Recommendations for the use of principal component-compressed observations from infrared hyperspectral sounders

Forecasting Research & Development
Technical Report No. 536

Fiona Hilton and Andrew Collard

Contents

Recommendations for the use of principal component-compressed observations from infrared hyperspectral sounders

Fiona Hilton and Andrew Collard

This report investigates the use of principal component (PC) compression for hyperspectral sounding data from the Atmospheric InfraRed Sounder (AIRS) and the Infrared Atmospheric Sounding Interferometer (IASI) in the context of assimilation in a Numerical Weather Prediction (NWP) model. PC compression is a way of reducing correlated spectra of many thousands of radiances into a few hundred independent pieces of information known as "PC scores", and as such may provide an efficient way to represent the spectra for assimilation. Two methods of using PC-compressed spectra are evaluated—the direct assimilation of PC scores and assimilation of radiances reconstructed from the PC scores. Reconstructed radiances retain some of the theoretical benefits of the PC scores themselves while reducing some of the practical difficulties in their use.

1 Principal component compression of hyperspectral sounders and the reconstruction of radiances

Sounding instruments such as the Atmospheric InfraRed Sounder (AIRS) and the Infrared Atmospheric Sounding Interferometer (IASI) have thousands of channels measuring in the infrared at spectral resolutions of a fraction of a wavenumber, in contrast to older instrumentation such as the High-resolution Infrared Radiation Sounder (HIRS) where there are tens of channels with a broad spectral response of 10-30cm^{-1}. This poses some problems for data assimilation in Numerical Weather Prediction (NWP) models, as with current computational technology, it is not possible to process all of the huge volume of data arriving in near-real time. Furthermore, the information content of each spectrum is much lower than the number of individual channels (e.g. Huang et al., 1992) and so assimilating each channel across the whole spectrum is inefficient. Correlations between the measurements in different channels, arising mainly from similarities in their atmospheric sensitivity and to a lesser extent from the instrument design (for example through the apodisation process applied to IASI), mean that there is a large degree of redundancy in the spectrum.

Current operational NWP assimilation schemes make use of only a small fraction of the channels available on the instrument - typically less than 200 channels are assimilated for each observation (e.g. Hilton et al., 2009a; Collard and McNally, 2009, Hilton et al., 2009b). Although this reduces the computational cost of processing the data, the channels must be carefully chosen to contain as much information as possible about the atmospheric profile. Collard (2007) describes one such method to choose 300 channels suitable for use in NWP. The channels chosen may be statistically representative of those which will add the most value to an assimilation system, but their choice is subject to the assumptions made during the process used to select them.

One proposed method for reducing the spectrum into a smaller number of measurements while retaining the full information content is principal component (PC) compression. PCs are also known as Empirical Orthogonal Functions (EOFs). Principal component analysis involves transforming the data from the highly-correlated spectral domain to one described by a set of (orthogonal) eigenvectors. The matrix of eigenvectors, **L**, and the associated eigenvalue matrix **Λ**, are calculated from the covariance matrix, **C**, of noise-normalised spectra held in matrix **X**, of size *NxM* where *N* is the number of channels and *M* is the number of observations

$$\mathbf{C} = \frac{1}{M}\mathbf{X}\mathbf{X}^T = \mathbf{L}\mathbf{\Lambda}\mathbf{L}^T \qquad (1)$$

The eigenvectors are known as PCs, and the weights of each PC within an observed spectrum are known as “PC scores”. The theory and process of constructing principal components from the covariance matrix of noise-normalised observations is described in Antonelli et al. (2004) and Turner et al. (2006) and will not be considered here further.

PC compression is a lossy compression technique which uses a subset of the PCs to represent the spectrum. The number of PCs is equal to the number of channels in the spectrum, but the atmospheric information is typically contained within the first few hundred vectors only (e.g. Turner et al., 2006; Goldberg et al., 2003). In other words, the atmospheric variability in the dataset can be represented in a smaller number of pieces of uncorrelated data (the PC scores) than by using the full spectrum which has a high level of correlation between channels.

The spectral information can be reconstructed from the PC scores using the matrix of eigenvectors, **L**. The uncorrelated noise in the observed spectrum is mapped into higher-order PCs which can be discarded, increasing the signal to noise ratio of the reconstructed spectrum. In this case only the leading eigenvectors which should contain the bulk of the atmospheric signal are retained and the dimensionality of **L** is reduced. The reconstructed radiance spectrum, $\tilde{\mathbf{y}}$ is derived from the full spectrum, **y**, via the following equation:

$$\tilde{\mathbf{y}} = \mathbf{L}_P \mathbf{L}_P^{\mathbf{T}} \mathbf{y} \qquad (2)$$

where P is the number of leading eigenvectors retained. To reconstruct only a subset of channels the following equation applies:

$$\tilde{\mathbf{y}}_N = \mathbf{L}_{N,P} \mathbf{L}_P^{\mathbf{T}} \mathbf{y} \qquad (3)$$

where N is the number of required channels and $\mathbf{L}_{N,P}$ contains only the eigenvector score contributions for the required channels. It is common to monitor the PC compression by determining how faithfully the original radiance spectrum can be reconstructed from the retained PC scores. A reconstruction score (e.g. Goldberg et al., 2003) usually measures whether the spectrum can be reconstructed to within instrument noise, as it is only random instrument noise that we hope to suppress with PC compression.

The use of PC-compressed data in NWP may thus allow the use of more spectral information, but there are other reasons for considering the use of these data in an operational context. This method of data compression has been proposed for dissemination of hyperspectral sounder data in near-real time. The huge bandwidth required for transmission of the full spectra is currently expensive, and although it is possible that costs will decrease in the long term, we should be ready to adopt cheaper methods of dissemination if data quality is not degraded, particularly in the context of hyperspectral data from geostationary sounders. The dissemination of PC-compressed radiances is already planned for 2010 for the EUMETSAT Advanced Retransmission Service (EARS) which will transmit locally received observations to NWP centres with increased timeliness (EUMETSAT, 2009).

