assuming no spatial error correlations from observations in the case of Hollingsworth and Lönnberg (1986), or assuming weights in the assimilation system consistent with true weights in the case of Desroziers et al. (2005). Both methods also assume that there are no correlations between background and observation errors. As these assumptions may not be strictly valid, the error estimates will have some uncertainty. Nevertheless, the consistency of the results from different centres with different methods by different authors adds credibility that these diagnostics yield meaningful estimates of the true observation error. As a result, the natural question arises: to what extent can the diagnostics be used to improve our specification of the observation error covariance for IASI, and in particular, what role do error correlations play for the assimilation of IASI observations?

In the following report, we will investigate these two questions, that is we investigate the use of an observation error covariance matrix for IASI that is based on departure diagnostics and includes inter-channel error correlations. The work is similar to that of Weston et al. (2014), who used diagnosed matrices to update the observation error covariance matrix for IASI in the Met Office system. They found considerable benefits over an earlier ad-hoc specification of the observation error, albeit needed to make substantial modifications to the diagnosed matrices. These modifications were attributed to conditioning problems that otherwise led to poor convergence.

The structure of this memorandum is as follows. We first provide an overview of the IASI instrument and its use in the ECMWF system. We then introduce the observation error covariance matrix investigated in the present memorandum. This matrix will first be used to study the role of error inflation versus taking inter-channel error correlations into account. We then discuss the impact of the observation error covariance upgrade in assimilation trials. Initially this is studied using experiments with a lower spatial resolution without taking into account feedback on the specification of the background errors. This is followed by a discussion of results with higher resolution experiments that also investigate the effect of the observation error upgrade on background error estimates. Finally, overall conclusions are provided in the last section.

## 2 IASI instrument

IASI is an infrared interferometer with 8,461 channels flown on the METOP series of polar orbiters. It covers the spectral interval from  $645 - 2,760 \text{ cm}^{-1}$  with a spectral sampling of  $0.25 \text{ cm}^{-1}$  (Chalon et al. 2001). Currently, two such instruments are operational in space, on Metop-A and Metop-B.

In the present study, up to 191 IASI channels are assimilated, and these are summarised in Table 1 and Fig. 1. They include temperature-sounding, window, ozone, and humidity channels. The initial assimilation choices for IASI are outlined in Collard and McNally (2009). The bulk of the assimilated data are observations unaffected by clouds, identified using the scheme of McNally and Watts (2003) which looks for cloud-contamination based on evaluating FG-departure signatures. The scheme has been subsequently refined, for instance, by taking into account information on clouds from a collocated imager (Eresmaa 2014). The cloud detection scheme is applied to temperature-sounding channels; for the water-vapour and ozone band, the cloud-screening is linked to the results from the temperature-sounding

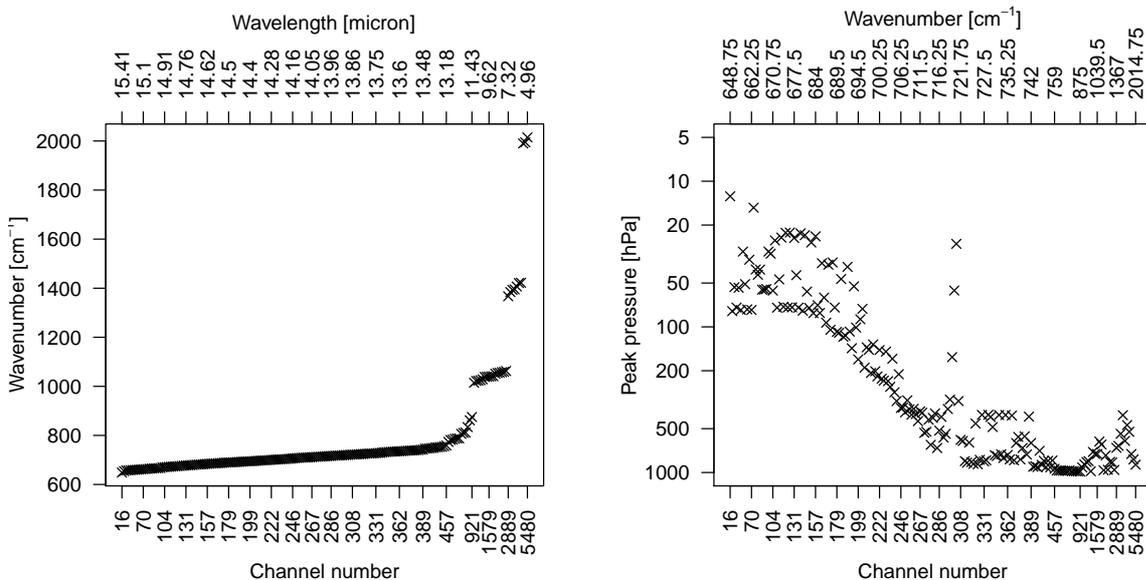


Figure 1: Wave lengths [ $\mu\text{m}$ ] (left) and pressures of the Jacobian peaks [hPa] (right) as a function of channel number for the 191 IASI channels used in this study. The upper x-axis also gives wavenumbers [ $\text{cm}^{-1}$ ]. Note that the x-axis is linear in the channel index and not the channel number, as is done throughout this report.

Table 1: Groups of IASI channels.

Group	Number of channels	Channel numbers	Wave-numbers $\text{cm}^{-1}$
Upper temperature sounding	80	16 - 241	648.75 - 705.00
Mid temperature sounding	28	246 - 306	706.25 - 721.25
Lower temperature sounding	33	308 - 386	721.75 - 741.25
Window	24	389 - 921	742.00 - 875.00
Ozone	16	1479 - 1671	1014.50 - 1062.50
Humidity sounding	10	2889 - 5480	1367.00 - 2014.75

channels. Cloud-affected data originating from completely overcast scenes are assimilated as well, using the methods described in McNally (2009). No IASI radiances are currently used over land.

Systematic errors between observed and modelled IASI observations are removed through variational bias correction (e.g., Dee 2014). The bias correction models are similar to those used for other sounder radiances at ECMWF. They consist of a linear model for the air-mass bias, with a constant component and four layer thicknesses calculated from the FG as predictors (1000-300 hPa, 200-50 hPa, 50-5 hPa, 10-1 hPa). Scan biases are modelled through a 3rd-order polynomial in the scan-angle. No air-mass bias correction is used for some window and lower sounding channels (380 - 1180 and 1820 - 2200), to avoid unwanted interaction between the cloud detection and the variational bias correction (e.g., Auligné and McNally 2007).

Further details on the assimilation of IASI data can be found in Collard and McNally (2009), with updates in McNally (2009), Han and McNally (2010), Dragani and McNally (2013), and Eresmaa (2014).

### 3 New observation error covariance matrix

#### 3.1 Errors and correlations

The observation error covariance matrix used in this study is shown in Figures 2 and 3 in terms of the error standard deviation ( $\sigma_0$ ) and a correlation matrix. This matrix has been derived using the departure-based diagnostic methods applied in Bormann et al. (2010), with some further adjustments. The derivation and the adjustments are described in more detail in Appendix A. The unscaled diagnosed matrix shows the features common to similar departure-based estimates (e.g., Garand et al. 2007, Bormann et al. 2010, Stewart et al. 2013), that is: 1) error standard deviations close to an average instrument noise estimate for upper tropospheric and stratospheric temperature sounding channels, with little error correlations; 2) error standard deviations much larger than the instrument noise for water vapour channels, combined with significant inter-channel error correlations; 3) error standard deviations larger than the instrument noise for lower temperature sounding, window, and ozone channels together with weaker, but still significant inter-channel error correlations. Error correlations introduced through apodisation are also apparent for neighbouring channels or near-neighbours, albeit somewhat reduced compared to theoretical values as a result of the adjustments described in Appendix A. It should be noted here that the instrument noise estimate shown in Fig. 2 has been converted from radiance to brightness temperature space using brightness temperatures for a standard atmospheric profile. As this conversion is non-linear and the instrument noise is only constant in radiance space, the actual instrument noise in brightness temperature space is instead dependent on the scene temperature. This effect is not considered throughout this memorandum, neither for the diagnostics nor the instrument noise estimate, and instead only globally averaged statistics

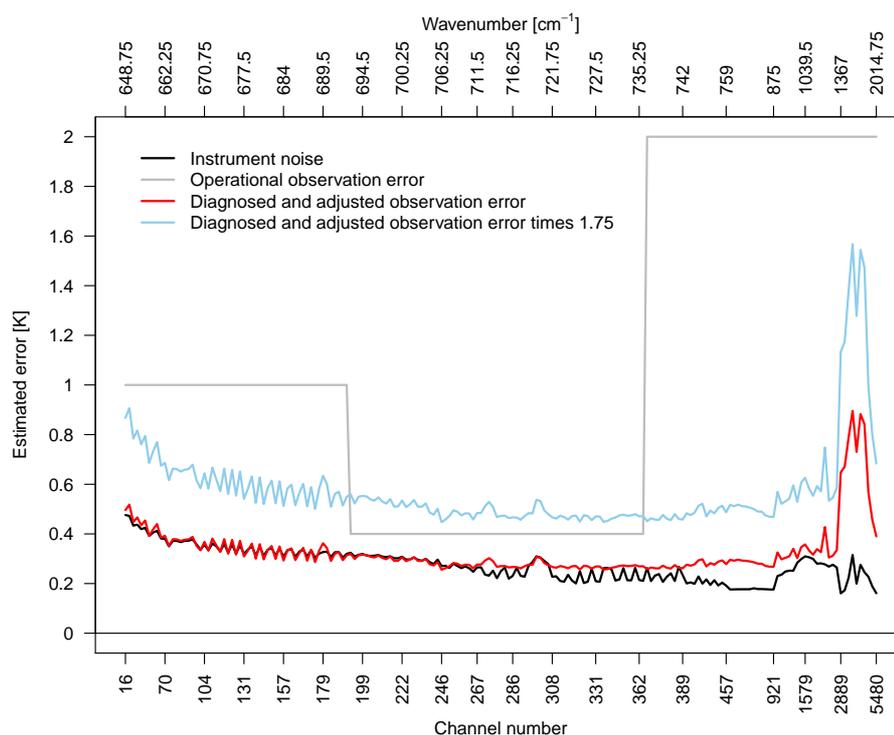


Figure 2: Diagnosed and adjusted observation error ( $\sigma_O$ ) for assimilated IASI channels (red), together with an estimate for the instrument noise (black) and the observation error currently assumed in the operational ECMWF system (grey). Also shown is the diagnosed observation error times an inflation factor of 1.75 (blue). The instrument noise has been converted from radiances to brightness temperatures using a mean scene temperature per channel. See the main text and Appendix A for further details about the derivation of the diagnosed observation errors, and the adjustments applied to them.

are shown. Figure 2 also includes the currently assumed observation error for IASI, which is significantly larger than that suggested by these diagnostics, albeit does not take into account any error correlations.

As evident from Fig. 2, the diagnostics suggest a rather large contribution from observation errors other than instrument noise for many channels. It is beyond the scope of this paper to investigate the origin of these errors, even though such an investigation would of course provide further insights that may well be useful to optimise the assimilation of IASI data in various ways. Depending on the spectral region, leading contributors are expected to be representativeness error, cloud screening error and radiative transfer error. Based on diagnostics from assimilation systems run at different resolutions, Weston et al. (2014) argue that the error correlations for the water vapour channels and some window/lower sounding channels are largely due to representativeness error, primarily a mis-match between the spatial scales represented in the forecast model and the observations. In addition, diagnostics from the ECMWF system show reduced error correlations when the cloud detection improvements described in Eresmaa (2014) were implemented, giving some indication that residual cloud contamination is another significant source.

It should be noted that the observation error covariance matrix shown here has undergone an adjustment to improve its use in the data assimilation system. The adjustment modifies the smallest eigen-values of the diagnosed matrix, and the motivation for the adjustment and its impact are described in detail in Appendix A.

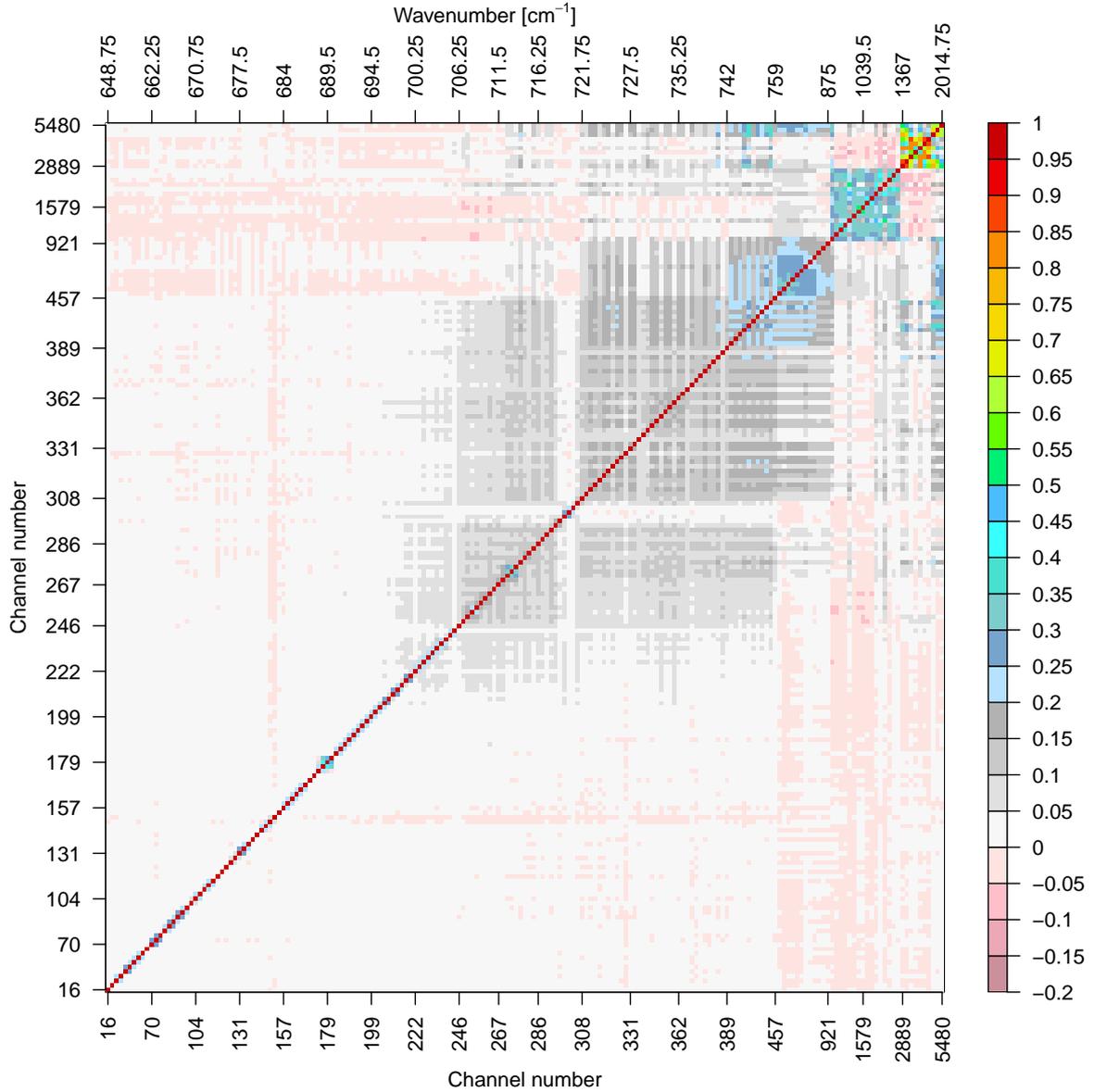


Figure 3: Observation error correlations used in this study for assimilated IASI channels. See main text and Appendix A for further details.

### 3.2 Properties

It is worthwhile to explore what the use of the inter-channel error correlation matrix implies for the assumed errors in IASI observations. To do that, let us consider the observation cost function, and rewrite the observation error covariance matrix  $\mathbf{R}$  in terms of a diagonal matrix  $\mathbf{\Sigma}_O$  with the error standard deviations on the diagonal and a correlation matrix  $\mathbf{C}$ :

$$J_O = \mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}$$

$$\begin{aligned}
&= \mathbf{d}^T (\boldsymbol{\Sigma}_O \mathbf{C} \boldsymbol{\Sigma}_O)^{-1} \mathbf{d} \\
&= \mathbf{d}^T (\boldsymbol{\Sigma}_O \mathbf{E} \boldsymbol{\Lambda} \mathbf{E}^T \boldsymbol{\Sigma}_O)^{-1} \mathbf{d} \\
&= \mathbf{d}^T \boldsymbol{\Sigma}_O^{-1} \mathbf{E} \boldsymbol{\Lambda}^{-1} \mathbf{E}^T \boldsymbol{\Sigma}_O^{-1} \mathbf{d} \\
&= (\boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T \boldsymbol{\Sigma}_O^{-1} \mathbf{d})^T \boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T \boldsymbol{\Sigma}_O^{-1} \mathbf{d}
\end{aligned} \tag{1}$$

Here,  $\mathbf{d}$  is the vector of background departures,  $\boldsymbol{\Lambda}$  is the diagonal matrix of the eigen-values of the error correlation matrix, and  $\mathbf{E}$  is the orthogonal matrix whose columns are the eigen-vectors of  $\mathbf{C}$ .

It can be seen that a non-diagonal  $\mathbf{C}$  is equivalent to an adjustment of the observation error that depends on the spectral structures of the departures: the departures are first normalised by the error standard deviations as usual (term  $\boldsymbol{\Sigma}_O^{-1} \mathbf{d}$ ), but this is followed by a projection onto the eigen-vectors of  $\mathbf{C}$ , and a normalisation by the square root of the associated eigen-values (see the term  $\boldsymbol{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T$ ). For a given eigen-vector, this additional normalisation is equivalent to an inflation or deflation of the observation errors compared to the case of using a diagonal  $\mathbf{R}$ , depending on whether the associated eigen-value is larger or smaller than 1. The square root of the associated eigen-values of  $\mathbf{C}$  are therefore providing inflation/deflation factors for the observation errors. In the assimilation system, these act in a situation-dependent way according to the spectral signatures of the observation departures.

The first and last few eigen-vectors and the square roots of their eigen-values are shown in Fig. 4. As can be seen, the leading eigen-vector represents structures with broad spectral features, similar to features expected from residual cloud contamination. Errors for such structures will be increased most heavily, with

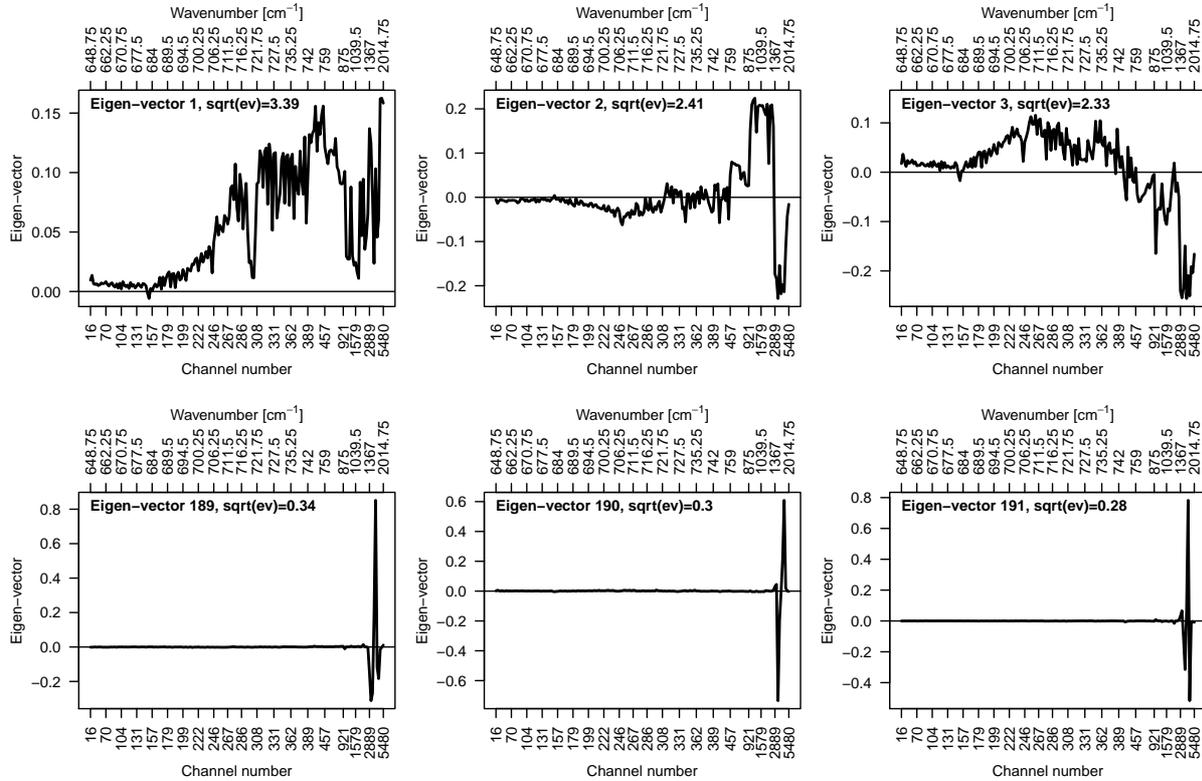


Figure 4: The first three (top) and last three (bottom) eigen-vectors of the error correlation matrix shown in Fig. 3. Also given are the square root of the eigen-values in each panel.

an inflation factor of 3.39 for the first eigen-vector. That is, the error for such structures is much larger than would otherwise be assigned if the error correlations were neglected (and the same  $\sigma_0$  was used). In contrast, the highest-order eigen-vectors represent differences between groups of humidity channels, and these are assigned a much smaller error compared to when the error correlations are ignored, by less than a third. Eigen-values less than one are also found for other eigen-vectors that represent differences between groups of temperature-sounding channels, for instance neighbouring channels affected by apodisation (not shown in Fig. 4).

The analysis of the eigen-vectors and eigen-values of the error correlation matrix has two main implications: Firstly, introducing error correlations will modify the weighting of the observations in a situation-dependent way, dependent on the structure of the departures. For departures that project primarily onto the leading eigen-vectors of the correlation matrix, taking error correlations into account will lead to a down-weighting of the data. In contrast, when departures project primarily onto the higher-order eigen-vectors, taking the error correlations into account will lead to an increase in the weight assigned to these spectra. This effect has been demonstrated already in single-spectra or 1DVAR experiments (e.g., Bormann and Collard 2012, Weston et al. 2014). Secondly, as the information on vertical resolution contained in IASI data originates from the differences between channels peaking at different altitudes, the structure-dependent representation of the errors through the error correlations will affect the ability to retrieve vertical resolution from IASI data. We will get back to both of these aspects later in the report.

## 4 Inflation vs accounting for error correlations

We will now investigate the role of error inflation and accounting for inter-channel error correlations for the assimilation of IASI data. We use the term “error inflation” to describe the method of assuming observation errors that are significantly larger than the true observation errors. Error inflation is a method commonly applied to counter-act some effects of neglected error correlations. Inflation factors of 2-3 are not uncommon, and Fig. 2 suggest, for instance, inflation factors of around 1.5-8.0 relative to the diagnosed observation error for the operationally used observation error for IASI at ECMWF. The approach taken is the same as in Bormann and Collard (2012), but the results are obtained with a more recent configuration of the ECMWF assimilation system.

### 4.1 Experiments

To investigate the role of error inflation and inter-channel error correlations, we perform two series of assimilation experiments with different assumed observation errors for IASI. In the first series (“NoCor”), the assumed observation error is equal to the diagnosed  $\sigma_0$ , but multiplied with a scaling factor ranging from 1.0 - 4.0. In this series, the diagnosed error correlations are ignored, and a diagonal matrix is assumed. In the second series (“Cor”), we also assign  $\sigma_0$  to be scaled versions of the diagnosed values, but in this case we also take the diagnosed inter-channel error correlations into account. This is achieved in the assimilation system without the explicit inversion of the observation error covariance matrix, using a Cholesky decomposition. The added computational cost of this is negligible in the context of 4DVAR assimilation experiments. To put our results into context, we also perform a “Denial” experiment, in which no IASI data is assimilated at all, and a “Control” experiment, in which the current operational observation error is used.

All these experiments have been run for 3 months for the period 5 February 2014 - 4 May 2014, with a model resolution of  $T_L511$  ( $\approx 40$  km), a final incremental analysis resolution of  $T_L255$  ( $\approx 80$  km),

and 137 levels in the vertical. We use ECMWF's hybrid 4DVAR system in which a flow-dependent background error covariance is provided through an Ensemble of Data Assimilations (EDA, Isaksen et al. 2010). The effect of altering the observation errors or denying IASI data has not been taken into account in the background modelling and instead the same background error specification is used for all experiments. All experiments use a 12-hour assimilation window, and assimilate all other operationally used observations. The experiments that include IASI data make use of IASI from Metop-A and Metop-B, and the same observation error covariance is used for both instruments.

## 4.2 Results

The performance of the two series of assimilation experiments is primarily assessed through the influence on departure statistics for other assimilated observations, in particular the standard deviations of background departures. These give an impression of the performance of the short-range forecast, without the problems typically encountered when using analysis-based verification of short-range forecasts (e.g., Geer et al. 2010). Past experimentation has shown that these statistics are very robust, and they give a more reliable initial assessment over a relatively short 3 month experimentation period than medium-range forecast scores.

Figures 5 to 7 show the characteristics of background departures as a function of the scaling factor for the two series of experiments for a range of selected observations. The results are rather consistent for different observations, so only a representative sample is shown here.

