Principal component and reconstructed radiance based assimilation techniques

Marco Matricardi, Tony McNally and Niels Bormann

ECMWF, Shinfield Park, Reading, UK
marco.matricardi@ecmwf.int

ABSTRACT
In this paper we present a discussion of the assimilation techniques based on the use of Principal Component Analysis. Using these techniques we have adapted the ECMWF operational ECMWF 4D-Var system 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. Results suggest that the PC assimilation system (in clear sky only) performs as well as – and in some respects marginally better than – the current ECMWF operational IASI radiance assimilation that also uses radiances peaking above clouds and overcast scenes. This demonstratesthe viability of an alternative route to radiance assimilation for the exploitation of data from high spectral resolution infrared sounders in NWP. We also discuss the handling of clouds in PC space.

1. Introduction

The assimilation oh high resolution radiances measured by the Infrared Atmospheric Sounding Interferometer (IASI) has produced a significant positive impact on forecast quality (Collard and McNally, 2009). The operational use of IASI radiances at ECMWF is currently restricted to a selection of temperature sounding channels in the long-wave and short-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. Currently, 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 because 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. It should be noted that the use of radiances reconstructed from PC scores (Collard *et al*., 2010) provides an alternative methodology for the efficient assimilation of high-resolution infrared sounder spectra. In this article we document the development and the functionality of a global four-dimensional variation (4D-var) assimilation system based on the direct use of PC data. The primary aim of this development is towards an efficient use of the entire measured IASI spectrum that could not be achieved by traditional radiance assimilation.

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

2.1 Principal component scores

PCA is a method that allows the reduction of the dimensionality of a data set by exploiting the interrelations between all the variables contained in the data set. 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 characteristic of the data.

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.

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). Optimal noise filtering can be achieved by noise normalising the spectra because this will ensure that the noise is distributed evenly among all eigenvectors. For interferometer instruments whose radiances have been apodised, it is important that spectra are normalised using the full instrument error covariance matrix because if we use a diagonal error covariance matrix we could lose signals with spectral signatures on the scale of the instrument resolution.

When compared to spectral radiances, the physical interpretation of PC score observations is less intuitive. This is illustrated in Figure 1 which shows the temperature Jacobians for the US Standard Atmosphere for the first ten PC scores of a portion of the IASI spectrum that comprises 165 long-wave channels whose primary sensitivity is to temperature and the surface although they also convey some humidity information. 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).

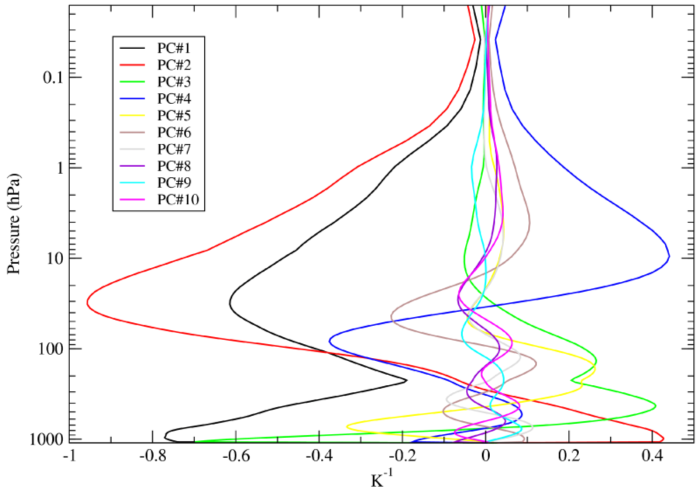

Figure 1. The temperature Jacobian for the first 10 PCs for the US Standard Atmosphere.

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, PC1 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. This is particularly true for PC1 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 PC1 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 PC1 score departures. Although not shown here, the ten PCs in Figure 1 are also sensitive to humidity. It should be stressed that the behaviour and the nature of the PCs is affected by the choice of the channel set used for the PC generation. By choosing a different channel set for the PC generation, Jacobians may become more or less localized, signals from different spectral regions may become more or less separable and an intuitive effect like that of PC1 behaving like a radiance window channel may be lost. Finally, it should be noted that the non-locality of the PC score Jacobians makes impossible to find a PC score observation insensitive to the presence of clouds.