A potential problem with the data is that the PCs are determined from statistics, i.e. a climatology of either real or simulated hyperspectral data is used as the training set, depending on the application. If the data used to form the statistics do not contain examples of a particular type of atmospheric structure, then this structure will not be represented in the leading PCs and it would instead be mapped into the lower-order PCs and discarded as “noise”. It is possible that rare but important structures not well represented in NWP models may also not be available in the PC-compressed data, potentially reducing its usefulness in certain meteorological settings.

The advantage of using simulated data for training the PCs is that one can ensure that the signature of particular atmospheric structures is present in the training set. Simulated data, however, may not be fully representative of reality, as the simulations are limited by the accuracy of the radiative transfer model and the range of atmospheres used to construct the training spectra. This is particularly true where cloudy radiances are being used. Atmospheric features whose signatures are present in real spectra but which are not modelled in the simulated data may become mapped into the lower-order PC scores and discarded. For this reason, the use of an extensive quantity of real data is usually preferred for the efficient distribution and storage of observations. Experience with eigenvectors calculated at the Met Office from a representative range of atmospheric profiles (Chevallier, 2001) forward-modelled using RTTOV-8 (Saunders et al., 2005) proved not to allow reconstruction of real IASI spectra to within instrument noise and so real spectra were used for training PC compression for data storage. A further advantage of real data is that the training set can be improved upon easily by adding spectra which are not well represented by the PCs, as they are found.

2 Proposed methods for assimilation of PC-compressed data

PC-compressed data could be assimilated into NWP models in one of two forms: either via direct assimilation of PC scores or through reconstruction of the radiance spectra followed by channel selection. Until recently, the former approach was not possible because a fast accurate radiative transfer model capable of simulating PC scores is required to use the scores directly. Recent advances mean that there are now several radiative transfer models capable of forward-modelling PC scores from hyperspectral sounders to a degree of accuracy comparable to that of conventional radiative transfer models (e.g. Liu et al, 2005; Havemann et al., 2009). Assimilation of PCs will require a number of new strategies for practical implementation, which are discussed below.

The assimilation of reconstructed channel radiances is in many ways more straightforward, as the reconstructed radiances (RRs) can be treated in the same way as raw radiances in an assimilation scheme. It may, however, be desirable to modify the scheme slightly for the use of RRs, for example, in terms of channel selection or observation errors. Some experiments in the assimilation of reconstructed IASI radiances have been performed at the Met Office, and the results of these trials are presented below, with results from trials of reconstructed AIRS radiances at ECMWF.

2.1 Direct assimilation of PC scores

One of the main disadvantages of PC compression is that the PC scores are not simple functions of the atmospheric state and spectral response, so monitoring and quality control of the observations is less intuitive in PC space. We are not yet able to treat each observation from a hyperspectral satellite instrument equally during assimilation. This may be because the radiative transfer is not yet good enough in all situations (e.g. in modelling solar contributions) or because NWP models are not yet capable of providing sufficiently accurate inputs (e.g. land surface emissivity and temperature), or both (e.g. cloudy atmospheres). In order to make as much use of each observation as possible, we exclude parts of the spectrum that are affected by these factors, for example excluding channels that are affected by the surface over land, or those affected by cloud.

Figures 1 and 2a and b, show the ten leading PCs of the IASI spectrum and their temperature and humidity Jacobians. The PC Jacobians each have contributions from both temperature and humidity (as well as other trace gas species not shown) and are extremely non-localised in the vertical. Although many of the IASI channels themselves are to some extent sensitive to both temperature and humidity, the sensitivity does at least tend to be localised and each channel can be quality controlled separately. Past experience at the Met Office has shown that AIRS channels with highly non-localised temperature jacobians have caused problems with the analysis in the past and were removed from the assimilation system when this was discovered. It is possible that one could construct a set of PCs which separates out the atmospheric contributions to a greater extent, for example by selection of subsets of the full spectrum in the training set, but the authors are not aware of any research that has attempted to resolve this issue.

Although the assimilation of PC scores has been demonstrated to a reasonable standard in standalone retrieval schemes (e.g. Liu et al., 2007), it has not yet been evaluated in a systematic way on a global dataset to understand any differences in behaviour against schemes which use the data in the form of channel radiances. There has been no testing in an operational context and under the more rigorous constraints of 4D-variational analysis. The assimilation of cloudy data into an NWP model is, in particular, likely to require significant further investigation. Cloud causes a highly non-linear response in the measured radiance, and the use of cloudy radiances in NWP is an area of active research. The inability to separate out parts of the observation affected by cloud will mean that PCs can presently only be used for clear scenes, which is at odds with current strategies to increase the use of hyperspectral data in cloudy regions.

