Direct assimilation of Principal Component data for operational Numerical Weather Prediction

Marco Matricardi, Tony McNally and Niels Bormann

ECMWF, Shinfield Park, Reading, UK

Abstract

The ECMWF operational ECMWF 4D-Var has been adapted to allow the direct assimilation of principal component (PC) scores derived from high spectral resolution infrared sounders. The primary aim of this development is towards an efficient use of the entire measured spectrum that could not be achieved by traditional radiance assimilation. We present a system that uses 50 PC scores derived from 305 IASI channels obtained by augmenting the 191 operational IASI channels with additional surface, ozone, and water vapour channels. The new scheme has been extensively tested in a full data assimilation system that uses all operational observations (satellite and conventional). Testing over a three-month summer period suggests that the quality of the analyses produced by the assimilation of 50 IASI PCs is almost identical to that obtained when the operational 191 IASI radiances are assimilated. The verification of forecasts launched from these test analyses further confirms that there is no loss of skill from the assimilation of IASI PCs compared to that of radiances. This result is all the more important in light of the fact that while the use of PC data is currently restricted to fully clear spectra, in the operational radiance assimilation system the use of IASI data extends to channels unaffected by clouds and to fully overcast scenes. In addition, the 50 PC score system based on 305 radiances uses ~20% less computer resources (during the 4D-var minimization) compared to the system that assimilates 191 radiances. This figure represents a significant saving inside the time critical processing path for NWP centres, but could potentially be improved even further by changing the setting of the tunable accuracy of the PC based fast radiative transfer model (PC_RTTOV) used for the assimilation of PC scores. To summarise, the results obtained from the direct assimilation of IASI PC scores are extremely significant and encouraging. They demonstrate the viability of an alternative route to radiance assimilation for the exploitation of data from high spectral resolution infrared sounders in NWP. Progress in this area is very timely - at the time of writing there were four such instruments in space (IASI on METOP-A and B, AIRS on AQUA and CrIS on NPP). Work is now urgently needed to take this system forward to a stage where it can be considered as an option for the safe and efficient operational exploitation of these crucial instruments.

1. Introduction

The operational use of Infrared Atmospheric Sounding Interferometer (IASI) radiances at ECMWF is currently restricted to a selection of temperature sounding channels in the long-wave region of the spectrum and to a small number of ozone and humidity sounding channels. In principle, to exploit the full information content of IASI, the number of channels used in the assimilation could be increased to cover the full spectrum. NWP users are limited to assimilating less than the full IASI spectrum by the prohibitive computational cost, but it is also known that the independent information on the atmosphere contained in an IASI spectrum is significantly less than the total number of channels (Huang *et al*., 1992). There is thus a need to find a more efficient way of communicating the measured information to the analysis system than simply increasing the number of channels.

Principal Component Analysis (PCA) is a classical statistical method for the efficient encapsulation of information from voluminous data (Joliffe, 2002). As such, it has been proposed as a

solution to the above problem. To investigate the feasibility of using PCA for the assimilation of satellite data, the operational ECMWF 4D-Var has been adapted to allow the direct assimilation of Principal Component (PC) scores derived from high spectral resolution infrared sounders. The primary aim of this development is towards an efficient use of the entire measure spectrum that could not be achieved by traditional radiance assimilation.

In the methodology adopted for the direct assimilation of PC scores, the observed IASI spectra are first screened for the presence of clouds and contaminated spectra are discarded. This must be done before assimilation as the PC training has been performed with only completely clear data and none of the eigenvectors correspond to cloud signals. The clear spectra are then projected into a fixed eigenvector basis derived from n channel radiances to produce a vector of observed PC scores. Each vector of observed PC scores has length $l=n$, but crucially we assimilate only the first m of these (where $m < n$ in ranked order). In truncating the vector of observed PC scores we make the assimilation highly efficient, while preferentially retaining highest rank PC scores (1,2,3 ..m) that convey most information about the atmospheric state. In addition to reducing the dimension of the observed information, the value of m can also be tuned to achieve filtering of the observations, using PCA to separate variations of the atmospheric signal from variations of the random instrument noise.

Trajectory estimates of the atmospheric state are used as input to the observation operator PC_RTTOV (Matricardi, 2010) to compute model equivalents of the m PC scores. During the minimization, perturbations of the atmospheric state are mapped into the observation (PC) space by the tangent linear of the observation operator PC_RTTOV_TL. Likewise, gradients of the cost function with respect to the PC score observations are evaluated and mapped into gradients with respect to the atmospheric state by the adjoint of the observation operator PC_RTTOV_AD.

The short wave spectral region covered by IASI band 3 (2000-2760 cm^{-1}) contains excellent temperature sounding channels which could in principle be exploited for assimilation in operational NWP. For instance, the lower-tropospheric peaking channels located in the spectral region between 2380 and 2400 cm^{-1} have the sharpest possible weighting functions of any part of the infrared spectrum. In addition, compared to equivalent long wave channels, short wave temperature sounding channels are less contaminated by water vapour and ozone absorption. Previous PC score assimilation investigation have been carried out based on the use of short wave data and results documented in a series of technical reports (e.g. Matricardi and McNally, 2012). However, IASI short wave channels are not currently used operationally at ECMWF for a number of reasons, which include day-night variations in data usability due to non-local thermodynamic equilibrium (LTE) effects and high instrument noise. Indeed the latter was the incentive of initial PCA investigation. Consequently, we have shifted our focus towards the assimilation of PC scores generated from the long-wave region of the IASI spectrum. This choice is made for two reasons: Firstly we avoid the day / night sampling issues associated with short wave data. Secondly, we are able to compare the PC assimilation system with a parallel radiance system based upon a selection of the operationally used IASI long wave channels.

A prototype long-wave PC score assimilation scheme (Matricardi and McNally, 2013b) has been initially tested in a baseline environment where conventional observations and atmospheric motion vectors (AMV's) are assimilated, but IASI observations are the only satellite sounding data used (either in the form of PCs or radiances). The absence of other satellite data amplifies the influence of IASI and allows changes to the analyses to be more directly attributable. Testing over two separate one month periods (winter and summer) suggests that the quality of the analyses produced by the assimilation of 20 IASI PCs is almost identical to that obtained when equivalent 165 IASI radiances are assimilated. Indeed in some respects - specifically the fit to radiosonde observations - the analyses based on the assimilation of IASI PCs are marginally improved although this may be related to different tuning. The verification of forecasts launched from these test analyses further confirms that there is no loss of skill from the assimilation of IASI PCs compared to that of radiances.

Based on the impressive performance of the PC assimilation prototype, in this paper we have given priority to the testing of the IASI PC approach in a full data assimilation system that contains all operational observations (satellite and conventional). Testing in this extremely demanding new baseline environment will allow us to check that the conclusions reached for the prototype system study are robust. Furthermore, following the investigation into the use of PCs to represent the IASI

short-wave and long-wave spectrum, we have taken the logical step of considering the extraction of information from the dedicated IASI water vapour and ozone bands.

Another subject discussed in the paper is the generation of new regression coefficients for the RTTOV and PC_RTTOV fast radiative transfer models. These coefficients are based on a new version of the underlying LBLRTM line-by-line (LBL) model. Within the same context, we have also revised the profile set used for the training of the RTTOV variable trace gas transmittances, the rationale being that the current training profiles no longer reflect present day concentrations. To allow pre-launch simulation studies of IASI-NG data, we have added this instrument to the list of available RTTOV and PC_RTTOV coefficients.

The outline of the paper is as follows. In Sections 2 and 3 we review the theory of PCA and discuss the channels selected for use in the current and future assimilation trials. In section 5 we present a physical interpretation of the PC scores. Section 6 discusses the assimilation methodology whereas results from the direct assimilation of PC scores using the ECMWF Integrated Forecasting System (IFS) are discussed in section 7 and 8. Finally, in Section 9 we present a summary of the results and discuss possible future directions.

2. A brief review of the theory of Principal Component Analysis

PCA is a method that allows the reduction of the dimensionality of a problem by exploiting the linear relationship between all the variables contained in a multivariate dataset. The reduction of the dimension of the dataset is obtained by replacing the original set of correlated variables with a smaller number of uncorrelated variables called *principal components*. Because the new derived variables retain most of the information contained in the original data set, PCA theory provides a tunable mechanism to efficiently represent the information in the dataset.

Our dataset consists of a sample of *l* spectra of *n* radiances arranged into an *l* by *n* data matrix ***R***. The dataset can then be represented by the vector population $\mathbf{r} = (r_1, r_2, ... r_n)^T$ (here ^{T} denotes the transpose). If ***C*** is the *n* by *n* covariance matrix of the data matrix ***R***, and ***A*** is the *n* by *n* matrix formed by the eigenvectors of the covariance matrix arranged as row vectors in descending order according to the magnitude of their eigenvalues, the PCs, ***p***, of the vector population can be written as:

$$\mathbf{p} = \mathbf{A}\,\mathbf{r} \tag{1}$$

The eigenvectors represent the directions of maximum variance in the data; consequently, each PC gives the linear combination of the variables that provides the maximum variation. The PCs are orthogonal, hence uncorrelated (although this does not imply that they are statistically independent), and the values associated to each spectrum are known as PC scores. If λ_i is the eigenvalue associated with the i^{th} eigenvector, then the value of $\lambda_i / \sum_{i=1}^{n} \lambda_i^2$ gives the proportion of variation explained by the i^{th} PC. Because the matrix ***A*** is orthogonal, its inverse is equal to its transpose and we can write:

$$\mathbf{r} = \mathbf{A}^T\mathbf{p} \tag{2}$$

Equations (1) and (2) can be written in discrete notation form as:

$$p_{i,j} = \sum_{k=1}^{n} A_{i,k} r_{k,j} \tag{3}$$

$$r_{i,j} = \sum_{k=1}^{n} A_{k,i} p_{k,j} \tag{4}$$

where i=1,n represents the i^{th} value and j=1,l is the j^{th} spectrum. A number of PCs, m, fewer than n can often represent most of the variation in the data. We can then reduce the dimension of the problem by replacing the n original variables with the first m PCs. In many applications, the choice of the number of dimensions is based on the total variation accounted for by the leading PCs and it will in general depend on specific aspects of the original dataset.