2.2 Reconstructed radiances

If required, the PC scores may be used to reconstruct a new vector of reconstructed radiances:

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

The reconstructed radiances are noise filtered and they generally have channel-correlated errors even in the case when the original radiance errors are not correlated. 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 and crucially

$r_i^{rec} \neq r_i^{obs}$ (i.e. PCA is not a lossless technique when truncating). In fact, any subset of n_r reconstructed radiances will contain the same information present in the *m* PC scores as long as $n_r \geq m$. If we select a subset of $n_r = m$ reconstructed radiances we can then achieve the same reduction in data volume achieved with the direct use of *m*PC scores. However, if $n_r \geq m$ the covariance matrix of the reconstructed radiances is in general not invertible (Masiello et al., 2009). Consequently, the use of a full error covariance matrix in the assimilation of reconstructed radiances can potentially cause numerical problems. It should be stressed that although it should be theoretically possible to find a subset of *m* reconstructed radiances whose covariance matrix can be inverted, this might not be achievable in practice for any choice of *n*. Of course, to avoid ill conditioning one can assimilate a subset of $n_r < m$ reconstructed radiances. A subset of reconstructed radiances should then be selected such that the number of radiances, n_r, is as close as possible to the number of PC scores, *m*, to reduce to a minimum the consequent loss of information.

Arguably, the use of reconstructed radiances is simpler to implement than the direct use of PC scores because NWP centres already know to deal with raw radiances. The use of reconstructed radiances would not require any of the significant technical and scientific investment needed to develop a system to directly assimilate PC scores. The techniques developed for handling clouds in assimilation systems based on raw radiances should be in principle applicable to reconstructed radiances. Although at a theoretical level the assimilation of PC scores or reconstructed radiances can be considered equivalent (if we do everything correctly), the successful introduction of either of these approaches in an operational NWP environment will eventually depend on how well the various elements of the assimilation system can be practically implemented and tuned. For instance, the observation operator used for the simulation of the reconstructed radiances should reproduce the true (i.e. multichannel) nature of each reconstructed radiance channel. This could be done using a conventional forward model but it would be prohibitively expensive. The question then arises whether we should approximate the simulated reconstructed radiances with simple calculations of the corresponding real channels or perhaps use a PC based forward model.

To date, the assimilation of reconstructed radiances has been studied at ECMWF by Collard et al. (2010) and at the Met Office by Hilton and Collard (2009) using AIRS and IASI data respectively. In these studies, reconstructed radiances were used as proxy of raw radiances. A diagonal error covariance matrix was used although an inflated observation noise was utilized to reduce the influence of errors arising from unmodelled inter-channel correlations. No attempt was made to reduce the dimension of the data using a subset of reconstructed radiances. The difference in forecast impacts between experiments using raw radiances and reconstructed radiances was found to be statistically neutral.

3. PC assimilation methodology

3.1 Overall architecture of the assimilation system

For the direct 4D-Var assimilation of PC scores derived from IASI fully clear spectra we use the methodology 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 fixed basis of synthetic eigenvectors used for the training of the PC based observation operator PC_RTTOV (Matricardi 2010) to produce a vector of observed PC scores Y_{OBS}^{PC} . Each vector of observed PC scores has length *n*, but crucially only the first *m* of these are assimilated (where *m*<*n* in ranked order).In truncating the vector of observed PC scores the assimilation is made 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 to compute model equivalents of the *m* PC scores, $Y_B^{PC}(X)$.Ignoring the time integration of the forecast model to the observations, the cost function to be minimized is essentially:

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

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 **R**. 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. It should be noted that within the same framework it should be possible to replace the observed PC scores Y_{OBS}^{PC} with the PC score data generated operationally by data providers as a solution to the dissemination problem. This would require a re-projection from the truncated real data eigenvector basis used by the data providers to the synthetic PC_RTTOV eigenvector basis, but otherwise no major obstacles are foreseen.