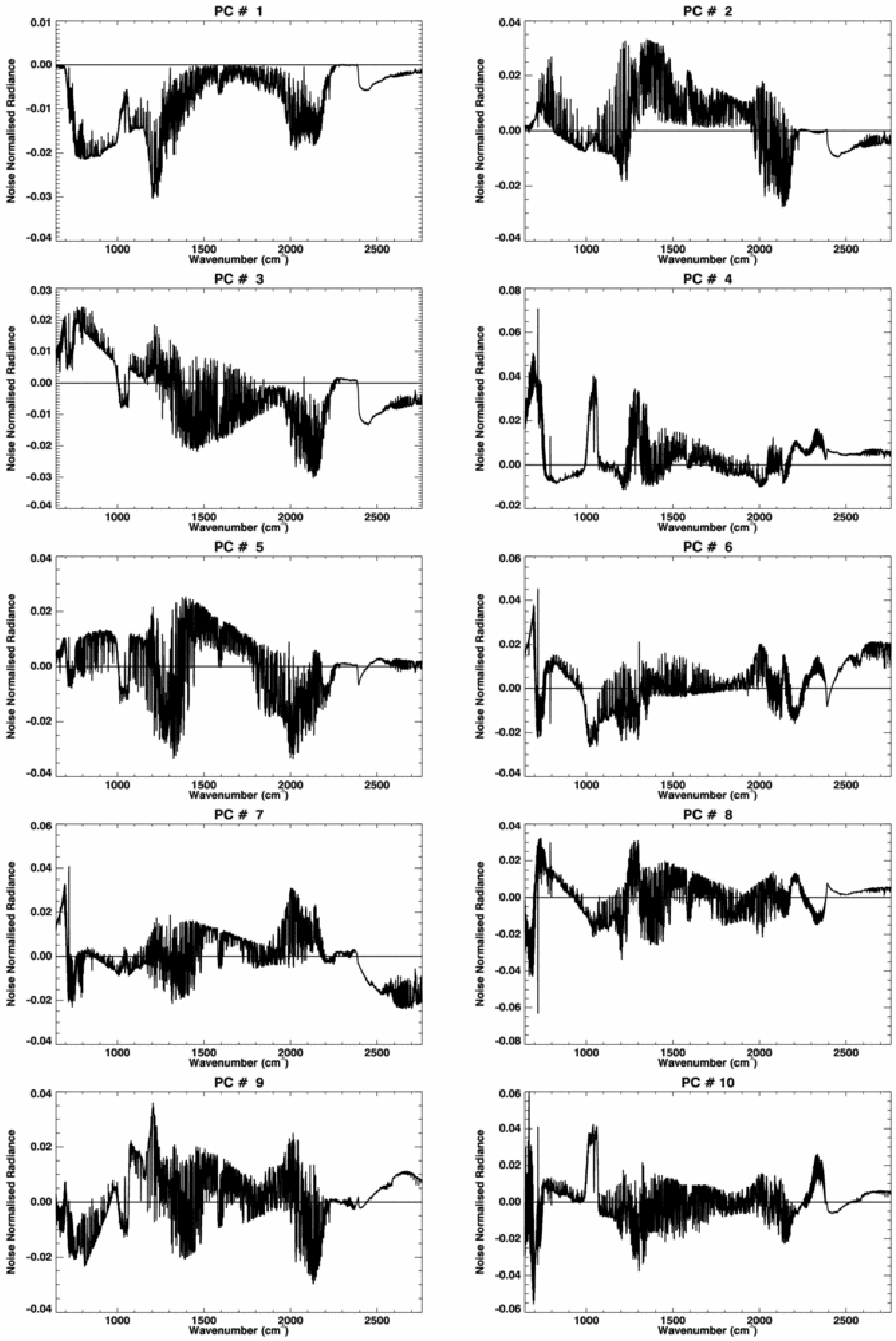

Figure 1: The 10 leading principal components of the IASI spectrum generated from 57383 noise-normalised IASI observations from the period between June and November 2007