For any new observed radiance spectrum, $\mathbf{r}^{obs}$, we can compute the equivalent PC scores by projecting the radiances upon the full set of eigenvectors derived from the covariance matrix of the training dataset. As discussed above, less than n eigenvectors are typically required to reproduce most of the information in the observed spectra. Therefore, we can compute a vector of m truncated observed PC scores, $\mathbf{p}^{obs}$:

$$p_i^{obs} = \sum_{k=1}^{n} A_{i,k} r_k^{obs} \tag{5}$$

where i=1,m. The truncated PC scores may be regarded as an efficient encapsulation of the original observation that may be used for storage, transmission or indeed assimilation. If required, the PC scores may be used to reconstruct a new radiance vector

$$r_i^{rec} = \sum_{k=1}^{m} A_{i,k} p_k^{obs} \tag{6}$$

Even though a radiance vector containing all n channels may be reconstructed from the m truncated PC scores, it should be stressed that the n reconstructed radiances only contain m independent pieces of information (the n reconstructed radiances are formally equivalent to the m PC scores used in the reconstruction) and crucially $r_i^{rec} \neq r_i^{obs}$ (i.e. PCA is not a lossless technique when truncating). The reconstructed radiances will also have different error characteristics.

In addition to reducing the dimension of the observed information, the value of m can also be tuned to achieve filtering of the observations, using PCA to separate variations of the atmospheric *signal* from variations of the random instrument *noise*. It is argued that the atmospheric signal is more highly correlated across the spectrum and as such is represented by the high rank eigenvectors (i.e. those with larger eigenvalues). Conversely, the random instrument noise is spectrally uncorrelated and is thus represented by low rank eigenvectors. In principle we may attempt to exploit this separation (in ranked eigenvector space) to retain only eigenvectors related to atmospheric signal and discard those eigenvectors describing instrument noise. Of course great care must be taken if truncating the PC scores for this specific purpose. Small scale and small amplitude atmospheric features can be important sources of rapid forecast error growth in NWP. However, such features may not be strongly correlated across the measured spectrum and could potentially be confused with noise (and removed if the truncation is too severe).

3. Spectral bands and PC generation

For the purposes of the demonstration of PC score assimilation it is assumed that we have access to the full IASI measured spectrum and that we are only investigating the suitability of PCA as a mechanism to efficiently present this information to an assimilation system. As such we are deliberately separating this from the potential application of PCA to the logistical issue of compressed data dissemination.

As discussed in the introduction, based on the impressive performance of the PC assimilation prototype, in this paper we intend to test the IASI PC approach in a full data assimilation system that contains all operational observations (satellite and conventional). Testing in this extremely demanding new baseline environment will allow us to check that the conclusions reached for the prototype

system study are robust. Furthermore, following the investigation into the use of PCs to represent the IASI short-wave and long-wave spectrum, we have taken the logical step of considering the extraction of information from the dedicated IASI water vapour and ozone bands. It should be noted that to be able to compare the PC assimilation system with a parallel radiance system based upon a selection of the operationally used IASI long wave channels we had to update the PC system to the latest version of the ECMWF libraries (i.e. cycle 38R2).

3.1 The subsets of IASI channels used in the PC assimilation

The assimilation of PC data requires the conversion of IASI radiances into PC scores. The two PC systems studied in this paper are based on 165 and 305 IASI channels respectively. The 165 IASI channels cover the region between 645 and 875 cm^{-1}. While the primary sensitivity is to temperature (and surface), PCs calculated from this channel set will also convey some humidity information. This is due to the presence of absorption from weak water vapour lines originating from the edge of the water vapour pure rotational band and to the water vapour continuum absorption in the more optically transparent window region (i.e. past 750 cm^{-1}). The 165 channel set is a subset of the 191 channel set used operationally at ECMWF and has been obtained by removing from the latter the short wave sounding channels and channels with a strong sensitivity to water vapour and ozone. This is a choice we have inherited from the prototype PC assimilation system where it was made to run assimilation experiments in more controlled conditions. For consistency reasons, we have considered appropriate to retain the above choice. The 165 channels (hereafter referred to as the B1 channel set) are represented by the red dots in figure 1 (the black curve is the full IASI spectrum).

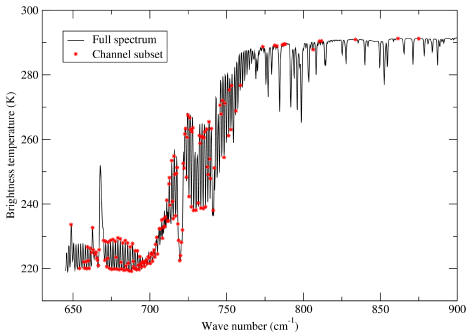

Figure 1. The 165 channels selected for use in the long wave assimilation trials.

Regarding the 305 channels, they have been obtained by supplementing the above 165 channels with 140 additional channels located in window region, in the 1042.08 cm^{-1} $\nu 3$ ozone band and in the P-branch of the 1594.75 cm^{-1} $\nu 2$ water vapour band. These 305 channels (hereafter referred to as the B1_B2 channel set) are identical to those selected by Collard and McNally (2009) and are represented by the red dots in figure 2. Compared to the equivalent full resolution IASI spectrum (i.e. the black curve in figure 2), the use of these channels should minimize the total loss of information for a NWP state vector consisting of temperature, humidity, ozone and carbon dioxide (Collard, 2007).

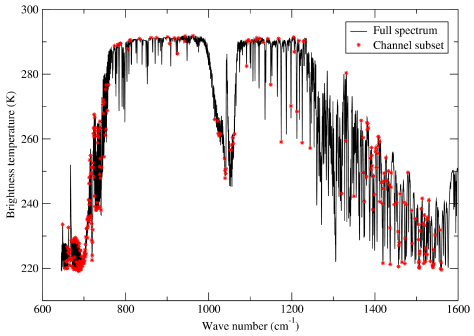

Figure 2. The 305 channels selected for use in the PC assimilation trials.

The conversion of IASI observed radiances into PC scores is carried out using Eq. 5, i.e. by projecting the radiance vector on the fixed basis of synthetic eigenvectors utilized in the PC_RTTOV fast radiative transfer model. Note that each channel set is associated to a different PC_RTTOV eigenvector basis. The simulation of PC scores derived from the B1 and B1_B2 radiances required the generation of dedicated PC_RTTOV regression coefficients. This task has been carried out during the course of the previous work package and the interested reader can refer to Matricardi and McNally (2013a) for further details. The same reader can also refer to Matricardi and McNally (2013a) for details regarding the physical interpretation of the PC scores derived from the above channel sets.

4. PC assimilation methodology

4.1 Overall architecture of the assimilation system

The methodology adopted for the direct 4D-Var assimilation of PC scores is shown schematically in figure 3. The observed IASI spectra are first screened for the presence of clouds and contaminated spectra are discarded. This must be done before assimilation as the PC training has been performed with only completely clear data and none of the eigenvectors correspond to cloud signals. The clear spectra are then projected on to the *n* channel basis described previously, to produce a vector of observed PC scores Y_{OBS}^{PC}. Each vector of observed PC scores has length *n*, but crucially we assimilate only the first *m* of these (where $m < n$ in ranked order). In truncating the vector of observed PC scores we make the assimilation highly efficient, while preferentially retaining highest rank PC scores (1,2,3 ..*m*) that convey most information about the atmospheric state.

The *m* observed PC scores are then provided as input to the 4D-Var. Trajectory estimates of the atmospheric state (**X)** are used as input to the observation operator PC_RTTOV (Matricardi, 2010) to compute model equivalents of the *m* PC scores, $Y_B^{PC}(X)$. If we ignore the time integration of the forecast model to the observations, the cost function to be minimized is essentially:

$$J(X) = [X - X_B]^T B^{-1} [X - X_B] + [Y_{OBS}^{PC} - Y_B^{PC}(X)]^T O^{-1} [Y_{OBS}^{PC} - Y_B^{PC}(X)] \quad (7)$$

where the accuracy of the background estimate of the atmospheric state $\mathbf{X_B}$ is described by the error covariance **B** and the accuracy of the observations and associated observation operator is described by the error covariance **O**. The specification of **O** is very important and is described separately in section 6.3. During the minimization, perturbations of the atmospheric state are mapped into the observation (PC) space by the tangent linear of the observation operator PC_RTTOV_TL. Likewise, gradients of the cost function with respect to the PC score observations are evaluated and mapped into gradients with respect to the atmospheric state by the adjoint of the observation operator PC_RTTOV_AD. The atmospheric state $\mathbf{X_A}$ that minimizes the above cost function is referred to as the *analysis* and the departures of this from the background atmospheric state $\mathbf{X_B}$ are referred to as analysis increments defined at the start of the 4D-Var window.

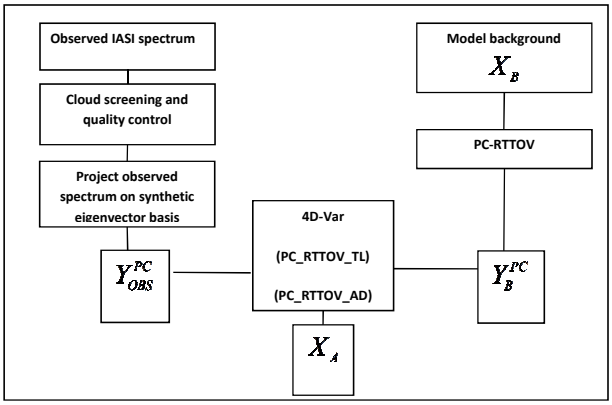

Figure 3. The flow diagram of the direct PC score assimilation.

4.2 The specification of the error covariance matrix R

A considerable amount of attention has been focussed on the specification of the PC score observation error covariance matrix **O**. The matrix **O** should describe the combined error of the observations (PC scores) and forward operator (PC_RTTOV). Departure statistics have been accumulated over long periods to obtain an initial estimate of the elements of **O**, computing the standard deviation of the observed minus background (O-B) departures. Of course these values are not optimal in that they contain a contribution from the uncertainties in the background state and as such can only be regarded as an upper bound upon the required error.

To separate the contribution of the observation error and the background error in the departure statistics, Hollingsworth and Lönnberg (1986) and Desroziers *et al.* (2005) have proposed different techniques. In the Hollingsworth/Lönnberg method pairs of background departures are used to compute statistics as a function of the separation. To estimate the observation error, the values of the covariances are extrapolated to zero separation. It is then assumed that the spatially uncorrelated component of the background departures is largely dominated by the observation error.