The NoCor series shows a consistent behaviour for observations sensitive to tropospheric temperature and humidity (see the red lines in Figures 5a-c and 6): if the un-scaled diagnosed observation errors are used (scaling factor 1.0) and correlations are neglected we see a degradation of the short-range forecast compared to not using the data at all. This is evident from standard deviations of background departures that are larger than those for the Denial. It is a clear indication that in this case we are under-estimating the size of the observation error, at least for certain spectral structures represented in the assimilated IASI channels. By scaling the observation errors we can mitigate this effect, and with an appropriately chosen scaling factor we can achieve a positive impact on short-range forecasts. Optimal scaling factors are mostly in the range of 2.5-3.0 for the NoCor series, and this optimal scaling factor is somewhat dependent on the level and geophysical quantity the observation is sensitive to. Most likely this reflects the relevance of the neglected error correlations for certain IASI channels. For the optimal scaling factor, the assigned observation error is undoubtedly far from the true error, as can be inferred from standard deviations of background departures. It is interesting to note that the optimal scaling factor is similar to the square root of the leading eigen-values of the error correlation matrix, which characterise the implicit error inflation for structures associated with the leading eigen-vectors when error correlation were taken into account. The optimal error inflation in the NoCor series thus assigns an observation error that is broadly consistent to the error implied by the full matrix for the structures associated with the leading eigen-vectors. Inflation beyond a factor 3.0 mostly does not appear beneficial, and the standard deviations of background departures then start to approach values of the Denial experiment.

In contrast, the WithCor series shows a very different behaviour for observations sensitive to tropospheric temperature and humidity (see the blue lines in Figures 5a-c and 6): even when using the un-scaled diagnosed observation error covariance, the observation departure statistics for many observations do not indicate a clear degradation compared to not using the IASI data at all. Nevertheless, a scaling factor larger than 1.0 also appears beneficial for this series, with an overall optimal scaling factor of around 1.75. This again leads to an observation error covariance matrix with error standard deviations that are larger than the true errors, as these are larger than the standard deviations of background departures. However,

the optimal scaling factor for the series that takes error correlations into account is considerably lower than for the NoCor series. Also, the minima around the optimal scaling factor are rather shallow, and relatively small changes to the scaling factor (e.g., < 15 %) have relatively little effect. This provides an estimate of the sensitivity of our results to the choice of scaling factor.

At the optimal scaling factors, the WithCor series shows a clearly better performance compared to the NoCor series for humidity in all geographical regions and for tropospheric temperature in the tropics. It appears that the diagnosed error correlations better reflect the error characteristics of the IASI data and hence lead to a more appropriate weighting of the IASI observations especially in these areas. For tropospheric temperature and humidity, the results from the WithCor series at the optimal scaling factors are overall also better than those from the Control experiment (indicated in grey in Figures 5 and 6). This aspect will be discussed in more detail in the next section.

For stratospheric temperature, the results for different observing systems are less consistent. AMSU-A

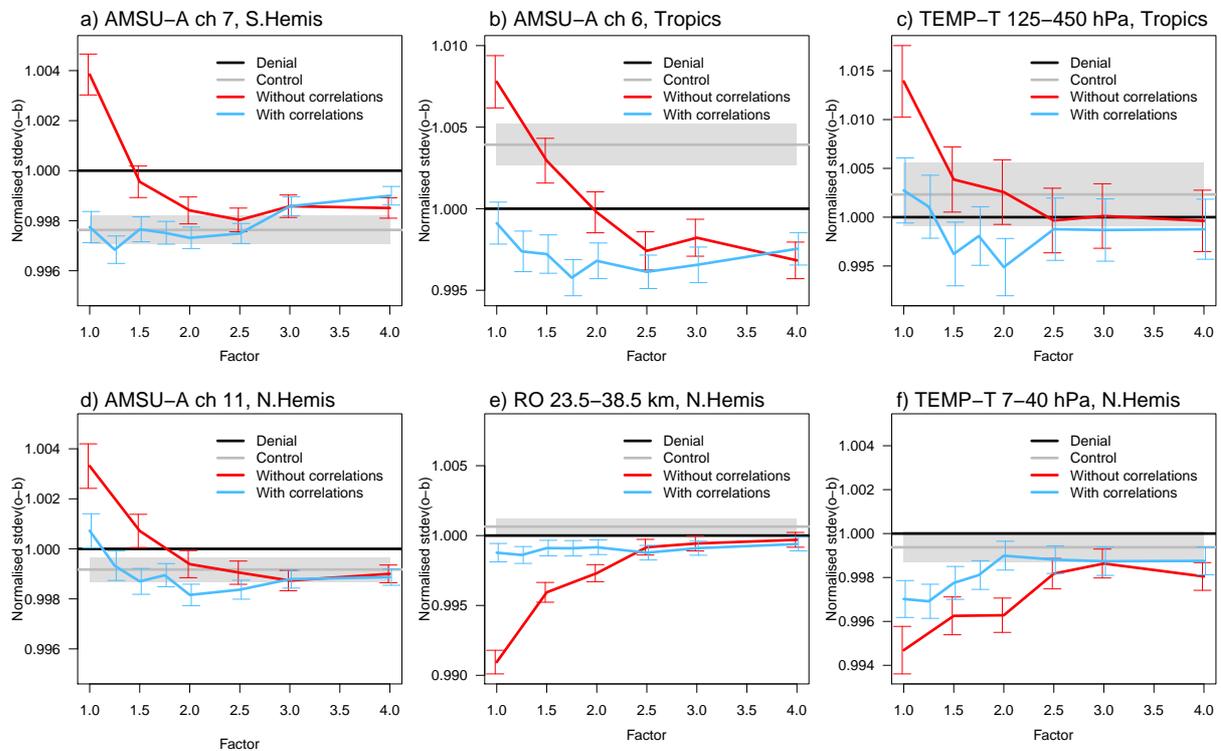


Figure 5: a) Normalised standard deviations of background departures (after bias correction) as a function of the scaling factor for AMSU-A observations in channel 7 over the Southern Hemisphere extra-tropics. The results have been normalised by the values from the Denial. Red and blue lines indicate the results for the NoCor and WithCor series of experiments, respectively, and error bars give statistical significance intervals for differences to the Denial at the 95 % level. Also shown is the performance of the Control experiment in grey, with significance intervals as shaded region. Results from all used AMSU-A instruments have been combined here. Channel 7 of AMSU-A is primarily sensitive to upper tropospheric temperature. b) As a), but for AMSU-A channel 6 over the tropics. This channel is primarily sensitive to mid- to upper tropospheric temperature. c) As a), but for radiosonde temperature observations between 125 and 450 hPa over the tropics. d) As a), but for radio occultation observations with an impact parameter of 23.5–38.5 km from all used satellites over the Southern Hemisphere extra-tropics. e) As d), but over the Northern Hemisphere extra-tropics. f) As d), but for radiosonde temperature observations between 7 and 40 hPa over the Northern Hemisphere extra-tropics.

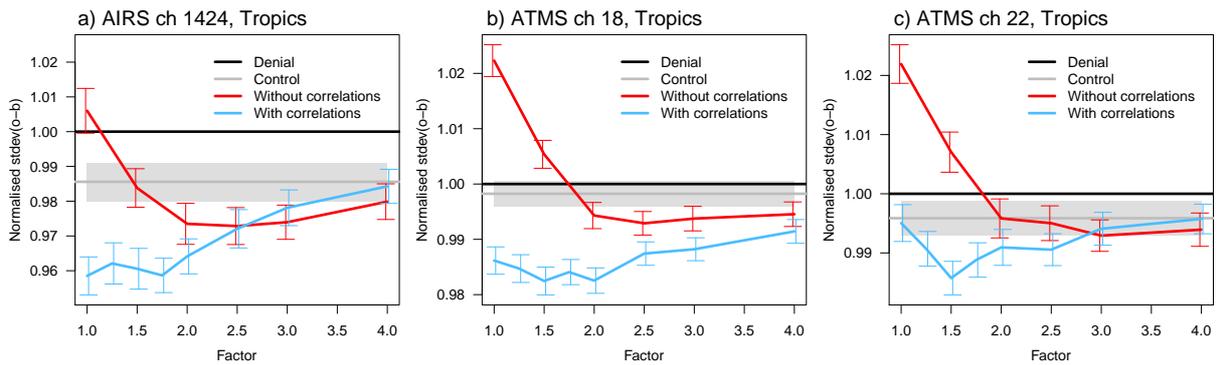


Figure 6: a) As Fig. 5a, but for the lower mid-tropospheric humidity sounding channel 1424 from AIRS over the Tropics. b) As a), but for the mid-tropospheric humidity sounding channel 18 of ATMS. c) As a), but for the upper-tropospheric humidity sounding channel 22 of ATMS.

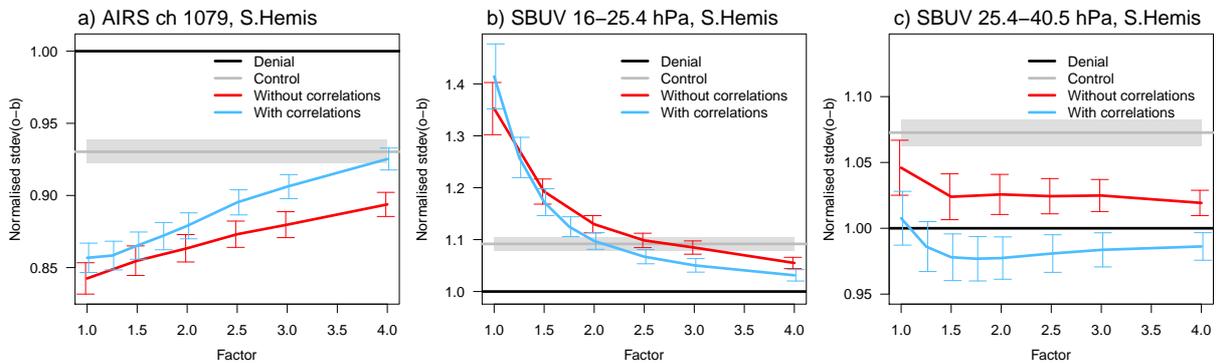


Figure 7: a) As Fig. 5a, but for a typical AIRS ozone channel (1079) over the Southern Hemisphere extra-tropics. b) As a), but for SBUV ozone retrievals in the layer 16-25.4 hPa. c) As a), but for SBUV ozone retrievals in the layer 25.4-40.5 hPa.

observations show a behaviour mostly similar to that described above for the tropospheric temperature, that is, a better performance of the WithCor series compared to the NoCor series and benefits from inflation for both (e.g., Fig. 5d). In contrast, for radio occultation observations or radiosonde temperature measurements over the Northern Hemisphere and, to a lesser extent, over the Southern Hemisphere the experiments without error correlations perform better (Fig. 5e,f). Here, the experiment without inflation and no error correlations even performs best. At the same time, it is worth pointing out that both series are nevertheless performing better than the Control experiment, most likely a result of the lower and more appropriate  $\sigma_0$  for the stratospheric temperature channels.

Ozone also shows a less consistent behaviour for the two series (Fig. 7). For ozone channels from infrared instruments such as AIRS or IASI, standard deviations of background departures are smallest when the un-scaled diagnosed  $\sigma_0$  are used, with the NoCor series of experiments giving slightly better results (Fig. 7a). Both series perform substantially better than the Control experiment, probably as a result of the large reduction in  $\sigma_0$  common to both. In contrast, for SBUV ozone retrievals, even the largest scaling factor still leads to a degradation compared to the Denial experiment for many layers, especially above the ozone maximum (Fig. 7b), and some improvement can be found only for a few

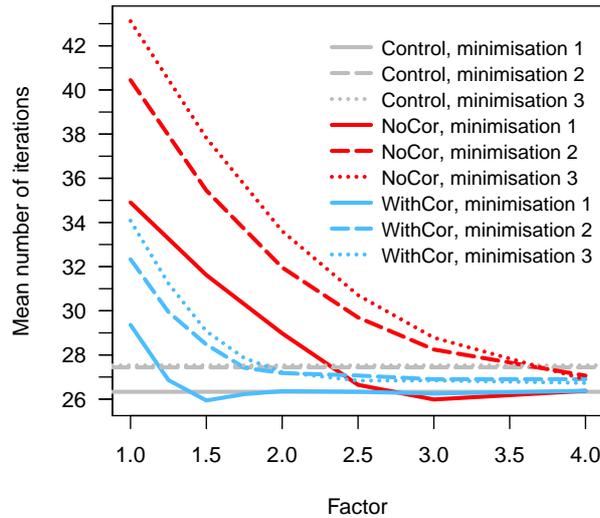


Figure 8: Mean number of iterations as a function of the inflation factor for the NoCor (red) and the WithCor (blue) series of experiments, for the three minimisations performed for these experiments, as indicated in the legend. Also shown is the number of iterations for the Control experiment as grey horizontal lines.

layers below around 25 hPa when some inflation is applied (Fig. 7c). This, to a large extent, reflects the primary sensitivity for channels in the long-wave infrared ozone band. It is clear that the observation error changes have a very substantial impact on the ozone analysis, and this aspect will be discussed in more detail in section 5.6.

The choice of scaling factor, and the choice of error correlation matrix, also significantly affects the number of iterations performed in each experiment (Fig. 8). The number of iterations is obviously important for the computational performance, as a larger number of iterations is more costly. In ECMWF's incremental 4DVAR system, three minimisations are performed, with a re-linearisation of the observation operators around a full-resolution non-linear trajectory between each of them. The number of iterations in the minimisation is determined through reaching certain convergence criteria. As can be seen in Fig. 8, the number of iterations mostly decreases with the scaling factor, but the NoCor series of experiments requires significantly more iterations, even at relatively large scaling factors. It can also be seen that this is not directly a property of using a diagonal matrix - the Control experiment needs a similar number of iterations as the WithCor experiment at the optimal scaling factor of 1.75, despite using a diagonal observation error covariance matrix for IASI.

### 4.3 Partial correlations

Following the results of the previous section, it is interesting to investigate which block of inter-channel correlations is most relevant to take into account. To address this, we ran another series of assimilation experiments parallel to the NoCor and WithCor series in which we use inter-channel error correlations only for certain sub-sets of channels. Starting from the 10 humidity channels with the clearest error correlations, we gradually introduce error correlations in the 6 groups of channels given in Table 1. That is, the first experiment takes into account error correlations only for the humidity channels and neglects error correlations elsewhere; the second experiment takes into account error correlations only for the humidity and ozone channels, and so on. In all these experiments  $\sigma_O$  is set to the diagnosed values,

scaled by the optimal scaling factor of 1.75. Again, we primarily assess our results by considering background departure statistics for other observations.

The results suggest that the impact seen from accounting for inter-channel error correlations for all assimilated channels does not originate primarily from one or two groups of channels. Instead, the benefit is achieved from a combination of several groups (see, for instance, Fig. 9a, b, d-f). This is most evident for the improvements in humidity seen in the reduction of the standard deviations of background departures for humidity channels of infrared or microwave instruments (e.g., Fig. 9d-f). Here, the introduction of error correlations for the IASI humidity channels alone makes little difference compared to neglecting the error correlations altogether. Significant benefit is only achieved when we take inter-channel error correlations into account also for the IASI window channels, with further benefits when the error correlations are extended to the lower temperature-sounding channels as well. One interpretation of this finding is that an over-weighting of IASI observations is only avoided when error correlations are taken into account for most channels. Another related interpretation is that inter-channel error correlations for several different groups of channels are required to get the full benefit of the situation-dependent weighting of IASI data discussed in section 3.2. Several groups of channels with error correlations taken into account will allow a better identification of spectral signatures due to representativeness or cloud screening error as described by the leading eigen-vectors of the error correlation matrix. There will be

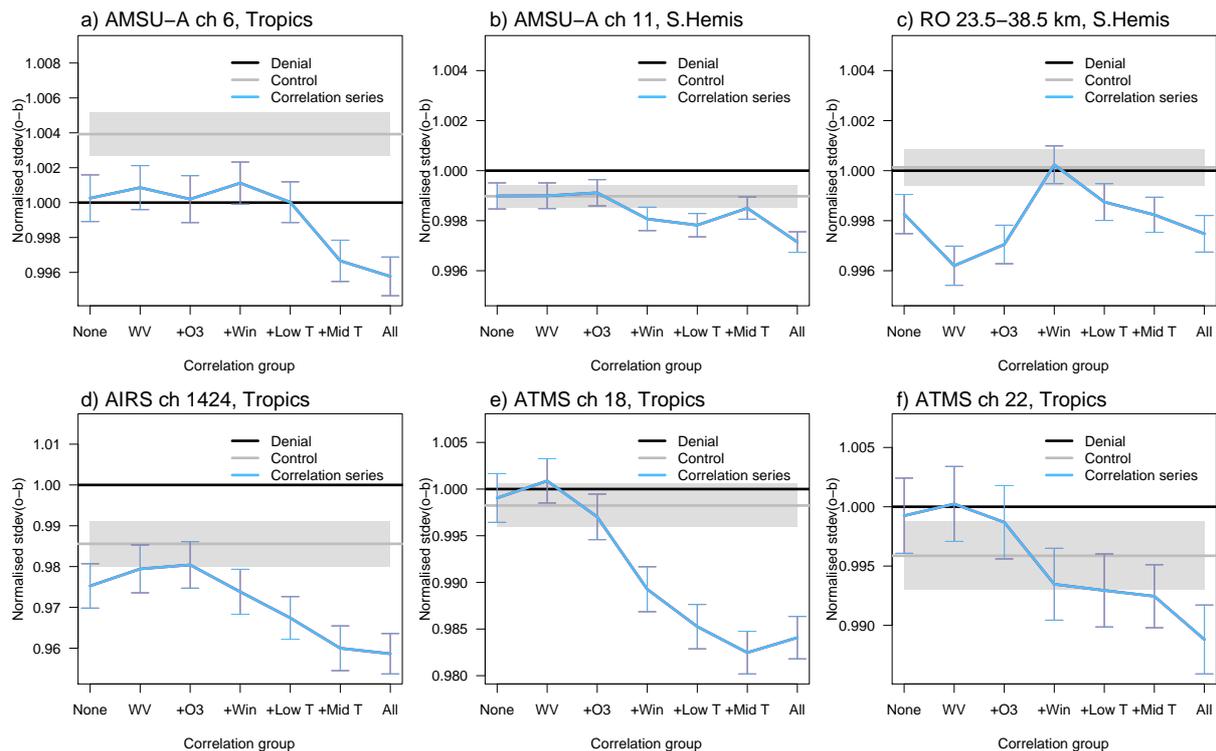


Figure 9: a) As Fig. 5b, but with the blue line indicating the results for the series of experiments with inter-channel error correlations introduced for different groups of channels as described in the main text. b) As a), but for AMSU-A channel 11 over the Southern Hemisphere. These observations are sensitive to mid-stratospheric temperature. c) As a), but for radio occultation bending angle observations with an impact parameter in the range 23.5–38.5 km over the Southern Hemisphere. d) As a), but for the lower mid-tropospheric humidity channel 1424 of AIRS over the tropics. e) As a), but for the mid-tropospheric humidity channel 18 of ATMS. f) As a), but for the upper-tropospheric humidity channel 22 of ATMS.

more ambiguity if this information about the statistical properties of the error is restricted to a sub-set of channels only. In any case, the finding suggests that the benefits seen for humidity are the result of an interaction between different types of channels, rather than just a better treatment of observation errors for the humidity-sounding channels.

For temperature in the stratosphere the results of the experiments with partial error correlations are again less conclusive (cf, Fig. 9b, c). This is where we previously noted a less consistent behaviour when comparing the effects of taking error correlations into account and error inflation. Interestingly, for radio occultation observations in the stratosphere, the best results are achieved when error correlations are taken into account only for the IASI humidity channels (Fig. 9c), and a similar result can also be found for radiosonde temperature observations (not shown). There is a clear reduction of the benefit in the stratosphere when the error correlations are extended to the IASI window channels, even though this appears to be a key factor for achieving better impact for the tropospheric humidity. The mechanisms behind this interaction are not fully understood and could be investigated further.

#### 4.4 Discussion

In summary, we are finding that accounting for inter-channel error correlations for IASI allows the use of observation errors that are more consistent with departure statistics, whereas relatively large error inflation has to be used if these are ignored. Inflating diagnosed values nevertheless appears beneficial even if inter-channel error correlations are taken into account, with an overall optimal scaling factor of 1.75. The results presented here confirm earlier work by Bormann and Collard (2012) who found similar optimal scaling factors in experiments run at slightly lower spatial resolution with a configuration that did not include IASI ozone channels. The best results for humidity are achieved when error correlations are taken into account for all assimilated channels.

It is important to stress that the optimal inflation factor is an empirical, ad-hoc adjustment to the diagnosed matrices to improve their use in the assimilation system, but it is not considered an adjustment that leads to a better estimate of the true observation error characteristics. Statistics of background departures suggest that the resulting assumed observation error after inflation is in fact considerably larger than the true error.

The question arises, why inflating the diagnosed values is beneficial in assimilation trials even when inter-channel error correlations are taken into account. For some spectral structures, it is possible that this counteracts deficiencies of the diagnosed matrices, resulting from only partially valid assumptions made during their derivation. However, it is likely that the inflation also addresses remaining sub-optimality in our assimilation of IASI data, and most likely this is the dominant factor. For instance, there are observation error characteristics that are not accounted for through the globally-constant observation error covariance matrix with inter-channel error correlations, but no spatial error correlations assumed here. For instance, Bormann et al. (2010) found some indications of spatial error correlations, and such error correlations are still neglected. Also, the characteristics of the true observation error are scene-dependent, and again this aspect is neglected in our experimentation. A range of aspects contribute to this scene-dependence: for instance, the size of the contribution from the instrument noise is scene-dependent when assimilating brightness temperatures, as the instrument noise is constant only in radiance space; in addition, cloud screening errors will have scene-dependent characteristics, both geographically and when only a sub-set of channels is diagnosed as cloud-free. It is very likely that the inflation counter-acts these neglected effects, not unlike using inflation when a completely diagonal matrix is assumed. In this context, it is also worth pointing out that a range of error diagnostics clearly suggest that the true observation error for most channels is significantly larger than the true background error. In such cases, assimilating

the observation while assuming a too low observation error can lead to a degradation of the background, whereas a too large observation error will, in the worst case, only result in no improvement. Using an inflated observation error covariance is thus the safer choice that reduces the risk of a degradation of the background. Nevertheless, further work that determines why inflation is beneficial is likely to give further insights in the use of the observation error diagnostics and the assimilation of IASI data.

It is also important to note that the above experiments only explore the introduction of one inflation factor, applied to all assimilated channels. It is clear from these experiments that a single scaling factor applied to all channels cannot achieve the same effect as accounting for error correlations. This was already apparent given the situation-dependent changes to the weighting highlighted in section 3.2, which shows that the adjustment factor introduced through the error correlations depends on the spectral structures of the eigen-vectors. However, it is quite possible that different scaling factors are better for different spectral regions, as the relevance of the uncharacterised aspects mentioned above or the implicit inflation through the error correlations is likely to be different for different spectral regions. This aspect could be explored further, but is not pursued in this report, neither for the NoCor nor the WithCor series. However, the Control experiment can be viewed as an experiment of the NoCor series, but with different scaling factors selected for each channel. The performance shown for the Control experiment in Figures 5 and 6 suggests that there is clearly scope for finding scaling factors that optimise certain aspects of the assimilation of IASI data, while avoiding detrimental aspects such as the increase in the number of iterations observed with the single scaling factor. In this context, adjoint methods that investigate the sensitivity of the forecast error to the specification of  $\sigma_O$  could be beneficial (e.g., Daescu and Todling 2010, Lupu et al. 2015). However, such channel-specific inflation tuning is likely to be at the expense of making the assumed observation errors physically less plausible and less consistent with the statistical properties of the true observation errors.

## 5 Impact in extended assimilation experiments

### 5.1 Experiments

We will now discuss in more detail extended experiments with an updated observation error covariance matrix for IASI. As in the previous section, we consider a Denial experiment with no IASI data assimilated, and a Control experiment that is equivalent to the operational configuration with a diagonal observation error covariance matrix for IASI, with  $\sigma_O$  as shown by the grey line in Fig. 2. In the NewR experiment we use an updated observation error covariance matrix that takes inter-channel error correlations into account. The observation error correlation matrix is as in the previous section, and we chose 1.75 as the optimal inflation factor for  $\sigma_O$ , resulting in the blue line in Fig. 2. For the new observation error covariance,  $\sigma_O$  is significantly smaller than the one currently assumed operationally for the stratospheric temperature channels, the window, ozone and humidity channels. For the window channels,  $\sigma_O$  is now only around a quarter of the old value, a very substantial change. For most tropospheric temperature sounding channels  $\sigma_O$  is instead larger by around 20-40% compared to what is currently assumed operationally.

In addition to the observation error covariance update we also change the ozone channel used to anchor the variational bias correction in the NewR experiment. In the operational configuration, channel 1585 is assimilated without a bias correction, a pragmatic fix to avoid spurious drifts in the variational bias correction (Han and McNally 2010). In additional experimentation we found it beneficial to change this channel to 1574 to avoid a degradation of the bias of the ozone analysis, a result of the significantly

altered weighting of the ozone channels in the NewR experiment. This modification will be further discussed in section 5.6.

The three experiments discussed here cover the 7 month period 5 February 2014 to 4 September 2014, and they otherwise use the same configuration as the experiments described in the previous section, with a spatial model resolution of  $T_L511$  ( $\approx 40$  km), and an incremental analysis resolution of  $T_L255$  ( $\approx 80$  km).

## 5.2 Weighting of IASI data

The new R leads to a very different weighting of IASI in the assimilation system. This will be discussed in the following on the basis of departure statistics for IASI and some direct responses of the assimilation system resulting from the observation error upgrade.