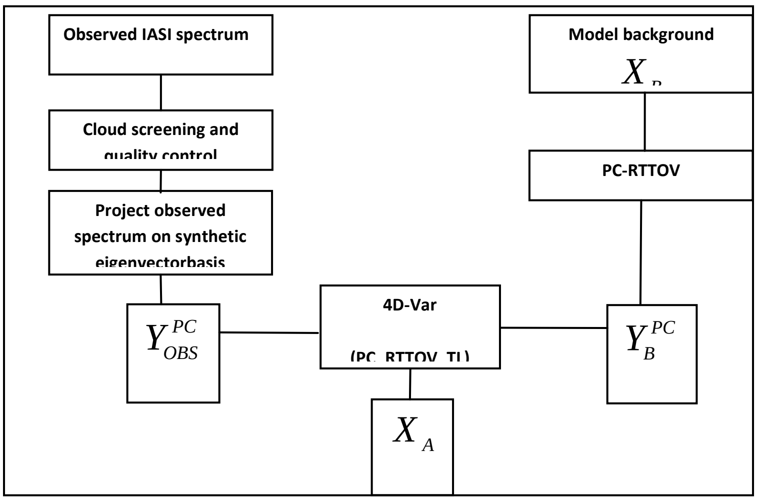

Figure 2.The flow diagram of the direct PC score assimilation.

The specification of the PC score observation error covariance matrix **R** requires some degree of attention. The matrix **R** should describe the combined error of the observations (PC scores) and forward operator (PC_RTTOV). An initial estimate of the diagonal elements of **R** can be obtained 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, we have used the techniques proposed by Hollingsworth and Lönnberg (1986) and Desroziers *et al.* (2005). 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 **R** are expressed as the expectation value

$$\mathbf{R} = E[\mathbf{d}_a \mathbf{d}_b^T] \tag{8}$$

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). It should be noted that both the Hollingsworth/Lönnberg and Desroziers method can be used to diagnose inter-PC score error correlations. Matricardi et al. (2014) have recently demonstrated how the neglect of the off-diagonal terms of the PC error covariance matrix can have a negative impact on the skill of the PC based assimilation and forecast system.

3.2 Cloud detection in radiance space and PC based quality control

The assimilation of PC scores at ECMWF is currently 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 version 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 2, if the principal components are derived from a set of channels which comprises channels sensitive to the surface, PC1 has similar characteristics to an infrared window channel – in particular a heightened sensitivity to the surface emission and the presence of clouds. Large positive departures of the observed PC1 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 PC1 is sufficient to reject most cases of residual cloud contamination. Note that the known accuracy of the model skin temperature over the ocean means that large PC1 departures cannot be due to skin temperature error. For instance, based on the values tabulated in Table I, an error of 1K in skin temperature could only account a PC1 departure of ~10 units.

3.3 Bias correction for PCs

In the ECMWF PC based assimilation system, biases in the PC observations or due to systematic errors in the PC based 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 two to three 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 a small number of PC scores that are slower to stabilize and tend to drift slightly over time. These particular PC scores have the strongest sensitivity to the surface and to the stratosphere and the slow drift of these bias corrections to a large extent mimics the behavior often seen in the corrections computed for window and stratospheric channel radiances. This suggests that same processes that cause drifts in radiance biases (time varying model error and feedback with quality control) could be responsible. While this slow variation of bias corrections is undesirable and certainly warrants further investigation, previous experience with radiances – confirmed by tests with PC scores – suggests that it is not a significant source of degradation in the assimilation.

3.4 Assimilation experiments

To quantify the performance of the PC score assimilation system we have designed a set of 4D-Var assimilation experiments that typically consist of a baseline experiment, a radiance assimilation control experiment and a PC score experiment. The baseline experiment uses all operational observations (satellite and conventional) with the exception of IASI data. The radiance control experiment and the PC score experiments are identical to the baseline but they additionally assimilate IASI radiances and truncated PC scores respectively. We should again stress that the use of PC data is currently restricted to fully clear spectra. This is in contrast to the radiance assimilation system where the use of IASI data extends to channels unaffected by clouds and to fully overcast scenes.

