TECHNICAL MEMORANDUM

690

The Direct Assimilation of Principal Components of IASI Spectra in the ECMWF 4D-Var

M. Matricardi and A.P. McNally

Research Department

December 2012

Submitted to *Q J Roy Meteor Roc*

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.

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

Series: ECMWF Technical Memoranda

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

Contact: library@ecmwf.int

© Copyright 2012

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire 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. 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 European Centre for Medium-Range Weather Forecasts (ECMWF) 4D-Var data assimilation system has been modified to allow the direct assimilation of Principal Component (PC) scores derived from spectra measured by the Infrared Atmospheric Sounding Interferometer (IASI). Testing of a prototype system where 165 IASI radiances are replaced by just 20 PC scores shows significant computational savings with no detectable loss of skill in the resulting analyses or forecasts. Indeed in some respects the assimilation of PC scores leads to marginal improvements over the traditional radiance based assimilation.

KEYWORDS: IASI; Data Assimilation; Satellite Data; Principal Component Analysis

1 Introduction

The assimilation of high resolution radiances measured by the Infrared Atmospheric Sounding Interferometer (IASI) (Chalon et al., 2001) on the MetOp-A platform has produced a significant positive impact on forecast quality (Collard and McNally, 2008). The operational use of IASI radiances at the European Centre for Medium-Range Weather Forecasts (ECMWF) and other NWP centres is currently restricted to a selection of temperature sounding channels in the long-wave region of the spectrum and to a small number of 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. Similarly, satellite agencies are seeking a more efficient means of near-real time data dissemination for instruments such as IASI – as the traditional practice of transmitting full spectral data at full spatial resolution is likely to become prohibitively expensive in the future (as instruments are flown on multiple polar and geostationary platforms).

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 problems although - while noting that the two issues are quite similar- the requirements are quite separate. There are strong indications that data providers will evolve to the dissemination of Principal Component (PC) scores to improve efficiency. It is thus timely and opportune to investigate the feasibility of directly assimilating PC scores into NWP models. Of course alternative methodologies are available for the efficient encapsulation and assimilation of high-resolution infrared sounder spectra. These include the use of radiances reconstructed from PC scores (Collard *et al*., 2009) and the use of data in the form of partial eigen-decomposition and null-space filtered retrievals (Joiner and Da Silva, 1998). We should stress that in this paper no attempt is made to compare the relative merits of these, which at a theoretical level can arguably be considered equivalent (see e.g. Masiello *et al*., 2011). Our main objective is to document the development and the functionality of a prototype global 4D-Var assimilation system based on the direct use of PC data.

The outline of the paper is as follows. In Sections 2 and 3 we review the theory of PCA and discuss the features of PC data derived from radiance spectra. In Section 4 we present the assimilation methodology. Section 5 discusses results from the direct assimilation of PC scores using the ECMWF Integrated Forecasting System (IFS). Finally, in Section 6 we present a summary of the results and discuss possible future directions.

2 A 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 $\boldsymbol{R}$. 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 $\boldsymbol{C}$ is the n by n covariance matrix of the data matrix $\boldsymbol{R}$, and $\boldsymbol{A}$ 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, $\boldsymbol{p}$, 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 $\boldsymbol{A}$ 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). 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 this prototype 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.

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.

In this study we only consider 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.

3.1 The long-wave channels and derived PC scores

The assimilation of PC data requires the conversion of IASI radiances into PC scores. The radiances used in this paper are a subset of those used operationally in cycle 36R1 of the ECMWF Integrated Forecasting System (IFS) (see Collard and McNally, 2008 for details on the channel selection). This subset comprises 165 long wave channels in IASI band 1 and has been obtained by removing from the original dataset ten channels located in the $\nu 2$ water vapour vibro-rotational band. Excluding radiances with a strong sensitivity to water vapour allows us to run assimilation experiments in more controlled conditions - using sounding channels covering the region between 645 and 875 cm^{-1} with a primary sensitivity to temperature and the surface. It should be noted however, that some of the remaining channels are still marginally affected by water vapour absorption corresponding to weak water vapour lines originating from the edge of the water vapour pure rotational band.

The conversion of radiances to 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 based fast radiative transfer model used in the assimilation (see section 4 for details). Note that the eigenvectors are derived from the 165 long wave radiances described above.

When compared to spectral radiances, the physical interpretation of PC score observations is less intuitive. This is illustrated in Figure 1 where we have plotted the temperature Jacobians (corresponding to a US Standard Atmosphere) for the 10 highest ranking PC scores. Radiance temperature Jacobians are broad, but relatively localized in a given part of the atmosphere (e.g. surface or stratosphere) whereas the PC score Jacobians are not localized and can have multiple maxima throughout the entire atmosphere (e.g. at the surface and in the stratosphere).