In the Desroziers method, the elements of the error matrix **O** are expressed as the expectation value

$$\mathbf{O} = E[\mathbf{d}_a \mathbf{d}_b^T] \quad (8)$$

where $\mathbf{d}_a$ and $\mathbf{d}_b$ are the analysis and background departures in the observation space. This relationship can be derived from the quasi-linear estimation theory used as the basis for variational assimilation schemes like 4D-Var. Assuming initial estimates of the weights are reasonable, the Desroziers algorithms produces a refined estimate of the observation error. A detailed description of the experimental set-up used to compute the tuned observation errors can be found in Bormann et al. (2010).

Observation errors for the first 100 PC scores derived from the channels in the B1 and B1_B2 sets are shown in figure 4 and 5 respectively. In these figures the green curve denotes results for the Desroziers case, the blue curve denotes results for the Hollingsworth/Lönnberg case and the green curve represents the standard deviation of the observed minus background departures. The error estimates given by the Hollingsworth/Lönnberg and Desroziers methods are very similar with the possible exception of the first three PCs. It is evident how the background departures of the low rank PCs have standard deviations approaching a value of 1 which is expected due to the noise normalization of the radiances.

Finally, it should be noted that both the Hollingsworth/Lönnberg and Desroziers method can be used to diagnose inter-PC score error correlations. Correlations between the errors of different PC scores are generally small, with some larger values between high rank PCs (PC#1, PC#2, and PC#3). However, the decision has been taken to avoid this complication and we have neglected the off-diagonal terms of the error covariance matrix.

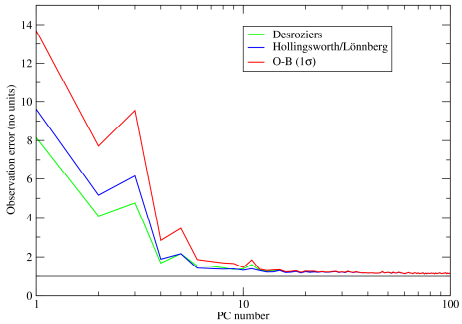

Figure 4. Estimates of observation errors for the PCs derived from the 165 IASI channels. The Red line presents the standard deviation of the background departures. The Desroziers and Hollingsworth/Lonberg errors are plotted as green and blue curves respectively.

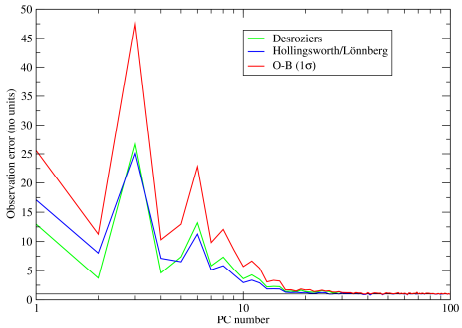

Figure 5. Estimates of observation errors for the PCs derived from the 305 IASI channels. The Red line represents the standard deviation of the background departures. The Desroziers and Hollingsworth/Lonberg errors are plotted as green and blue curves respectively.

4.3 Cloud detection in radiance space and PC based quality control

For the purposes of this prototype study, the assimilation of PC scores is restricted to clear sky conditions. In the ECMWF operational radiance assimilation clouds are detected using the algorithm described in McNally and Watts (2003). However, this scheme requires as input the computation of overcast radiance at the interface of each atmospheric layer and this quantity is not readily available from the current implementation of PC_RTTOV. To avoid an awkward hybrid system (where RTTOV is used for cloud detection and PC_RTTOV used for subsequent assimilation) an alternative cloud detection has been developed. It uses three separate tests applied to uncorrected radiance departures and seeks to identify only fully clear IASI scenes (for details see Matricardi and McNally (2011)).

In conjunction with the new cloud detection scheme, an additional PC based quality control is used and acts as an extra check for residual cloud contamination. As discussed in Matricardi and McNally (2013a), PC#1 derived from the long-wave 165 channel dataset has similar characteristics to an infrared window channel showing a heightened sensitivity to the surface emission and the presence of clouds. Warm O-B departures of the observed PC#1 score from the clear sky computed value are an indication that the observation is affected by clouds. Simulations show that PC#1 derived from the 305 channel dataset should behave in a similar manner. However, while warm O-B departures are indeed present in the long-wave PC#1 spectra, there is scant evidence of warm departures in the spectra of the 305 channel based PC#1. This could be tentatively explained by the fact that compared to the long-wave PC#1, the 305 channel based PC#1 has a heightened sensitivity to lower tropospheric water vapour (see Matricardi and McNally, 2013a for details). It is then possible that the small cloud signal is masked by errors in the first guess humidity fields. While we continue to investigate this issue, the pragmatic solution adopted in this paper is not to apply any quality control to the 305 channel based PC#1. Regarding the long-wave PC#1, using a visual inspection of AVHRR imagery overlaid with IASI pixels it was found that a threshold of 40 units applied to the departure in the long-wave PC#1 is sufficient to reject most cases of residual cloud contamination.

4.4 Bias correction for PCs

Biases in the observations or due to systematic errors in the radiative transfer model and cloud screening are removed using the variational bias correction scheme (VarBC) described by Dee, 2004. This is an adaptive correction algorithm used operationally at ECMWF for all satellite data including IASI radiances (and indeed some in situ observations such as aircraft) where the bias is expressed as a linear combination of pre-defined atmospheric predictors. These predictors account for air-mass variations of the bias correction, but also variations dependent upon the scan geometry. For consistency with radiance observations, but also because PC scores are likely to be influenced by rather similar sources of systematic error, we have applied the same multi-predictor bias correction scheme for the assimilation of the PC scores.

After an initial training phase of typically one to two weeks it is found that the adaptively computed bias corrections for PC scores perform extremely well - becoming very stable in time and removing almost all systematic differences between the observations and the analysis. An exception to this are the corrections computed for PC#1 and PC#2 that are slower to stabilize and tend to drift slightly over time. These particular PC scores have the strongest sensitivity to the surface and any residual cloud contamination - and the slow drift of these bias corrections to a large extent mimics the behaviour often seen in the corrections computed for window channel radiances. For radiances this drift is attributed to a feedback between the quality control steps and the adaptive bias correction. In the PC context, the feedback is between the bias correction and the PC#1 departure quality control. However, for the low rank PC scores we also have the additional complication of stratospheric sensitivity. A number of measures have been employed to alleviate this feedback process in the context of radiance assimilation including limiting the number of degrees of freedom of the bias correction (i.e. the number of predictors), anchoring with trusted uncorrected data and applying additional departure independent QC checks (e.g. from collocated imager information). Similar steps have been tested to limit the drift of the low rank PC score bias corrections and while they have indeed reduced the amplitude of variations over time - none have stopped the drift completely. While this slow variation of bias corrections is undesirable and certainly warrants further investigation, previous experience with radiances - confirmed by tests here with PC scores - suggests that it is not a significant source of degradation in the assimilation. This is mainly because the window channel radiances and their analogue low rank PC scores have little sensitivity to the free atmosphere and generally have rather low weight in the assimilation system (compared to other radiances and PC scores).

5. Assimilation experiments

5.1 Set up of the experiments

To quantify the performance of the PC score assimilation system we have designed a basic set of 4D-Var assimilation experiments that consist of a baseline experiment, a radiance assimilation control experiment and two PC score experiments. The baseline experiment (BASE) uses all operational observations (satellite and conventional) with the exception of IASI data. The radiance control experiment (RAD) is identical to BASE, but additionally assimilates the 191 IASI channels used in the operational 4D-Var system. The long-wave PC score experiment (PC_B1) is identical to BASE, but additionally assimilates 20 PC scores derived from the 165 IASI radiances. Finally, the PC score experiment PC_B1_B2 is identical to BASE but additionally assimilates 50 PC scores derived from the 305 IASI radiances. When assessing the performance of the PC assimilation system it is important to note that the use of PC data is currently restricted to fully clear spectra. This is in contrast to the RAD assimilation system where the use of IASI data extends to channels unaffected by clouds and to fully overcast scenes. Thus, the RAD system makes a more extensive use of the information provided by channels with weighting functions peaking higher in the atmosphere. The use of overcast scenes and channels unaffected by clouds has wide ranging implications for the assimilation system and we have dedicated section 6 to a detailed discussion of the problem of handling clouds in PC space. For the same reason, we have considered appropriate to supplement the RAD experiment

results with results obtained by running an additional radiance experiment (RAD_CLEAR) where the use of the operational IASI radiances is restricted to fully clear spectra. This will allow us to compare the PC assimilation system to a more closely matched radiance system and see how differences in the data coverage affect the assimilation results. Note that in the RAD_CLEAR experiment the methodology used to detect a clear scene is based on the McNally and Watts (2003) algorithm and uses the cloudy flag for IASI window channel #921 (875 cm^{-1}). All experiments have been run using a reduced horizontal resolution version (T511, ~40 km) of the ECMWF IFS cycle 38R2 with 137 vertical levels. To date we have tested the period from 1 June 2012 to 15 September 2012 assimilating IASI data over ocean.

The choice of PC score truncation threshold is based upon a set of short preliminary assimilation experiments. Starting from an initial number of 10, the number of PC scores assimilated in the PC_B1 and PC_B1_B2 systems was varied up to the full number of 165 and 305 scores respectively. It was found that beyond around 20 PC scores in the PC_B1 system and beyond around 50 PC scores in the PC_B1_B2 system there was no discernable improvement in performance (as measured by the fit of the analysis to other observations). Thus it was decided to retain only the first 20 and 50 PC scores for the main PC_B1 and PC_B1_B2 assimilation testing.

In a similar set of preliminary assimilation experiments it was found that both the Desroziers and Hollingsworth/Lönnberg refinements of the diagonal observation error for PC scores produced significantly better results than simply using the untuned standard deviation of observed minus background departures. However, the Desroziers error values gave an additional marginal improvement over the Hollingsworth/Lönnberg estimates so these have been adopted for the main assimilation testing presented here. For the observation error covariance matrix **O** of the RAD experiment we have chosen to use the same diagonal matrix used operationally at ECMWF (see Collard and McNally, 2009 for details). The reason for making this choice was to ensure we have a reliable control based on a sound operationally proven radiance assimilation system (rather than attempting to produce two matched systems with no heritage).