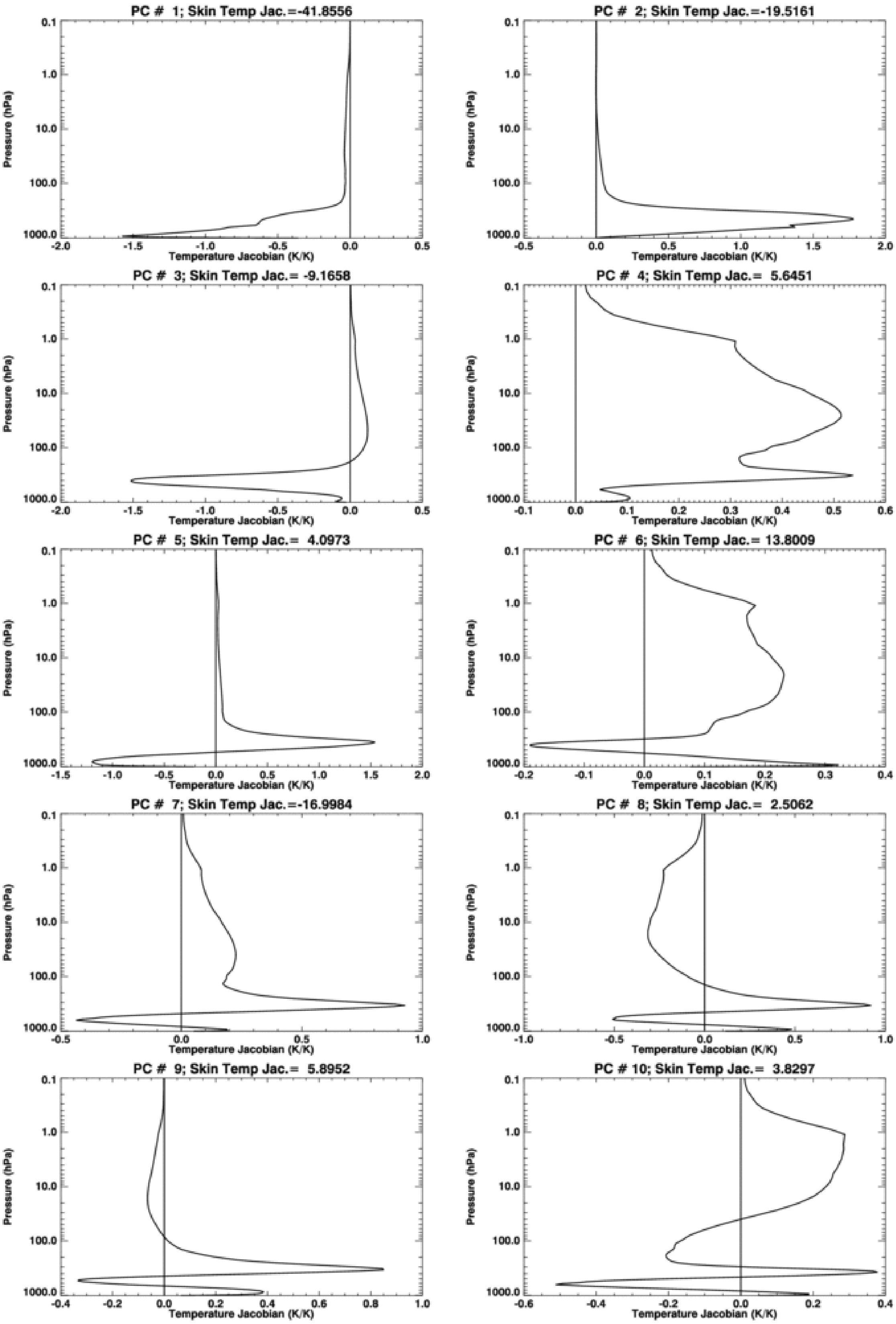

Figure 2a: Temperature Jacobians of the 10 leading principal components of the IASI spectrum shown in Figure 1. Unlike Jacobians for IASI channels, beyond the first two PCs, the Jacobians exhibit sensitivity to the temperature throughout the full vertical extent of the profile. The text at the top of each plot gives the value of the skin temperature jacobian for each PC.

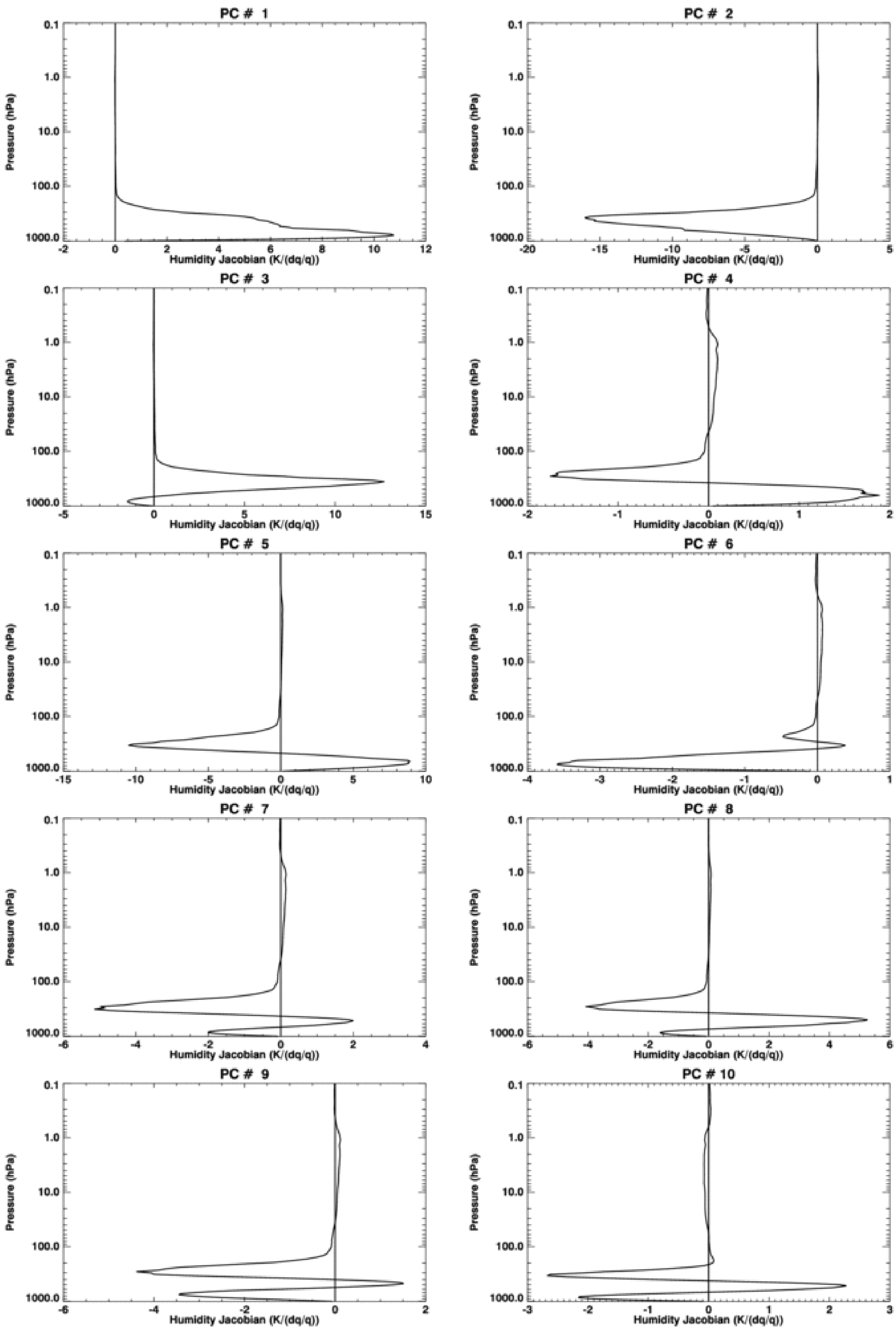

Figure 2b: Humidity Jacobians of the 10 leading principal components of the IASI spectrum as shown in Figure 1. As well as showing sensitivity to the full temperature profile (Figure 2a), the PCs all have sensitivity to the humidity structure of the atmosphere as well.

One of the main reasons for evaluating the use of PC-compressed hyperspectral data is that we expect transmission of PC scores in the near future. It would therefore seem efficient to assimilate the PC scores directly, rather than having to convert back to radiances then to brightness temperatures before assimilation. While, for data transmission and storage, real observations should be used to form the training data set, for the construction of training sets for PC radiative transfer, simulated observations are generally considered preferable.

Fast radiative transfer models require the construction of regression coefficients, derived from calculations by line-by-line models, which map atmospheric predictors into observation space. It is a common misunderstanding that the same PCs would be used for both dissemination and forward modelling—even if one wanted to do this there is some concern that there is insufficient accuracy in the use of NWP model profiles to represent the true atmosphere seen by the real observations to allow their use for the regression calculation (M. Matricardi, personal communication). Even if the NWP models are sufficiently accurate, there is no requirement for real atmospheric profiles to be used as long as the training set is fully representative, and in fact the regression relationship may be clearer with simulated data, In any case, there are practical reasons why the data distribution PCs would not be used for radiative transfer modelling—they may be periodically reviewed to include more spectra in the training set, which would require a recalculation of radiative transfer coefficients each time.