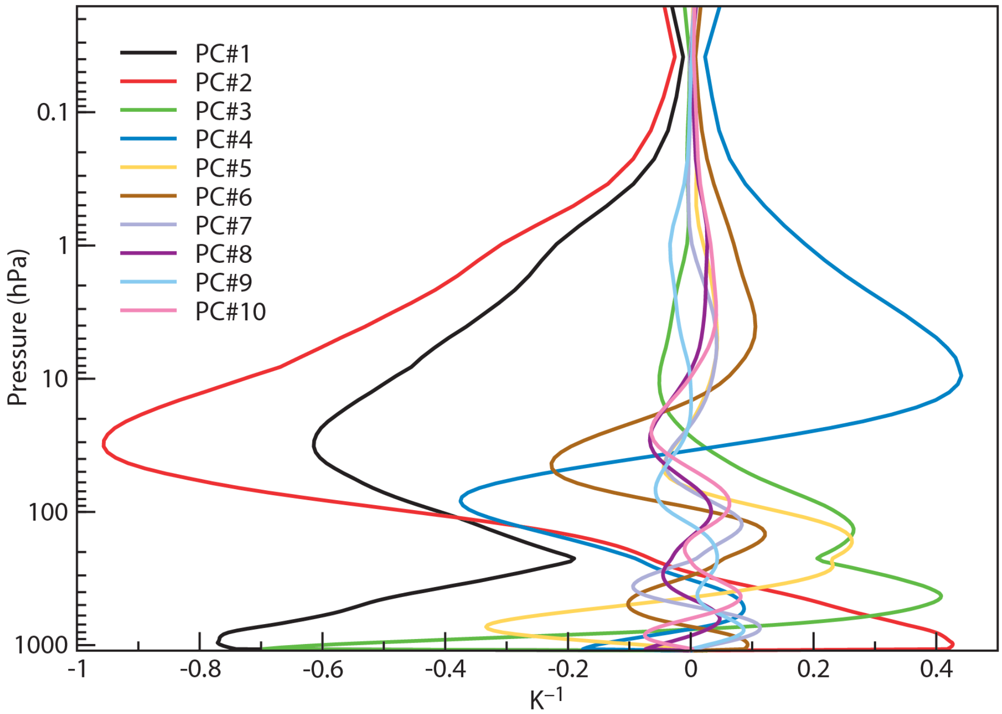

Figure 1. The temperature Jacobian for the first 10 PCs for the US Standard Atmosphere.

Table 1. The skin temperature Jacobian for the first 10 PCs for the U.S. Standard Atmosphere.

PC number	Jacobian (K^{-1})
1	-9.181
2	5.459
3	-11.783
4	-2.428
5	4.281
6	-0.896
7	-0.797
8	0.487
9	-1.357
10	0.565

Near-surface layers give a contribution to the signal that is significantly larger for the first three PCs. While a sensitivity to the stratosphere is apparent to a variable extent in a number of PCs, PC#2 has the largest contribution from that region. It should be noted that the highest ranking PCs also have a significant sensitivity to changes in the surface skin temperature (see Table I). This is particularly true for PC#1 whose behaviour is closely related to that of a window channel radiance and has the strongest response to the presence of cloud. However, in this case if the warm surface is obscured by a cold cloud, we expect a warming of the observed PC#1 score (opposite to the response of an infrared window channel that would cool). Cloud signals appear as an asymmetry (i.e. a warm tail) in the histogram of the observed minus computed PC score departures.

4 PC assimilation methodology

4.1 Overall architecture of the assimilation system

The methodology adopted in this study for the direct 4D-Var assimilation of PC scores is shown schematically in Figure 2. 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 165 long wave channel basis described previously, to produce a vector of observed PC scores Y_{OBS}^{PC}. Each vector of observed PC scores has length $n=165$, 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.

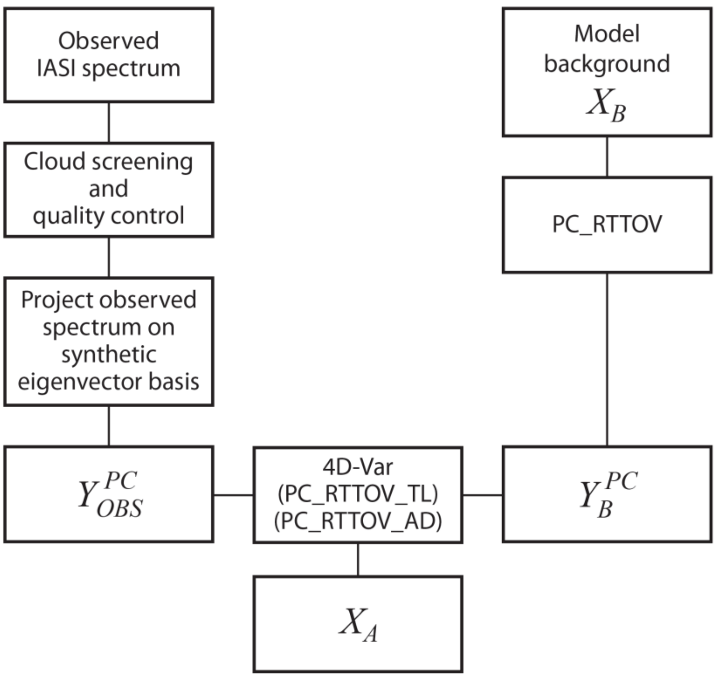

Figure 2. The flow diagram of the direct PC score assimilation.