5.2 Impact on the assimilation: analysis increments and the fit to radiosonde data

Figure 6 shows the difference between zonally averaged root-mean-square temperature analysis increments evaluated over three months of assimilation during the June-September 2012 period (e.g. $rms(RAD)^{june-september} - rms(BASE)^{june-september}$). Results for the RAD experiment are plotted in the top left panel, results for the RAD_CLEAR experiment are plotted in the top right panel while results of the PC_B1 and PC_B1_B2 experiments are plotted in the lower left and lower right panel respectively. Analysis increments (defined as the change to the initial conditions at the beginning of the 4D-Var analysis window) are a good indication of how much and where the background errors are corrected by the assimilation of observations. First of all, in these statistics there is no evidence of any anomalous or spurious behavior in the analysis increments produced by the PC score experiments. We should then note that the RAD and RAD_CLEAR experiments display slightly different patterns of analysis increments structures which should be attributed to the different sampling of IASI data. In the troposphere, the assimilation of 20 long-wave PC scores (PC_B1) produces larger corrections to the background errors than the assimilation of 191 IASI radiances. The results shown in the two top panels suggest that the different data coverage is unlikely to be at the origin of these larger corrections which we think should be attributed to the slightly grater weight assigned to the PC scores by virtue of the assigned observation error covariance. The assimilation of 50 PC scores (PC_B1_B2) produce even larger corrections to the background errors. Although some of the structures in the PC_B1 analysis increments (i.e. at Northern and Southern high latitudes and at the Tropics near the surface) are still present (albeit slightly inflated) in the PC_B1_B2 experiment, we observe a substantial and almost universal growth of the analysis increments in the troposphere. Because the data coverage in the PC_B1 and PC_B1_B1 is very similar, one could tentatively attribute the origin of these differences to the information conveyed by the additional IASI channels included in the PC experiment although the different weights assigned to the PC scores could still play a role.

Specific humidity analysis increments are shown in figure 7 where we have adopted the same plotting convention utilised in figure 6. We should note that specific humidity increments are plotted on model levels rather than as a function of the atmospheric pressure. Indicatively, level 75 corresponds to ~200 hPa. As for temperature, the assimilation of 50 PC scores produces the largest corrections to the background errors. A distinctive feature of the increments associated to the PC_B1_B2 system is the magnitude of these increments around and above the top of the troposphere where the RAD, RAD_CLEAR and PC_B1 experiments all produce very similar results. As shown in figure 2, the channel set used to derive the PC_B1_B2 scores contain a large number of humidity sounding channels whose weighting functions have tails that can reach well beyond the top of the troposphere. In principle, the information conveyed by these channels could be responsible for the larger specific humidity increments seen above the troposphere and elsewhere although it cannot be ruled out that these increments are in fact the result of changes made to the background temperature fields.

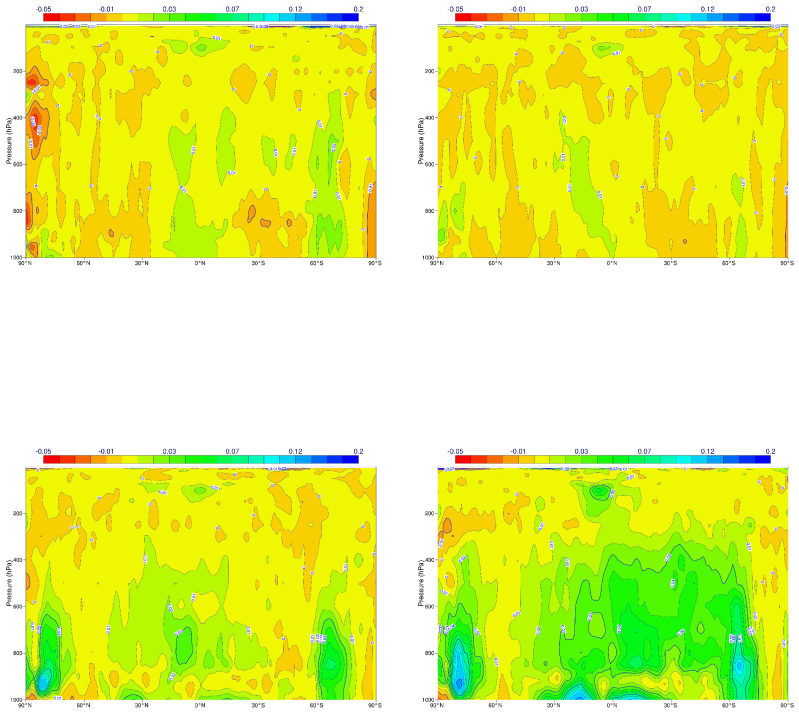

Figure 6. The difference between zonally averaged root-mean-square temperature analysis increments evaluated over three months of assimilation during the June-September 2012 period. Results are shown for RAD-BASE (top left panel), RAD_CLEAR-BASE (top right panel), PC_B1-BASE (lower left panel) and PC_B1_B2-BASE (lower right panel).

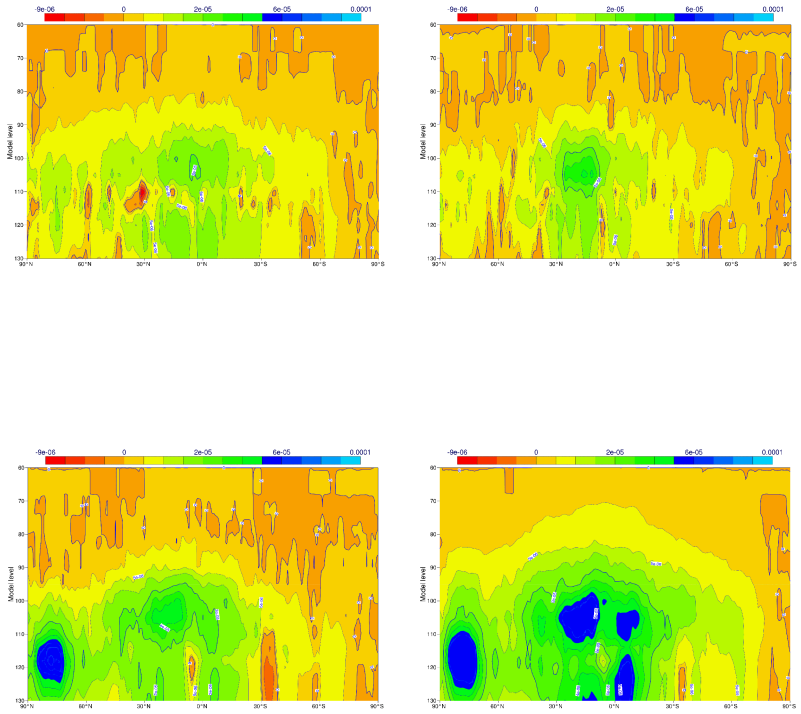

Figure 7. The difference between zonally averaged root-mean-square specific humidity analysis increments evaluated over three months of assimilation during the June-September 2012 period. Results are shown for RAD-BASE (top left panel), RAD_CLEAR-BASE (top right panel), PC_B1-BASE (lower left panel) and PC_B1_B2-BASE (lower right panel).

Statistics of the fit to radiosonde temperatures for the analysis (left panels) and the background (right panels) averaged over the three month test period are shown in figures 8, 9, and, 10 for the extratropical Northern Hemisphere (20ºN to 90ºN), the Tropics (20ºS to 20ºN) and the extratropical Southern Hemisphere (90ºS to 20ºN) respectively. In the two top panels of each figure we compare the PC_B1 (black curve) and the PC_B1_B2 (red curve) experiment to the RAD_CLEAR (green curve) experiment whereas in the lower panels we compare the same PC experiments to the RAD experiment. In each panel standard deviations are shown with respect to the BASE experiment (values are explicitly normalised by the BASE to improve visualization). Thus, reduced values indicate the extent to which the assimilation of the IASI satellite data (either in radiance or PC form) improves the fit to radiosonde data compared to the BASE assimilation. The error bars represent a 95% confidence interval and they can give an indication of the statistical significance of the results.

In the Northern hemisphere, the availability of conventional data means that radiosonde fitting statistics are less sensitive to changes in the use of satellite observations. Changes over the

BASE are in fact much smaller that changes observed in the other two geographical regions considered in the study. Differences between the standard deviations of background departures are typically not statistically significant with the possible exception of the 20 hPa level where the RAD experiment produces better results than the PC based experiments. In general, the PC_B1_B2 experiment tends to produce smaller standard deviations than the PC_B1 experiment above all below 100 hPa where the PC_B1_B2 experiment also matches results from the RAD experiment. From figure 8 it is evident that the use of clear scenes has a detrimental effect on the assimilation of IASI data in that this results in increased values of the standard deviations of the background departures. Finally, in the stratosphere, the assimilation of radiances or 20 long-wave PCs produce a better fit to radiosondes than the assimilation of 50 PC scores.

A similar stratospheric pattern is observed in the Tropics (figure 9). The main difference is that when compared to the RAD system, the PC_B1_B2 system displays larger background standard deviations that are also statistically significant (at least at 100, 1560, and, 200 hPa). Conversely, below 400hPa the PC_B1_B2 system outperforms all other systems. In particular, around 850hPa the better fit to radiosondes is also statistically significant.

A somewhat different picture emerges from the study of the fit to radiosonde data in the Southern Hemisphere (figure 10). The comparative lack of conventional data in the extratropical Southern Hemisphere means that radiosonde fitting statistics are particularly sensitive to changes in the use of satellite observations and provide a reliable quality metric. In figure 10 we note the absence of the dual tropospheric/stratospheric pattern observed in the Tropics and in the Northern Hemisphere. In the stratosphere, all systems are closely matched and none of the differences are statistically significant. In the troposphere, the RAD_CLEAR and the PC_B1 systems lag behind the RAD and PC_B1_B2 systems. It is worth noting that among the various systems considered in this study the RAD_CLEAR and the PC_B1 systems are the closest ones in terms of data and spectral coverage. The detrimental effect of restricting the use of IASI radiances to clear scenes is patently evident in the troposphere where the fit to radiosondes produced by the RAD system is better than that produced by the RAD_CLEAR system although differences are still not statistically significant. In the same atmospheric region it is also evident that the assimilation of the 50 PC scores produces a fit to radiosonde data that is similar to that produced by the RAD system.