Changing from one PC set to another is not necessarily a problem, as it can be effected by a matrix multiplication which will project the PCs used for data compression into those used for radiative transfer. However, it is not known what effect the truncation of the first set of PCs and subsequent projection into a second truncated set will have. It is possible that, rarely, some feature in the data may be obscured or exaggerated by the process.

2.2 Assimilation of reconstructed radiances

There are a number of advantages in using RRs over PC scores. RRs can be used as a proxy for the original raw radiances, and quality control can be carried out in much the same way as for raw radiances, with channels sensitive to the surface, clouds, etc screened out as required. Monitoring of observation-model biases can be carried out in brightness temperature or radiance space and can also be compared with monitoring of the raw radiances themselves. RRs can be forward modelled as raw radiances using conventional radiative transfer models, and although there is some small additional forward model error introduced because PC compression will map some atmospheric

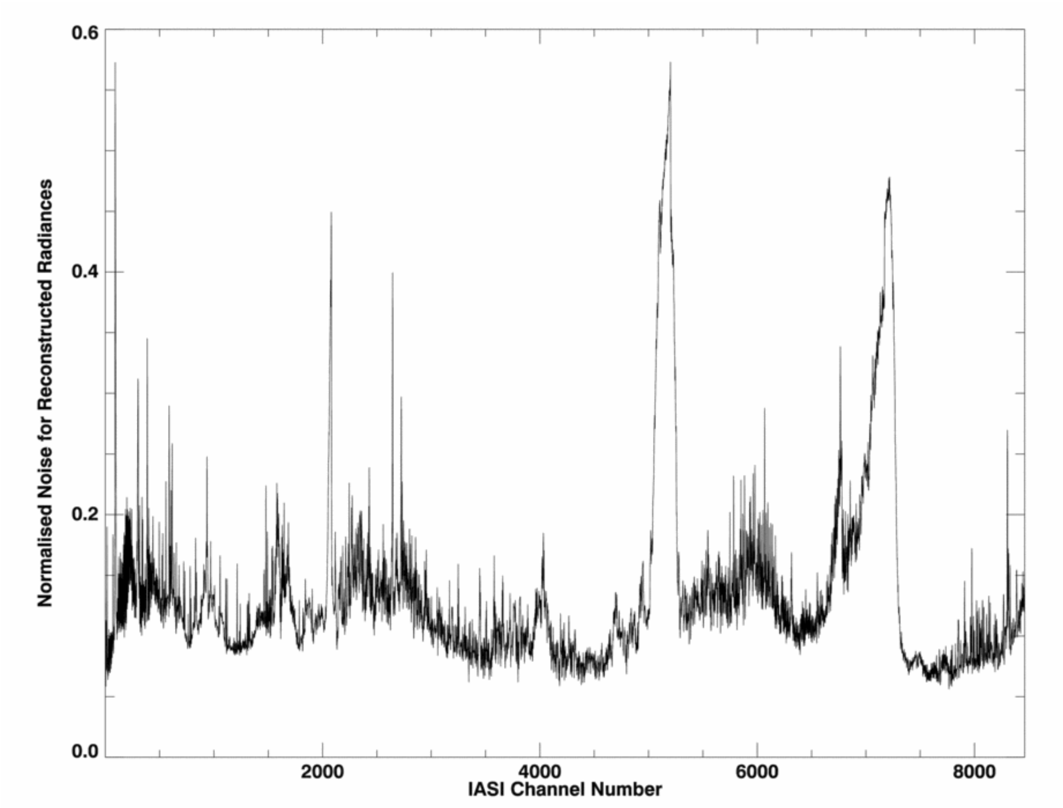

Figure 3: Diagonal of the normalised covariance of instrument noise for the IASI spectrum following reconstruction from 200 PC scores derived from real IASI data as in equation 2. The diagonal of the normalised noise covariance matrix for the original data is the identity matrix.

signal into the "noise" components which may be present in the forward-modelled NWP columns, this is expected to be small if the PCs are well chosen and representative.

The other main advantage of RRs is that they have the properties of a truncated set of PC scores, in that information from the whole spectrum is combined into a smaller number of pieces of information, but the jacobians are as localised as those of raw radiances. RRs have a theoretical advantage over raw radiances because, for each reconstructed channel, only the information which is correlated with other channels—mostly made up of the atmospheric signal—is retained, while the part of the measurement uncorrelated with other channels—mostly instrument noise—is reduced, and so the signal to noise ratio is increased. However, the process does increase inter-channel correlation as the signal in the RRs is derived from the full spectrum. Figure 3 illustrates the reduction of the variance of the normalised instrument noise for IASI RRs (i.e. the diagonal of the error covariance matrix for noise-normalised observations) while Figure 4 displays the error correlation matrix for the ECMWF IASI channel selection (Collard and McNally, 2009) for noise-normalised reconstructed radiances.