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:

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

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 4.4. 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.

4.2 The observation operator for PC scores

The observation operator utilized for the simulation of the PC scores (PC_RTTOV) is a PCA-based version of the RTTOV (Matricardi *et al*., 2004) fast radiative transfer model. For a detailed description of PC_RTTOV the reader can refer to Matricardi (2010). Here we will discuss only those aspects related to the specific nature of this study.

The PC_RTTOV fast radiative transfer model performs rapid and accurate simulations of PC scores of IASI radiances using a multiple linear regression scheme. In this scheme, the PC scores are expressed as a linear combination of a selected number of polychromatic radiances simulated by the conventional RTTOV fast model. The regression coefficients are computed using the PC scores obtained from the eigenvectors of the covariance matrix of a large dataset of synthetic noise-free clear sky radiances calculated using an accurate line-by-line model. Note that the principal component scores are calculated from radiance spectra that have been normalized using the diagonal elements of the IASI instrument noise covariance matrix. The PC scores can be computed for any state vector that includes variable profiles of temperature, water vapour, ozone, and surface parameters. The number of predictor variables used in the regression algorithm is a tunable parameter in the model. By varying this parameter, we can trade off computational efficiency against the accuracy of the simulations.

Regression coefficients for the PC scores derived from the long wave radiances used in this study are available based on 30, 35, 40, 45, and 50 predictors. The accuracy of the PC score fast model is illustrated in Figure 3 where we have plotted the standard deviation of the difference between exact PC scores (i.e. derived from line-by-line radiances) and fast model PC scores predicted with 35 and 50 variables respectively. The statistics in Figure 3 have been obtained applying the exact and fast calculations to an independent set of 5190 atmospheric profiles (i.e. independent of the population used to train the fast model). It is evident that the error can be reduced by increasing the number of predictors used in the regression. The systematic component of the error (not shown here) is a negligible fraction of the random component. Finally, it should be noted that the errors shown in Figure 3 only represent the ability of the fast PC_RTTOV to reproduce the *exact* PC scores computed from line-by-line radiances. In that respect, they do not represent the total error in the PC_RTTOV simulated PC scores as this is likely dominated by weaknesses in the underlying line-by-line model to describe the true radiative transfer of the atmosphere.

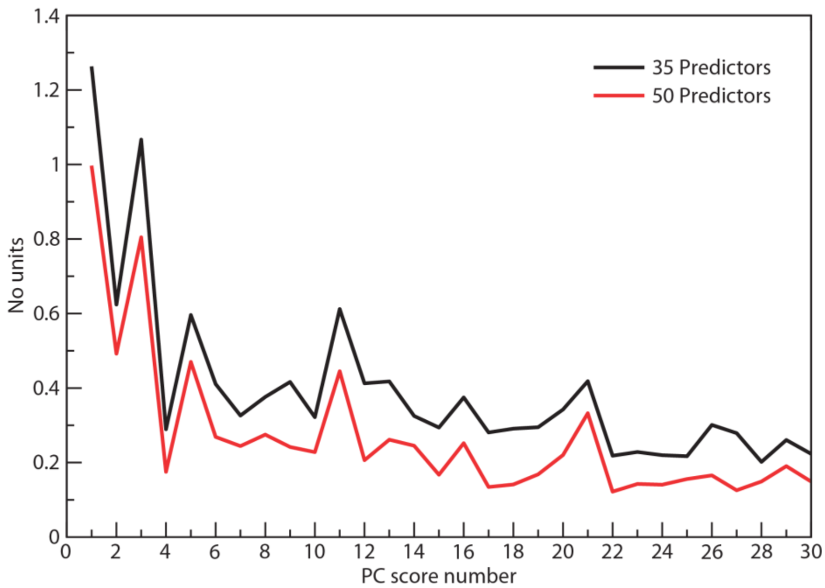

Figure 3. The standard deviation of the difference between exact and simulated PC scores. The black curve denotes results obtained using 35 predictors whereas the red curve denotes results obtained using 50 predictors.