The choice of PC score truncation threshold is typically 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 system is varied up to the full number of available scores. We then look for a number of truncated PC scores beyond which there is no discernible improvement in performance (as measured by the fit of the analysis to other observations). Thus we retain only these truncated PC scores for the main PC assimilation testing. In a similar set of preliminary assimilation experiments we have also found that both the Desroziers and Hollingsworth/Lönnberg refinements of the diagonal observation error for PC scores produces significantly better results than simply using the untuned standard deviation of observed minus background departures. However, the Desroziers error values tend to give an additional marginal improvement over the Hollingsworth/Lönnberg estimates so these have been adopted for the main assimilation testing. For the observation error covariance matrix **R** of the control radiance 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). The use of PCs has been investigated both in the short-wave and long-wave spectrum. Recently we have focused on maximising the spectral information of IASI in clear sky, assimilating 50 PCs generated from selected radiances in the IASI long-wave temperature band augmented with channels from the water vapour and ozone spectral bands (a total of 305 channels) using a full PC error covariance matrix. The current status (for details see Matricardi and McNally, 2013[a], Matricardi and McNally, 2013[b]; Matricardi and McNally, 2014) is that the PC assimilation system (in clear sky only) performs as well as – and in some respects slightly better than – the current ECMWF operational IASI radiance assimilation that also uses radiances peaking above clouds and overcast scenes. In addition, performance tests indicate that the use of PC scores in the 4D-Var minimization requires ~25% less computer resources (elapsed CPU time) compared to a system that assimilates the full number of equivalent radiances. This figure represents a significant saving inside the time critical processing path for NWP centres, but could potentially be improved even further by better tuning the computational efficiency of the PC based fast model simulations.

4. Handling of clouds in PC space

As discussed in the previous section, thus far the assimilation of PC data has only been performed in cloud free conditions. The restriction of using only infrared sounding data that can be predetermined as clear represents a major under-exploitation of very high cost instruments such as the IASI. Estimates of cloud cover vary in the literature, but an instrument with a footprint of order 10Km will typically only yield between 10 − 30% completely clear soundings. Furthermore, using radiances only in clear-sky has the potential to bias the assimilation system towards particular synoptic or climatological regions (e.g. areas with low humidity) and there is evidence to suggest that cloudy areas are meteorologically sensitive (McNally, 2002) such that constraining analysis errors in these regions (with observations) is important to limit forecast error growth.

Within the context of radiance assimilation a number of significant steps towards a better utilization of high spectral resolution infrared sounders have been taken in recent years. Cloud detection schemes have been improved beyond finding completely clear locations (e.g. English et al., 1999) towards identifying clear channels (i.e. those channels above the cloud) in potentially cloud affected locations (e.g. McNally and Watts, 2003). However, the broad vertical extent of radiance weighting functions – with significant sensitivity at altitudes far above and below their peak – results in radiance data use above clouds located in the mid to upper troposphere still being highly restricted.

Acknowledging that the observed radiance spectra contain potentially useful information on clouds – a number of approaches have been developed to explicitly treat cloud parameters simultaneously with other variables such as temperature and humidity. These range from the treatment of weakly affected radiances in a 1D-Var pre-processor and passing cloud parameters to the main assimilation step (e.g.

Pavelin et al., 2008) to the direct assimilation of homogeneous fully overcast spectra in a 4D-Var system (McNally, 2009).While these two special cases of cloud (weak and homogeneous overcast) yield very useful information – as yet there is no operational treatment infrared data in all cloudy conditions along the lines of the microwave all-sky generalised framework.

4.1 The construction and assimilation of PC scores above cloud

In this section we assess to what extent the approaches that have been developed for handling clouds in radiance assimilation can be adopted in PC space. To gauge the likely benefits (in terms of improved forecast skill) a number of radiance assimilation experiments have been run exercising the various options (currently available at ECMWF) for handling clouds in IASI spectra. Specifically we have measured the benefits with respect to a no-IASI baseline of the following: 1) IASI in fully clear scenes only, 2) IASI in fully clear scenes plus clear channels above clouds, 3) IASI in fully clear scenes plus clear channels above clouds plus full overcast spectra. The results are presented in Figure 3 for the Tropics and Southern Hemisphere (regions where the signals are largest).