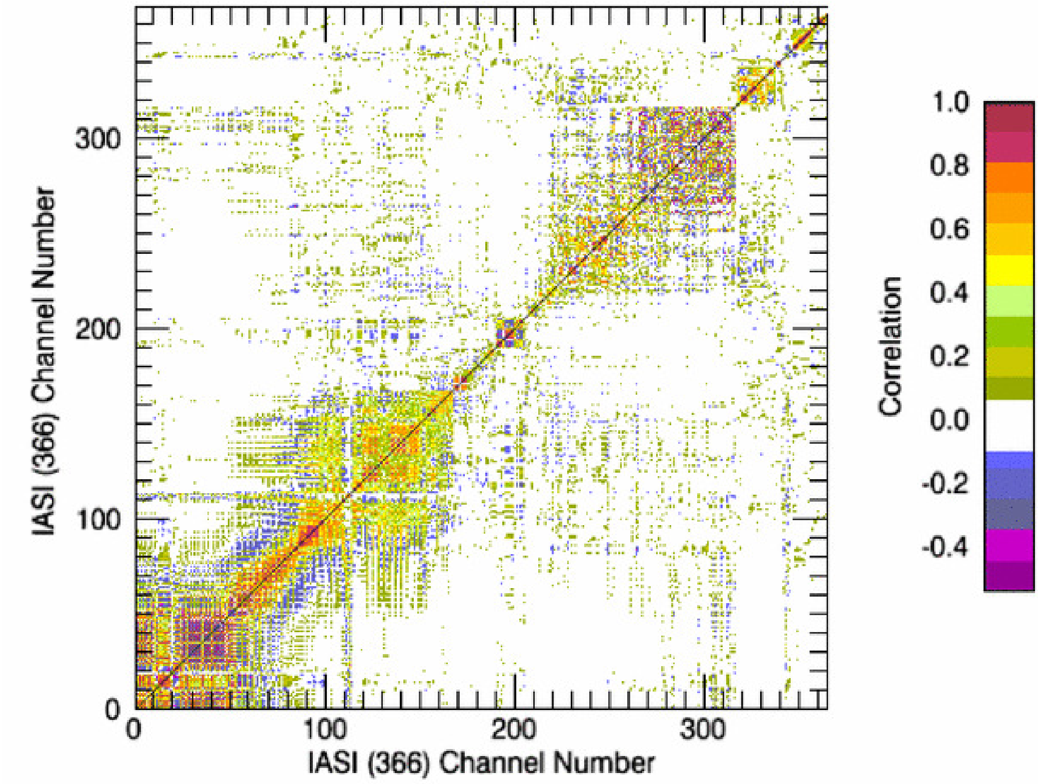

Figure 4: Error correlation structure of noise-normalised reconstructed IASI radiances shown in Figure 3. The channels shown are those monitored operationally at ECMWF (Collard and McNally, 2009). Channels 1-170 are the long wave CO_2 channels, 172-220 are window and ozone channels, 221-338 are water vapour channels and the remainder are shortwave CO_2 and window channels.

Figure 5 shows the matrix transformation mapping IASI raw radiances into reconstructed radiances. This assumes that the raw radiances are apodised (though in practice, we use deapodised raw radiances for PC compression). The matrix is not symmetrical, because the raw radiances are noise-normalised before calculation of PC scores in order to ensure the correct signal to noise ratio for each channel is taken into account. Without noise normalisation, the noisier channels are over-represented in the reconstruction and noise can propagate to other channels.

Figure 5 illustrates how the correlated atmospheric information is spread through the spectrum. Row *i* of the matrix represents the contributions of all channels to RR channel *i*, and column *j* represents the contribution of channel *j* to each channel in the RR spectrum. It is clear that the information from Band 1 in particular contributes heavily to the reconstruction of all channels in the spectrum. Although this plot suggests that Band 1 RRs themselves do not have much contribution from the other parts of the spectrum, it is in fact possible to reconstruct the RRs for the entire spectrum from RRs of the first 200 channels only, showing that the signal for other parts of the spectrum does contribute to the Band 1 RRs albeit at a very low level.