### 5.2.1 Departure statistics for IASI

The new observation error covariance means that IASI observations are fitted very differently in the analysis, as can be seen in statistics of analysis departures (Fig. 10). For many of the mid-tropospheric temperature sounding channels (channel numbers 246-432), the analysis departures for the NewR experiment are much larger, whereas for the window (434-921) and ozone channels (1479-1671) the analysis departures are much smaller. Several factors contribute to this, to a large extent the changes in  $\sigma_O$ , but also the introduction of error correlations together with a different response of the skin temperature retrieval included in 4DVAR, and these will be further discussed below. The reduction for the window and ozone channels primarily reflects that  $\sigma_O$  has been drastically reduced for these channels, so the analysis is forced to draw more closely towards these channels. Note that for the window channel with the strongest surface sensitivity (channel 921), the standard deviations of the analysis departures in the Control are actually larger than the background departures, a peculiar feature that is not present in the NewR experiment. For the humidity channels, there has been comparatively little change of the overall

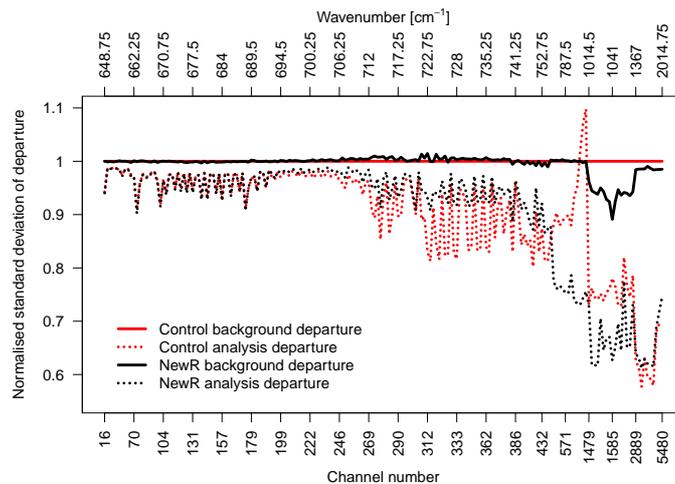


Figure 10: Standard deviations of background (solid) and analysis (dotted) departures for used IASI data over the Southern Hemisphere for the Control experiment (red) and the NewR experiment (black). For display purposes, the standard deviations have been normalised by the standard deviation of the background departures of the Control.

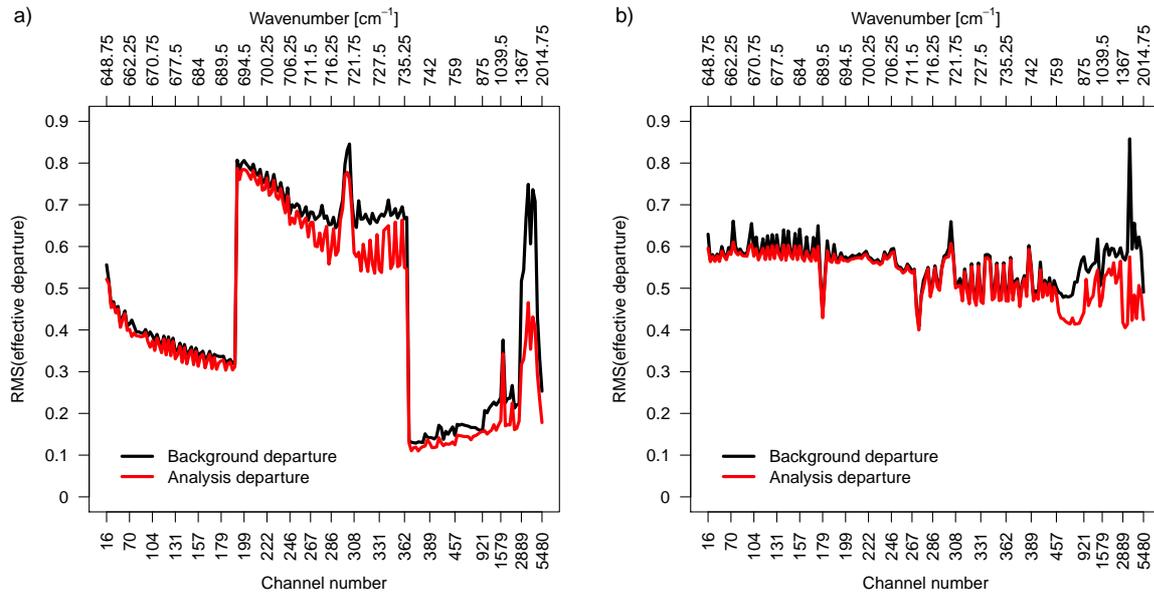


Figure 11: Root mean square (RMS) of the effective background (black) and analysis (red) departure, that is the departure normalised by the square root of the assumed observation error covariance matrix. Statistics have been calculated for the period 1-14 August 2014. a) Statistics for the Control. b) Statistics for the NewR experiment.

magnitude of the analysis departures, and this is the combined effect of lowering  $\sigma_0$ , but introducing quite significant inter-channel error correlations for these channels.

On their own, it is impossible to say whether these changes to the analysis departures are a positive or a negative aspect - they merely reflect the very different weight given to IASI in the analysis as a result of the changes to the observation error covariance matrix. However, the size of the background departures gives a first indication of whether the resulting short-range forecast is closer to observations before they are assimilated. The changes here are much smaller, with reduced background departures for the ozone and humidity channels, and increased departures for the mid-tropospheric temperature sounding channels in the channel number range 246-386. These statistics hence suggest significant improvements in terms of humidity and ozone, but potential degradations for the mid-tropospheric temperature. We will investigate these aspects more fully in the context of the entire observing system.

It is also illustrative to also examine the *effective* departures for IASI for the Control and the NewR experiment, that is, the departures normalised by the square root of the assumed observation error covariance matrix. Fig. 11 shows the root mean square of the effective departures for both experiments. The square of this quantity is the contribution by channel to the 4DVAR cost function. If the assumed  $\mathbf{R}$  was equal to the true observation error covariance one would expect these values to be above one for background departures, and around one for analysis departures. For the Control experiment, the root mean square (RMS) of the effective departures reflects the step-function used for the  $\sigma_0$  in this experiment (cf Fig. 2). Stratospheric channels (numbers 16-199) and the window and ozone channels show relatively small effective departures, far from the expected values. This is a result of the substantially inflated  $\sigma_0$  values used here, reflecting a very cautious assimilation of these more difficult channels that are not expected to be the leading contributors to forecast skill. In contrast, the size of the effective departures is much more even among all channels with the new observation error covariance matrix (Fig. 11b), and closer to the ideal values for the stratospheric, window, ozone, and humidity channels. As a result of the inflation dis-

cussed in the previous section, the root mean square of the effective departures for the NewR experiment is, however, still much lower than the ideal values, and for channels 199-362, the values are in fact less close to the ideal values than the Control.

These two Figures 10 and 11 reflect a very different behaviour of the analysis system regarding the assimilation of the mid-tropospheric temperature sounding channels of IASI. Two key effects are at play here: 1) the structure-selective weighting introduced in section 3.2, and 2) a very different response of the skin-temperature retrieval performed inside 4DVAR. Both effects affect the size of the analysis departures particularly for many tropospheric temperature sounding channels. The former effect does this through the situation-dependent weighting highlighted in the eigen-vector analysis in section 3.2: many of the departures for the lower tropospheric temperature sounding channels project onto the leading eigen-vectors and are hence down-weighted. This contributes to relatively small differences between the RMSs of the effective background and analysis departures seen in Fig. 11b for these channels. If the RMSs of effective departures are re-calculated without taking the error correlations into account the differences between the background and analysis departures are considerably larger (not shown). This effect is hence primarily the result of introducing the error correlations. The second effect, ie the impact on the skin-temperature retrieval, is the result of the  $\sigma_O$  as well as the error correlation changes and deserves a little further analysis.

### 5.2.2 Effect on skin-temperature retrieval

In the assimilation of IASI data in the ECMWF system, an independent skin temperature value is retrieved for each IASI field of view during 4DVAR. This so-called “sink-variable” is fitted during the minimisation, but subsequently discarded and does not influence the model forecast of skin temperature.

With the operational observation error for IASI, this skin temperature retrieval is strongly influenced by the sounding channels, as the window channels are down-weighted through a large observation error, such that the reduction in the observation cost function is not as large if the analysis attempts to fit these window channels. This can even lead to a situation as seen in Fig. 10, where the most surface-sensitive window channel shows a poorer fit to the analysis than the background. The danger in this case is that signal in the sounding channels is aliased erroneously into a skin temperature signal, instead of correcting errors in the atmospheric background.

In contrast, in the NewR experiment, the window channels dominate the retrieval of the skin temperature, and the analysis is less allowed to use the skin temperature variable to fit the sounding channels. This effect contributes to the much reduced analysis departures for the window channels, and the increased departures for the lower and mid-tropospheric temperature sounding channels seen in Fig. 10.

Reflecting the different constraint on the skin-temperature, the adjustments made to the skin temperature relative to the background value are very different in the two experiments (Fig. 12). They are smaller in the NewR experiment than in the Control in the extra-tropics, and larger than the Control in the tropics around the Inter Tropical Convergence Zone (ITCZ). Of course it is not guaranteed that the window channels will provide a better estimate of the skin temperature, and to what extent this is successful will depend, among other things, on the performance of the cloud detection and the appropriateness of all the errors assigned. However, the very different response is an important modification in the assimilation of IASI data between the NewR and the Control experiment. If the skin temperature retrieval with the window channels is reliable, this should provide an enhanced capability to retrieve information contained in surface-sensitive channels, for instance information on lower tropospheric humidity or ozone.

### 5.3 Short-range impact versus other observations

We will now assess the impact of the updated observation error covariance matrix on short-range forecasts. This will be done by examining more closely background departure statistics for other observations, as these provide the most robust means of assessing short-range forecast performance.

Background departures for other observations suggest overall a very significant improvement in the short-range forecast compared to the Control from using the updated observation error covariance for IASI. Most assimilated observations show significantly reduced standard deviations of background departures compared to the Control, indicating a more accurate short-range forecast from NewR that agrees better with the observations (see, for instance, Fig. 13). For many observations, these measures indicate roughly a doubling of the impact of IASI compared to the Control.

The improvement in the short-range forecast is particularly clear for humidity-sensitive observations (see, for instance, Figures 13b, d and e for the humidity channels of ATMS, radiosonde statistics, and the infrared humidity channels from AIRS around the  $7.6\mu\text{m}$  band, respectively), and it is particularly noticeable in the tropics (see Fig. 14). Over the tropics, the impact of IASI is more than doubled compared to the Control in terms of the departure statistics for ATMS humidity-sounding channels 18-22. The MHS instruments also show a clear improvements of up to 0.8 %, which is particularly notable as

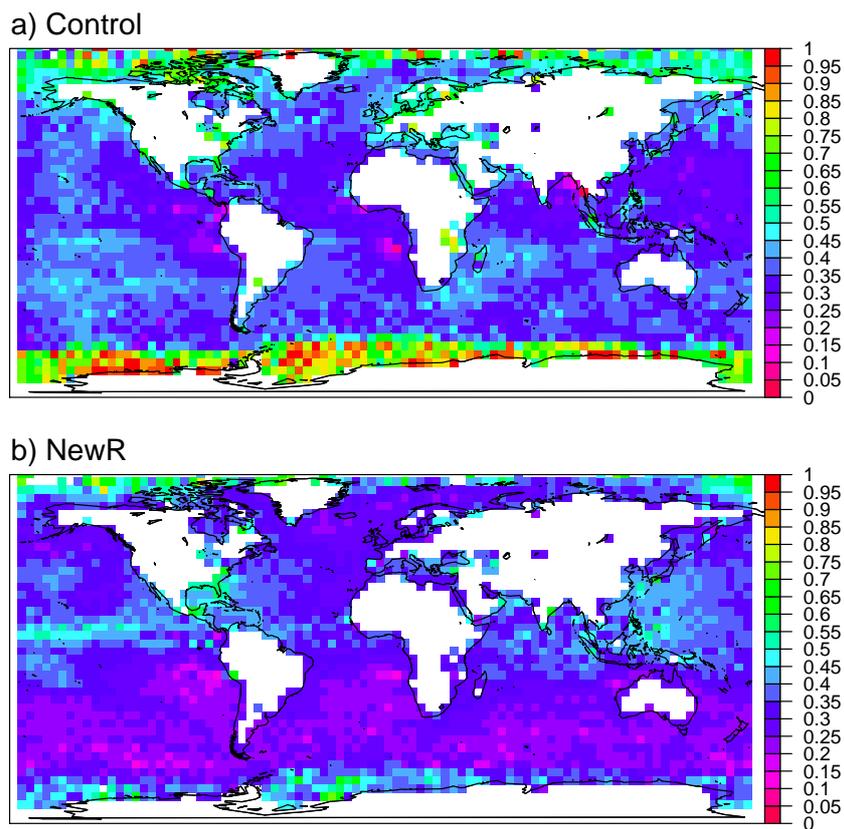


Figure 12: Root mean square of the skin temperature increment for (a) the Control experiment and (b) the NewR experiment, for July and August 2014. We consider only IASI spectra from Metop-A for which channel 921 (ie, the channel with the largest cloud sensitivity of the assimilated channels) has been diagnosed as cloud-free and is assimilated.

the Control experiment shows no statistically significant impact against this observing system. The MHS instruments are assimilated in the all-sky system in these experiments (Geer et al. 2014), in contrast to the ATMS humidity-sounding channels which are used in clear regions only. Further investigations show that the improvements for MHS in NewR relative to the Control or the Denial originate from clear-sky as well as cloudy regions (not shown). In contrast, the Control shows a small improvement over the Denial in clear-sky regions, but the impact in cloudy regions is neutral to slightly negative.

Following our earlier findings, the improvements in the humidity are most likely a result of the introduction of error correlations for the IASI humidity and lower sounding/window channels, together with a better constraint on the skin temperature through the increased weight on the window channels. The latter is expected to allow a better analysis particularly of low-level humidity. The improved representation of clouds in the short-range forecasts is probably also linked to an improved representation of the dynamics, as indicated through closer agreement between short-range forecasts and wind observations (e.g., Fig. 14c).

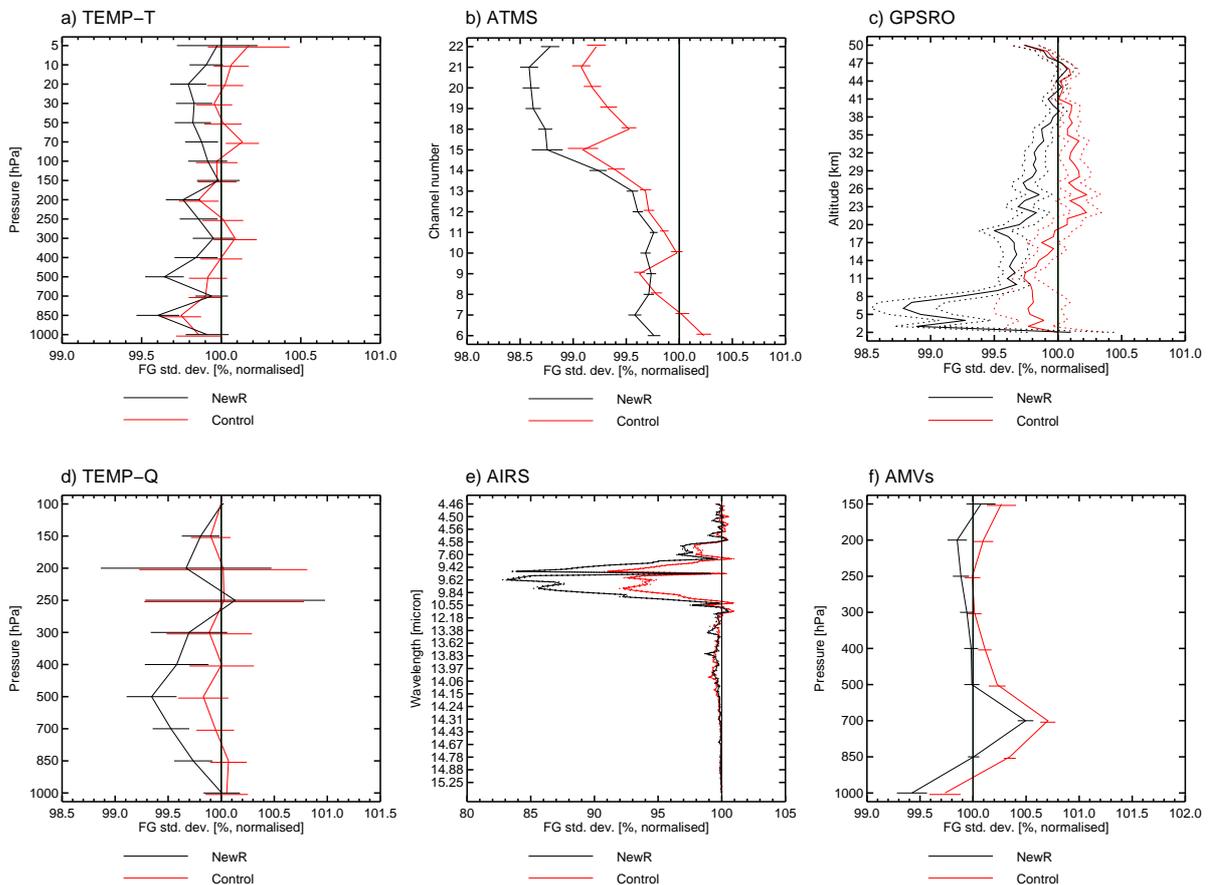


Figure 13: Global standard deviations of background departures for a selection of observing systems for the NewR (black) and the Control experiment (red). The values have been normalised by the standard deviations of background departures from the Denial experiment, such that values less than 100% indicate an improvement with respect to the Denial. The statistics have been calculated from 7 months of experimentation. Horizontal lines indicate 95% confidence intervals. The observing systems are: a) Radiosonde temperature observations, b) ATMS brightness temperatures (with temperature sounding channels 6-15, and humidity sounding channels 18-22), c) GPSRO bending angle observations from COSMIC, Metop-A and B and GRACE-A, d) Radiosonde humidity observations, e) AIRS brightness temperatures, and f) Atmospheric Motion Vectors from 10 satellites.

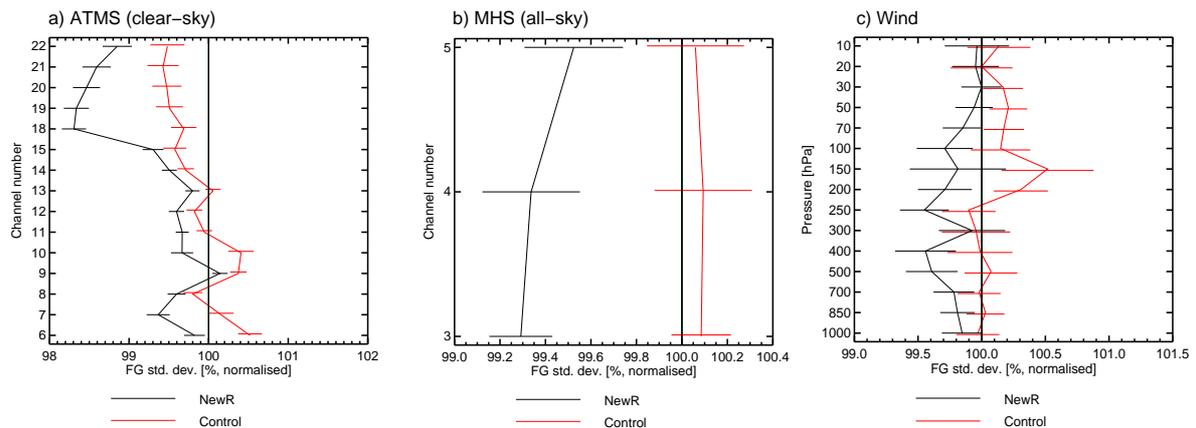


Figure 14: As Fig. 13, but for normalised standard deviations of background departures in the tropics for a) ATMS observations, b) MHS observations from 4 satellites, and c) radiosonde wind observations.

While most observation statistics suggest a positive impact from the observation error change, there are, however, a few noteworthy exceptions. Observation statistics for several satellite radiances suggest a small degradation compared to the Control for mid- to upper tropospheric temperature. AMSU-A channels 7 and 8 and ATMS channels 8 and 9 show a small increase in the standard deviations of background departures in the extra-tropics, at the most by 0.2%, which is nevertheless statistically significant. A more significant increase in the standard deviations of background departures appears for some mid- to lower tropospheric temperature-sounding channels in the infrared from AIRS, or HIRS (see, for instance, AIRS channels around  $14\mu\text{m}$  in Fig. 13e). These increases are consistent with the behaviour seen for IASI channels in the  $706\text{--}735\text{cm}^{-1}$  wavenumber range in Fig. 10. The increase primarily occurs in the extra-tropics, where it can be up to 0.7%. This increase has been investigated further, and it can be partly attributed to a different response of the McNally and Watts (2003) cloud detection, which is very sensitive to changes in the First Guess and the bias correction. This results in a very different sampling of the observations, which makes the comparison and interpretation of these statistics more difficult. For most of the affected observations, there is, however, still an improvement in the NewR experiment compared to the Denial. So, even if there is a degradation in these areas compared to the Control, assimilating the data still gives clear benefits. In this context it is worth mentioning that the sample of data used to derive the observation error covariance matrix is somewhat biased towards the tropics and sub-tropics, and this may also contribute to a better performance in the tropics. It should be noted here as well, that the Control also shows increased background departures for some other observing systems, most notably some AMSU-A and ATMS temperature-sounding channels in the tropics, GPS radio occultation measurements sensitive to temperature, and low-level AMVs (e.g., Figures 13b, c, f and 14a, c). On balance, neither the NewR nor the Control fare better in this regard.

## 5.4 Impact on analysis increments and mean analyses

The updated observation error covariance matrix for IASI leads overall to a large reduction in the analysis increments. This can be seen, for instance, in Fig. 15 which shows the normalised zonal mean differences in the RMS of the increments for wind between the Control and the Denial and the NewR and the Denial, respectively. Adding IASI in the Control experiment leads to a large increase in the wind increments compared to the Denial experiment. This is very different in the NewR experiment, where increments

are reduced compared to the Denial in some areas, but increased slightly in others. Other geophysical variables show a very similar behaviour. Overall, compared to the Control, the NewR experiment shows a very significant reduction in the size of the increments for all geophysical variables.

An increase in the size of the increments is often a feature of adding new observations, a reflection of adding further variability and reducing the correlations between background and analysis errors. Such an increase is therefore not necessarily an indication of a poor performance, especially when it is accompanied with a better agreement between the background and other assimilated observations (e.g., Bouttier and Kelly 2001, Geer et al. 2010). In the present experimentation, the large reduction in the size of the increments in the NewR experiment compared to the Control is accompanied with a significantly improved background fit for many other assimilated observations. This combined result suggests a clear improvement in the overall consistency of the background, the assimilated observations, and the analysis, and it is hence considered to be a significant improvement over the Control experiment.

The observation error upgrade also affects the mean analyses, particularly for temperature and humidity. The NewR as well as the Control experiment show considerably different mean analyses when compared to the Denial, and the differences exhibit distinct geographical pattern. For some levels, these changes are larger for the Control, whereas for others they are larger in the NewR experiment. The strongest changes in the NewR experiment are seen for the 850 hPa temperature and relative humidity field (Fig. 16). The NewR and the Control experiment both decrease the relative humidity in the ITCZ by around 1 % compared to the Denial, but the NewR experiment moistens and cools the subsidence regions more strongly than the Control (of the order of 1 % and 0.1 K, respectively), and it dries/warms the polar regions more strongly by a similar amount.

It is not clear whether these changes to the analysis biases are a positive or negative aspect. While data assimilation systems are designed to generate increments with a zero mean, such changes in the mean analysis are a frequent occurrence when adding observations, and it is often not clear whether these changes correct existing biases or introduce new ones. One way to evaluate such changes in the

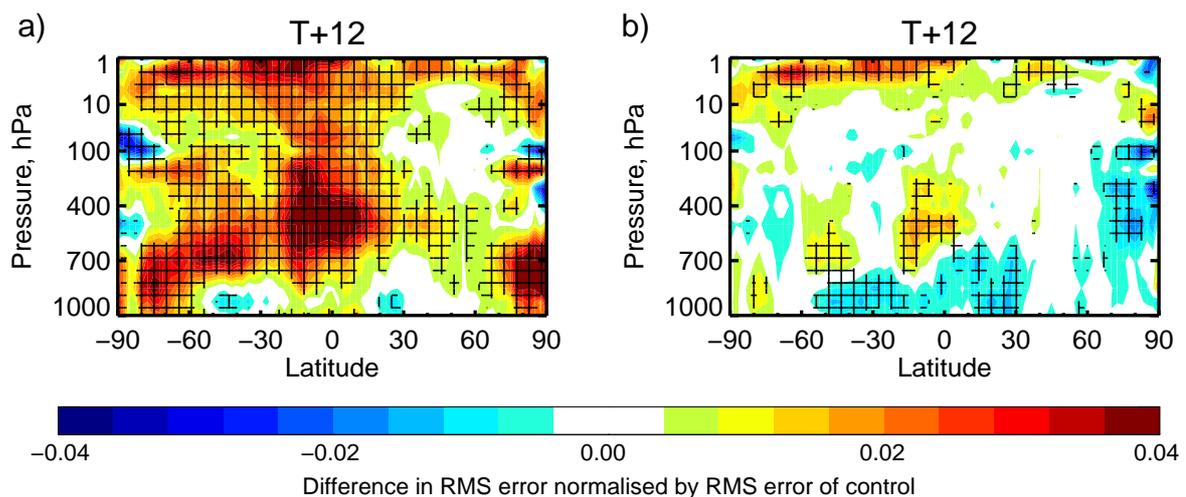


Figure 15: a) Zonal means of the normalised differences of the root mean square of the wind increments between the Control and the Denial experiment. A negative value means a reduction of the increments compared to the Denial. The results are based on 423 assimilation cycles over 7 months. Cross-hatching indicates statistical significance at the 95 % level. b) As a), but for the NewR experiment versus the Denial.