In light of the large additional number of water vapour sounding channels used in the PC_B1_B2 experiment, we want to conclude this paragraph by documenting the fit to radiosondes specific humidity data. Results for the extratropical Northern Hemisphere, the Tropics, and the Southern Hemisphere are shown in figures 11, 12, and 13 respectively, using the same plotting convention utilized for the temperature data. The results shown in these figures seem to suggest that the use of the additional IASI channels included in the PC_B1_B2 experiment brings and added value to the skill of the assimilation system. This is because, with the possible exception of the region above 200hPa, the PC_B1_B2 system consistently outperforms all other systems producing background departures with lower standard deviations, although, at many pressure levels, results are not statistically significant.

Regarding the apparent lack of skill of the PC assimilation systems in the stratosphere, it has been pointed out (Bormann, personal communication) that the neglect of the off-diagonal terms in the error covariance matrix used in PC score experiments is equivalent to the introduction of spurious correlations between channels that provide information from regions of the atmosphere spatially unrelated. The experience gained with the study of error correlations in radiance assimilation seem to suggest that this mechanism might be responsible for the behaviour observed in the PC score assimilation experiments. In the next phase of our study we intend to experiment with the introduction of off-diagonal terms in the PC score error covariance matrix although it should be stressed that the correct specification of the correlated observation errors is among the biggest challenges in data assimilation.

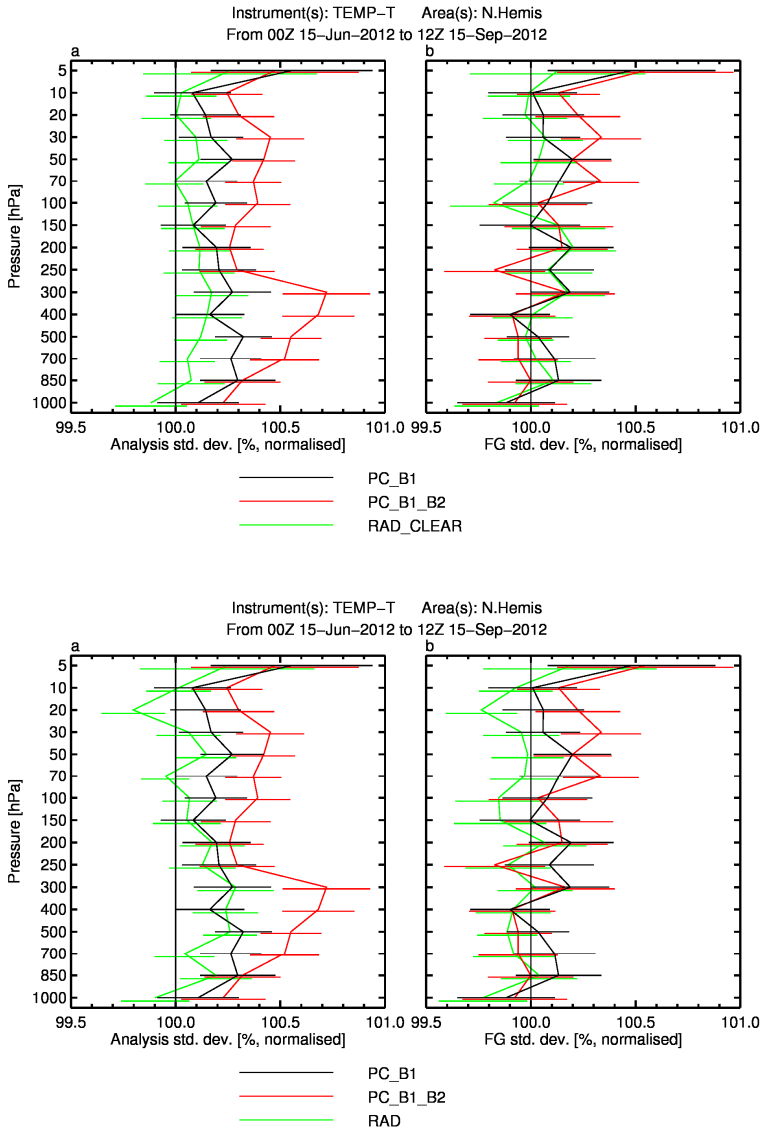

Figure 8. The fit to radiosonde temperature data of the background (right panel) and analysis (left panel) in the Northern Hemisphere. Standard deviations for the period 15 June 2012 - 15 September 2012 are shown for the PC_B1 (black line), PC_B1_B2 (red line), RAD_CLEAR (green line, top panels), and RAD (green line, lower panels) experiments.

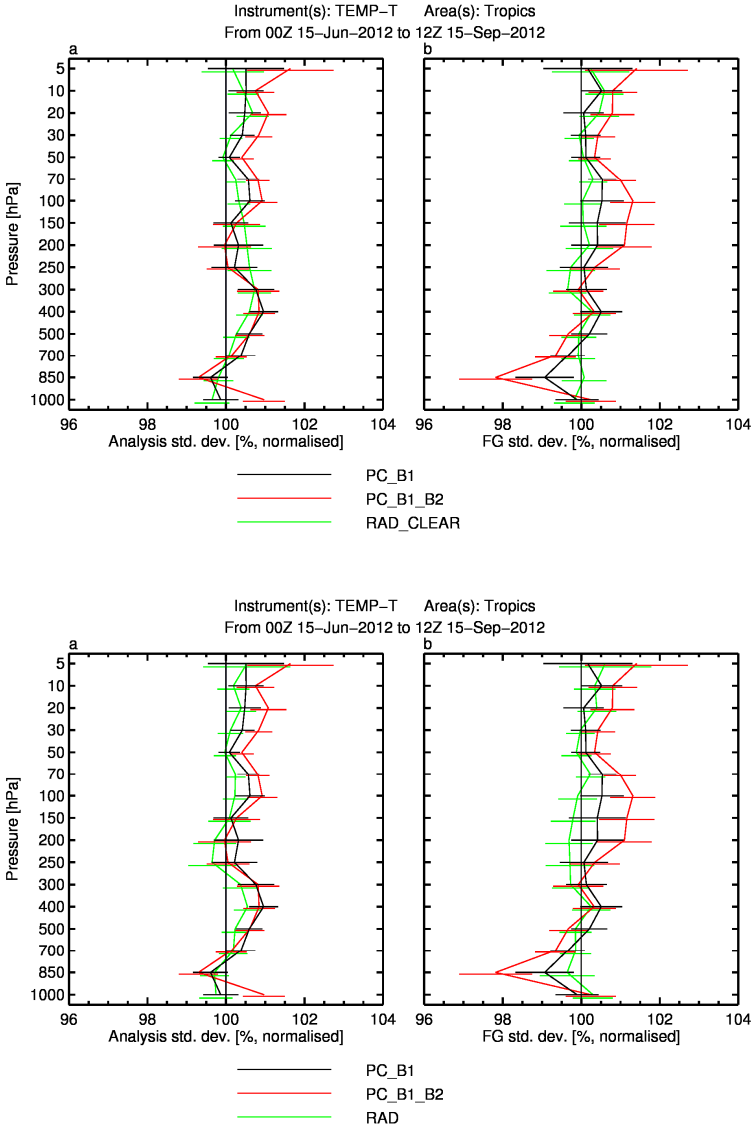

Figure 9. The fit to radiosonde temperature data of the background (right panel) and analysis (left panel) in the Tropics. Standard deviations for the period 15 June 2012 - 15 September 2012 are shown for the PC_B1 (black line), PC_B1_B2 (red line), RAD_CLEAR (green line, top panels), and RAD (green line, lower panels) experiments.

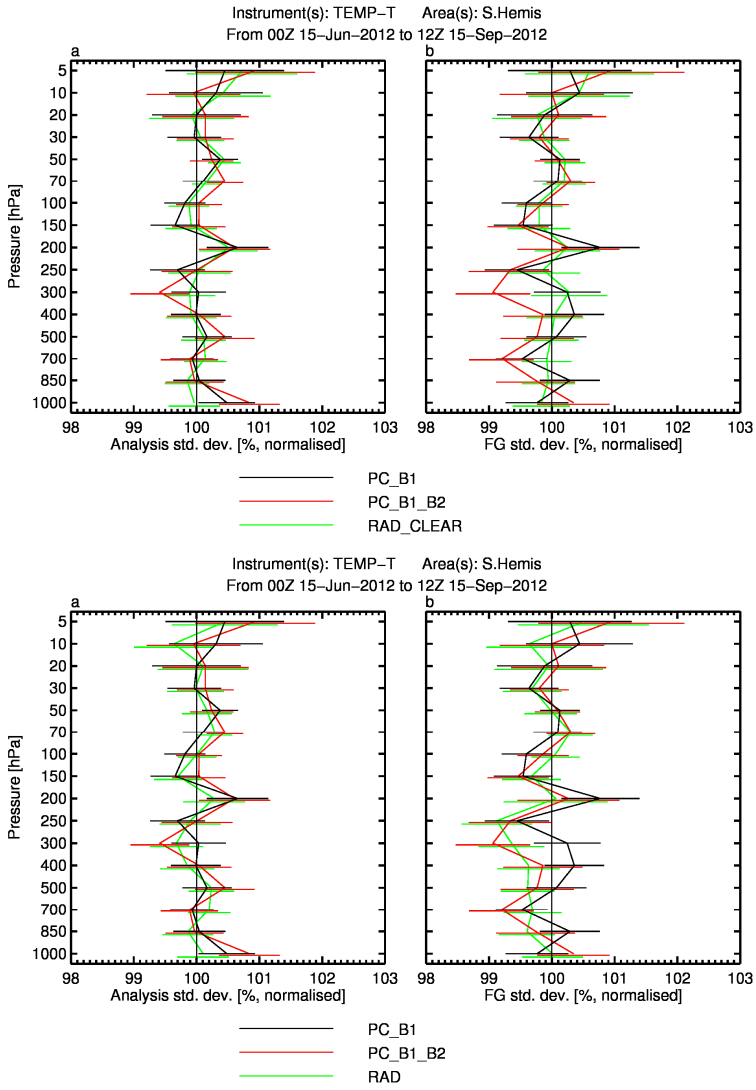

Figure 10. The fit to radiosonde temperature data of the background (right panel) and analysis (left panel) in the Southern Hemisphere. Standard deviations for the period 15 June 2012 - 15 September 2012 are shown for the PC_B1 (black line), PC_B1_B2 (red line), RAD_CLEAR (green line, top panels), and RAD (green line, lower panels) experiments.