4.3 The truncation of the PC scores in the assimilation

We have to decide how many PC scores (m) should be retained to replace the original n radiances to improve the efficiency of the assimilation system without incurring in a damaging loss of information. Depending on the application, several different methods have been proposed to determine the optimal value of m (e.g. Turner et al. 2006). In this paper, we have followed the approach taken by Antonelli *et al.* (2004). To this end, we have assembled a large ensemble of 10882 observed clear IASI spectra and have reconstructed them from a variable number, p, of truncated PC_RTTOV synthetic eigenvectors. We have then computed the following quantity:

$$d(p) = \frac{1}{n}\sum_{j=1}^{n}\left|(R_j(p) - E_j\right| \tag{8}$$

where $R_j(p)$ is the root-mean-square of the radiance reconstruction error and E_j is the root-mean-square of the instrument noise. Here p denotes the number of truncated PC scores, j is the IASI channel number, and n is the total number of channels.

In an ideal situation where we know exactly the instrument noise, the function $d(p)$ should attain a minimum in correspondence of the optimal number of eigenvectors (i.e. the reconstruction residuals mostly represent instrument noise). Although a minimum value is actually reached for 11 eigenvectors, this number is likely to not correspond to the optimal value. The reason for this is twofold. Firstly, the tabulated values of the noise used in Eq. (8) may differ from the true values. If for instance we alter the noise values by ± 10%, the minimum moves to 5 and 16 eigenvectors respectively. Secondly, by truncating the synthetic eigenvectors we introduce spectroscopic errors in the reconstructed radiances. In addition, because the variability of CO_2 is not represented in the synthetic eigenvector basis, the truncation will remove part of the CO_2 signal. In conclusion, although not optimal, the number of 11 eigenvectors can nevertheless still be taken as reasonable indication of the number of eigenvectors we need to retain.

4.4 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

$$\mathbf{O} = E\left[\mathbf{d}_a \mathbf{d}_b^T\right] \tag{9}$$

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 20 PC scores are shown in Figure 4 where the black curve denotes results for the Desroziers case, the red curves 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. From Figure 4 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 and PC#2). However, the decision has been taken to avoid this complication and we have neglected the off-diagonal terms of the error covariance matrix.

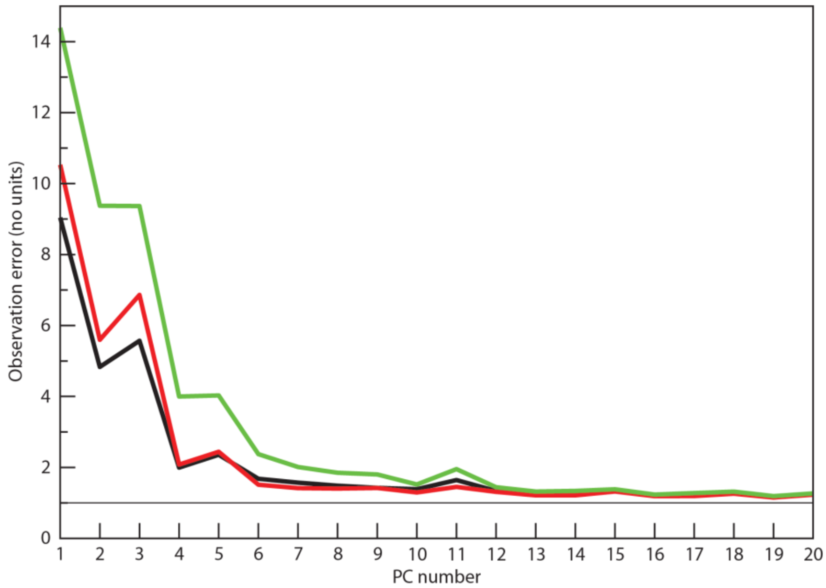

Figure 4. Estimates of observation errors for the PCs used in the 20 PC score assimilation experiment. The green line represents the standard deviation of the background departures. The Hollingsworth/Lönnberg and Desroziers errors are plotted as red and black lines respectively.

4.5 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 section 3.1, PC#1 has similar characteristics to an infrared window channel – in particular a heightened sensitivity to the surface emission and the presence of clouds. Warm departures of the observed PC#1 score from the clear sky computed value are an indication that the observation is affected by clouds. 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 PC#1 is sufficient to reject most cases of residual cloud contamination.

4.6 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 behavior 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 long-wave PC score assimilation system we have designed a set of 4D-Var assimilation experiments that consist of a baseline experiment, a radiance assimilation control experiment and a PC score experiment. The baseline experiment (hereafter referred to as the BASE) uses only conventional data plus atmospheric motion vectors from polar and geostationary satellites. The radiance control experiment (hereafter referred to as RAD) is identical to BASE, but additionally assimilates 165 IASI long wave radiances discussed in section 3.1. The PC score experiment (hereafter referred to as the PC) is identical to the BASE, but additionally assimilates IASI PC scores (derived from the 165 long wave radiances). All experiments have been run using a reduced horizontal resolution version (T511 ~ 40Km) of the ECMWF IFS cycle 36R1 with 91 vertical levels. Two periods in different seasons are tested - from 12 June 2010 to 15 July 2010 and from 15 December 2010 to 15 January 2011. The RAD and PC experiments utilize the same pixel based cloud detection scheme (described in section 4.5) and we assimilate IASI data (radiances or PC scores) only 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 was varied up to the full number of 165 scores. It was found that beyond the around 20 PC scores 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 PC scores for the main 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, 2008 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 PC score and Radiance Data Coverage