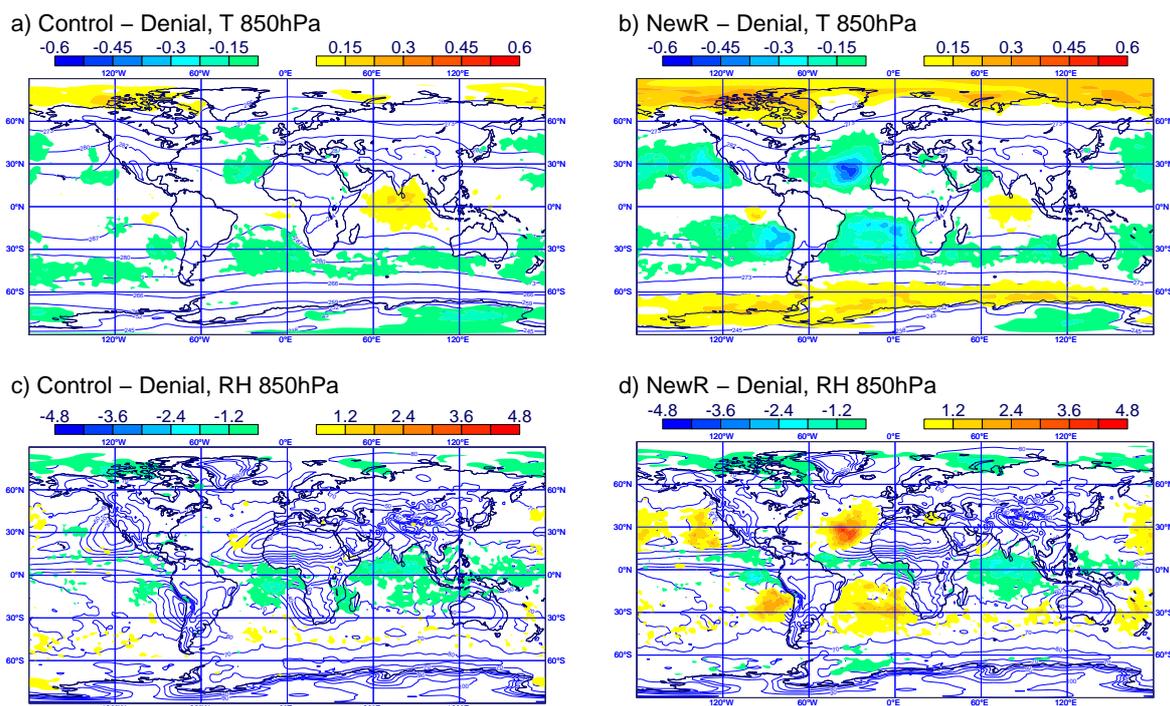


Figure 16: a) Difference in the mean 850 hPa temperature analysis [K] between the Control and the Denial (shading) together with the mean 850 hPa temperature field of the Denial (contours). b) As a), but for the difference between the NewR experiment and the Denial. c) As a), but for the 850 hPa relative humidity [%]. d) As c), but for the difference between the NewR experiment and the Denial.

analysis biases is by comparison against other observations not prone to biases. However, this is not straightforward in our case, as only a limited number of such observations are available in the lower troposphere over oceans in the affected areas, and their representativeness is also somewhat questionable. Nevertheless, the changes in the bias of the humidity analysis at 850 hPa as well as the warming in the polar regions are supported by radiosonde observations in the affected areas, and the changes reduce biases against these observations (not shown). However, the cooling in the subsidence regions increases a warm bias against radiosondes already present in the Denial in these areas.

## 5.5 Forecast impact

Verification of each experiment against its own analysis suggests a significant improvement in short-range forecasts. For this measure, forecast errors are reduced significantly up to day 3-4 in the NewR experiment compared to the Control for almost all variables and levels (e.g., Figures 17, 18 and 19), with significant reductions out to day 5 or further in the stratosphere and the lowest tropospheric levels. The impact in the stratosphere is likely to be a result of the increased and more appropriate weight given to the stratospheric temperature channels from IASI in the NewR experiment. The forecast impact appears stronger for temperature rather than the geopotential, possibly a reflection of adding vertical temperature structure through the more sophisticated weighting of IASI data, which may enhance the representation of vertical gradients as discussed earlier. In the medium-range in the mid- to upper troposphere, the impact of IASI is more similar in the Control and the NewR experiment, and both experiments show a

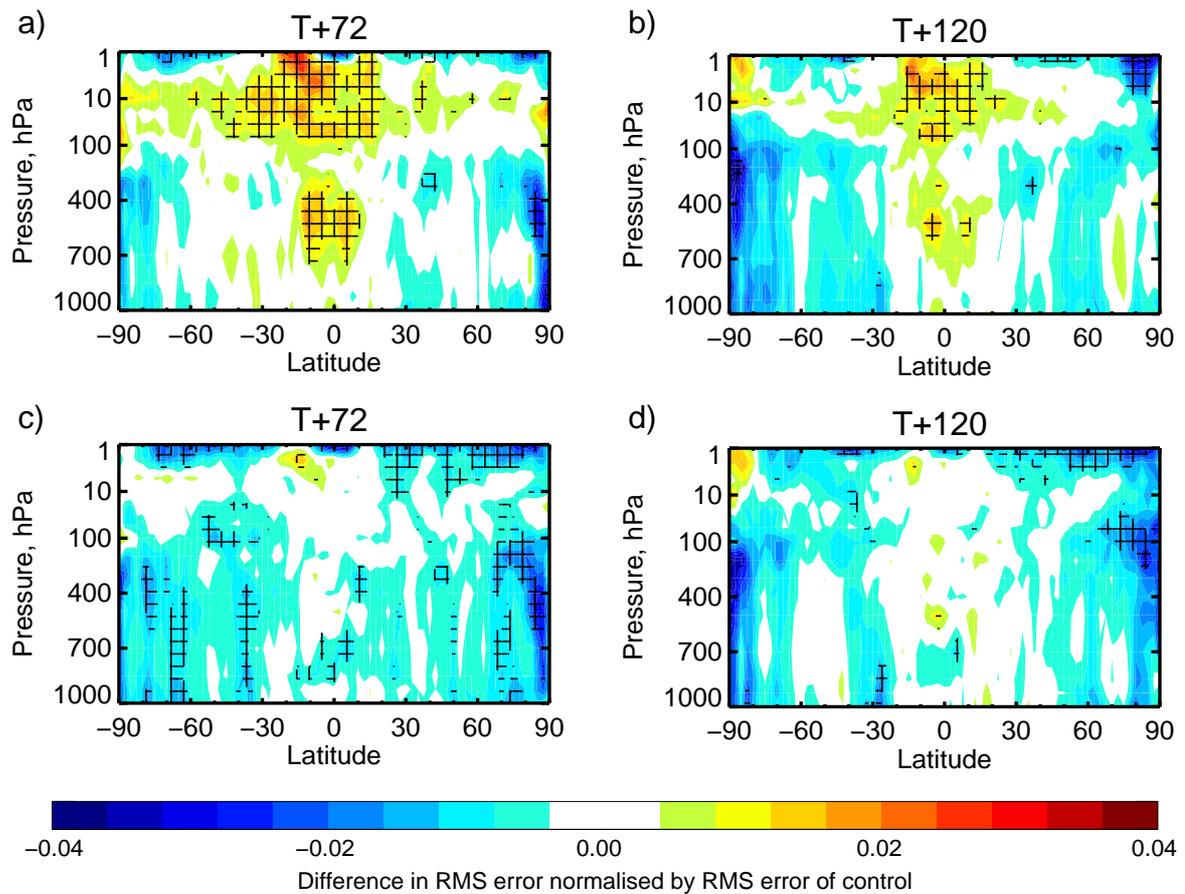


Figure 17: a) Zonal means of the normalised differences of the root mean squared forecast error of wind between the Control and the Denial experiment for the 72 h forecast. Each experiment has been verified against its own analysis, and a negative value means an improvement compared to the Denial. The results are based on up to 418 forecasts over 7 months. Cross-hatching indicates statistical significance at the 95 % level. b) As a), but for the 120 h forecast. c) As a), but for the normalised differences in the root mean squared forecast error of wind between the NewR and the Denial experiment. d) As c), but for the 120 h forecast.

statistically significant reduction of forecast errors by around 1 % for the 500 hPa geopotential in the extra-tropics (Fig. 18).

Short-range forecast impact when evaluated against analyses should be treated with some caution, as the contribution from analysis errors is more significant, and in particular the correlations between the analysis and short-range forecast errors play a more important role. It is therefore not surprising that the short-range impact presented in Fig. 18 is, in parts, sensitive to the choice of the verifying analysis. For instance, verification against the operational analysis shows a more neutral impact for the NewR experiment compared to the Control for the extra-tropics for the first two days, whereas the medium-range forecast scores are comparable to the ones shown here, as the choice of analysis is less important (not shown). Over the tropics and in the stratosphere, the NewR experiment also shows a significant improvement over the Control for the day 2-3 forecast when verified against the operational analysis, providing further indication of a robust forecast improvement in these areas. Choosing the operational analysis for verification arguably favours the Control, because the operational analysis uses the same ob-

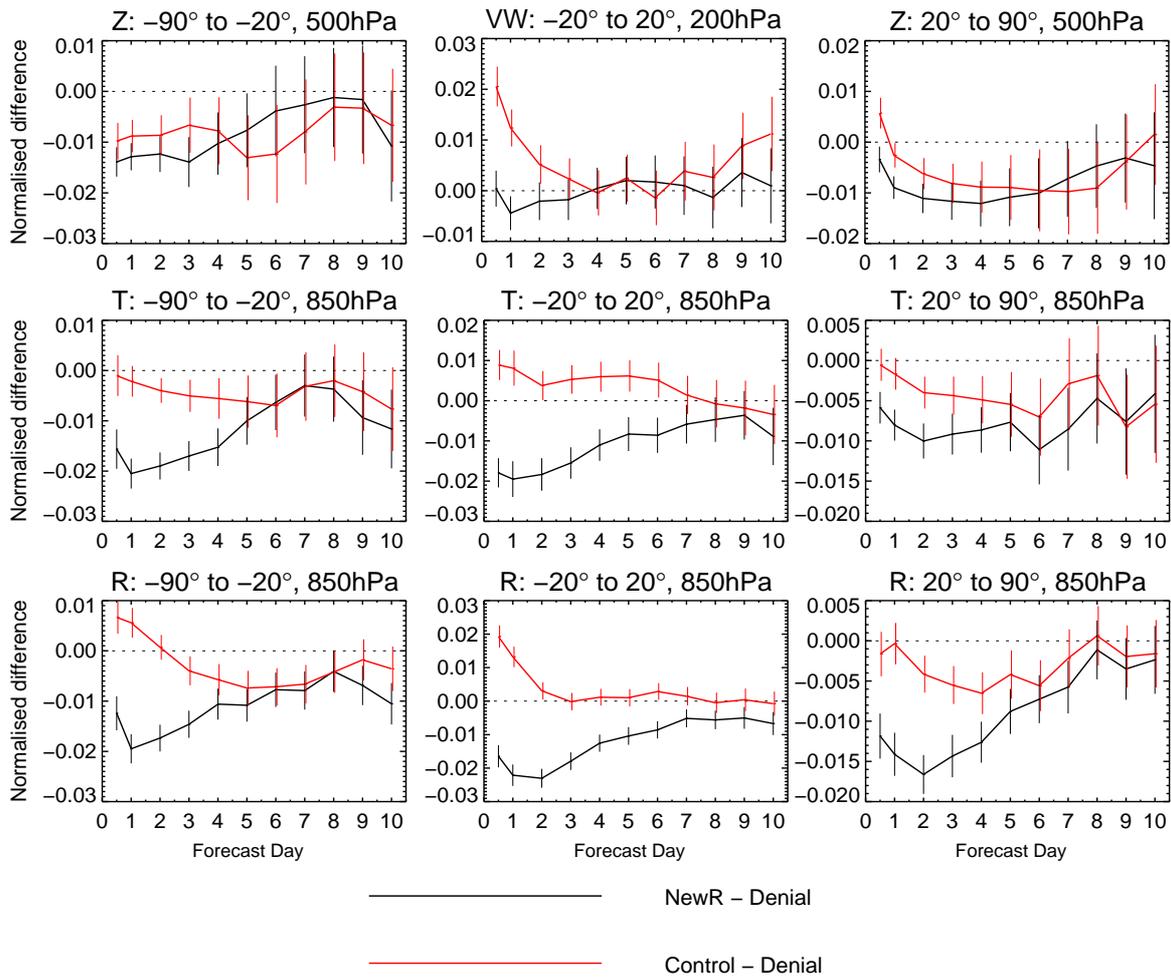


Figure 18: Normalised differences of the standard deviation of forecast errors compared to the Denial experiment as a function of forecast range in days, covering up to 423 forecasts over 7 months. Error bars indicate significance intervals for the differences to the Denial at the 95 % level. Black shows the results for the NewR experiment versus the Denial, whereas red shows the results for the Control experiment versus the Denial. Verification is against the own analysis. A negative value means an improvement compared to the Denial. The various panels show, from left to right: Results for the 500 hPa geopotential over the Southern Hemisphere extra-tropics, the 200 hPa wind in the tropics, and the 500 hPa geopotential over the Northern Hemisphere extra-tropics (top row). The 850 hPa temperature over the Southern Hemisphere extra-tropics, the tropics, and the Northern Hemisphere extra-tropics (middle row). The 850 hPa relative humidity over the Southern Hemisphere extra-tropics, the tropics, and the Northern Hemisphere extra-tropics (bottom row).

servation error covariance and it is therefore likely to show some similar structures in the analysis errors as the Control. These will lead to an under-estimation of the short-range forecast errors for the Control when this verifying analysis is used. These aspects have been discussed in detail, for instance, in Geer et al. (2010). They argue that for analysis-based verification choosing each experiment's own analysis is mostly the best compromise that does not prejudice one experiment setup over another. We therefore place more emphasis on the own-analysis scores presented here, in combination with the analysis of background departures for other assimilated observations as shown in Figures 13 and 14. The latter are usually a more reliable indicator of the short-range forecast impact than analysis-based verification, and

Domain	Parameter	Level	Anomaly correlation	RMSE
N.Hem	r	850hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
		30hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
	t	50hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
		100hPa	▲▲▲	▲▲▲▲▲▲▲▲
		500hPa	▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲	▲▲▲▲▲▲▲▲
		1000hPa	▲▲▲	▲▲▲▲▲▲▲▲
		200hPa	▲▲	▲▲▲▲▲▲▲▲
	vw	500hPa	▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲	▲▲▲▲▲▲▲▲
		100hPa	▲	▼▲▲▲▲▲▲▲
	z	500hPa	▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲	▲▲▲▲▲▲▲▲
		1000hPa	▲▲	▲▲▲▲▲▲▲▲
850hPa		▲▲	▲▲▲▲▲▲▲▲	
S.Hem	r	850hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
		30hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
	t	50hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
		100hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
		500hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲▲▲	▲▲▲▲▲▲▲▲
		1000hPa	▲▲▲	▲▲▲▲▲▲▲▲
		200hPa	▲	▲▲▲▲▲▲▲▲
	vw	500hPa	▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲	▲▲▲▲▲▲▲▲
		100hPa	▲▲	▼▲▲▲▲▲▲▲
	z	500hPa	▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲	▲▲▲▲▲▲▲▲
		1000hPa	▲▲	▲▲▲▲▲▲▲▲
850hPa		▲▲	▲▲▲▲▲▲▲▲	
Tropics	r	850hPa	▲▲▲▲▲▲	▲▲▲▲▲▲▲▲
		30hPa	▲▲▲▲▲▲	▲▲▲▲▲▲▲▲
	t	50hPa	▲▲▲▲▲▲	▲▲▲▲▲▲▲▲
		100hPa	▲▲▲▲▲▲	▲▲▲▲▲▲▲▲
		500hPa	▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲▲▲▲▲	▲▲▲▲▲▲▲▲
		1000hPa	▲▲▲▲▲▲	▲▲▲▲▲▲▲▲
		200hPa	▲▲	▲▲▲▲▲▲▲▲
	vw	500hPa	▲▲	▲▲▲▲▲▲▲▲
		850hPa	▲▲	▲▲▲▲▲▲▲▲
850hPa		▲▲	▲▲▲▲▲▲▲▲	

- Symbol legend: for a given forecast step... (*d*: score difference, *s*: confidence interval width)
- ▲ experiment better than control statistically highly significant (the confidence bar above zero by more than its height) ( $d/s > 3$ )
  - ▲ experiment better than control statistically significant ( $d/s \geq 1$ )
  - experiment better than control, yet not statistically significant ( $d/s \geq 0.5$ )
  - no any significant difference between control and experiment
  - experiment worse than control, yet not statistically significant ( $d/s \leq -0.5$ )
  - ▼ experiment worse than control statistically significant ( $d/s \leq -1$ )
  - ▼ experiment worse than control statistically highly significant (the confidence bar below zero by more than its height) ( $d/s < -3$ )

Figure 19: Score-cards for the NewR experiment compared to the Control for the full 7 month period. Verification is against each experiment’s own analysis, and the statistical confidence level is 95 %. See symbol legend for further explanations.

very clearly indicate an improved performance in the short-range for the NewR experiment.

Some of the most significant reductions in the forecast errors against the own analyses even out to the medium-range can be found for temperature and humidity at 850 hPa (e.g., lower two rows of Fig. 18).

Here, the Control experiment is showing only a small impact compared to the Denial, whereas in the NewR experiment standard deviations of the forecast errors are significantly reduced at least up to day 6, especially in the tropics. However, these large reductions should also be treated with some caution, bearing in mind the considerable changes to the mean analyses at these levels noted earlier, and bearing in mind that analysis-based forecast scores for humidity are generally relatively un-reliable. Relative humidity exhibits much more spatial and temporal variability and the verifying analysis is not as well constrained as, say, that of the geopotential. Nevertheless, the reductions seen here are also apparent in maps of standard deviations of the forecast errors (not shown), suggesting that the reductions are not (only) an artifact of the geographical pattern of the mean analysis differences. However, further investigation shows that the activity of the analysis (ie the standard deviation of the analysis anomalies to climatology) for these fields is fractionally reduced by 0.4-1.0 %, while the activity of the forecasts is gradually less affected. The reduction in the activity of the analysis brings it more in-line with the activity in the forecasts, and of course it is not clear what the “right” level of activity is for the analysis. Nevertheless, the reduction in activity may contribute to the apparent reduction in the forecast errors when own-analysis scores are considered. Note, however, that forecast scores calculated against the operational ECMWF analysis, and hence against an un-changed analysis, as well as against radiosonde observations still show significant reductions for relative humidity at 850 hPa in the tropics out to day 4. While the size of the impact at 850 hPa may be exaggerated in scores against the own analysis, it is hence likely that the reductions indeed reflect a real forecast improvement.

## 5.6 Ozone impact

The departure statistics for other observations discussed earlier have already indicated that some of the largest impact in the NewR experiment can be found for ozone. The ozone channels on the infrared instruments (AIRS, IASI, HIRS) suggest a very substantial improvement, showing up to 17% reduction in the global standard deviations of the background departures with respect to the Denial, compared to up to 8% reduction for the Control (e.g., Fig. 13e). However, the impact of IASI data on the ozone analysis is less clear when considering SBUV retrievals which are the only source of ozone profile information assimilated in these experiments. These show increased standard deviations of background departures for most layers below 10 hPa in the Control (Fig. 20). Mostly, the new observation error covariance for

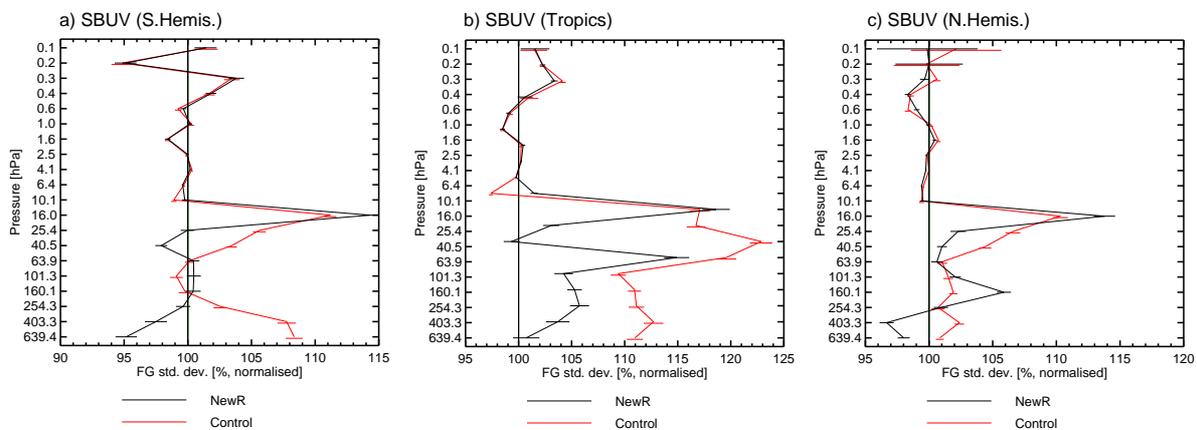


Figure 20: As Fig. 13, but for normalised standard deviations of background departures for SBUV ozone retrievals a) over the Southern Hemisphere extra-tropics, b) the tropics, and c) the Northern Hemisphere extra-tropics.

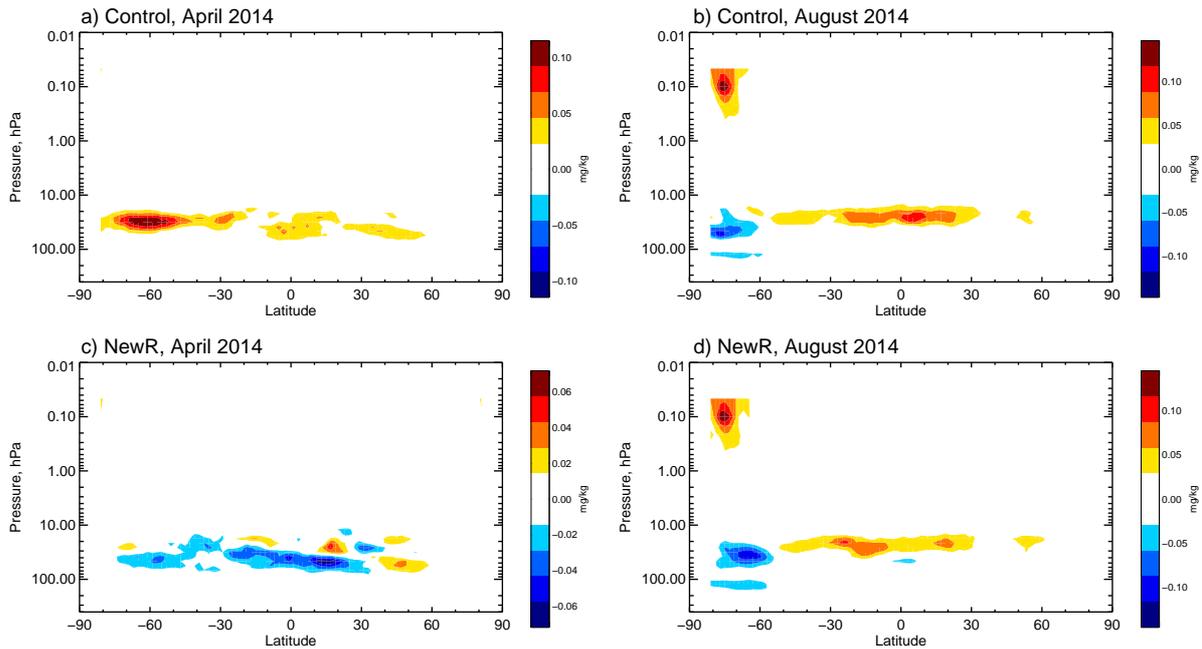


Figure 21: a) Zonal means of the changes in the standard deviation of differences between MLS retrievals and the analysis of ozone mixing ratio for the Control compared to the Denial for April 2014. Positive (red) values indicate an increase in the standard deviations in the Control. b) As a), but for August 2014. c) As a), but for the NewR experiment compared to the Denial. d) As c), but for August 2014. Please note that different colour scales are used in the four panels.

IASI reduces this degradation, particularly in the tropics, although the benefit compared to the Denial is less clear. The interaction between IASI radiances and ozone profile information from SBUV has been discussed in detail in Dragani and McNally (2013). The small degradations for some SBUV layers are primarily a reflection of the limited vertical resolution provided by IASI, which means IASI tends to cause relatively broad analysis increments, as the weighting functions of the ozone channels are very broad, but the increments primarily reflect the errors in the background at the levels that IASI radiances are most sensitive to. These increments may not be appropriate over the full range of levels that IASI ozone channels are sensitive to.

To investigate the performance for ozone further, the ozone analyses have been compared to ozone retrievals obtained from the Microwave Limb Sounder (MLS) on board the Aura satellite (e.g., Froidevaux et al. 2006). These retrievals are not assimilated in our experiments and hence provide an independent assessment of the ozone field. The comparison shows a clear improvement in the ozone analysis around the ozone maximum between 20–50 hPa for the NewR experiment compared to the Control in terms of the standard deviations of the differences between MLS and the ozone analysis (compare the top and bottom rows of Fig. 21). For most (but not all) months, the NewR experiments also show smaller standard deviations than the Denial (e.g., Fig. 21c), providing an independent confirmation that IASI is able to improve the ozone analysis. The situation is less clear for the Control, where improvements compared to the Denial are mainly confined to the Antarctic region and the Southern Hemisphere winter months, during which SBUV observations are not available in this region (e.g., Fig. 21b).