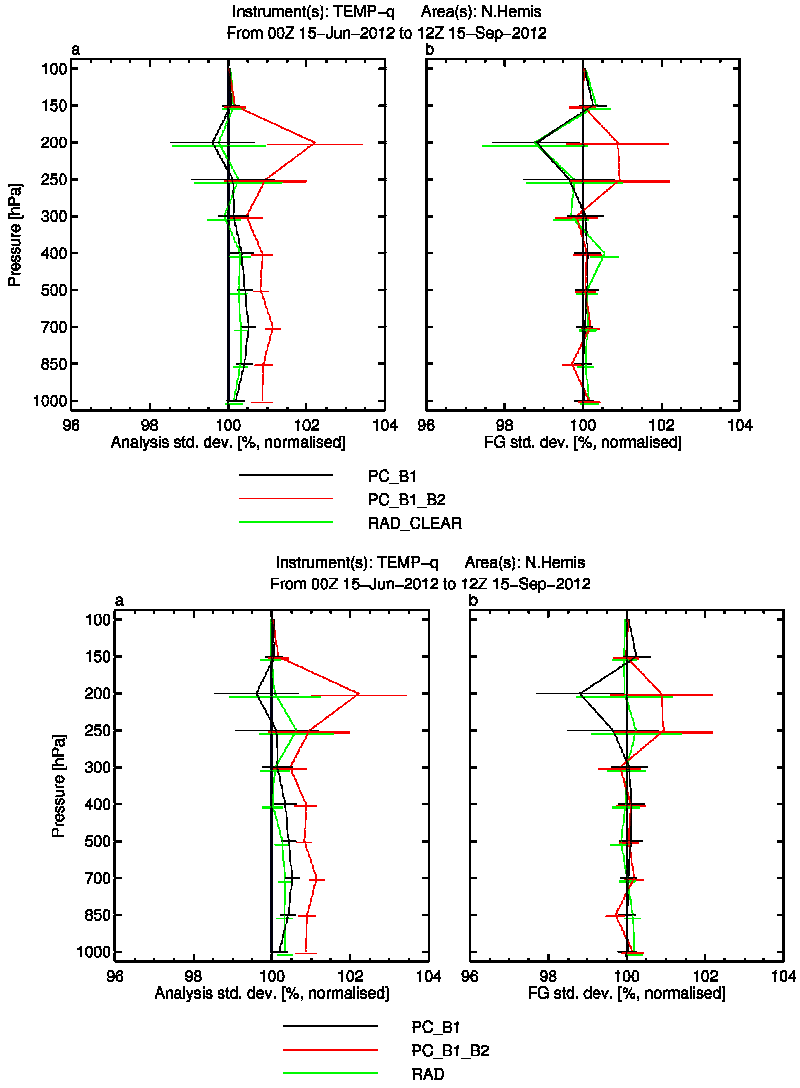

Figure 11. The fit to radiosonde specific humidity data of the background (right panel) and analysis (left panel) in the Northern Hemisphere. Standard deviations for the period 15 June 2012 - 15 September 2012 are shown for the PC_B1 (black line), PC_B1_B2 (red line), RAD_CLEAR (green line, top panels), and RAD (green line, lower panels) experiments.

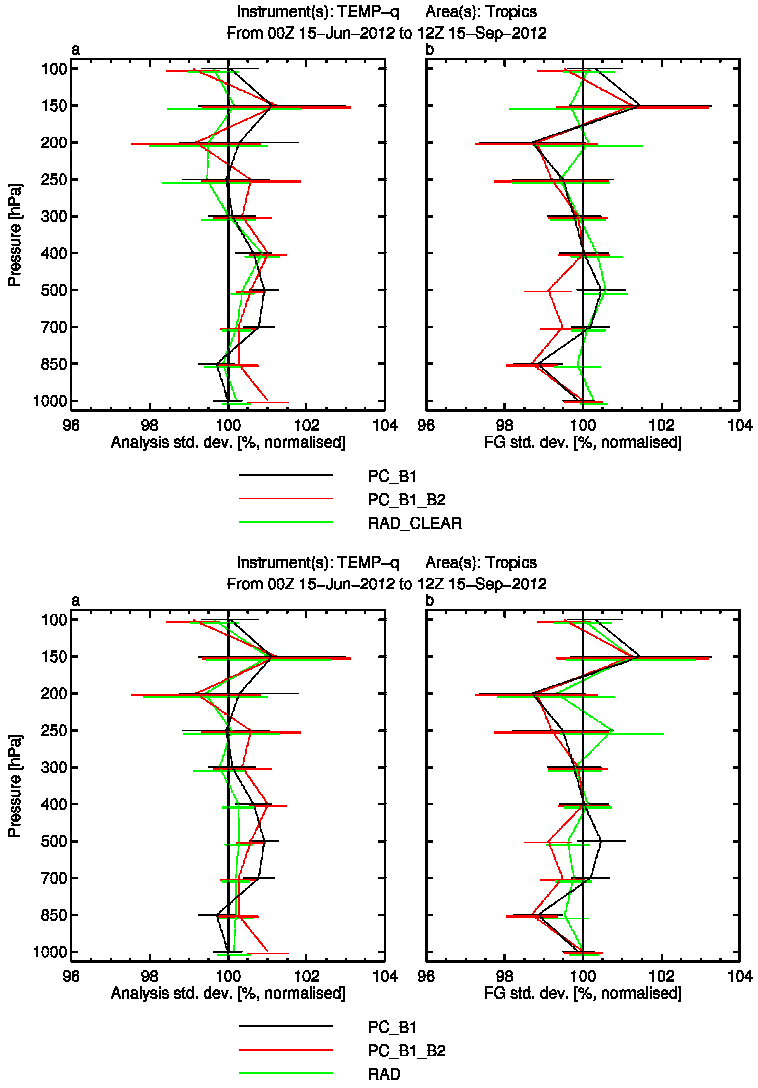

Figure 12. The fit to radiosonde specific humidity data of the background (right panel) and analysis (left panel) in the Tropics. Standard deviations for the period 15 June 2012 - 15 September 2012 are shown for the PC_B1 (black line), PC_B1_B2 (red line), RAD_CLEAR (green line, top panels), and RAD (green line, lower panels) experiments.

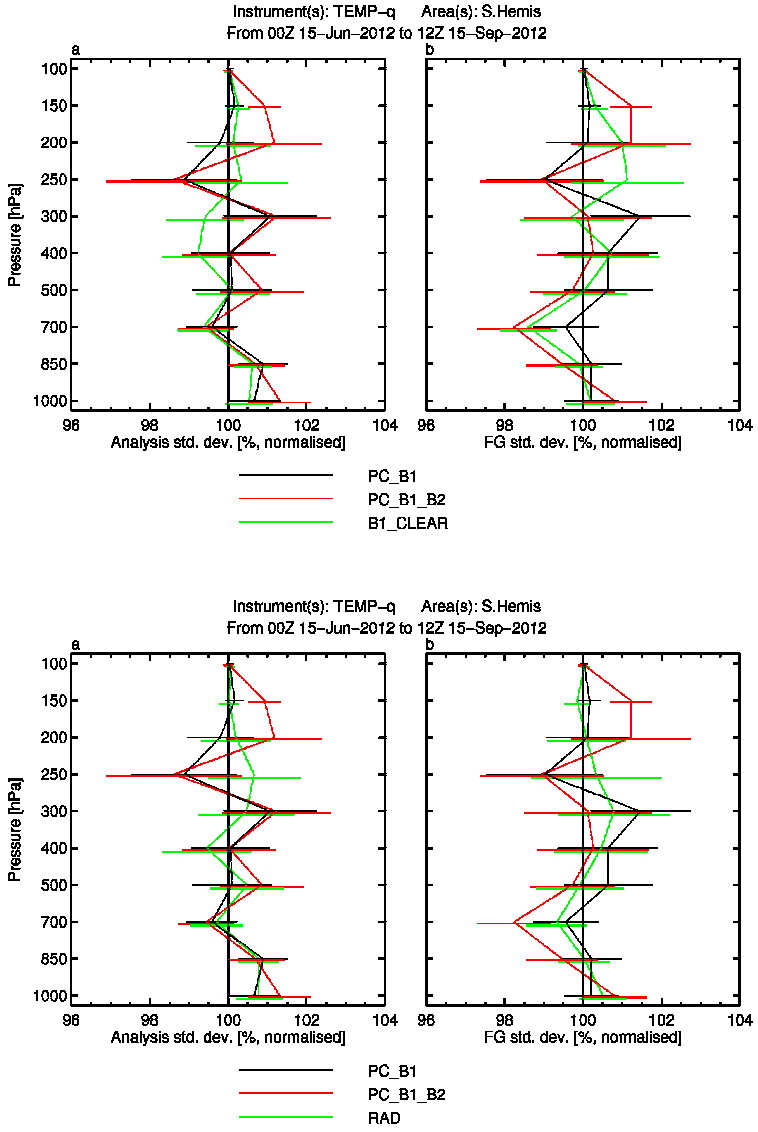

Figure 13. The fit to radiosonde specific humidity data of the background (right panel) and analysis (left panel) in the Southern Hemisphere. Standard deviations for the period 15 June 2012 - 15 September 2012 are shown for the PC_B1 (black line), PC_B1_B2 (red line), RAD_CLEAR (green line, top panels), and RAD (green line, lower panels) experiments.

5.3 Impact on Forecasts

Forecasts have been run from analyses generated by the BASE, RAD, RAD_CLEAR, PC_B1 and PC_B1_B2 assimilation systems and verified using ECMWF operational analyses, which contain information from the full global observing system including the operationally assimilated IASI radiances. Forecast scores over the June-September test period have been computed as the change in the root mean square error compared to the BASE and the differences normalized by the forecast error of the BASE experiment. While this normalisation is arguably the best way to illustrate the impact on forecast errors in the medium range, it can result in the amplification of small differences in the shorter range (0 to 72 hours) where errors in the verifying analyses may be important. To verify the forecast, the geopotential is traditionally used a representative field because it provides a very good measure of the skill to predict large-scale flows and the general weather type. In the Tropics however, because of the different nature of the atmospheric circulation, the geopotential height (and indeed temperature) is not suitable to describe the predictive skill of the forecasting system and it is better to verify the wind vector. To facilitate the visualisation of the data, results for the RAD_CLEAR and RAD experiments are presented in two separate figures. Results for the geopotential have been plotted in figures 14, 15 whereas results for the wind vector have been plotted in figures 16, and, 17. For the reasons discussed above, in figures 14 and 15 the reader should ignore the middle panels (i.e. the Tropics). A negative value of the forecast score means that the use of IASI data improves forecast accuracy compared to the BASE. Error bars indicate a statistical significance of 95%.