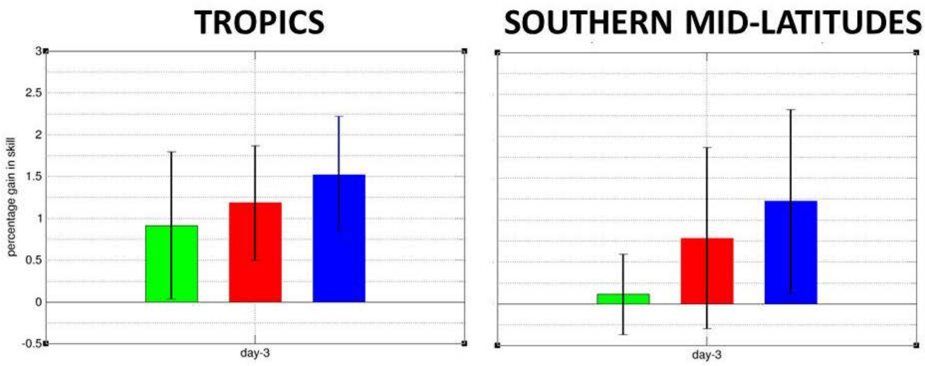

Figure 3: The forecast impact associated with different approaches to handling clouds in IASI radiance data. The vertical bars show the percentage reduction of average forecast errors (for 500hPa geopotential height) with respect to a no IASI baseline when: only fully clear data are used (green), additional use is made of clear channels above cloud (red) and additional use is made of clear channels over cloud plus fully overcast scenes.

It can be seen that in the Tropics there is a significant impact of assimilating just the fully clear spectra and a modest stepwise improvement in skill when varying degrees of additional cloudy data are retained (either using clear channels above cloud or overcast scenes). However, in the Southern Mid-latitudes the benefit of assimilating only fully clear IASI spectra (compared to no IASI) is very small and much larger gains are made if additional cloudy data can be retained. These results demonstrate that the importance of exploiting cloudy data is much larger in the mid-latitudes where fully clear scenes are rare compared to the Tropics.

To consider a potential PC analogue of the use of clear channels above clouds we must examine the Jacobians of the PC score observations (see Figure 1 for a selection of the leading eigenvectors). For any given cloud condition there may be channels with Jacobians that peak sufficiently high in the atmosphere that we may reasonably assume that they are unaffected by the cloud. Unfortunately –

while it is possible to identify PC scores that are predominantly sensitive to the stratosphere or predominantly sensitive to the troposphere – the Jacobians are non-local with multiple peaks throughout the troposphere and stratosphere. For a given cloud condition it is thus impossible to find a PC score observation insensitive to the presence of the cloud.

The non-locality of the PC score Jacobians is a direct consequence of the way the eigenvectors and PC score projections have been computed. If a particular PC score has contributions from a wide variety of IASI channels (e.g. from the stratosphere and the troposphere) it will – to a varying degree – retain the sensitivities of those channels. In this case a PC score which is predominantly sensitive to the stratosphere will nonetheless contain contributions from window channels that are highly sensitive to the presence of clouds. However, we can construct a PC score that is only sensitive to the upper atmosphere by constructing it from only IASI channels that peak in the upper atmosphere. Furthermore, for a given cloud condition we can construct a set of PC scores that are insensitive to the cloud by only allowing contributions from IASI channels that are insensitive to the presence of the cloud. There are two options to achieve this in practice:

Option 1: We predetermine a set IASI channels that are always unaffected by cloud as the eigenvector training basis. Of course this channel basis will be restricted to the uppermost stratospheric sounding channels of IASI and in the case of low stratiform cloud this may still represent a gross under-exploitation of the available IASI data. However, it has the technical advantage of allowing us to pre-compute a fixed set of eigenvectors on to which PC scores can be projected when cloud is detected.

Option 2: We run the radiance based cloud detection scheme as a pre-processor to identify what subset of channels can be used in a particular cloud affected scene. Using the climatological radiance training set for just these channels we re-compute (on the fly) a dedicated set of eigenvectors and project the measured spectrum to produce cloud free PC scores for this scene. While this optimises the use of IASI clear-sky information for each cloudy scene – it is a technically demanding option as it requires running the full radiance cloud detection scheme (to identify clear channels) and the calculation of different eigenvectors for each IASI pixel.

4.2 The assimilation of PC data in overcast scenes

To assimilate PC scores in overcast situations we require a forward model that can compute cloud affected PC scores. The PC radiative transfer model PC_CLD_RTTOV (Matricardi and McNally, 2011) is capable of simulating cloud affected PC scores for any given input cloud condition. Of course the eigenvector basis upon which the model is trained will be completely different for clear-sky and cloudy-sky calculations. This has implications when deciding upon an appropriate truncation for assimilation. For clear-sky spectra the leading eigenvectors describe the largest sources of variance – namely the surface emission, humidity and the stratosphere. However, when the eigenvectors are required to efficiently represent the variability of cloudy spectra the leading (highest ranking) eigenvectors describe the cloud signal itself – and the comparatively smaller sources of variance from atmospheric temperature and humidity reside in a range of lower order PC scores. This can be seen in the Jacobians calculated by PC_CLD_RTTOV for a cloud free scene, using a clear eigenvector basis and a fully cloudy eigenvector basis (Figure 4). Thus great care must be taken on choosing an appropriate truncation for the assimilation so as not to exclude lower order PC scores that contain useful atmospheric information.