In this context it should be mentioned that the choice of the ozone channel used to anchor the variational bias correction has an important effect on the performance of the ozone analysis. As mentioned earlier,

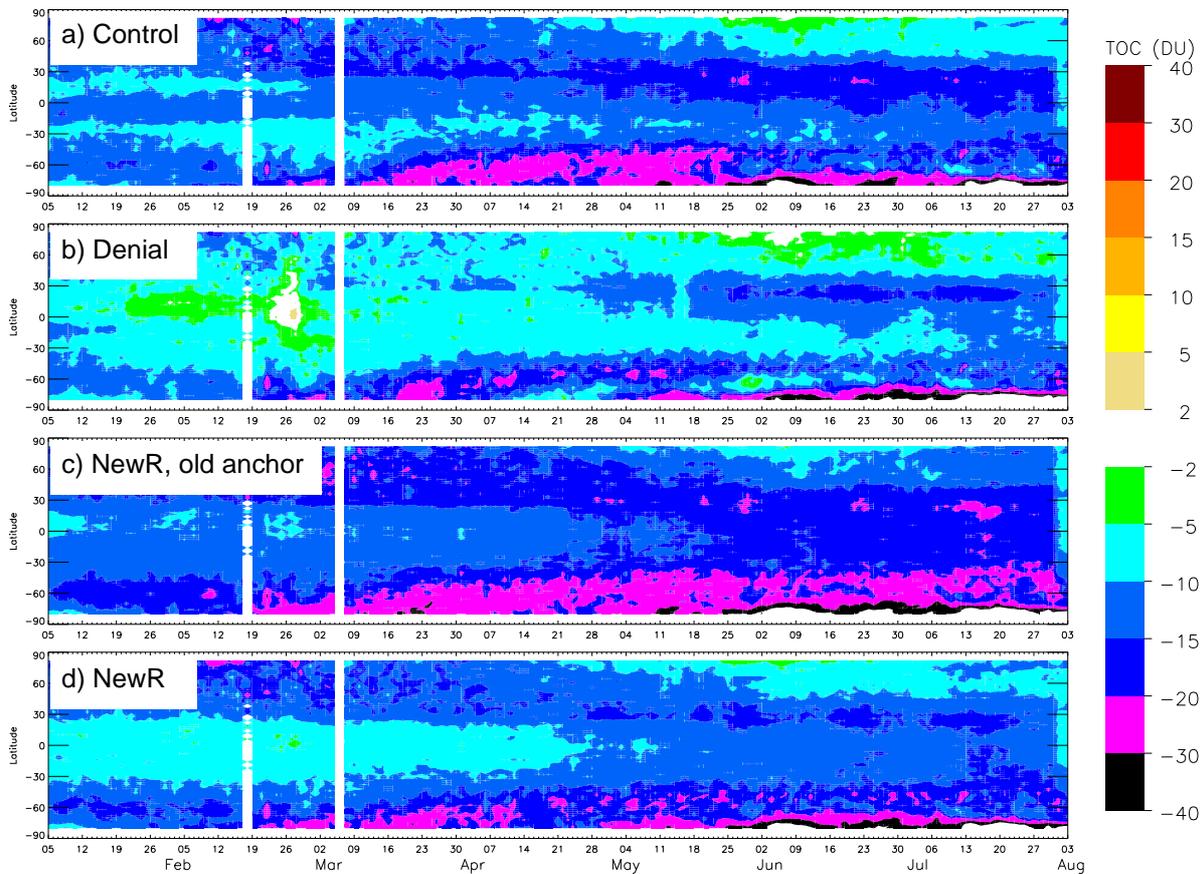


Figure 22: a) Hovmoeller plot of the mean differences in total column ozone between MLS and the Control analysis [DU]. b) As a), but for the Denial. c) As a), but for an experiment with the updated observation error covariance for IASI and the same anchor channel as used in the Control. d) As a), but for the NewR experiment.

in the NewR experiment, we use channel 1574 without a bias correction, whereas channel 1585 is used in the Control. The main reason for this change is that initial experiments with the old anchor channel showed a significant degradation in the ozone bias, as indicated by ozone profiles from SBUV or MLS, or OMI total column ozone values (compare, for instance, Fig. 22a and c). This change in bias is a result of the increased weight given to ozone channels in the NewR experiments, combined with a non-zero analysis bias for the anchor channel in the Control. The increased weight of the ozone channels means the analysis draws more closely to the ozone channels, including any biases in these. As the anchor channel has a non-zero bias against the analysis in the Control, this means a modification to the bias of the ozone analysis if the same anchor channel is used. To ameliorate this, we chose instead an ozone anchor channel for IASI which has a near-zero bias correction in the Control (and, by design of VarBC, a near-zero bias against the analysis after bias correction). This successfully restores the ozone biases to levels similar to the previous values, with a somewhat better performance over the tropics (compare Fig. 22a and d). Obviously, simply using a single channel to anchor the variational bias correction remains a pragmatic solution, and the choice needs to be re-evaluated whenever bias characteristics change, for instance when upgrading spectroscopic parameters. Further work would be beneficial to better restrict the size of bias corrections obtained by VarBC, possibly linked to estimates of plausible sizes of the biases, for instance

based on the uncertainty in the spectroscopy for ozone channels.

The choice of the ozone anchor channel also has an effect on other aspects of the assimilation and is not only confined to the ozone analysis. For instance, the change in the channel used for anchoring restored the size of the standard deviations of background departures for short-wave channels of AIRS, and also led to improvements in the day 5-10 forecast scores for temperature at 100 hPa over the Northern Hemisphere (not shown). The sensitivity to the anchor channel may be increased as a result of the stronger weight given to the ozone channels in the NewR experiment, but it is probably present in the Control experiment as well. This is an unsatisfactory situation that warrants further study.

## 6 Use of the new observation error covariance in the EDA

So far we have considered only the effect of updating the observation error covariance matrix for IASI in a deterministic experiment with 4-dimensional variational data assimilation. However, in the ECMWF system, the observation error covariance is currently also used to determine the size of the perturbations applied to observations in the Ensemble of Data Assimilations (EDA) which is used to produce flow-dependent estimates of background errors in 4DVAR (e.g., Bonavita et al. 2012). In the current operational configuration of the EDA, 25 independent 4DVAR experiments are run with different random perturbations to the observations, as well as perturbations to the physics parameterisations and sea surface temperatures. As a result, the update in the observation error covariance for IASI potentially also has an effect on the estimation of the background errors used in 4DVAR. This aspect has been neglected in the experiments presented in the previous sections, in which the same background error has been used for all experiments.

We will now investigate the effect of also updating the observation error covariance for IASI in the EDA. First, we will characterise the influence of the change on the spread and correlations statistics of the EDA. After that, we will investigate to what extent an updated EDA affects the performance of the modified observation error covariance when used to specify background errors in assimilation experiments.

### 6.1 Experiments

To investigate the impact of the observation error covariance update we performed two EDA experiments: in the Control experiment, the old observation error specification for IASI is used as in operations. In contrast, in the NewR EDA experiment we use the updated observation error covariance matrix including inter-channel error correlations to define the perturbations applied to assimilated IASI observations, as well as for the assimilation of the IASI data. In contrast to the NewR experiments presented in the previous section, the NewR EDA experiment uses the same ozone anchor channel as the Control, as these experiments were conducted earlier. Both EDA experiments cover the two periods 5 February to 4 May 2014 and 16 July to 30 September 2014. The spatial model resolution of the EDA is  $T_L399$ , with an incremental analysis resolution of  $T_L255$  and 137 levels in the vertical. We use 25 ensemble members, as in the current operational configuration.

### 6.2 Influence on the spread statistics of the EDA

The update of the observation error covariance matrix for IASI has a notable influence on the spread of the EDA, that is, the standard deviation of the ensemble members (e.g., Fig. 23). The spread in temper-

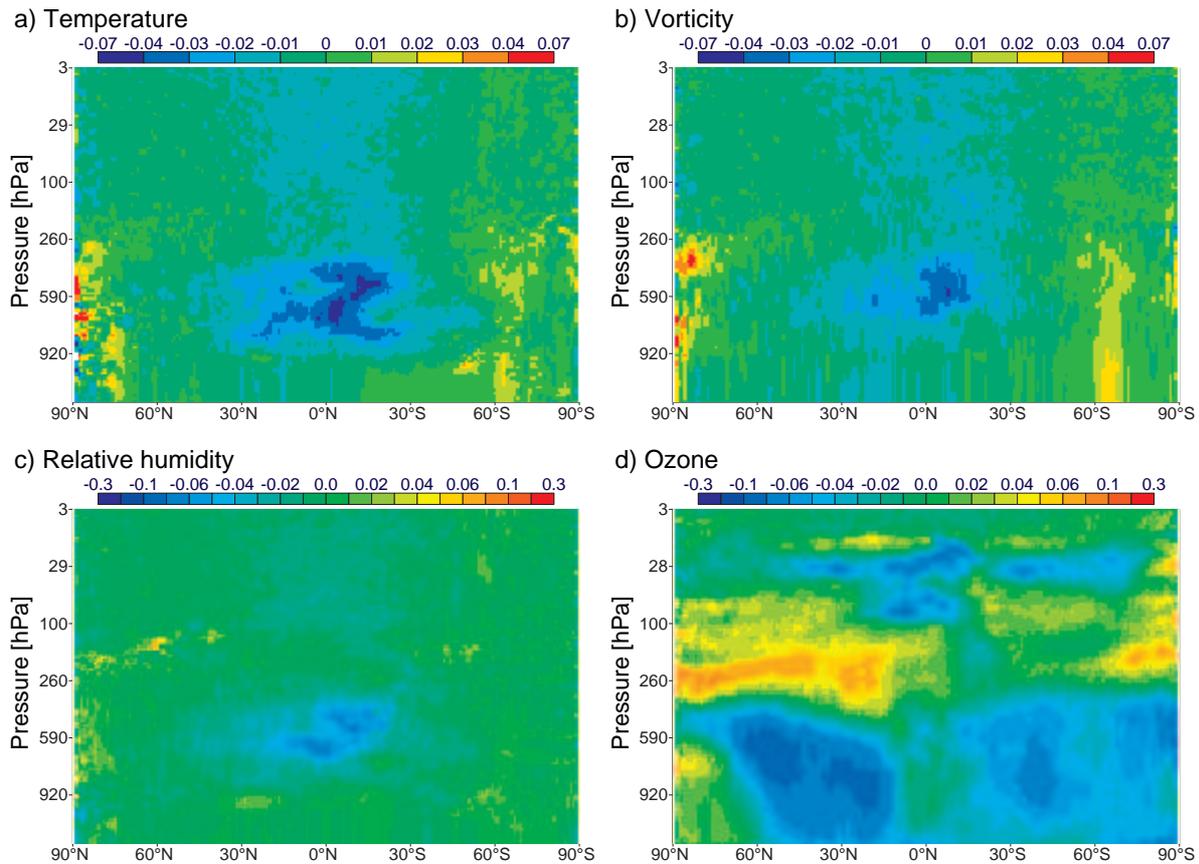


Figure 23: a) Zonal mean of the relative difference in the temperature spread of the EDA for August 2014. Negative values indicate a reduction of the spread in the NewR experiment compared to the Control. b) As a), but for vorticity. c) As a), but for relative humidity. d) As a), but for ozone.

ature, vorticity and relative humidity is typically decreased by around 3 % in the mid-troposphere in the tropics (a little more for relative humidity), whereas some increases in the spread can be found at higher latitudes. The EDA spread is an indication of the size of the background errors, under the assumption that the applied perturbations to observations, sea surface temperatures and physics parameterisations reflect the true errors. The parameters and regions with the largest reduction in the spread are also the areas where our investigation of the 4DVAR experiments found the clearest improvements in the observation fit statistics or forecast impacts. This is reassuring as a reduction in the spread can be seen as a further indication of an overall improvement of the assimilation system in these regions. At the same time, it is difficult to determine whether the NewR EDA also provides an improved estimate of the size of the background error. This is because the EDA spread is the result of the inter-play of not only the observation perturbations, but also the stochastic physics and perturbations of the sea surface temperatures, all of which may have their deficiencies. A detailed analysis of this aspect is beyond the scope of the present paper and is hence left for future work.

The update of the observation error covariance also has a small effect on the vertical correlations of the background errors derived from the EDA. This can be seen, for instance, in globally averaged statistics derived from the two EDAs and displayed in Fig. 24 for temperature and humidity. These are used in the ECMWF system to estimate the background error correlations. The changes are relatively small and mainly confined to the lower and mid-troposphere, below model level 80 (around 250 hPa). For

temperature, the modification leads to a slight broadening of the correlations, combined with less strong anti-correlations on the off-diagonals. For humidity, the correlations are slightly sharper around model level 120, which is at 920 hPa for a 1013 hPa surface pressure. This may reflect a better representation of vertical detail for humidity. The mid- to lower tropospheric levels with the largest changes in the correlations are levels where we have previously noted clear signals in the forecast evaluations.

The results show that the different error characteristics assumed for IASI affect both the size of the EDA spread as well as the vertical correlations between differences of the EDA members. The latter gives a further indication that the updated observation error covariance matrix indeed affects IASI's ability to resolve vertical features. This aspect could be studied further, for instance in simpler 1DVAR simulations that estimate the theoretical information content of IASI in an idealised framework (e.g., Collard and Healy 2003). Such studies tend to assume rather simple observation errors that do not take

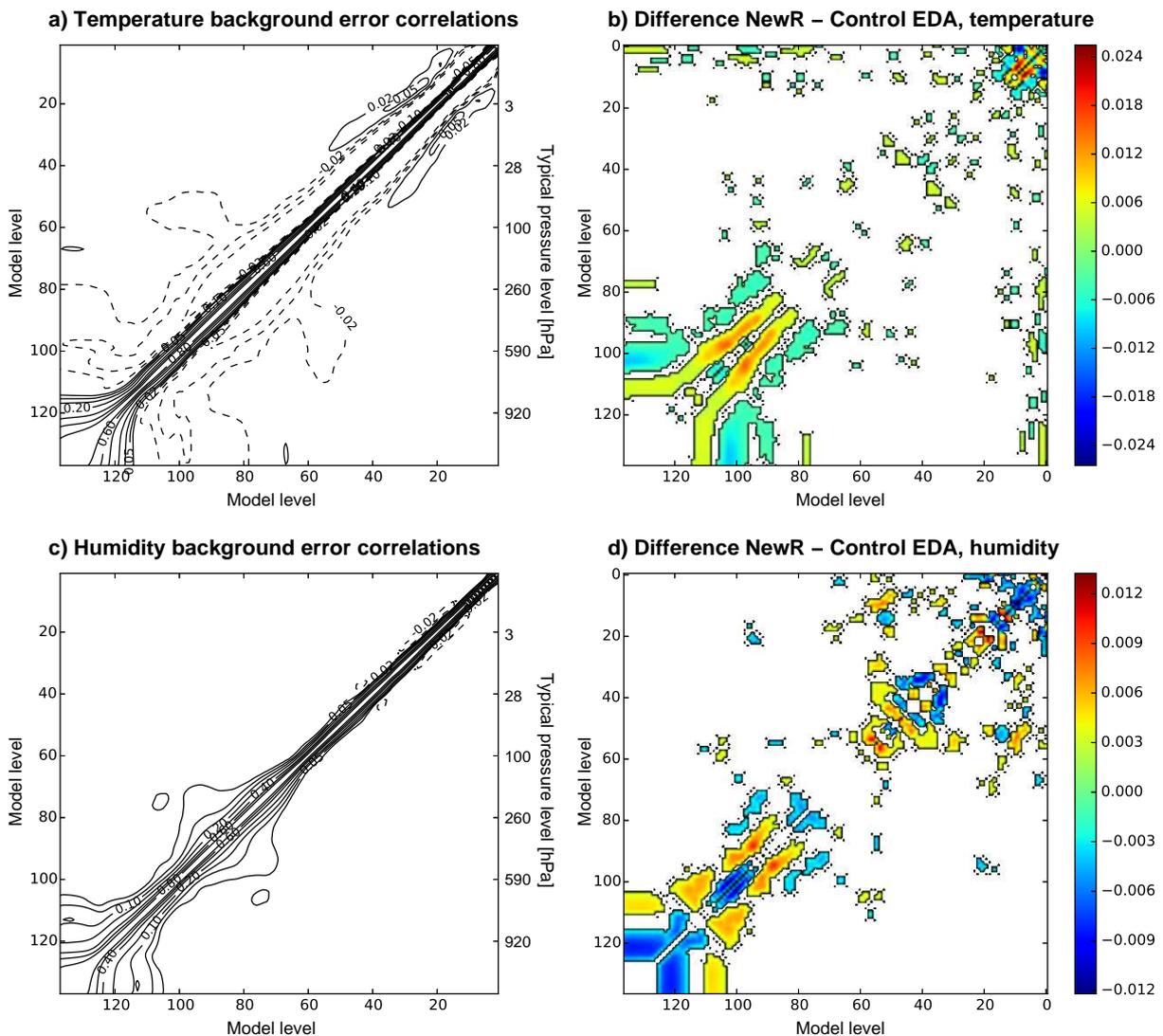


Figure 24: a) Global average of the vertical correlations of background errors for temperature derived from the Control EDA experiment over the July–September period. b) Difference in the temperature background error correlations derived from the NewR and the Control EDA. c) As a), but for humidity. d) As b), but for humidity.

into account correlated contributions to the error. In the EDA context, the overall effect of assuming a more realistic observation error appears fairly small, but it is possible that it is more significant in specific meteorological conditions, and this aspect is not covered here. Further investigations of the EDA spread and correlation characteristics are beyond the scope of the present memorandum and these are left for future work.

### 6.3 Analysis and forecast impact in high-resolution 4DVAR experiments

The influence of the upgrade of the observation error covariance matrix for IASI on the specification of the background errors has been tested further by running 4DVAR experiments that include or neglect the observation error upgrade in the EDA that is used for the background error specification. Three experiments will be discussed here: the Control experiment uses the operational observation error specification in the EDA and in 4DVAR. In contrast, the NewR experiment uses the updated observation error covariance matrix for IASI, but with a background error specification that is the same as in the Control (equivalent to what has been done in section 5). The ozone anchor channel is also updated as in earlier experiments. In the NewR+EDA experiment, we use the same setup as in the NewR experiment, but also apply the updated observation error covariance matrix in the EDA which is used to specify the situation-dependent component of the background errors. Note that in the latter experiment the overall magnitude

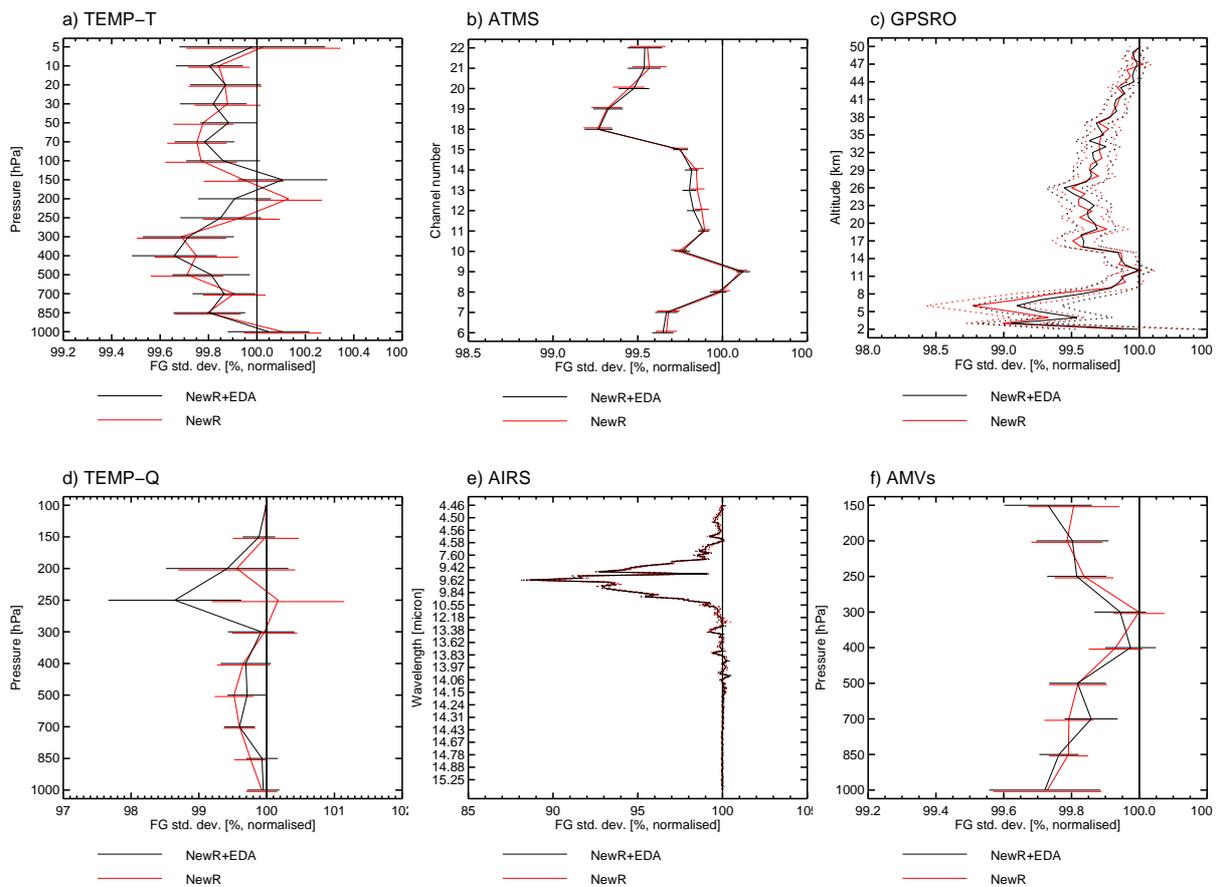


Figure 25: As Fig. 13, but for the NewR (red) and NewR+EDA (black)  $T_L1279$  experiments compared to the Control, rather than a Denial. Statistics over 5  $\frac{1}{2}$  months of experimentation have been combined.

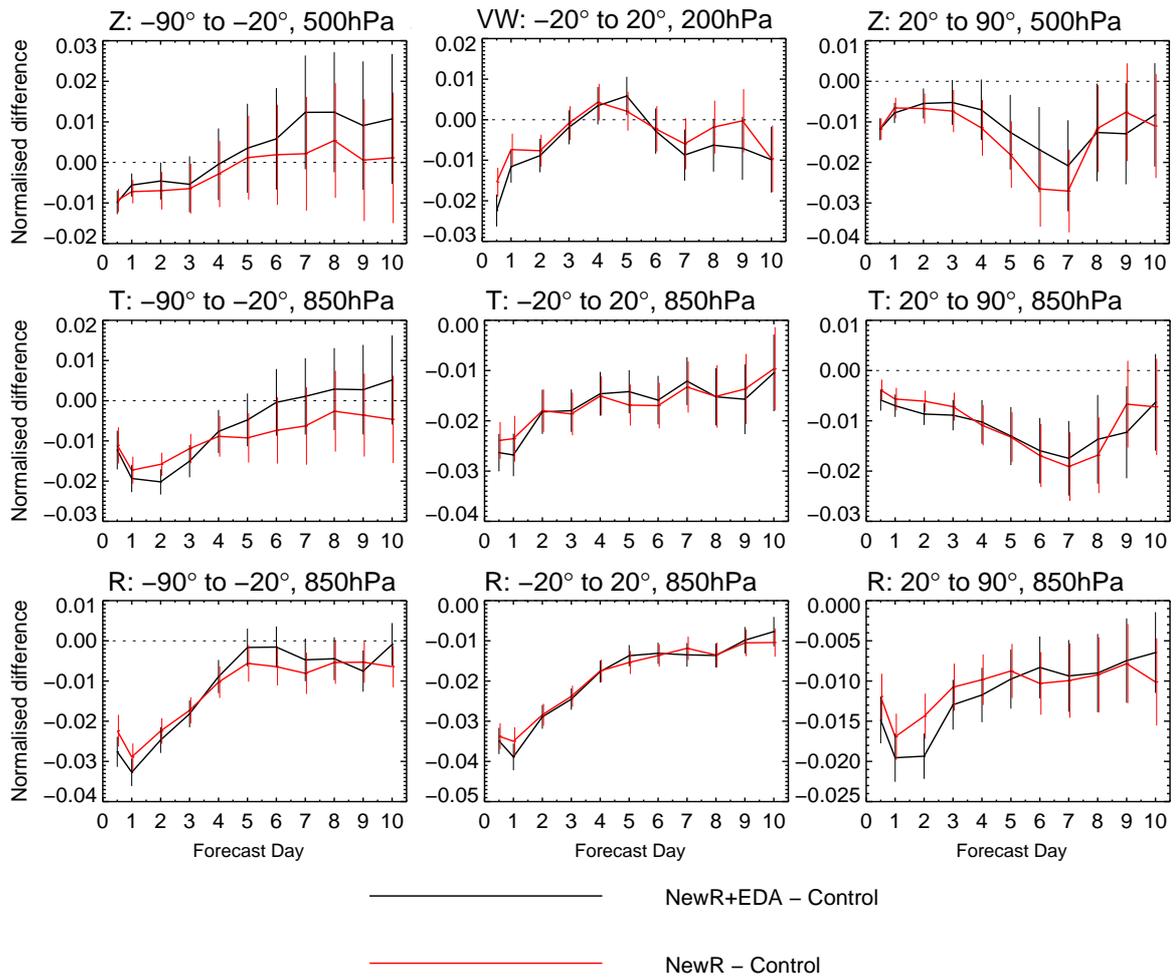


Figure 26: As Fig. 18, but for the NewR (red) and NewR+EDA (black)  $T_L1279$  experiments compared to the Control, rather than a Denial. Statistics over  $5 \frac{1}{2}$  months of experimentation have been combined. A negative value means an improvement compared to the Control.

of the background error has not been altered - this overall magnitude is determined by a climatological background error that is the same for all experiments. All three experiments are run at the higher spatial model resolution of  $T_L1279$  ( $\approx 16$  km), that is, the spatial resolution currently used operationally at ECMWF, combined with an incremental analysis resolution of  $T_L255$  ( $\approx 80$  km, ie, the same as in earlier experiments). This also tests whether there is some resolution-dependence of the results presented in section 5. The experiments cover the periods 5 February to 4 May 2014 and 16 July to 30 September 2014, a total of  $5 \frac{1}{2}$  months. They use the full observing system assimilated operationally at the time, including two IASI instruments.