Based on the study of the results one might argue that the PC_B1_B2 system tends to produce better forecasts than the PC_B1 system strengthening the evidence that the information conveyed by the additional channels is beneficial to the skill of the assimilation system. Likewise, the restriction of the use of radiances to clear scenes seems to be detrimental to the accuracy of the forecasts. However, if we exclude the shorter range, differences are generally not statistically significant. Differences between the PC_B1_B2 and the RAD system are also not statistically significant suggesting that the assimilation of 50 PC scores performs at least as well as the operational assimilation of 191 radiances. Thus, within the limits of these assimilation trials, despite the use of PC observations is limited to clear scenes only, the PC assimilation based on the use of more IASI channels seems to produce a level of performance that is very close to that produced by the operational radiance assimilation system which is based on the use of fully overcast scenes and on channels unaffected by clouds.

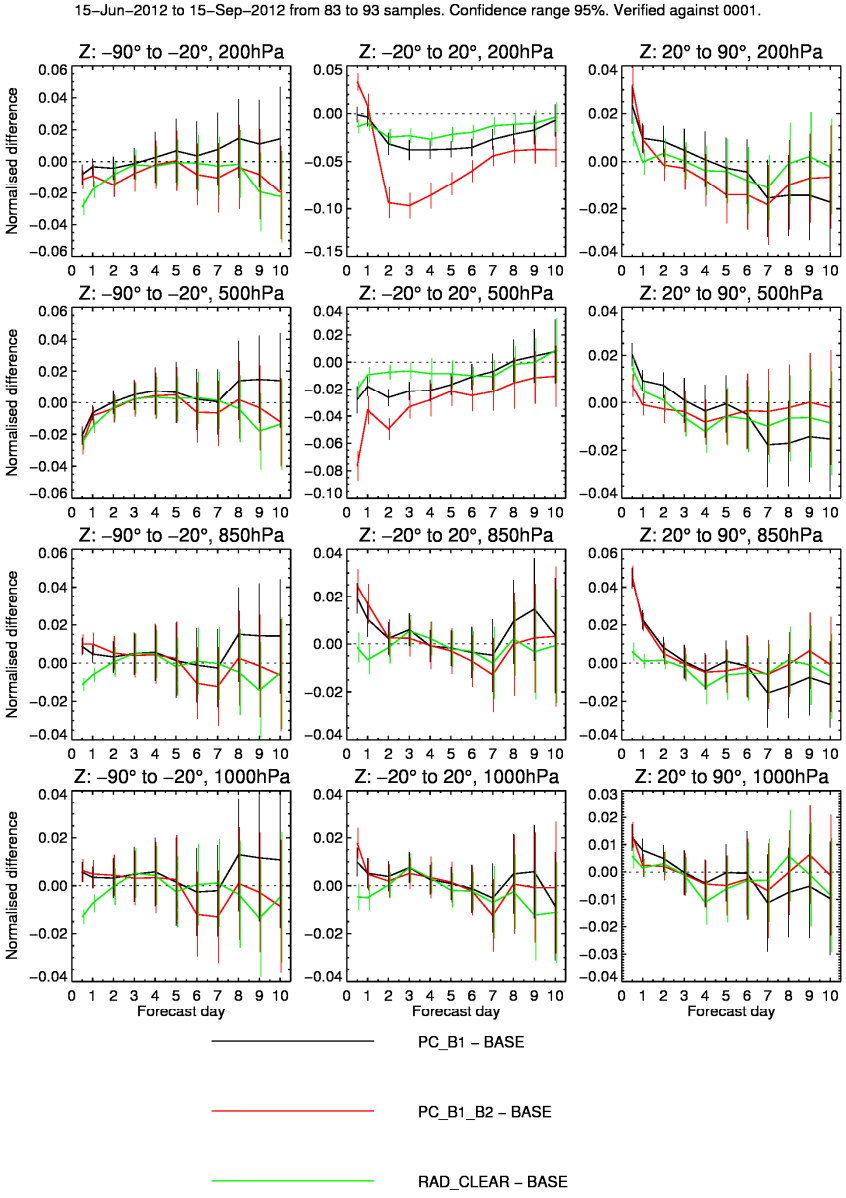

Figure 14. Normalized root-mean-square error difference for geopotential forecasts verified versus the operational analysis for the period 15 June 2012-15 September 2012. Negative values indicate a positive impact from the inclusion of IASI data.

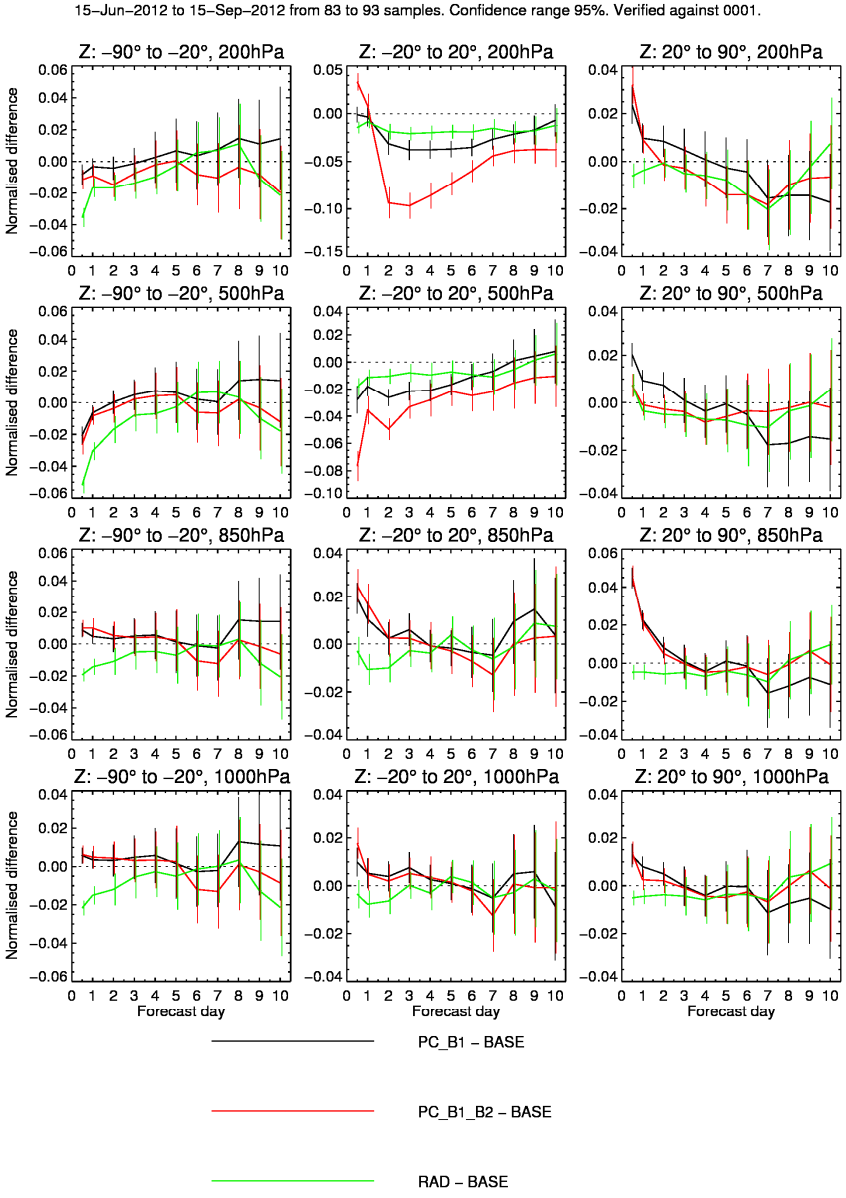

Figure 15. Normalized root-mean-square error difference for geopotential forecasts verified versus the operational analysis for the period 15 June 2012-15 September 2012. Negative values indicate a positive impact from the inclusion of IASI data.

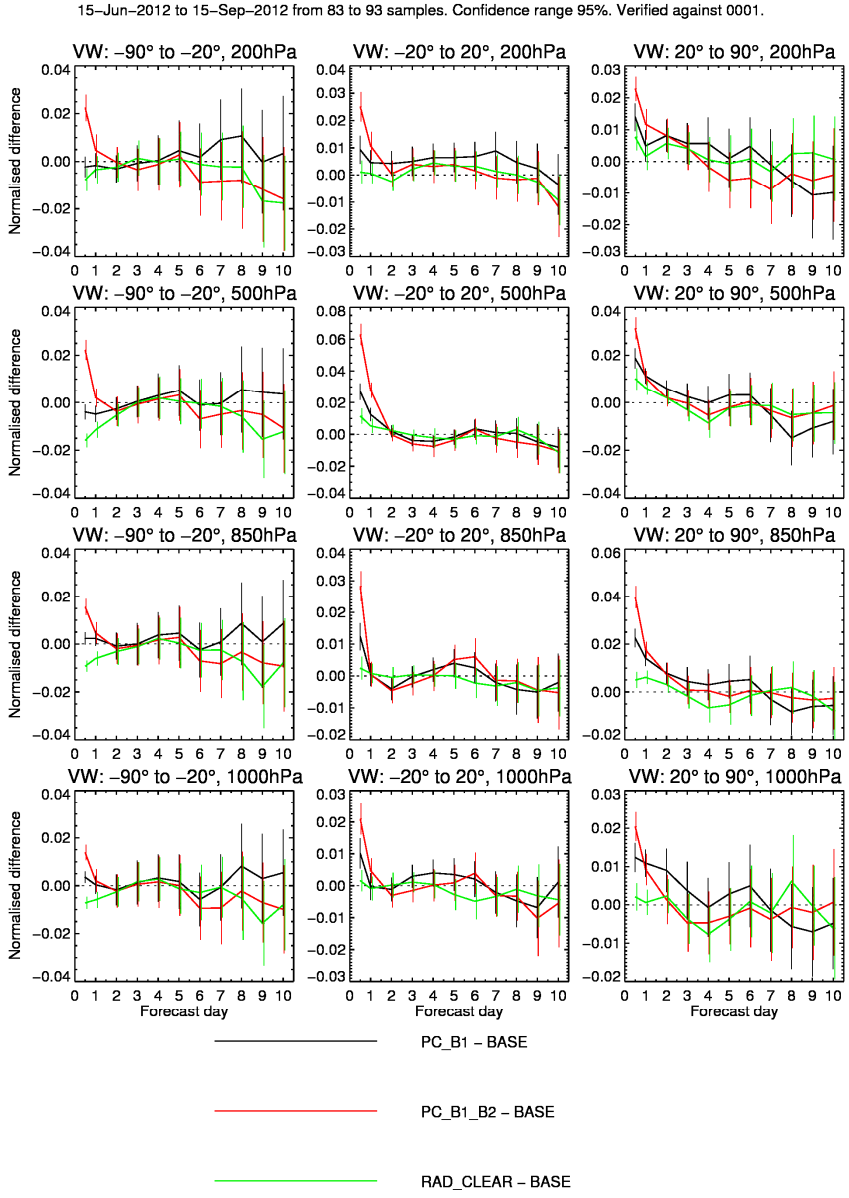

Figure 16. Normalized root-mean-square error difference for wind vector forecasts verified versus the operational analysis for the period 15 June 2012-15 September 2012. Negative values indicate a positive impact from the inclusion of IASI data.