In the assimilation of overcast radiances an estimate of the cloud conditions is obtained before the main assimilation to a) determine if the pixel is overcast and b) provide a linearization point for the subsequent forward calculations. For convenience there is no reason why this step should not still be done in radiance space for the overcast PC score assimilation. For a single level homogeneous overcast cloud the Jacobians of the cloudy PC scores terminate at the cloud top in the same way as the equivalent radiance Jacobian**.** This termination produces high resolution temperature information at the cloud top that in principle can be exploited by the assimilation scheme. This has been demonstrated in the radiance context, but not yet with PC scores. A technical consideration is that the current version of PC_CLD_RTTOV does not allow the user to input a simple single layer cloud fraction and amount. This functionality would have to be built in to PC_CLD_RTTOV or the input cloud profile variables could be constructed to act as a proxy for the simple single layer cloud (i.e. an optically opaque cloud must be formed with the required cloud top and fraction). Other than this no significant obstacles are foreseen to reproducing the functionality and benefits of overcast radiance assimilation in PC space.

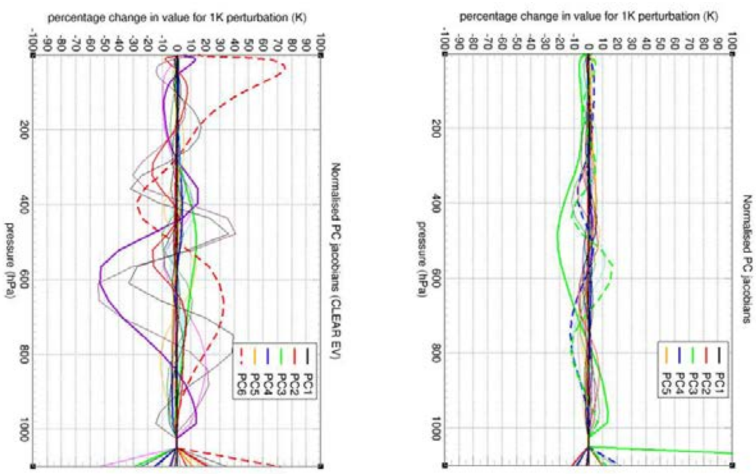

Figure 4: The temperature Jacobians for the six highest ranking PC scores computed by PC-RTTOV trained on: clear-sky spectra (left panel) and on all sky spectra (right panel). In both cases zero cloud is fed as input to the RT calculation.

4.3 The all-sky assimilation of cloud affected PC data

The assimilation of radiance data in any cloud conditions (the so called all-sky approach) is arguably the most comprehensive exploitation of infrared data, but is equally the most challenging and complex. For any given model state we compute fully cloudy radiance spectra and compare these to observed cloudy radiances. Differences between the observed and simulated radiances are mapped back through a chain of adjoint operators (including those of the model physical parameterizations related to cloud) to adjustments of the model state vector. This approach is successfully implemented at ECMWF for the assimilation of microwave radiance data, but as yet the application to infrared data is only at the prototype stage. Challenges related to the extreme nonlinearity of the observation operators (and their adjoints) and the difficulty in reconciling the spatial scales of the NWP model and the observations could mean that an operational implementation is still several years away. Thus we consider it premature to comment in this report on the likely success of the all-sky approach in PC space. However, it has already been stated that a based PC based fully cloudy fast model exists as such and if a solution in radiance space can be found – a PC implementation would not be far behind.

5. Summary

In this paper we have presented a discussion of NWP assimilation techniques based on the use of Principal Component Analysis and we have documented the development and the functionality of a global four-dimensional variation (4D-var) assimilation system based on the direct use of principal component data derived from IASI spectra. 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 direct assimilation of principal component data typically allows an 8 fold reduction in data volume and a 25% reduction in the overall cost of assimilation. The new principal component scheme has been extensively tested in a full observing system where IASI observations used either in the form of principal components or radiances. Results suggest that the quality of the analyses produced by the assimilation of principal components is almost identical to that obtained when 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.The results obtained from the direct assimilation of IASI principal component data 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.