The  $T_L1279$  NewR experiment shows the same improvements as the earlier  $T_L511$  experiments, whereas the differences between the NewR and the NewR+EDA experiment are very minor. This can be seen in statistics for other observations (e.g., Fig. 25, compare to Fig. 13) as well as in forecast scores of verification against the own analysis (e.g., Fig. 26, compare to Fig. 18). The former show overall the same marked reductions in the standard deviations of background departures for the NewR experiment compared to the Control as seen previously in  $T_L511$  experiments, but little statistically significant changes

for the NewR+EDA experiment compared to the NewR experiment. Short-range forecast scores are also similarly improved in the NewR experiment compared to the Control, with particularly noticeable changes for lower level humidity and temperature (e.g., Fig. 26). Again, there is no clear advantage for the NewR+EDA experiment compared to the NewR experiment in terms of forecast scores.

The lack of impact from using the updated EDA for the background error specification is somewhat disappointing, given the clear effect on the ensemble spread statistics. However, two aspects of the presented experimentation are worth bearing in mind here. Firstly, in the present experiment, only the situation-dependent aspects of the background error have been updated, whereas the overall magnitude of the underlying static climatological background has not been altered. This will reduce the impact of the background error update, as overall changes to the size of the background error will not be taken into account. Secondly, the derivation of the background error from the EDA statistics includes a calibration step, which uses the operational ECMWF analysis to estimate short-range forecast errors (e.g., Bonavita et al. 2012). Ideally, this calibration step should have also used the analysis from the NewR+EDA experiment, but this is currently technically not possible. The relevance of both of these aspects is not clear, but if addressed they may contribute to a more appreciable difference in the impact.

The results suggest that the forecast impact found earlier in  $T_L511$  experiments is also representative for experiments with much higher spatial model resolution. This is an interesting finding, as it has been argued that some of the error correlations are due to (spatial) representativeness errors (e.g., Stewart et al. 2013, Weston et al. 2014). Observation error covariances diagnosed from  $T_L1279$  experiments with the Desroziers et al. (2005) method indeed show some small differences, for instance, for humidity channels (not shown). But it appears that the matrix derived from  $T_L511$  nevertheless is a significant improvement over the currently used observation error. It is worth pointing out in this context that the  $T_L511$  experiments and the  $T_L1279$  experiments have been run with the same incremental analysis resolution. This may contribute to the finding that the error derived from a lower-resolution experiment is adequate for the higher resolution experiment. It remains to be seen what impact changing the spatial resolution of the incremental analysis resolution has on the present results.

## 7 Summary and conclusions

This memorandum has investigated the use of an updated observation error covariance matrix for IASI. The matrix is derived from observation space diagnostics (Hollingsworth and Lönnberg 1986, Desroziers et al. 2005), and uses different  $\sigma_0$  values compared to the currently assigned observation error and takes inter-channel error correlations into account. The currently assigned observation error consists of a diagonal matrix with constant diagonals over three wave-number bands, with substantial inflation for stratospheric, window, humidity, and ozone channels. The additional computational cost of taking the inter-channel error correlations into account is negligible. The main findings are:

- The new observation error covariance allows the use of an observation error that is more consistent with departure statistics for many assimilated channels. It leads to significant benefits in terms of forecast skill over the first 3-4 days compared to the observation error currently used. The improvements are clearest in the lower troposphere, the stratosphere, and in the tropics, and for humidity.
- The new observation error covariance matrix leads to a different weighting of IASI in the analysis, with reduced increments and an overall much improved consistency with other observations.

- The update has a particularly notable impact on the ozone analysis, improving the fit to other assimilated as well as independent ozone observations.
- Using diagnosed observation errors while neglecting the diagnosed error correlations leads to a significant degradation of the analysis compared to not using IASI data at all. Error inflation can be used to partially compensate for this effect, with optimal inflation factors of around 2.5-3.0.
- Some inflation of the diagnosed  $\sigma_O$  values is beneficial also if inter-channel error correlations are taken into account, but the optimal factor is smaller than when error correlations are neglected (around 1.75). Using diagnosed values without inflation does not lead to a similarly strong degradation as seen without error correlations, suggesting the impact is more robust.
- Using the updated observation error covariance in the EDA leads to notable changes in the ensemble spread used to specify background errors in 4DVAR, and minor changes to the vertical correlations. However, the benefits from using the updated background error from this EDA in 4DVAR are minor.

The observation error upgrade leads to an overall more consistent assimilation of IASI data. A number of diagnostics suggest that the short-range impact of the observation error upgrade is comparable in magnitude to the introduction of IASI observations in the current system. However, for the medium-range forecasts, the impact is not as strong, and the upgrade is mostly neutral for a forecast range beyond day 5. The reasons why the short-range benefits do not translate to stronger improvements in the longer medium-range are not clear, but the finding may reflect IASI's particular role as an instrument that has the ability to provide additional vertical information.

The present study provides another example that a more sophisticated treatment of observational uncertainty in the assimilation of satellite data can lead to very significant improvements in the use of these observations. It is also another example that highlights that observation-space diagnostics and the inter-channel error correlations provided by them give useful information for such a more sophisticated treatment of the observation errors. Similar results have been obtained by Weston et al. (2014) for IASI data, and by other observation error re-specification studies for AMSU-A or GPS-RO data at ECMWF. While the observation-space diagnostics will have short-comings and provide misleading results when their assumptions are egregiously violated, it is clear that they provide useful input to the observation error specification that enables us to improve on the current observation error specifications for satellite data.

At the same time, a number of questions remain about the use and applicability of the observation-space diagnostics. For instance, our study found considerable benefits from scaling the error standard deviations estimated with the diagnostics, and only small benefits from methods that adjusted the conditioning of the diagnosed matrix. In contrast, Weston et al. (2014) emphasises the need for improving the conditioning of the diagnosed matrix. The finding that an adjustment is beneficial is similar in the two studies, as is the finding that benefits are obtained with  $\sigma_O$  values that are very likely to be larger than the true observation errors. However, the chosen approaches to make adjustments and the severity of the encountered problems when using un-adjusted matrices are very different.

The finding that adjustments to the diagnosed matrices are beneficial raises the question why these adjustments are necessary and what adjustments are most plausible. When considering this question it is important to bear in mind that these are adjustments made in order to specify observation errors in an assimilation system, and they thus reflect the characteristics of the assimilation systems and the degree of sophistication used for the observation error specification. For instance, the adjustment may merely reflect that we still use only a globally constant observation error covariance with inter-channel error

correlations, but we continue to neglect other aspects of the true observation error characteristics. We continue to neglect spatial error correlations, the scene-dependence in brightness temperature space of the instrument noise contribution, and the situation-dependence of other aspects of the observation error, such as cloud screening or radiative transfer errors. It should therefore be not surprising that some adjustments are necessary, and that we continue to need to assume observation errors that are most likely larger than the true observation errors. On the other hand, it is also possible that the adjustments are addressing short-comings in the diagnostic methods, arising from assumptions made during their derivation that are likely to be not strictly true (e.g., assumptions on the weights used in the assimilation system, or the assumption that there are no error correlations between the background and the observations). While we cannot rule out the latter possibility, we consider the remaining sub-optimality the dominant reason, at least for the majority of the channels considered here. Future work on further refinements to the assumed observation errors is likely to be beneficial to further understand which adjustments are most appropriate.

To better interpret the results from the observation error diagnostics we expect that an enhanced understanding of the statistical properties of the main uncertainties in the assimilation of IASI data will be very beneficial. Studies that aim to build up an inventory of all error contributions will also help to highlight where the diagnostics are plausible and where there may be short-comings. Such studies will provide further input to the observation error specification, and they are also likely to help identify areas where improvements in the assimilation of IASI can be obtained in general. At the same time, compiling a complete error inventory that includes reliable estimates for contributions such as the instrument noise, radiative transfer error, cloud screening and representativeness error is a challenging task, and it is likely that such an inventory will also be prone to assumptions that are not strictly valid. Nevertheless, such studies are currently under-way at ECMWF, and they are expected to shed further light on the assimilation of IASI data.

The current results are highly relevant for the assimilation of other hyper-spectral IR sounders, such as AIRS or CrIS (e.g., Aumann et al. 2003, Han et al. 2013), or future geostationary hyperspectral infrared sounders planned for the European Meteosat Third Generation satellites or the Chinese FY-4 series. Observation error contributions from sources other than instrument noise are even more relevant for low-noise instruments such as CrIS, and the use of more sophisticated observation errors that include the correlations of these errors is therefore considered important. The findings also contribute to a better use of alternative representations of hyperspectral IR data, such as Principal Component Scores or reconstructed radiances (e.g., Collard et al. 2010, Matricardi and McNally 2014). In the case of reconstructed radiances, very low noise is achieved through linear combinations of the observations, leading to error correlations in the transformed instrument noise. Achieving the full benefit from these alternative representations is likely to be dependent on the use of a full observation error covariance matrix.

Our experiments show particular benefits for the humidity analysis, an area where the use of IASI and other hyperspectral IR instruments has so far been relatively limited in the ECMWF system. It appears that accounting for the relatively strong inter-channel error correlations for the humidity channels, together with an improved weighting of surface-sensitive channels enables a more robust impact in this respect. In past experimentation, it has often been difficult to successfully add further humidity sounding channels with the present assimilation setup, without an increased observation error. The more sophisticated treatment of humidity channels in the present experiments may allow a more extensive use of the water vapour band. Preliminary experiments in this direction with 114 additional humidity channels from IASI have been conducted, using the same channel selection as in Matricardi and McNally (2013) and Matricardi and McNally (2014). These preliminary experiments indeed show further improvements in terms of background fits to other observations when the additional channels are assimilated with an observation error covariance matrix derived and scaled by 1.75 in an equivalent way as reported in the

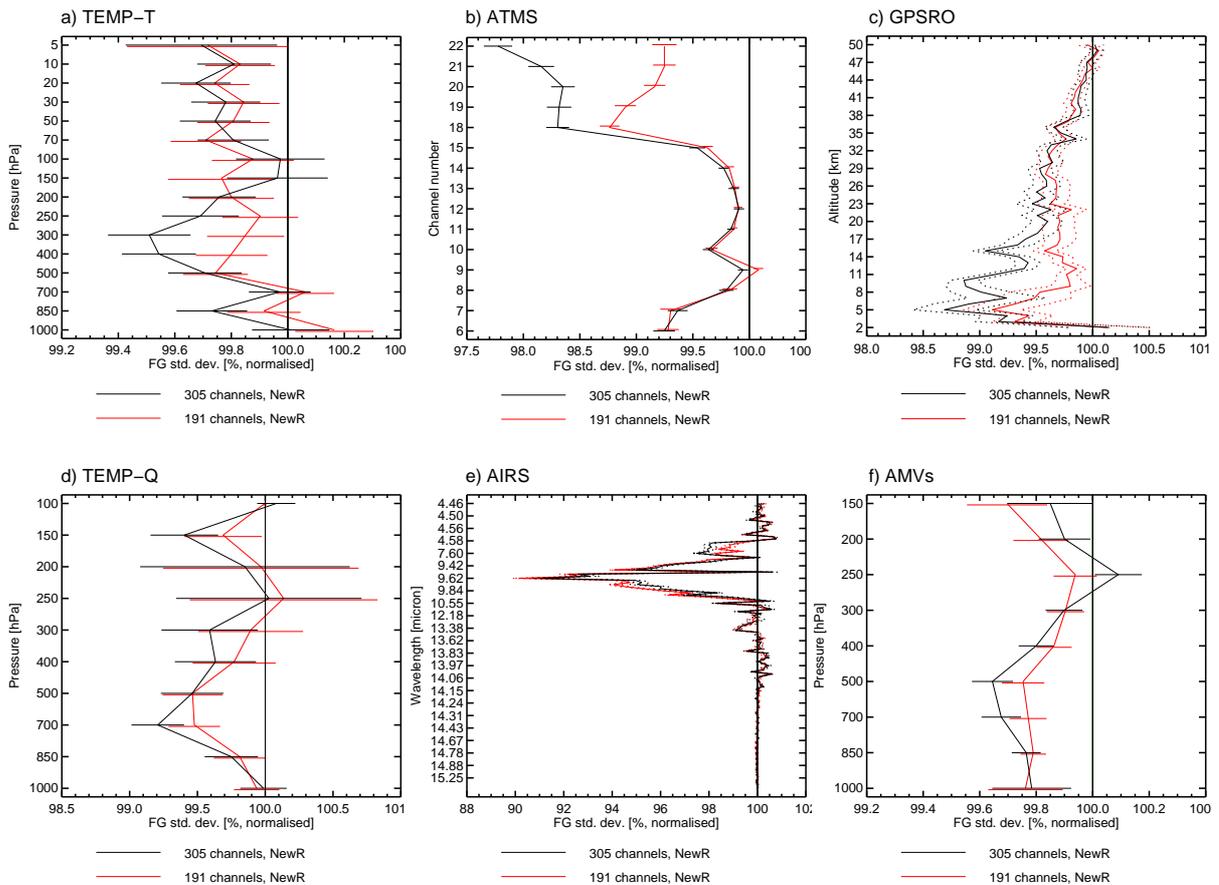


Figure 27: Global standard deviations of background departures for a selection of observing systems for two experiments that use an observation error covariance matrix as constructed in this Technical Memorandum, but assimilated a different number of IASI channels. Red shows results for using 191 IASI channels (similar to the NewR experiment described in section 5, but without an update to the ozone anchor channel), whereas black shows results for using 305 channels, with the additional channels primarily located in the humidity band. The experiments use a  $T_L511$  spatial model resolution and cover the  $6\frac{1}{3}$ -month period 5 February – 14 August 2014. The values have been normalised by the standard deviations of background departures from a Control experiment that uses the operational observation error and 191 channels. Horizontal lines indicate 95% confidence intervals. The observing systems are: a) Radiosonde temperature observations, b) ATMS brightness temperatures (with temperature sounding channels 6-15, and humidity sounding channels 18-22), c) GPSRO bending angle observations from COSMIC, Metop-A and B and GRACE-A, d) Radiosonde humidity observations, e) AIRS brightness temperatures, and f) Atmospheric Motion Vectors from 10 satellites.

present memorandum (e.g., Fig. 27), although with neutral impact on medium-range forecast scores (not shown). Further work in this area is likely to be beneficial to enhance our understanding of the use of the water vapour band.

The finding of particular benefits for humidity also has wider implications for the general use of humidity-sensitive radiances. An increased use of humidity channels is expected to be particularly useful in preparation for future geostationary hyperspectral sounders, which should allow a better representation of the humidity field at frequent time intervals, and hence enable better estimation of dynamical information through tracing effects in 4DVAR. If representativeness or radiative transfer errors are indeed dominant contributors to the correlated errors for humidity channels then the findings reported here are likely to be

also relevant for the assimilation of microwave humidity sounders. These have been found in the past to also exhibit inter-channel error correlations when assimilated in clear-sky conditions (e.g., Bormann and Bauer 2010), and such error correlations are likely to be even stronger when the data is assimilated in all-sky conditions, where representativeness errors are a much larger contribution (e.g, Geer et al. 2014). A more sophisticated treatment of such observation error contributions and their full characteristics is likely to lead to further benefits.

## Appendix A: Derivation of the observation error covariance matrix

### Diagnosing the observation error covariance

The updated observation error covariance matrix used for IASI in the present study has been derived based on methods described in Bormann and Bauer (2010) and Bormann et al. (2010). The estimates use a combination of observation-space diagnostics, applied to samples of IASI data for which all 191 assimilated channels are used in the assimilation system. The diagnostics have been applied in the following way:

- First, we use the Hollingsworth and Lönnberg (1986) method to obtain an initial estimate of the observation errors including their inter-channel correlations. These could have been derived from an experiment that monitors IASI data passively, but for convenience we chose an experiment with active assimilation of IASI.

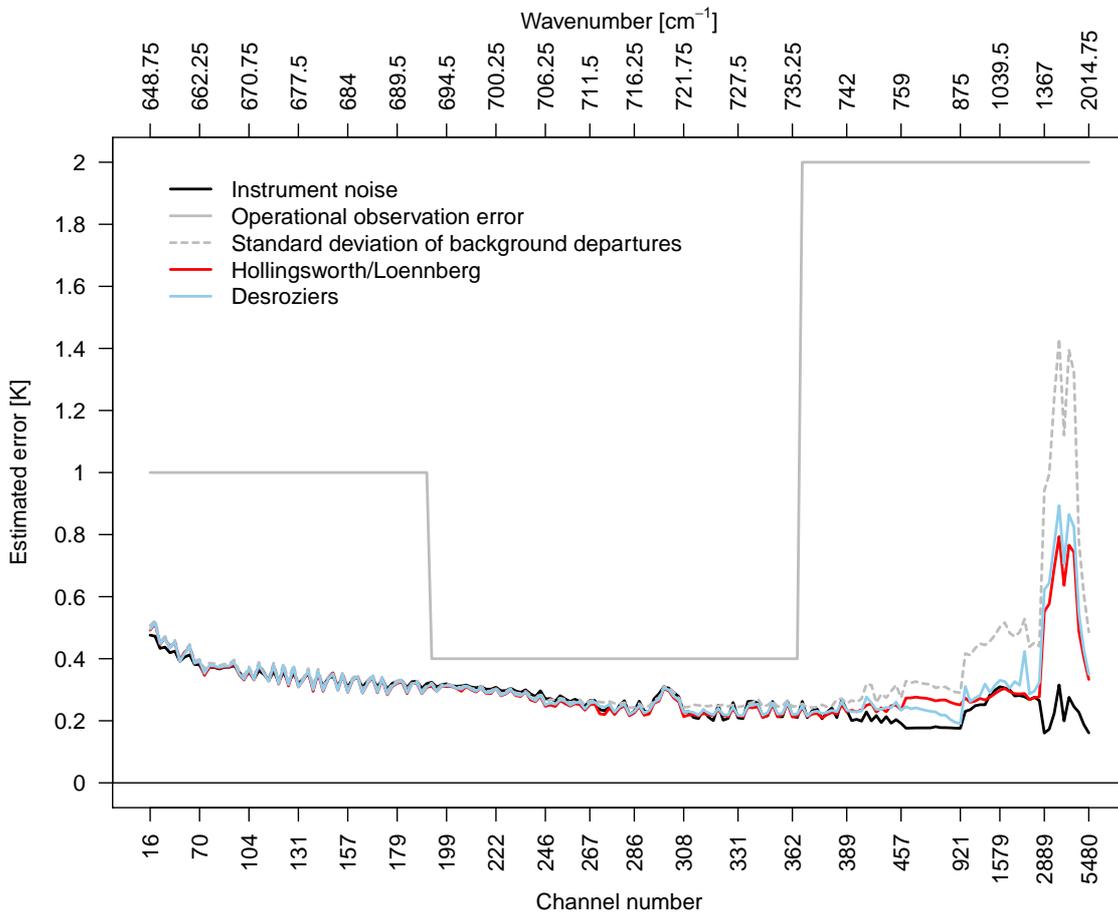


Figure 28: Raw diagnosed observation error ( $\sigma_0$ ) for assimilated IASI channels. The red line shows the initial values obtained with the Hollingsworth/Lönnberg method from an experiment that assumes the operationally used observation errors (grey). The blue line gives the results from the Desroziers diagnostic applied to an experiment that assumes an observation error covariance matrix based on the Hollingsworth/Lönnberg matrix (with inflation of  $\sigma_0$  by 75 %) as described in the main text. Also shown is an estimate of the instrument noise (black).

- Next, we use these diagnosed covariance matrices in a new assimilation experiment, with  $\sigma_O$  scaled by 1.75 (see section 4), and apply the Desroziers et al. (2005) diagnostic to re-estimate the observation error covariances and their inter-channel error correlations. This re-diagnosed matrix is the basis for our further investigations.

The rationale behind this approach is as follows: the first step allows an initial estimate of the observation error covariance that could in principle be independent of earlier assimilation choices and avoids the very strong inflation of the observation error for some channels (e.g., window or ozone channels). Note that this initial estimate does not make an assumption on the size of the background error used in the assimilation system. At the same time, the Hollingsworth/Lönnberg method makes the strict assumption that observation errors are spatially uncorrelated, and all spatial error correlations are due to background error. This is questionable for radiance observations, as radiative transfer or representativeness error may show spatial error correlations, as has been found by Bormann et al. (2010) particularly for window and humidity channels. The assumption of negligible spatial error correlations is relaxed in the second step by applying the Desroziers diagnostic. When applying the Desroziers diagnostic, the assumption is made that the weights specified in the assimilation system are consistent with the true weights, thus introducing a potential dependence on the assumed background error. Further experimentation has found this sensitivity to the specification of the background error to be relatively small (see Appendix B).

All diagnostics have been derived from one month of Metop-A IASI data (15 March - 14 April 2014), based on experiments with a  $T_L511$  spatial model resolution and the full operational observing system. The resulting  $\sigma_O$  and error correlation matrices are shown in Figures 28 and 29, respectively. The adjustment made by the second step is relatively small, but there are nevertheless noticeable differences for window, ozone, and humidity channels. Statistics have also been derived for Metop-B IASI, which show slightly smaller estimates of  $\sigma_O$ , consistent with independently estimated lower instrument noise for the given period.

The re-calculation of the observation error covariance using the Desroziers diagnostic gives some small benefits in terms of forecast impact. To test this, two experiments have been run with the  $T_L511$  configuration used in section 5, covering the 6 month period 5 February - 4 August 2014. One experiment uses the initial Hollingsworth/Loennberg matrix, with  $\sigma_O$  scaled by 1.75, and one uses the re-calculated Desroziers matrix, with  $\sigma_O$  again scaled by 1.75. The impact on geopotential scores between the two experiments is relatively small, but there are nevertheless some apparent benefits in terms of forecast skill for low-level temperature and humidity (Fig. 30). In contrast to this, changes in the FG-fit for other assimilated observations are mostly minor and often not statistically significant. Nevertheless, some small reductions in the standard deviations of FG departures can be seen for some observations and these are vaguely consistent with the short- and medium range forecast scores (Fig. 31). One of the clearer benefits is found for ozone, where standard deviations of FG departures are reduced when the re-calculated Desroziers covariance is used. The improvements are consistent with the finding that the largest changes in the updated covariance matrix are occurring for low-level temperature, window, and ozone channels.

## Adjustments to the observation error covariance

It is illustrative to examine the diagnosed observation error covariance matrices a little further. To do so, we present in Figure 32 the eigenvalues of the Desroziers-diagnosed matrix together with the instrument noise. The instrument noise has been diagnosed independently from the instrument monitoring, and a full matrix with error correlations due to the apodisation is used here. There are several aspects worth noting:

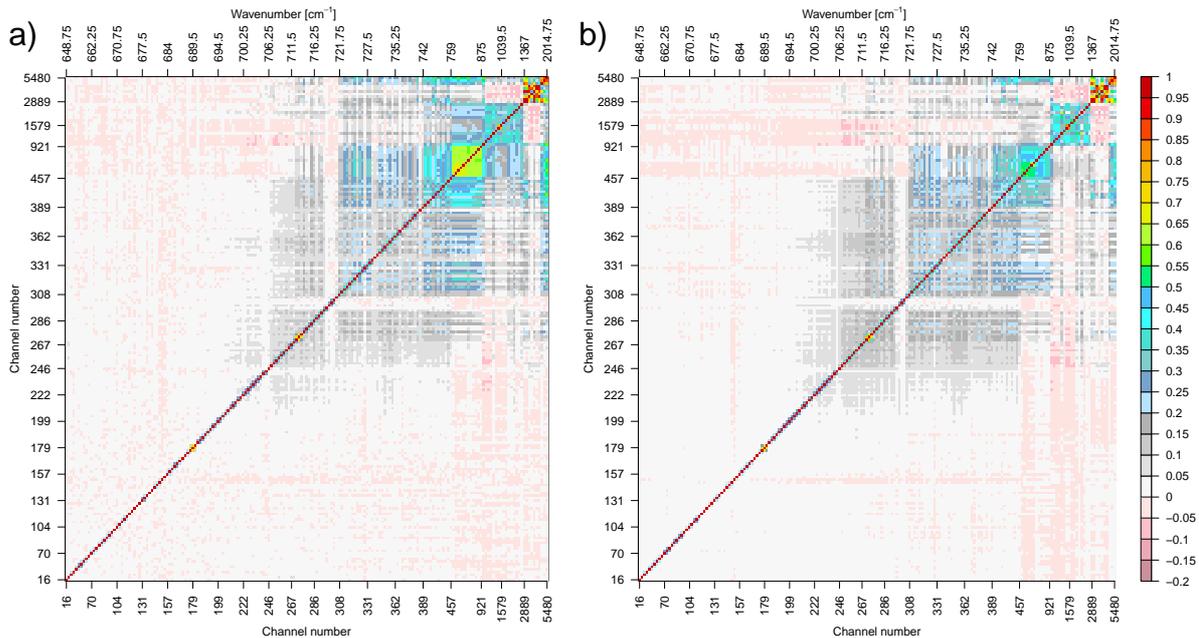


Figure 29: a) Raw diagnosed inter-channel error correlations obtained with the Hollingsworth/Lönnerberg method from an experiment that assumes the operationally used observation errors. b) Raw diagnosed inter-channel error correlations obtained with the Desroziers diagnostic applied to an experiment that assumes an observation error covariance matrix based on the Hollingsworth/Lönnerberg matrix as described in the main text.

- For the largest eigen-values, the Desroziers diagnostic gives values that by far exceed the instrument noise, and it appears that for these structures other error contributions dominate.
- Many of the higher-order eigen-vectors have, in contrast, eigen-values that are smaller than corresponding eigen-values calculated from the instrument noise. For eigen-values 100-170, the instrument noise matrix is smaller by a rather constant 20 %, implying that our diagnosed matrix suggests errors for these spectral structures that are below the instrument noise.
- For the very smallest eigen-values, there is a sharp drop, both for the diagnosed matrices and the instrument noise. For the instrument noise, the shape of this sharp drop-off is largely determined through the error correlations due to apodisation, which are well-known as they originate from a mathematical manipulation of the measured spectrum. The drop-off for these smallest eigen-values starts earlier for the diagnosed matrices than for the instrument noise.