An example of the data coverage for a single assimilation window (i.e. 12 hours) is illustrated in Figure 5. In the left panel we have plotted the coverage for the long wave PC score assimilation system (PC) whereas in the right panel we have plotted the coverage for the long wave radiance assimilation system (RAD). As expected the majority of fully clear pixels are identified in the Tropics and it can be seen that both systems suffer data loss in regions with extensive and frequent cloud cover – in particular the mid-latitude storm tracks and the equatorial convection zones. Crucially however, the two panels show that the data coverage for the two systems is almost identical (differences in sample size are typically less than 1%). Small discrepancies are inevitable: after a number of days the PC and RAD systems evolve different background fields of temperature and humidity; radiance calculations used in the cloud detection are performed by different fast models (i.e. RTTOV and PC_RTTOV) and the PC system employs an additional cloud check based on the PC# departure. However, the differences are sufficiently small that we can eliminate data coverage as a source of any change between the PC and RAD systems.

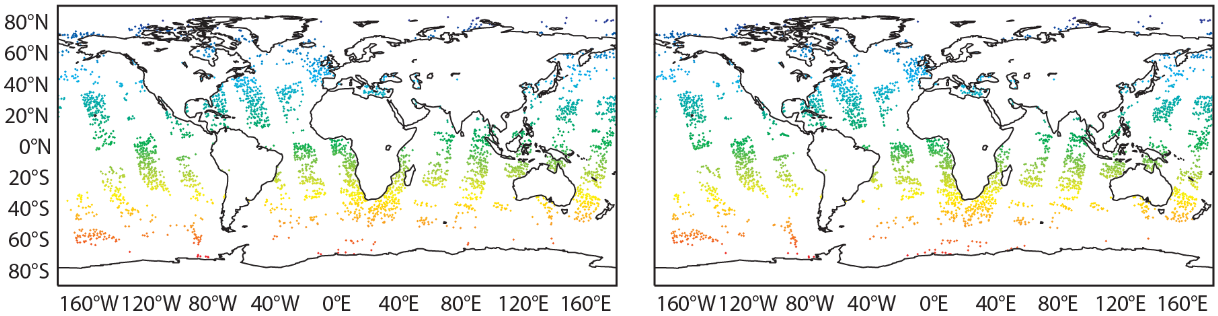

Figure 5. The 12-hour data coverage for (left panel) the radiance and (right panel) the PC score assimilation system.

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

Figure 6 shows zonally averaged root-mean-square temperature analysis increments evaluated over one month of assimilation during the June-July 2010 period. Results for the reference IASI long wave radiance experiment (RAD) are plotted in the left panel while results of the IASI long wave PC score experiment (PC) are plotted in the right panel. 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 being corrected by the assimilation of observations. It can be seen in figure 6 that the two experiments display largely similar patterns of analysis increments structures – indicating that (at least statistically) the assimilation of either radiances or PC scores results in similar adjustments of the background fields. In some areas there are small differences in magnitude of increments, for example in the Tropics between 300 and 600 hPa where the assimilation of IASI PC scores produces slightly larger adjustments than the IASI radiances. As the coverage of assimilated data is almost identical, it is likely that the differences can be attributed to the slightly greater weight assigned to the PC scores by virtue of the assigned observation error covariance. However, in these statistics (and those of the winter period which are not shown) there is certainly no evidence of any anomalous or spurious behavior in the analysis increments produced by the PC score experiment. As such we conclude that the IASI PC scores are conveying broadly similar information to the assimilation as the IASI radiances.