While the performance of the principal component assimilation system is impressive it is important to highlight that the use of PC data is currently restricted to fully clear spectra. This is an important limitation to the use of the PC system in an operational environment. We have suggested that it is reasonable to conclude that the steps that have been taken to handle clouds in infrared radiances can indeed be reproduced in PC space. However, to date little progress has been made in implementing the options discussed in this paper. This is due to the fact that dealing with clouds in PC space is technically demanding within the context of a global assimilation scheme, requiring the recalculation and storage of eigenvector projections and retaining a facility to identify clouds in radiance space before the PC analysis. While these technical demands are not insurmountable in the longer term, they have been considered beyond the demands of the research projects allocated to the study of PC assimilation in NWP. Nonetheless, the results of impact radiance assimilation trials with different cloud treatments demonstrate the importance of handling cloud affected scenes (particularly in the extra-tropical mid-latitudes) and could arguably preclude an operational implementation of PC assimilation until a solution is found. A pragmatic hybrid approach is conceivable where PC scores are assimilated in completely clear sky and, if cloud is detected, the assimilation system could be reverted to radiance assimilation. While this is far from elegant it is technically feasible.

In the paper we have also argued that the use of radiances reconstructed from principal components is perhaps simpler to implement than the direct use of principal component data because NWP centres already know to deal with raw radiances. The use of reconstructed radiances would not require any of the significant technical and scientific investment needed to develop a system to directly assimilate PC scores. The techniques developed for handling clouds in assimilation systems based on raw radiances should be in principle applicable to reconstructed radiances. Although at a theoretical level the assimilation of PC scores or reconstructed radiances can be considered equivalent (if we do everything correctly), the successful introduction of either of these approaches in an operational NWP environment will eventually depend on how well the various elements of the assimilation system can be practically implemented and tuned. Work is now needed to study the assimilation of reconstructed radiances and see whether this system can be considered as an alternative option for the safe and efficient exploitation of high spectral resolution data.

References

Bormann, N. and P. Bauer, 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*., **649**, 1036–1050.

Collard, A.D. and A.P. McNally, 2009: Assimilation of IASI radiances at ECMWF. *Q. J. R. Meteorol. Soc*., **135**, 1044–1058.

Collard, A.D., A.P. McNally, F.I. Hilton, S.B. Healy and N.C. Atkinson, 2010: The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Q. J. R. Meteorol. Soc*., **653**, 2038–2050.

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., L. Berre, B. Chapnik and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.,* **131**, 3385–3386.

English, S.J., J.R. Eyre and J.A. Smith,1999: A cloud-detection scheme for use with satellite sounding radiances in the context of data assimilation for numerical weather prediction. *Q. J. R. Meteorol. Soc*., **125**, 2359–2378.

Joliffe, I.T., 2002:*Principal Component Analysis*, Springer, New York.

Hilton, F. and Collard, A., 2009: Recommendations for the use of principal component-compressed observations from infrared hyperspectral sounders. *Met Office Forecasting Research & Development Technical Report No.536*.

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

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

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

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

Matricardi, M. and T. McNally, 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. and T. McNally, 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. and T. McNally, 2013[a]: 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 A.P. McNally, 2013[b]: The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc*., **679**, 573–582.

Matricardi, M. and T. McNally, 2014: The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var and the generation of RTTOV and PC_RTTOV new regression coefficients for AIRS, IASI, and IASI-NG. *EUMETSAT Contract No. EUM/CO/07/460000101/PS*.

McNally, A.P.,2002: A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. *Q. J. R. Meteorol. Soc*., **128**, 2551–2556.

McNally, A.P. and P.D. Watts, 2003: A cloud detection algorithm for high-spectral-resolution infrared sounders. *Q. J. R. Meteorol. Soc*., **129**, 3144–3423.

McNally, A.P., 2009: The direct assimilation of cloud-affected satellite infrared radiances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc*., **135**, 1214–1229.

Pavelin, E.G., S.J. English, and J.R. Eyre, 2008: The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction. *Q. J. R. Meteorol. Soc*., **134**, 737–749.