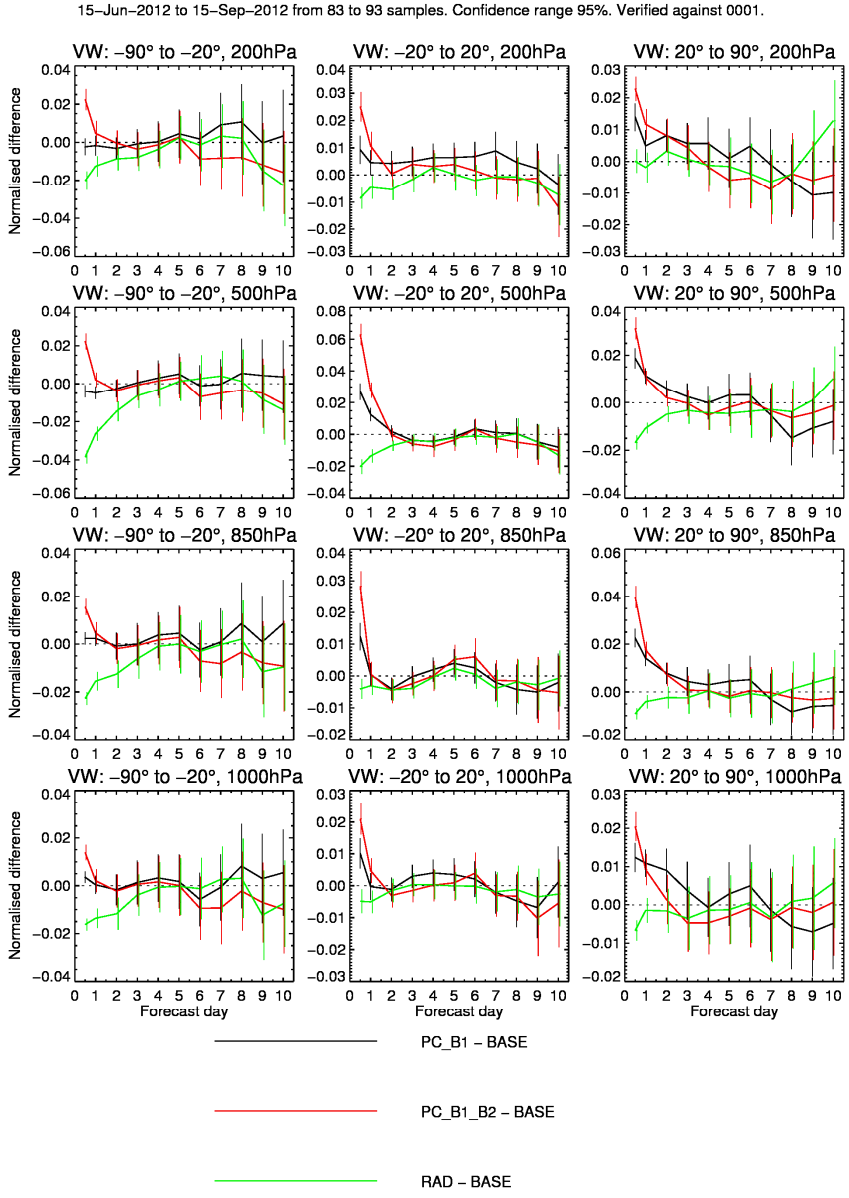

Figure 17. Normalized root-mean-square error difference for wind vector forecasts verified versus the operational analysis for the period 15 June 2012-15 September 2012. Negative values indicate a positive impact from the inclusion of IASI data.

9. Summary and future directions

In this study we have continued the work of the direct assimilation of Principal Component Scores derived from high spectral resolution infrared sounders. In particular we have built upon the results obtained with a prototype long-wave PC score assimilation scheme (Matricardi and McNally, 2013b) which was initially tested in a baseline environment where conventional observations and atmospheric motion vectors, AMVs, are assimilated, but IASI are the only satellite sounding data used (either in the form of PCs or radiances). Based on the impressive performance of this PC assimilation prototype, in this study we have given priority to the testing of the IASI PC approach in a full data assimilation system that contains all operational observations (satellite and conventional). Testing in this extremely demanding new baseline environment has allowed us to check that the conclusions reached for the prototype system study are robust. Furthermore, following the investigation into the use of PCs to represent the IASI short-wave and long-wave spectrum, we have taken the logical step of considering the extraction of information from the dedicated IASI water vapour and ozone bands. This is a logical step towards the exploitation of the full IASI spectrum. To this end, we have designed a new experiment where we have assimilated 50 PC scores derived from 305 IASI channels obtained by augmenting the 191 IASI operational channels with additional surface, ozone and water vapour sounding channels. Because the assimilation of PC scores is performed for clear scenes only, we have also designed a new radiance assimilation experiment where the use of IASI radiances is restricted to clear scenes only. This has allowed us to compare the PC system to a more closely matched radiance assimilation system and to see how the restriction to clear scenes only affects the assimilation of IASI radiances.

The study of the zonally averaged temperature analysis increments (defined as the change to the initial conditions at the beginning of the 4D-Var analysis window) indicate that below 200 hPa the assimilation of 50 PC observations produce larger corrections to the background errors than the assimilation of 191 operational IASI radiances. This effect is universal and is larger in the Tropics near the surface and at Northern high latitudes in the lower troposphere. Although the data coverage is not identical and the observations are assigned different weights, one could tentatively attribute the origin of these differences to the information conveyed by the additional IASI channels included in the PC experiment. When we examine how the assimilation of either IASI PC scores or radiances affects the fit to radiosonde temperature observations, we see that in the Northern Hemisphere, Southern Hemisphere and Tropics, differences are generally not statistically significant. When the differences are statistically significant, the most salient features seen in the statistics is the slightly better fit produced in the troposphere by the assimilation of PC scores and the worse fit produced by the assimilation of PC scores in the stratosphere. It has been suggested that the worst performance of the PC score system in the stratosphere might be associated with the neglect of the off-diagonal terms in the error covariance matrix used in PC score experiments. We intend to tackle this issue in the next phase of the study but it should be stressed that the correct specification of the correlated observation errors is among the biggest challenges in data assimilation.

Forecasts have been run from analyses generated by the radiance and PC assimilation systems and verified using the ECMWF operational analyses. Forecast scores over the three month test period have been computed as the change in the root-mean-square error compared to a baseline radiance system where we do not assimilate IASI radiances. Differences between the forecasts produced by the PC and RAD system are generally not statistically significant suggesting that the PC score assimilation performs at least as well as the radiance system. Thus, despite the use of PC observations is limited to clear scenes only, the PC assimilation based on the use of more IASI channels seems to produce a level of performance that is very close to that produced by the operational radiance assimilation system which is based on the use of fully overcast scenes and on channels unaffected by clouds. This result is all the more important in light of the fact that the 50 PC score system based on 305 radiances uses ~20% less computer resources (during the 4D-var minimization) compared to the system that assimilates 191 radiances.

References

Bormann, N. and Bauer, P. (2010), Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. Q.J.R. Meteorol. Soc., 136: 1036–1050. doi: 10.1002/qj.616

Collard, A. D. 2007**.** Selection of IASI channels for use in numerical weather prediction. Q.J.R. Meteorol. Soc., **133**: 1977–1991. doi: 10.1002/qj.178.

Collard AD, McNally AP. 2009. Assimilation of IASI radiances at ECMWF, *Q. J. Roy. Meteorol. Soc.*, **135**: 1044-1058.

Dee D. 2004. Variational bias correction of radiance data in the ECMWF system. Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, 28 June-1 July 2004, Reading, UK.

Desroziers G., Berre L., Chapnik B., Poli, P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Quarterly Journal of the Royal Meteorological Society, **131**, 3385-3386.

Joliffe IT. 2002. Principal Component Analysis. Springer: New York.

Hollingsworth A, Lönnberg P. 1986. The statistical structure of short-range forecast errors determined from radiosonde data. Part I: The wind field. Tellus, **38A**, pp. 111-136.

Huang H-L, Smith WL, Woolf HM. 1992. Vertical resolution and accuracy of atmospheric infrared sounding spectrometers. J. Appl. Meteorol. 31: 265–274.

Matricardi M. 2010. A principal component based version of the RTTOV fast radiative transfer model. Q.J. Roy. Meteorol. Soc., **136**, pp. 1823-1835.

Matricardi M, McNally T. 2011. The direct assimilation of IASI short wave principal component scores into the ECMWF NWP model. EUMETSAT Contract No. EUM/CO/07/4600000475/PS.

Matricardi M, McNally T. 2012. The direct assimilation of IASI short-wave principal component scores into the ECMWF NWP model. EUMETSAT Contract No. EUM/CO/07/4600000475/PS.

Matricardi M, McNally T. 2013a. The direct assimilation of principal component of IASI spectra in the ECMWF 4D-VAR: extension to IASI band 1 and IASI band 2. EUMETSAT Contract No. EUM/CO/07/4600000475/PS.

Matricardi, M. and McNally, A. P. 2013b, The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var. Q.J.R. Meteorol. Soc.. doi: 10.1002/qj.2156

McNally AP, Watts PD. 2003. A cloud detection algorithm for high-spectral-resolution infrared sounders. Q. J. Roy. Meteorol. Soc., **129**: 3144-3423.