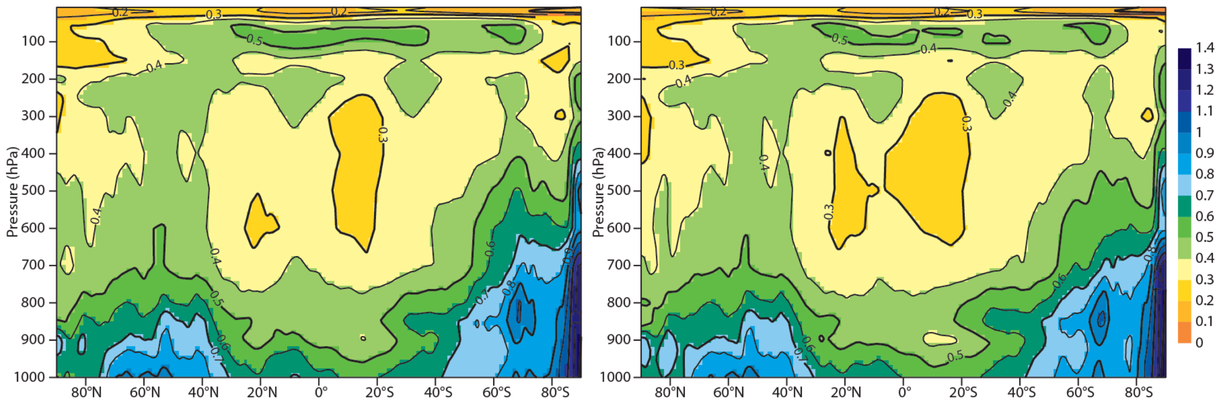

Figure 6. Cross section of root-mean-square of temperature analysis increments for the period 15 June 2010-15 July 2010. Results are for (left panel) the RAD assimilation experiment, (right panel) the PC assimilation experiment.

Further evidence of the similarity between the two systems is found by examining how the assimilation of either IASI PC scores or radiances modifies the fit of the analyses to radiosonde temperature observations. The comparative lack of conventional data in the extratropical Southern Hemisphere (-20ºS-90ºS) means that radiosonde fitting statistics are particularly sensitive to changes in the use of satellite observations and provide a reliable quality metric. Statistics of the fit to radiosondes temperatures for the background (left panels) and analyses (right panels) averaged over the two test periods are shown in figures 7 and 8. In each panel the standard deviations are shown with respect to the BASE experiment (background values are explicitly normalized by the BASE values to improve visualization). Thus, reduced values for either the RAD system (red line) or the PC system (green line) indicate the extent to which the assimilation of the IASI satellite (in either form) improves the fit to radiosonde data compared to the BASE assimilation (black line). The most salient feature of the results plotted in Figures 7 and 8 is that the assimilation of IASI data in either form adds considerable value over the BASE system, with PC scores producing radiosonde data fits that are very close to those generated by the radiance assimilation experiment. Differences between the PC and RAD systems are much smaller than the improvement of both over the BASE, but for some regions of the atmospheric column (in both the troposphere and stratosphere) the assimilation of PC scores produces a slightly better fit to radiosondes than the radiance assimilation. Consistent features are seen in the statistics of the Northern Hemisphere and Tropics (not shown) – although the magnitude of the improvements over the BASE is much smaller with either IASI system, reflecting the comparatively smaller influence of satellite data. Statistics of the fit to other observed parameters such as radiosonde wind and humidity further confirm that the assimilation of 20 IASI PC scores produces results at least comparable with those obtained from the assimilation of 165 radiances – or in some respects marginally better. As with the analysis increments the very small differences between the IASI PC and radiance assimilation systems are most likely attributable to the observation error covariances assumed.

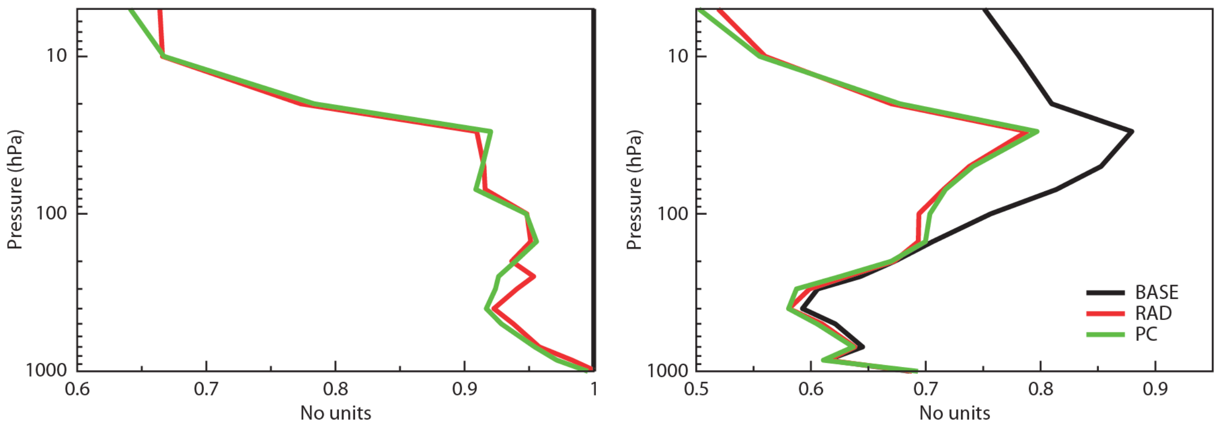

Figure 7. The fit to radiosonde temperature data of the background (left panel) and analysis (right panel) in the Southern Hemisphere. Standard deviations for the period 15 June 2010 - 15 July 2010 are shown for the BASE (black line), RAD (red line), and PC (green line) experiments.