The reasons for the finding that a range of eigen-values of the diagnosed matrices is smaller than those from the instrument noise are not clear. The comparison is somewhat hampered by the fact that the instrument noise is not constant in brightness temperature space. The instrument noise is constant in radiance space, but the non-linearity of the conversion to brightness temperature space adds a scene-dependence to the instrument noise estimate in brightness temperature space. The conversion of the instrument noise used here is performed at a channel-dependent brightness temperature derived for a standard atmospheric profile, whereas we compare the results to the diagnostic estimates obtained from a sample of IASI data with global coverage. This obviously does not allow a very strict comparison, and differences may simply be an artifact of this. Also, the instrument noise shows significant temporal variations, depending on the instrument decontamination which is performed every few years. The lowest

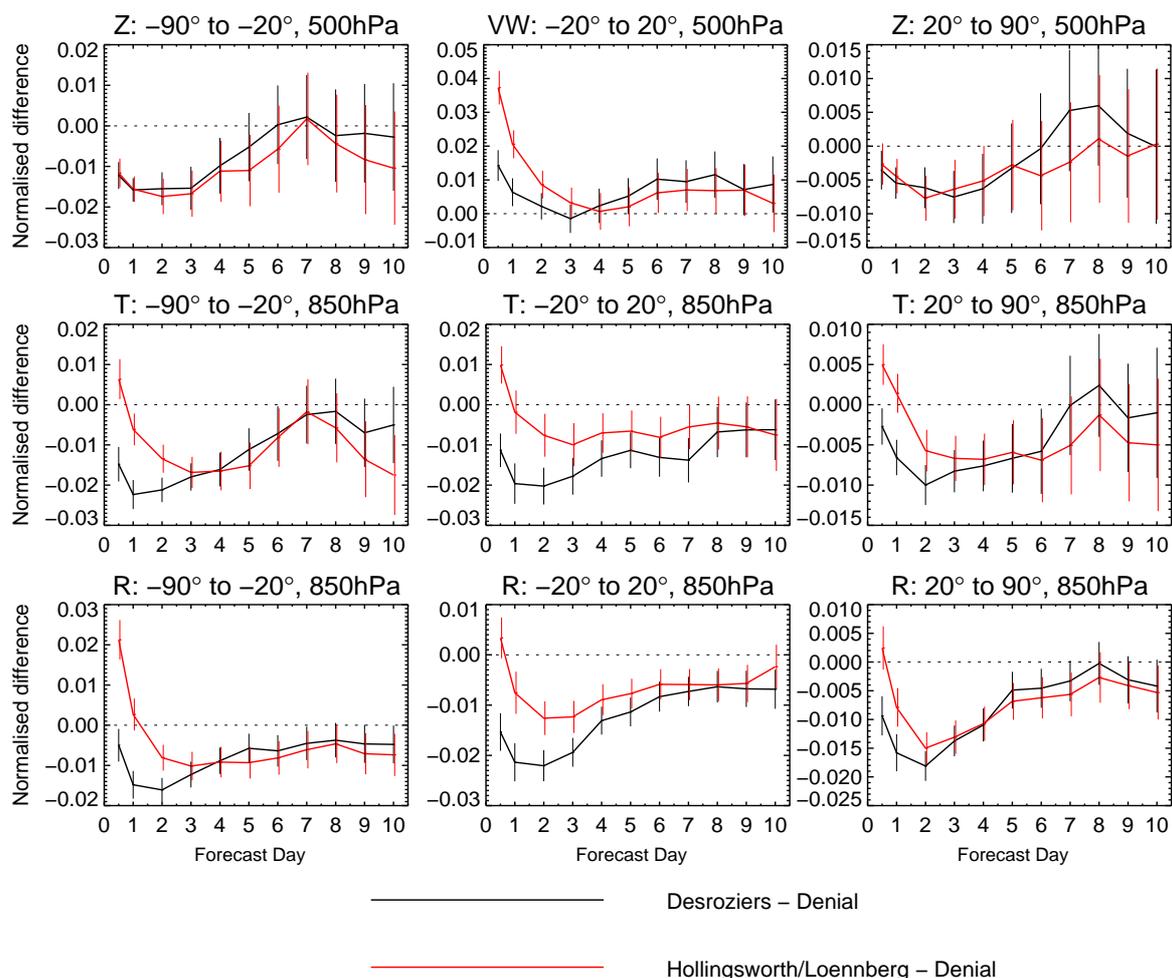


Figure 30: Normalised differences of the standard deviation of forecast errors compared to the Denial experiment as a function of forecast range in days, covering up to 361 forecasts over 6 months. Error bars indicate significance intervals for the differences to the Denial at the 95 % level. Red shows the results for the Hollingsworth/Lönnberg experiment versus the Denial, whereas black shows the results for the Desroziers experiment versus the Denial (see main text for further details). Verification is against the own analysis. A negative value means an improvement compared to the Denial. The various panels show, from left to right: Results for the 500 hPa geopotential over the Southern Hemisphere extra-tropics, the 500 hPa wind in the tropics, and the 500 hPa geopotential over the Northern Hemisphere extra-tropics (top row). The 850 hPa temperature over the Southern Hemisphere extra-tropics, the tropics, and the Northern Hemisphere extra-tropics (middle row). The 850 hPa relative humidity over the Southern Hemisphere extra-tropics, the tropics, and the Northern Hemisphere extra-tropics (bottom row).

values are obtained just after decontamination of the instrument and the highest values just before. The noise estimate used here is not for the same period, but for a different period where the noise was at a similar level. This mis-match further contributes to the differences seen. Of course, short-comings in the observation error diagnostics may also contribute as these may provide incorrect estimates if some of the assumptions inherent in these diagnostics are not valid. For instance, the weights assigned to IASI in the assimilation system may well not be consistent with true weights, as assumed for the Desroziers diagnostic. Or the cloud detection may introduce error correlations between background errors and cloud-screening errors, which are neglected in either observation space diagnostic. Neglecting

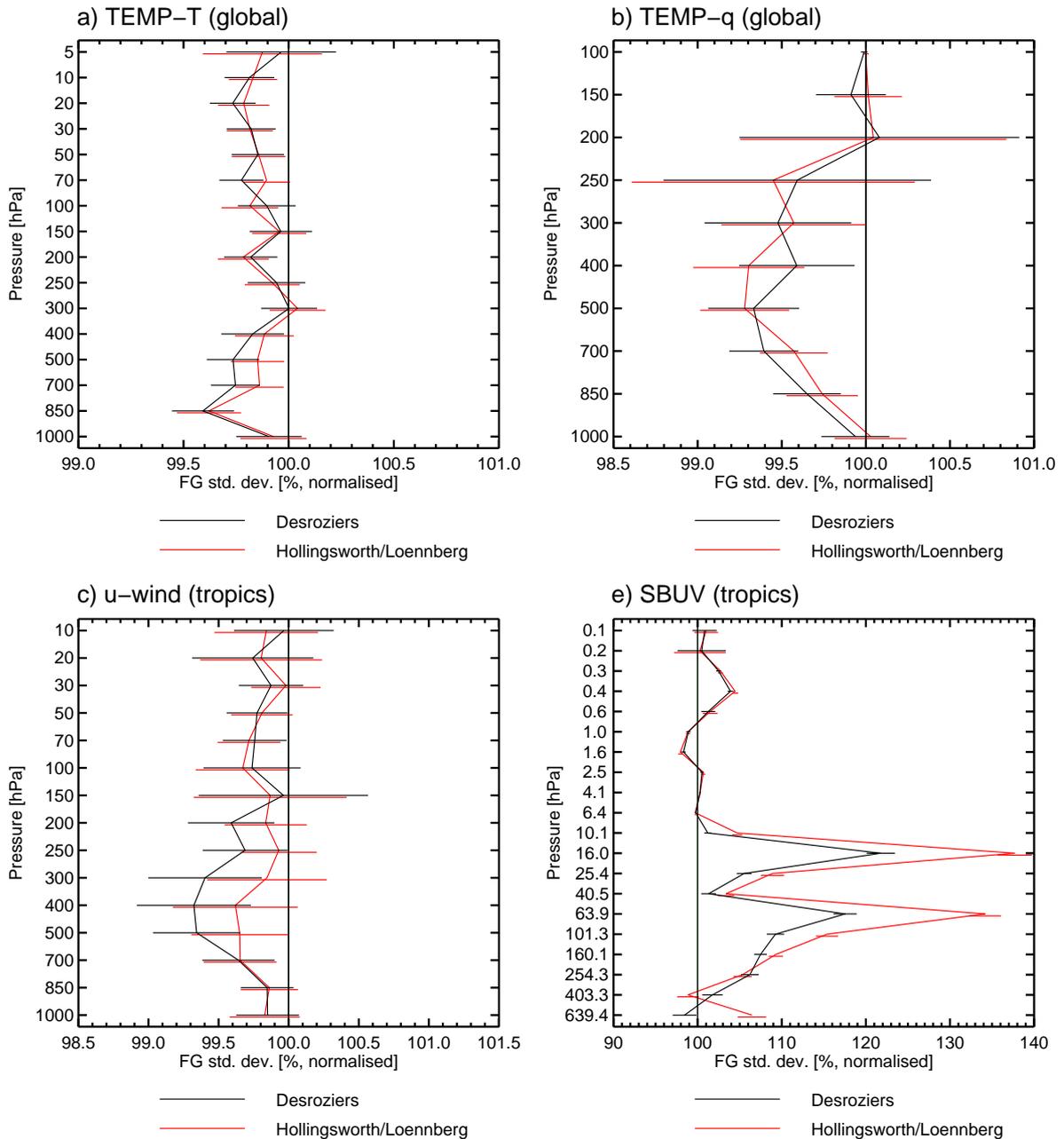


Figure 31: Standard deviations of FG-departures, normalised by the Denial experiment for the 6 month experimentation period covering 361 assimilation cycles. Error bars indicate significance intervals at the 95 % level. Red shows the results for the Hollingsworth/Lönnberg experiment, whereas black shows the results for the Desroziers experiment (see main text for further details). The four panels show results for different observing systems: a) Temperature observations from radiosondes over the globe. b) As a), but for humidity. c) Observations of the u-component from aircrafts and sondes over the tropics. d) SBUV retrievals of ozone over the tropics.

these is likely to result in an under-estimation of the observation error, as these are most likely positive error correlations. In any case, these caveats suggest that the instrument noise as well as the diagnostic estimates should be treated with some caution.

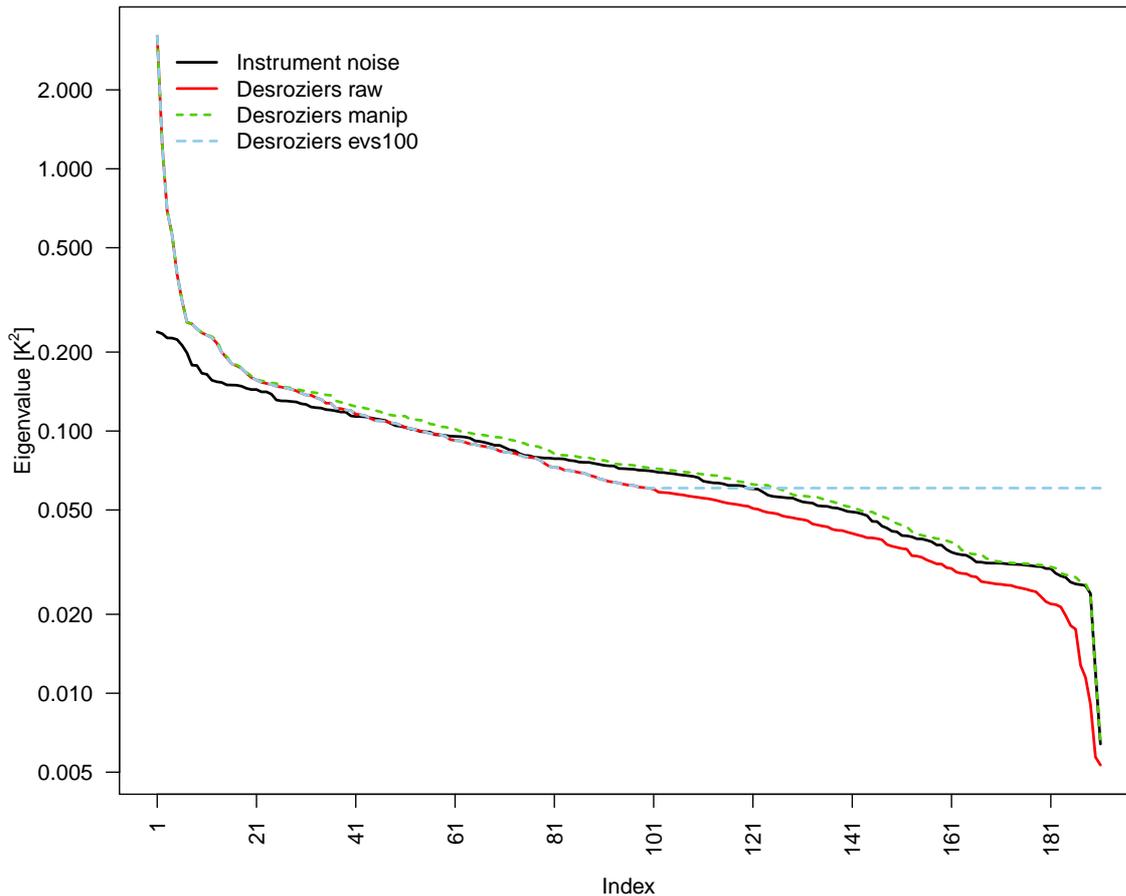


Figure 32: Eigen-spectra of the instrument noise matrix (black) and the Desroziers-diagnosed observation error covariance matrix. Also shown are the eigen-spectra of the Desroziers-diagnosed observation error covariance matrix after re-setting the smallest 91 eigen values to the value of the 100th eigenvalue (dashed blue) and after making the resulting matrix larger than the instrument noise matrix (dashed green).

The above considerations prompted us to investigate two methods to modify the diagnosed matrix, and the results of this are also included in Fig. 32. In the first one, referred to as “instrument noise method”, we adjust the diagnosed error covariance matrix such that the resulting matrix has errors that are larger than the instrument noise estimate for all eigenvectors. This is done by first subtracting the instrument noise matrix from the diagnosed matrix, then setting all negative eigenvalues of the resulting matrix to zero, and then adding the instrument noise matrix again. This approach assumes that the instrument noise matrix estimate is reliable, and it uses the diagnostics primarily to estimate the additional error contributions from the cloud screening, forward model, representativeness, etc. In the second approach, termed “eigen-value method” we discard the smallest 91 eigenvalues of the diagnosed matrix and re-set them to the value of the 100th eigenvalue. This means that we trust only the leading eigen-values of the diagnosed matrix, and do not rely on either the instrument noise estimate or the smallest eigen-values of the diagnosed matrix. The choice of keeping the first 100 eigenvalues is to some extent arbitrary and

could be investigated further. The value is a little larger than what could be suggested by the cross-over point of the eigenspectra of the diagnosed matrix and the instrument noise values.

The resulting adjustments to the diagnosed matrices are displayed in Figures 33 and 34 in terms of the  $\sigma_O$  and the error correlation matrix. As we primarily adjust the smallest eigen-values, the adjustments are relatively small. Both methods lead to an increase in the estimated observation error, mostly for the lower temperature sounding and window channels, with a stronger increase for the eigen-value method. The eigen-value method also leads to stronger adjustments to the error correlations (cf Figures 29b and 34a, b). This includes a dampening of the correlations for neighbouring channels, which is combined with an increase in the  $\sigma_O$  for these channels. In contrast, these features are preserved when the instrument noise method is used (Fig. 34a).

When applied in assimilation experiments, the differences in forecast performance between an experiment that uses the raw diagnosed matrix and experiments that use the adjusted matrices is small compared to the effect of the scaling factor discussed in section 4 (see black and red lines in Figures 35 and 36). While some statistically significant changes can be detected, either in terms of differences in forecast

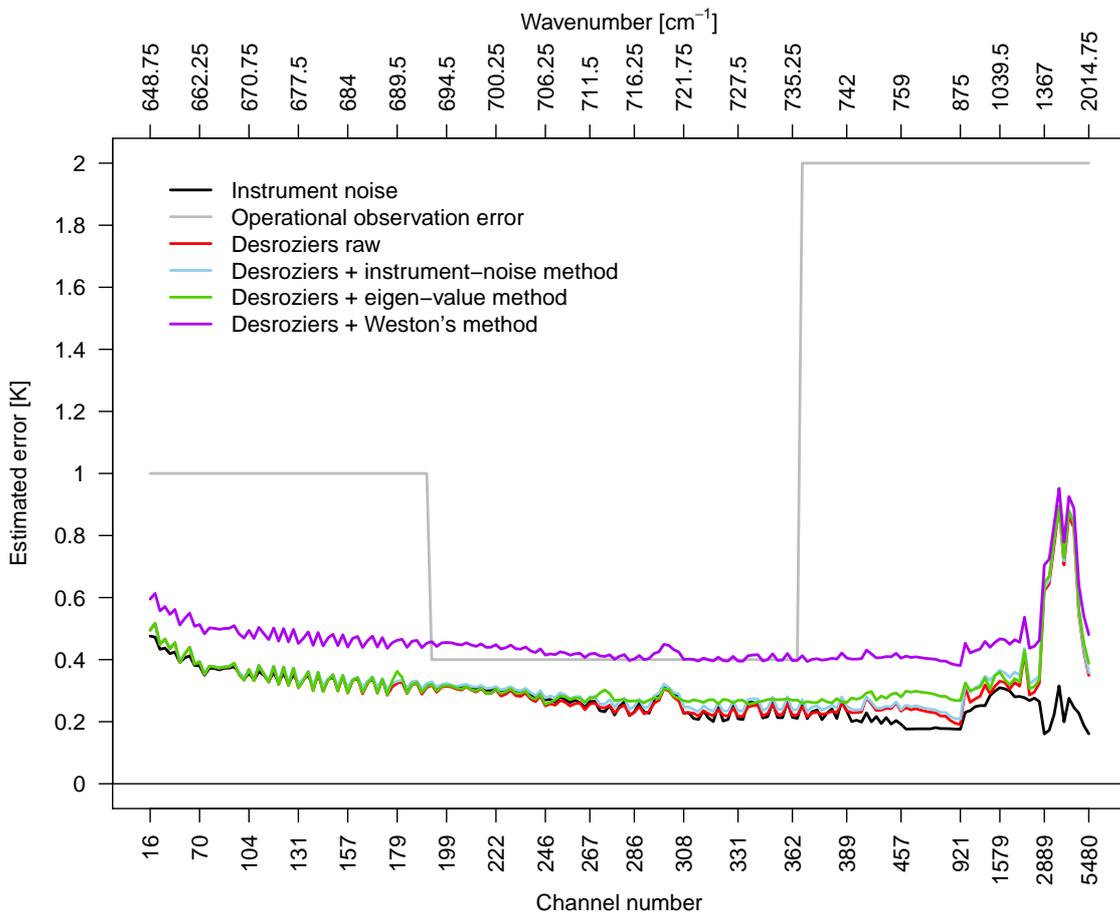


Figure 33: Raw diagnosed observation error (red) for the assimilated IASI channels, together with the resulting values after adjustment with the instrument-noise method (light blue) and the eigen-value method (green). Also shown is an estimate of the instrument noise (black) and the operationally assumed observation error (grey). In addition, we provide values resulting from the adjustment used by Weston et al. 2014 (dark blue).

scores for some parameters in certain areas, or in terms of FG-fits for other observations, these are generally small and no clear overall benefit is apparent for any one of the considered experiments. The Figures shown here are for experiments in which the 1.75 scaling of the diagnosed  $\sigma_0$  has been applied, but results for un-scaled matrices are similar, with a small benefit for the matrix adjusted with the eigen-value method. The relevance of the adjustments is expected to be specific to the present experimentation, and the adjustments may be more relevant for other instruments or channel selections.

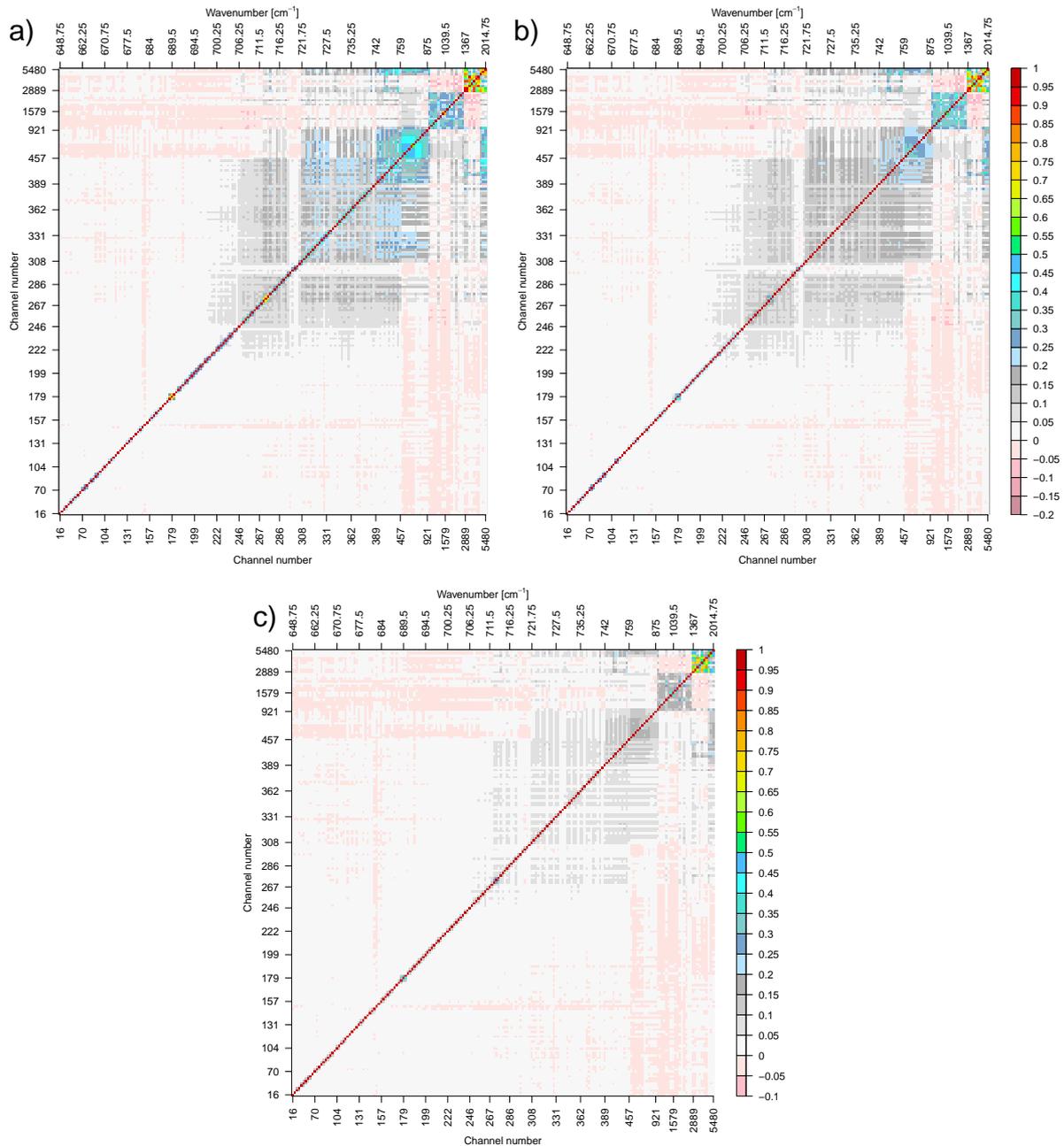


Figure 34: a) As Fig. 29b, but after applying the instrument noise method to adjust the diagnosed observation error covariance matrix. b) As a), but after applying the eigen-value method to adjust the diagnosed observation error covariance matrix. c) As a), but after applying the re-conditioning method proposed by Weston et al. (2014).

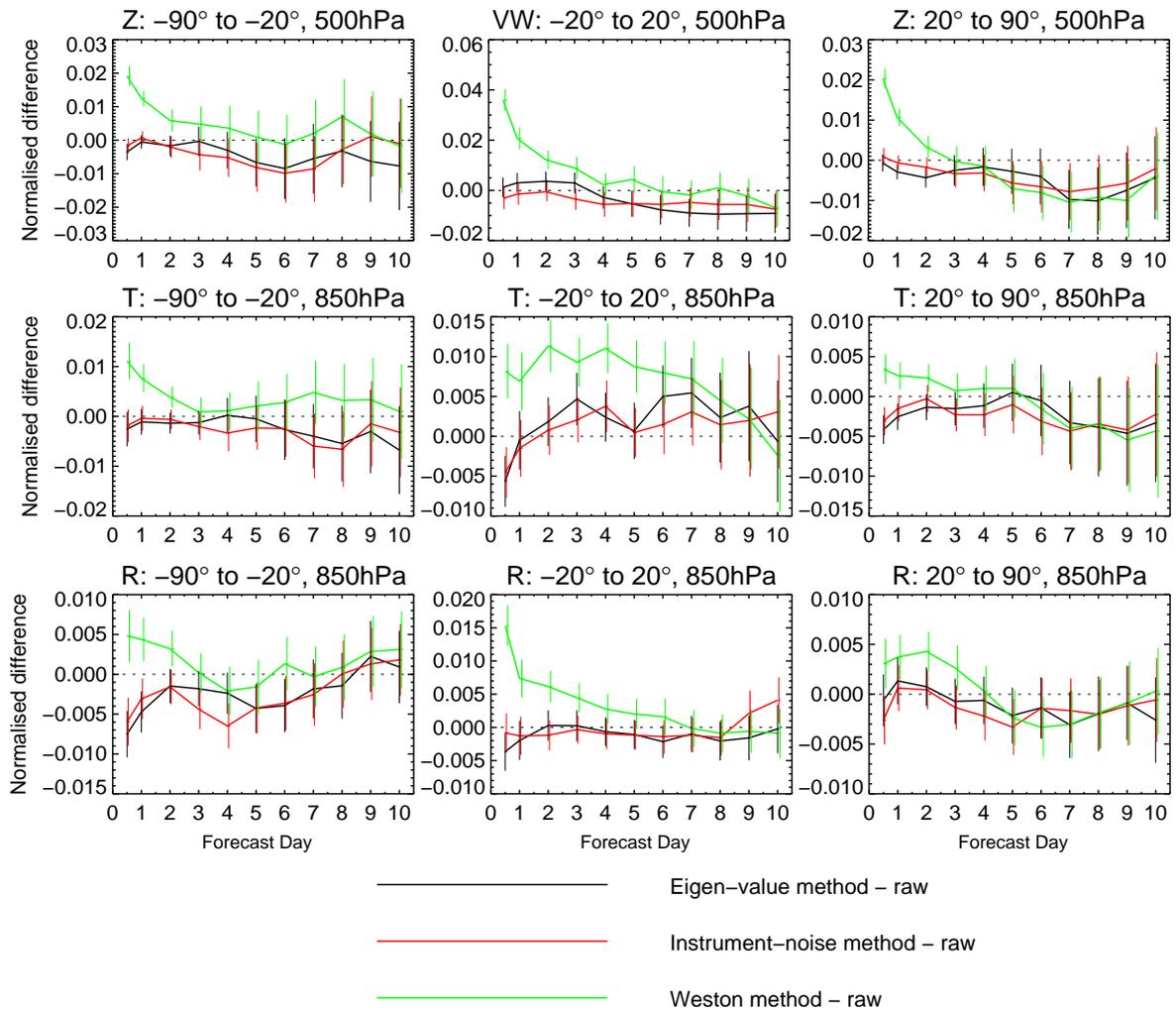


Figure 35: As Fig. 30, but for the normalised differences of the standard deviation of forecast errors relative to an experiment that uses the unadjusted observation error covariance matrix for IASI, re-derived using the Desroziers diagnostic and with  $\sigma_O$  scaled by 1.75. The black lines indicate the performance of an experiment for which the IASI observation error covariance matrix has been adjusted using the eigen-value method, the red line one where instead the instrument-noise method has been used, and green an experiment where the method used in Weston et al. (2014) has been used to modify the original, un-scaled matrix.