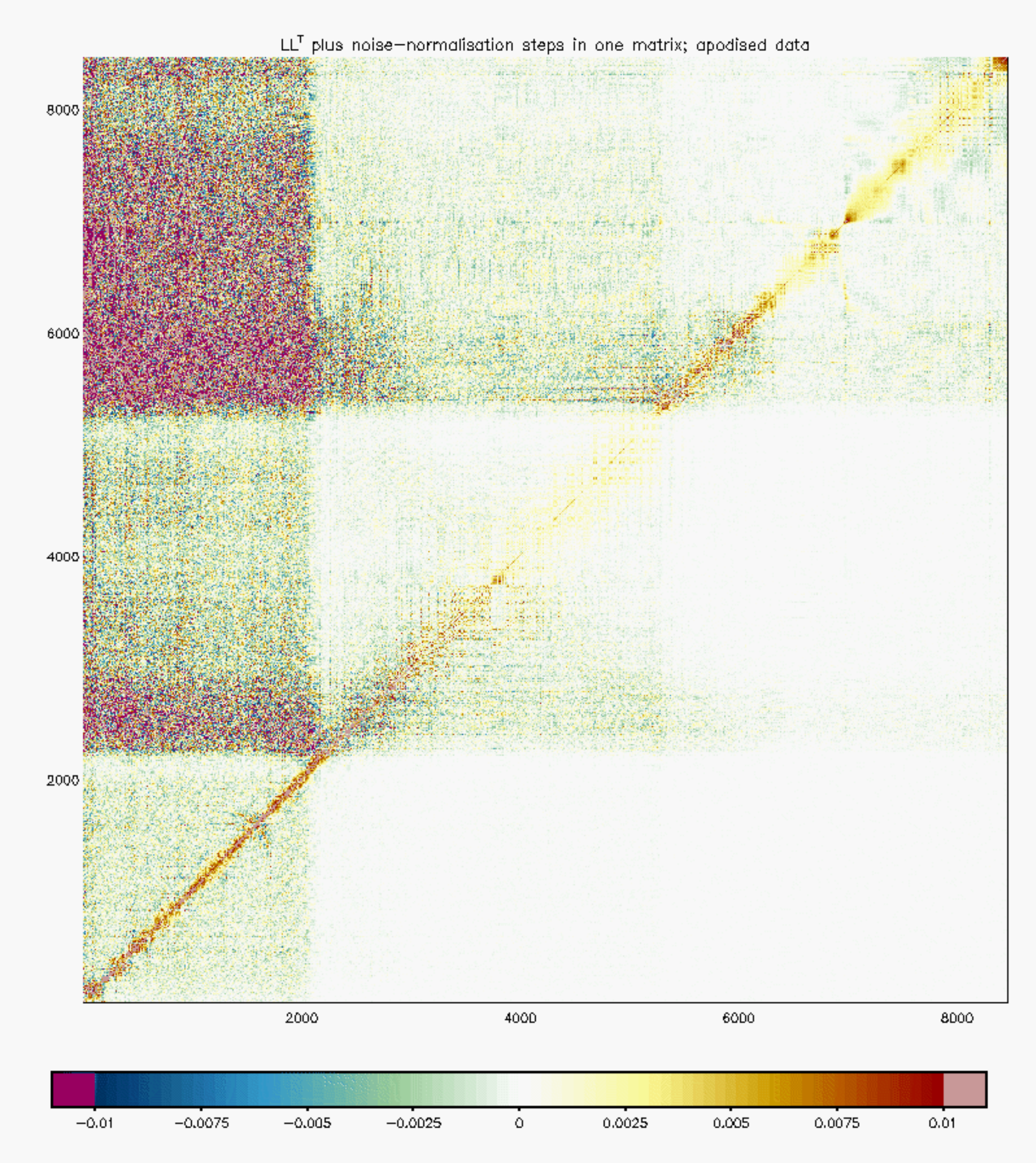

Figure 5: The matrix transformation converting apodised IASI spectra into reconstructed radiances including the noise-normalisation step. PCs derived from real IASI data as described in Section 3.1

Conversely, it appears that a large portion of the reconstructed signal from Band 2 and in, particular Band 3, comes from Band 1. For some applications, it may be desirable to compress each band separately to avoid this, but using the full spectrum does allow full use to be made of the correlations between channels with similar atmospheric sensitivity, and the increased signal to noise ratio of certain parts of the spectrum.

The introduction of significant inter-channel correlations may cause problems for assimilation under the current assumption of a diagonal error covariance matrix. The common practice of inflating the observation error variances in 4D-Var will, however, compensate to some extent for the lack of treatment of correlations. There are already correlations that we ignore in assimilation of raw radiances, such as those resulting from the apodisation function of IASI, from correlated forward model errors and from representativeness errors. Nevertheless, the increased levels of inter-channel correlation will make it difficult to benefit from lowering the observation error variances in line with the expected reduction in instrument noise. Furthermore, the effects of assuming an incorrect error covariance matrix are not yet well understood, although information content studies have shown that the use of an uninflated diagonal matrix when the true matrix is not diagonal is likely to introduce errors (Collard, 2009). Recent studies indicate that retaining at least some correlation structure is likely to be important—experiments with various non-diagonal approximations to a second order auto-regressive (SOAR) covariance matrix allowed retention of significantly more of the information content than a diagonal matrix (L. Stewart, personal communication).

Although the error correlation matrix in Figure 4 shows significant off-diagonal elements, it may be possible to produce a channel selection for IASI which penalises the selection of highly correlated channels which might allow the use of a diagonal error covariance matrix. However, this has not yet been demonstrated and it is not known whether it would be possible to select enough channels to approach the information content of the current channel selection without significant off-diagonal contributions.

Even if it were possible to do this, we may still not be able to benefit from the reduction of the instrument error to any great degree. Figure 6 shows the fit of the model background to raw and reconstructed IASI brightness temperatures. The channels which show the greatest benefit from the reconstruction process in terms of a reduction in standard deviation of observation-model fit are those peaking in the stratosphere, where the instrument error forms the dominant part of the observation error, but because the background model error is small, the fit is already good. The water vapour band is to a great extent dominated by NWP model error terms and we do not see a significant improvement in fit for reconstructed radiances. It is possible that Band 3 would show the biggest beneficial effect, but these channels currently cannot be used during the day because of solar effects, and it has not been investigated whether the use of a mixed day/night training dataset for the PCs would cause any deterioration of the night time observations. As expected, the mean fit between observation and model is barely affected by the reconstruction process, with the exception of some small differences for some water vapour channels with sensitivity to stratospheric temperatures, for which the change in bias has not been investigated.