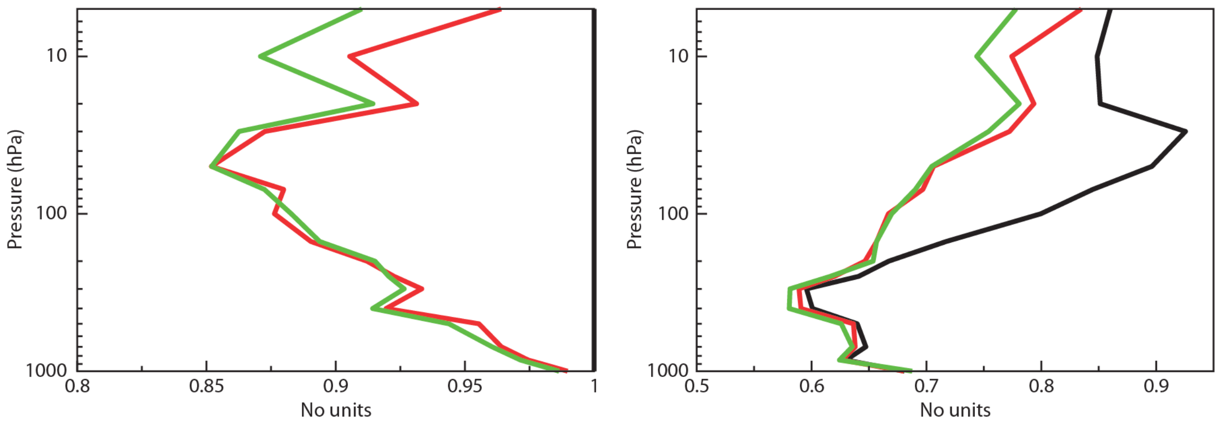

Figure 8. The fit to radiosonde temperature data of the background (left panel) and analysis (right panel) in the Southern Hemisphere. Standard deviations for the period 15 December 2010 - 15 January 2011 are shown for the BASE (black line), RAD (red line), and PC (green line) experiments.

5.4 Computational efficiency of the data assimilation

The primary objective of developing a PC score assimilation system is to improve computational efficiency. Performance tests indicate that the 4D-Var minimization requires 25% less computer resources (elapsed CPU time) when 20 PC scores are used compared to the system that assimilates 165 IASI radiances. This figure represents a significant saving inside the time critical processing path for NWP centres, but could potentially be improved even further. The cost of assimilating a given observation is strongly influenced by the numerous calls to the observation operator and its adjoint (which for PC scores is PC_RTTOV and PC_RTTOV_AD). In this study we have used 50 predictors inside PC_RTTOV, which is the maximum accuracy, but lowest efficiency setting of this tunable parameter. Offline tests have shown that the number of predictors could be reduced from 50 to 35 without significantly affecting the accuracy of the PC score simulation. This would help further improving the computational efficiency of the PC score assimilation system.

5.5 Impact on Forecasts

Forecasts have been run from analyses generated by the BASE, RAD and PC 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 two test periods 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 normalization 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. Forecast scores are tabulated in Tables II and III for the June-July and December-January periods respectively. Results are presented in terms 500hPa geopotential height for the extratropical Northern and Southern Hemispheres (20ºN-90ºN / 20ºS-90ºS), and in terms of the 850 hPa wind vector for the Tropics (20ºS-20ºN). A positive value of the forecast score means that the use of IASI data improves forecast accuracy compared to the BASE and values with a statistical significance of 95% or better are indicated in bold italic.

Tables II and III show that in all regions the inclusion of IASI data has a positive impact on the forecasts accuracy compared to the BASE, with, as expected the largest impact in the Southern Hemisphere. Most importantly, however, the PC score assimilation system is producing forecasts of similar accuracy to those from the radiance assimilation system. In fact, if we consider only the statistically significant figures, we see that the use of PC scores results in a marginally improved forecast skill for some regions and ranges compared to the radiance assimilation. The opposite is seen in the statistics for the extratropical Southern Hemisphere during December-January case when the PC score system exhibits less, albeit still positive, skill. Despite these differences, the overall conclusion from the forecast scores confirms that of the analysis diagnostics – suggesting the PC score assimilation performs at least as well as the radiance system.

Table 2. Normalized root-mean-square error difference for 500 hPa geopotential and 850 hPa wind vector forecasts verified versus the operational analysis. Positive values indicate a positive impact from the inclusion of IASI radiance (RAD) or PC (PC) data. Figures in bold indicate a signal significant at the 95% level or better. Results cover the period from 15 June 2010 to 15 July 2010 for extratropical Northern Hemisphere (N), Tropics (Tr), and extratropical Southern Hemisphere (S).