The adjustments made to the covariance matrix do, however, affect the number of iterations required to reach convergence in the minimisation. The number of iterations is smallest for the eigen-value method, and largest for the raw matrices (see, for instance, Table 2). The number of iterations required for convergence obviously has a very significant impact on the computational expense, and a smaller number of iterations is preferred. For this reason, all experimentation presented in the main part of this Technical Memorandum has been performed with matrices that have been adjusted using the eigen-value approach.

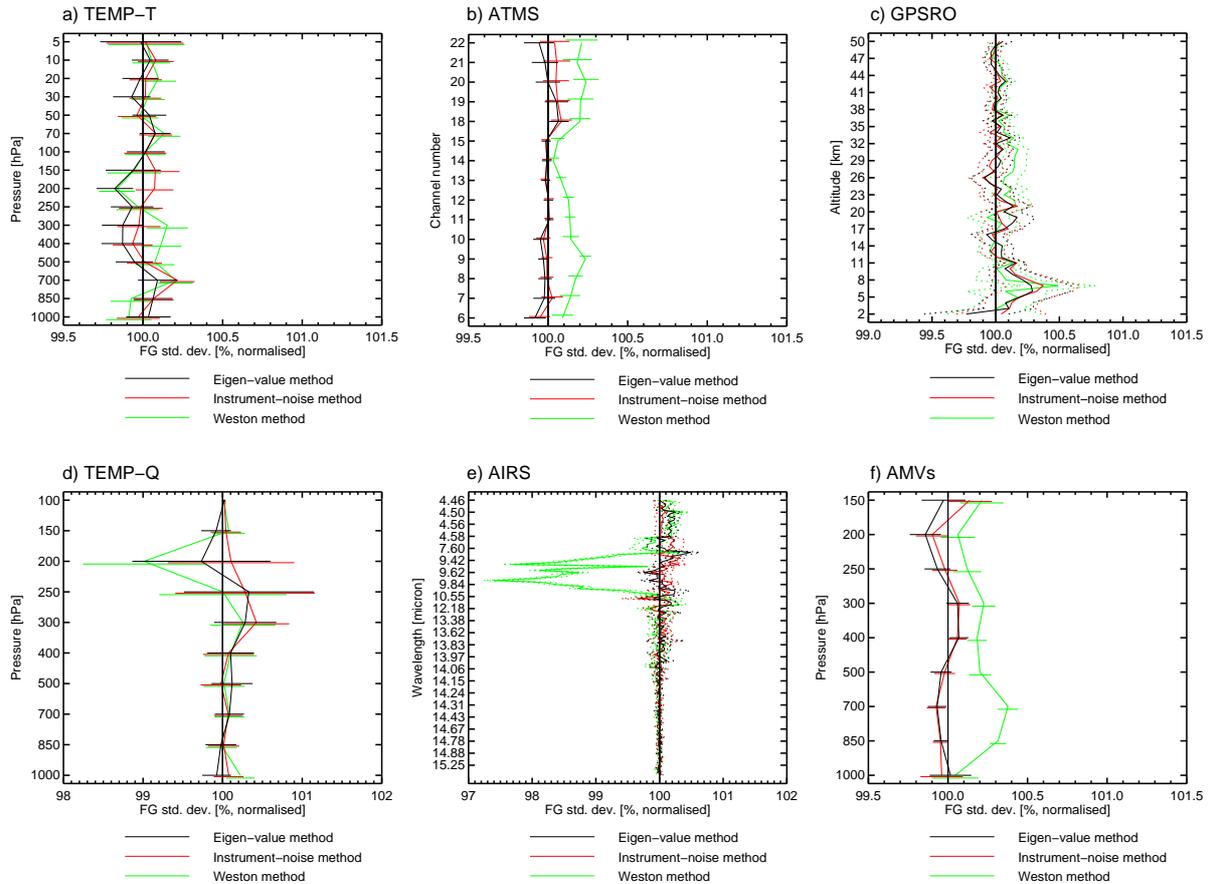


Figure 36: As Fig. 31, but for experiments with different adjustments of the diagnosed matrix, normalised by values from an experiment that uses the unadjusted observation error covariance matrix for IASI with  $\sigma_O$  scaled by 1.75. The black lines indicate the performance of an experiment for which the IASI observation error covariance matrix has been adjusted using the eigen-value method, the red line one where instead the instrument-noise method has been used, and green an experiment where the method used in Weston et al. (2014) has been used to modify the original, un-scaled matrix. The six panels show global statistics for: a) Temperature observations from radiosondes, b) ATMS observations (channels 6-15 are lower-upper temperature-sounding channels; 18-22 are humidity-sounding channels), c) GPS radio occultation bending angles, d) humidity observations from radiosondes, e) AIRS observations, and f) AMVs.

## Adjustments and reconditioning

The adjustment performed by the eigen-value method could be viewed as a reconditioning of the diagnosed matrix to reduce its condition number. The condition number is the ratio between the largest and the smallest eigenvalues of the matrix. In experiments that account for inter-channel error correlations using diagnosed matrices at the Met. Office, manipulations that reduce the condition number were found very important, as the raw matrices led to severe convergence problems (Weston et al. 2014). In their work, this reconditioning is achieved by adding an error equivalent to 0.33 K to the diagonal of the diagnosed matrix for all channels. This gives a condition number of 67 for the matrix used in their study. Condition numbers encountered for the raw and adjusted matrices for our study are given in Table 2.

In our study we have found the magnitude of  $\sigma_O$  to be the more important factor for the performance of the assimilation system, rather than the condition number, in terms of the forecast impact and the number

Table 2: Condition number of  $\mathbf{R}$  and average number of iterations for the experiments with differently adjusted observation error covariances for IASI, using the eigen-value method, the instrument-noise method, or the method proposed by Weston et al. (2014). Statistics are based on the three month period 5 February - 4 May 2014.

	Control	Un-scaled			$\sigma_O$ scaled by 1.75			Weston method
		Raw	Eigen-value method	Noise method	Raw	Eigen-value method	Noise method	
Condition number of $\mathbf{R}$	5	604	54	493	604	54	493	29
Number of iterations								
1st minimisation	26.3	30.5	29.4	30.0	26.0	26.2	26.0	26.6
2nd minimisation	27.4	34.2	32.3	33.6	28.1	27.3	27.7	29.4
3rd minimisation	27.5	36.3	34.1	35.3	28.4	27.6	28.1	30.6

of iterations needed. While the condition number does have a small effect, it is considerably smaller than that of introducing a scaling factor (e.g., Table 2 or Fig. 8). It can therefore be speculated whether the change in the size of the assumed  $\sigma_O$  was actually the more important factor in solving the convergence problems encountered by Weston et al. (2014), rather than the size of the conditioning number of  $\mathbf{R}$ . The results may also reflect different pre-conditioning approaches used in the two assimilation systems. Interestingly, in the ECMWF system the adjustment proposed by Weston et al. (2014) leads to a somewhat poorer convergence (see Table 2) and a poorer performance than using the scaled raw diagnosed matrices (see the green line in Figures 35 and 36), with the exception of stronger benefits shown by ozone-sensitive infrared channels.

## Appendix B: Sensitivity of the Desroziers diagnostic to the background error specification

A key assumption of the Desroziers et al. (2005) diagnostic is that the weights applied in the assimilation system are consistent with the true weights. This assumption inherently introduces some dependence of the Desroziers estimates on the background error specification. It is therefore of interest to characterise the sensitivity of the Desroziers estimates used in this study to the background error specification.

In the following, we will investigate the background error sensitivity by diagnosing the observation error covariances from a set of assimilation experiments that use different specifications of the background errors. Apart from the experiment used to diagnose the Desroziers-based matrix presented in Appendix A, we also conducted two further experiments in which a degraded background error specification is used. In the first experiment, we use only a static background error, and exclude the situation-dependent aspects used in the present operational assimilation system. This mimics the background error specification used in the ECMWF system prior to the move to a hybrid EDA/4DVAR system. In the second experiment, we multiply the background error standard deviations by a factor of 3. This simulates a situation where the background error is grossly over-estimated, by an amount that is purposely unlikely to be realistic for current operational NWP systems. Both experiments are otherwise identical to the experiment used to derive the raw Desroziers-based estimate in Appendix A. In particular, they use an observation error covariance matrix for IASI that is based on a Hollingsworth/Lönnerberg estimate scaled by 1.75. The diagnostics are again applied to the 1-month period 15 March to 14 April 2014.

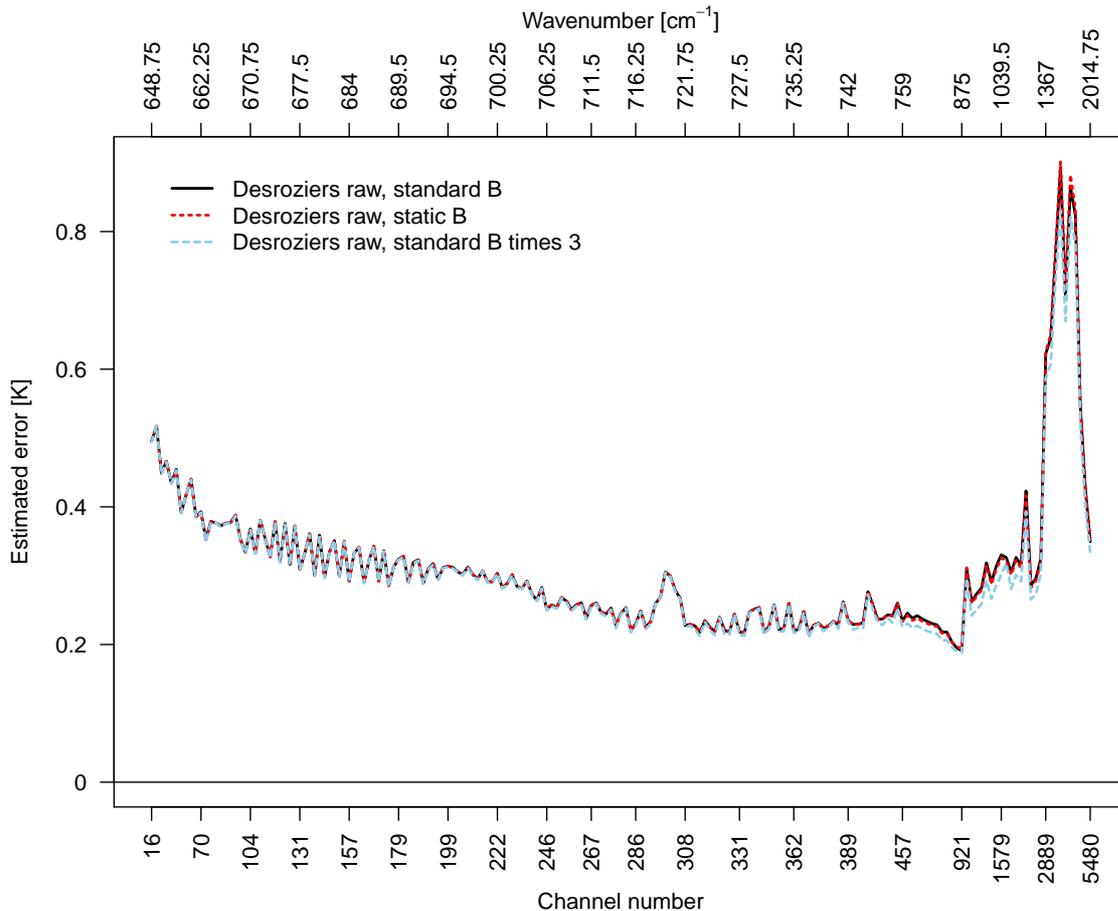


Figure 37: Observation error ( $\sigma_0$ ) for the assimilated IASI channels diagnosed with the Desroziers method from experiments that use a different background error specification: the black line shows the estimate obtained from the experiment that uses the standard background error (the same as the Desroziers estimate shown in Fig. 28), the red dotted line shows the results from an experiment in which instead a static background error has been used, whereas the blue dashed line shows the estimate from an experiment in which the standard background error standard deviation has been multiplied by a factor 3.

The differences between the diagnosed observation error covariance matrices from the three experiments considered here are rather small. The diagnosed error standard deviations are generally within 10 % (Fig. 37) when comparing the results from the experiments with the grossly inflated background error and the original experiment, and they are even smaller when we compare the results from the experiment with the static background error is compared to the original raw matrix. Similarly, the differences in the diagnosed correlations are typically within 0.1 (cf Figures 38 and 29b), and they are hence small compared to the adjustments discussed in Appendix A.

The results confirm that there is some sensitivity of the Desroziers diagnostics to the background error specification, but they also suggest that the influence is rather minor for relatively well-tuned assimilation systems. While we have not conducted assimilation experiments with the matrices derived from the experiments with the degraded background errors, the finding that the differences are minor compared to the adjustments investigated in Appendix A suggests that the impact of the differences is likely to be small. Currently, it appears that the background error dependence of the observation error diagnostics is

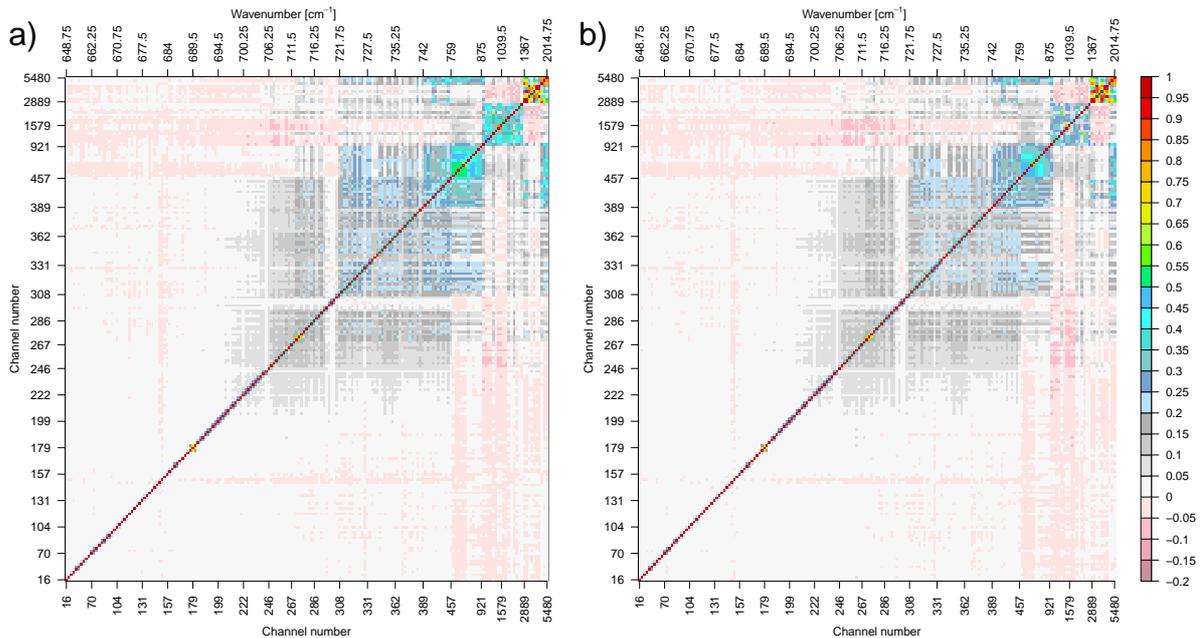


Figure 38: a) Raw diagnosed inter-channel error correlations obtained with the Desroziers diagnostic applied to an experiment that uses a static background error specification. b) Raw diagnosed inter-channel error correlations obtained with the Desroziers diagnostic applied to an experiment for which the background error estimate has been multiplied by a factor 3.

not a primary concern when the diagnostics are used to specify observation errors for IASI.

## References

- Auligné, T., and A. McNally, 2007: Interaction between bias correction and quality control. *Q. J. R. Meteorol. Soc.*, **133**, 643–653.
- Aumann, H., M. Chahine, C. Gautier, M. Goldberg, E. Kalnay, L. McMillin, H. Revercomb, P. Rosenkranz, W. Smith, D. Staelin, L. Strow, and J. Susskind, 2003: AIRS/AMSU/HSB in the Aqua mission: Design, science objectives, data products, and processing system. *IEEE Trans. Geosci. Remote Sens.*, **41**, 253–264.
- 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.
- Bormann, N., and A. Collard, 2012: Experimentation with inter-channel error correlations with AIRS and IASI at ECMWF. In Proceedings of the 18th international TOVS study conference, Toulouse, France.
- Bormann, N., A. Collard, and P. Bauer, 2010: Estimates of spatial and inter-channel observation error characteristics for current sounder radiances for NWP, part II: Application to AIRS and IASI. *Q. J. R. Meteorol. Soc.*, **136**, 1051–1063.

- Bouttier, F., and G. Kelly, 2001: Observing-system experiments in the ECMWF 4-DVAR assimilation system. *Q. J. R. Meteorol. Soc.*, **127**, 1469–1488.
- Chalon, G., F. Cayla, and D. Diebel, 2001: IASI: An advanced sounder for operational meteorology. In Proc. 52nd Congress of IAF, Toulouse France, 1-5 Oct. 2001, [http://smc.cnes.fr/IASI/A\\_publications.htm](http://smc.cnes.fr/IASI/A_publications.htm) [Accessed August 2009].
- Collard, A., and A. McNally, 2009: The assimilation of Infrared Atmospheric Sounding Interferometer radiances at ECMWF. *Q. J. R. Meteorol. Soc.*, **135**, 1044–1058.
- Collard, A., A. McNally, F. Hilton, S. Healy, and N. Atkinson, 2010: The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **136**, 2038–2050. doi: 10.1002/qj.701.
- Collard, A. D., and S. B. Healy, 2003: The combined impact of future space-based atmospheric sounding instruments on numerical weather-prediction analysis fields: A simulation study. *Q. J. R. Meteorol. Soc.*, **129**, 2741–2760.
- Daescu, D., and R. Todling, 2010: Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Q. J. R. Meteorol. Soc.*, **136**, 2000–2012.
- Daley, R., 1993: *Atmospheric data analysis*. Cambridge University Press, Cambridge, UK, 460 pp.
- 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.
- Dragani, R., and A. McNally, 2013: Operational assimilation of ozone-sensitive infrared radiances at ECMWF. *Q. J. R. Meteorol. Soc.*, **139**, 2068–2080, doi: 10.1002/qj.2106.
- Eresmaa, R., 2014: Imager-assisted cloud detection for assimilation of infrared atmospheric sounding interferometer radiances. *Q. J. R. Meteorol. Soc.*, **140**, 2342–2352. doi: 10.1002/qj.2304.
- Forsythe, M., and R. Saunders, 2008: AMV errors: A new approach in NWP. In Proceedings of the 9th International Winds Workshop, Annapolis, USA, EUMETSAT, [available from: [http://www.eumetsat.int/website/home/News/ConferencesandEvents/PreviousEvents/DAT\\_2042907.html](http://www.eumetsat.int/website/home/News/ConferencesandEvents/PreviousEvents/DAT_2042907.html)].
- Froidevaux, L., N. Livesey, W. Read, Y. Jiang, C. Jimenez, M. Filipiak, M. Schwartz, M. Santee, H. Pumphrey, J. Jiang, D. Wu, G. Manney, B. Drouin, J. Waters, E. Fetzer, P. Bernath, C. Boone, K. Walker, K. Jucks, G. Toon, J. Margitan, B. Sen, C. Webster, L. Christensen, J. Elkins, E. Atlas, R. Lueb, and R. Hendershot, 2006: Early validation analyses of atmospheric profiles from EOS MLS on the Aura satellite. *IEEE Trans. Geosci. Remote Sens.*, **44**, 1106–1121.
- Garand, L., S. Heilliette, and M. Buehner, 2007: Interchannel error correlation associated with AIRS radiance observations: Inference and impact in data assimilation. *J. Appl. Meteor.*, **46**, 714–725.
- Geer, A., F. Baordo, N. Bormann, and S. English, 2014: All-sky assimilation of microwave humidity sounders. Technical Memorandum 741, ECMWF, Reading, U.K., 57 pp.
- Geer, A., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Q. J. R. Meteorol. Soc.*, **137**, 2024–2037. doi: 10.1002/qj.830.
- Geer, A., P. Bauer, and P. Lopez, 2010: Direct 4D-Var assimilation of all-sky radiances. Part II: Assessment. *Q. J. R. Meteorol. Soc.*, **136**, 1886–1905.
- Guidard, V., N. Fourrié, P. Brousseau, and F. Rabier, 2011: Impact of IASI assimilation at global and convective scales and challenges for the assimilation of cloudy scenes. *Q. J. R. Meteorol. Soc.*, **137**, 1975–1987, doi:10.1002/qj.928.
- Han, W., and A. McNally, 2010: The 4D-Var assimilation of ozone-sensitive infrared radiances measured by IASI. *Q. J. R. Meteorol. Soc.*, **136**, 2025–2037, doi: 10.1002/qj.708.

- Han, Y., H. Revercomb, M. Crompton, D. Gu, D. Johnson, D. Mooney, D. Scott, L. Strow, G. Bingham, L. Borg, Y. Chen, D. DeSlover, M. Esplin, D. Hagan, X. Jin, R. Knuteson, H. Motteler, J. Predina, L. Suwinski, J. Taylor, D. Tobin, D. Tremblay, C. Wang, L. Wang, L. Wang, and V. Zavyalov, 2013: Suomi NPP CrIS measurements, sensor data record algorithm, calibration and validation activities, and record data quality. *J. Geophys. Res. Atmos.*, **118**, 12,734–12,748, doi:10.1002/2013JD020344.
- Hilton, F., N. Atkinson, S. English, and J. Eyre, 2009: Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments. *Q. J. R. Meteorol. Soc.*, **135**, 495–505, doi:10.1002/qj.379.
- Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136.
- 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.
- Lupu, C., C. Cardinali, and A. McNally, 2015: Adjoint-based forecast sensitivity applied to observation error variances tuning. *Q. J. R. Meteorol. Soc.*, **141**, accepted, doi: 10.1002/qj.2599.
- Matricardi, M., and A. McNally, 2013: The direct assimilation of principal component of IASI spectra in the ECMWF 4D-VAR: extension to IASI band 1 and IASI band 2. Report for EUMETSAT Contract No. EUM/CO/07/460000475/PS, ECMWF.
- Matricardi, M., and A. McNally, 2014: The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.*, **140**, 573–582, doi: 10.1002/qj.2156.
- Matricardi, M., and A. McNally, 2014: The direct assimilation of principal components of IASI spectra in the ECMWF 4D-VAR and the generation of RTTOV and PC\_RTTOV new regression coefficients for AIRS, IASI, and IASI-NG. Report for EUMETSAT Contract No. EUM/CO/07/4600001011/PS, ECMWF.
- McNally, A., 2009: The direct assimilation of cloud-affected satellite infrared radiances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.*, **135**, 1214–1229, doi: 10.1002/qj.426.
- McNally, A., and P. Watts, 2003: A cloud detection algorithm for high-spectral-resolution infrared sounders. *Q. J. R. Meteorol. Soc.*, **129**, 3411–3423.
- Salonen, K., and N. Bormann, 2012: Atmospheric motion vector observations in the ecmwf system: Second year report. EUMETSAT/ECMWF Fellowship Programme Research Report 28, ECMWF, 41 pp, available online: <http://old.ecmwf.int/publications/library/do/references/show?id=90665>.
- Stewart, L., S. Dance, N. Nichols, J. Eyre, and J. Cameron, 2013: Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Q. J. R. Meteorol. Soc.*, **140**, 1236–1244, doi: 10.1002/qj.2211.
- Ventress, L., and A. Dudhia, 2014: Improving the selection of IASI channels for use in numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **140**, 2111–2118, doi: 10.1002/qj.2280.
- Weston, P., W. Bell, and J. Eyre, 2014: Accounting for correlated error in the assimilation of high-resolution sounder data. *Q. J. R. Meteorol. Soc.*, **140**, 2420–2429, doi: 10.1002/qj.2306.