Region		**Day 1**	**Day 3**	**Day 5**
RAD experiment				
N	z	**0.048**	0.022	**0.037**
Tr	w	**0.016**	**0.034**	0.019
S	z	**0.228**	**0.154**	**0.114**
PC experiment				
N	z	**0.068**	**0.029**	0.028
Tr	w	**0.034**	**0.047**	**0.034**
S	z	**0.188**	**0.155**	**0.134**

Table 3. Normalized root-mean-square error difference for 500 hPa geopotential and 850 hPa wind vector forecasts verified versus the operational analysis. Positive values indicate a positive impact from the inclusion of IASI radiance (RAD) or PC (PC) data. Figures in bold indicate a signal significant at the 95% level or better. Results cover the period from 15 December 2010 to 15 January 2011 for extratropical Northern Hemisphere (N), Tropics (Tr), and extratropical Southern Hemisphere (S).

Region		**Day 1**	**Day 3**	**Day 5**
RAD experiment				
N	z	**0.044**	**0.044**	0.024
Tr	w	**0.012**	**0.019**	**0.024**
S	z	**0.286**	**0.201**	**0.131**
PC experiment				
N	z	**0.065**	**0.059**	**0.044**
Tr	w	**0.034**	**0.033**	**0.025**
S	z	**0.252**	**0.160**	**0.100**

6 Summary and future directions

The operational ECMWF 4D-Var has been adapted to allow the direct assimilation of Principal Component 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. The prototype system presented in this study uses 20 PCs instead of 165 IASI longwave radiances - achieving an 8 fold reduction in data volume and a 25% reduction in the overall cost of assimilation. These figures have been achieved with a rather conservative setting of the tunable accuracy of the PC_RTTOV radiative transfer model and further computational savings could be achieved.

The new scheme has been extensively tested in a baseline environment where conventional observations and AMVs are assimilated, but IASI 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 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. 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.

While the performance of the PC assimilation prototype is very impressive - it is important to highlight its current limitations. Arguably the most important of these is the restriction to fully clear spectra. It is known that completely clear scenes are rather rare - estimates vary, but typically they account for less than 15% of the data measured by an IASI instrument with a 12Km field of view. Thus the next step is to investigate how well the PCs assimilation methodology extends to partly cloudy or fully overcast scenes. While no fundamental obstacle to this is envisaged, the training of Principal Components specifically in cloudy skies will be required.

Another short term priority is to test the IASI PC approach in a full data assimilation system that contains all operational observations (satellite and conventional). While testing in the baseline environment is extremely demanding and we are confident that the conclusions reached in this study are robust - it is known that the background error characteristics in a densely observed system are different. In particular background errors tend to be smaller and spatially localised such that the empirically tuned truncation threshold (20 PCs) and observation error covariance R (including correlations) may have to be revisited to account for this.

Finally, while this study considered only data in the IASI long wave region, it follows a previous investigation in to the use of PCs to represent the IASI shortwave spectrum. A logical future step is to consider the extraction of information from the dedicated IASI water vapour and ozone bands towards the exploitation of the full IASI spectrum.

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 prototype system forward to a stage where it can be considered as an option for the safe and efficient operational exploitation of these crucial instruments.

Acknowledgments

The work of Marco Matricardi has been funded by EUMETSAT, contract No. EUM/CO/07/4600000475/PS. We want to gratefully thank Niels Bormann (ECMWF) for providing the Desroziers and the Hollingsworth/Lönnberg PC score observation error estimates and for the helpful suggestions on the implementation of the PC_RTTOV code in the ECMWF Integrated Forecasting System.

References

Antonelli P, Revercomb HE, Sromovsky LA, Smith WL, Knuteson RO, Tobin DC, Garcia RK, Howell HB, Huang H-L, Best FA. 2004. A principal component noise filter for high spectral resolution infrared measurements. *Journal of Geophysical Research*, Volume **109**, 2004, Doi:1029/2003JD004862.

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

Chalon G, Cayla F, Diebel D. 2001: ‘IASI: An Advanced Sounder for Operational Meteorology’. Proceedings of the 52th Congress of IAF, Tolouse, France.

Collard, A. D., McNally, A. P., Hilton, F. I., Healy, S. B. and Atkinson, N. C. (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 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.

Joiner, J. and Da Silva, A. M. (1998), Efficient methods to assimilate remotely sensed data based on information content. Q.J.R. Meteorol. Soc., 124: 1669–1694. doi: 10.1002/qj.49712454915

Jolliffe 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.

Masiello, G., Serio, C. and Antonelli, P. (2012), Inversion for atmospheric thermodynamical parameters of IASI data in the principal components space. Q.J.R. Meteorol. Soc., 138: 103–117. doi: 10.1002/qj.909

Matricardi M, Chevallier F, Kelly G, Thepaut J-N. 2004. An improved general fast radiative transfer model for the assimilation of radiance observations. *Q. J. R. Meteorol. Soc.,* **130**: 153-173.

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

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

Turner DD, Knuteson RO, Revercomb HE, Lo C, Dedecker RG. 2006. Noise Reduction of Atmospheric Emitted Radiance Interferometer (AERI) Observations Using Principal Component Analysis. *J. Atmos. Oceanic Technol.*, **23**, 1223-1238.