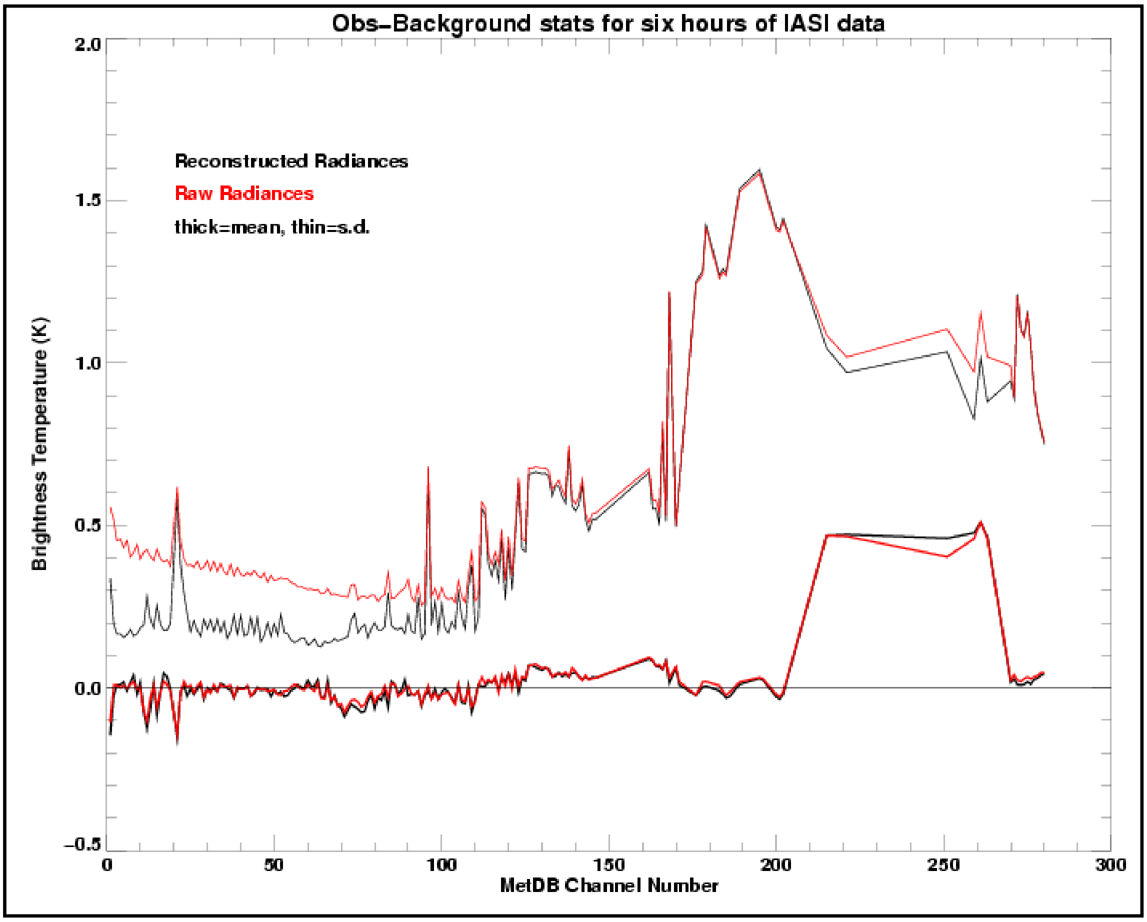

Figure 6: Mean and standard deviation of observation minus forward modelled NWP background for raw radiances and reconstructed radiances. The channels plotted are those used operationally in the 1D-Var retrieval in the Met Office observation processing system (OPS). MetDB channel number refers to the number within the Met Office channel selection (Hilton et al, 2009a)

3 Experiments in the assimilation of PC-compressed data

To date, no assimilation experiments have been run with principal component scores and work has focussed instead on RRs. Experiments in the assimilation of RRs have so far attempted to replace the raw radiances we currently assimilate with the RRs, rather than to design a system to optimise their use. As such, they have used the same channel selection as for assimilation of raw radiances, and have generally started with the same assumed observation errors.

There are two main reasons for starting with the same assumptions as for raw radiances. Firstly, to verify that we can actually use RRs in place of the raw radiances we use operationally without degrading the quality of the forecast. Secondly, it is a common assumption that RRs will provide forecast benefit over using raw data, and this can only be tested by treating them in the same way. Only when we understand any differences in behaviour can we begin to work on an optimised assimilation system for

RRs. Although there may be benefit to choosing a new channel selection for RRs, any work on optimising the assimilation has so far concentrated on the observation error covariance matrix.

3.1 The assimilation of IASI reconstructed radiances at the Met Office

At the present time, PC compression is not used for the dissemination of IASI radiances. Compression of the radiances is performed at the Met Office using AAPP (Atkinson et al., 2006) and 150 PC scores are stored alongside the 314 radiances used operationally at the Met Office (Hilton et al., 2009a). The training set for the PCs used for compression in AAPP was constructed from data from 11 July 2007 to 31 December 2007. The data was thinned in the following manner:

- 1 day in 5
- 1 granule in 4 (a granule contains 3 mins of data—22 or 23 scan lines)
- 1 footprint in *n* scanlines, where *n* is 4 at the equator but thinned further proportional to the secant of latitude at high latitudes.
- One spectrum is chosen at random from each selected scanline.

In total 15736 spectra were used in the training dataset. The data are deapodised and noise-normalised before the PCs are calculated. The eigenvectors are then combined with an apodisation matrix so that the reconstruction process will give Gaussian apodised radiances.

An assimilation experiment was run using the PC-compressed IASI data for the period 1 June to 30 June 2008. The experiment was compared against a control system assimilating raw IASI data. From the 150 PC scores, the same 314 channel radiances were reconstructed and converted to brightness temperatures. Observations were rejected if the reconstruction score indicated that the reconstructed radiances were different from the original spectrum by more than 1.25 times the instrument noise, typically rejecting <0.01% of observations. The RRs were then processed and assimilated as though they were raw radiances and assimilated with standard deviation observation errors of 0.5K for tropospheric temperature channels, 1K for stratospheric temperature channels and window channels, and 4K for water vapour channels.

The verification of this trial was statistically neutral, -0.026 on the NWP Index against observations and -0.170 against analyses. There were no marked changes to any fields. The neutral result confirms that if the supply of IASI data from EUMETSAT were to be changed to PC scores rather than radiances, we would be able to assimilate the data without any major effect on the NWP index or indeed any significant changes in processing.

A second assimilation experiment was then run for the period 1 June to 20 June 2008. In this experiment, the observation errors for tropospheric temperature sounding channels were reduced to 0.3K, stratospheric temperature sounding channels to 0.5K and those for the water vapour channels from 4K to 3K. The experiment was somewhat flawed in that the change in observation error for the water vapour channels was not justified by a change in fit between model and observation (Figure 6) because the instrument noise is not the dominant term in the observation error. However, the results of this experiment were also neutral, +0.109 on the NWP index against observations and -0.479 against analyses. The small change in index is likely to be the result of the heavier weighting to the water vapour channels.

These results are to be expected—the information assimilated is not greatly increased by using the same channels from the reconstructed spectrum as from the raw spectrum. The improvement in standard deviation of the model background fit to the observations in the temperature sounding channels is such that the bias in the fit becomes significant.

Although Figure 6 shows a very small mean global bias for the temperature sounding channels, Figure 7 shows that at certain times of year, residual biases in such channels can be up to 0.5K with a significant difference in bias between northern and southern hemispheres which cancels out when a global average is taken. A possible cause is seasonal and spatial variability in the distribution of carbon dioxide and ozone which are not accounted for in the radiative transfer modelling. These biases must be addressed before any benefit can be gained from assimilating RRs with a reduced observation error. One way in which it may be possible to benefit from the use of PC-compressed data is if we were to alter the channel selection to increase the information content of the assimilation system.

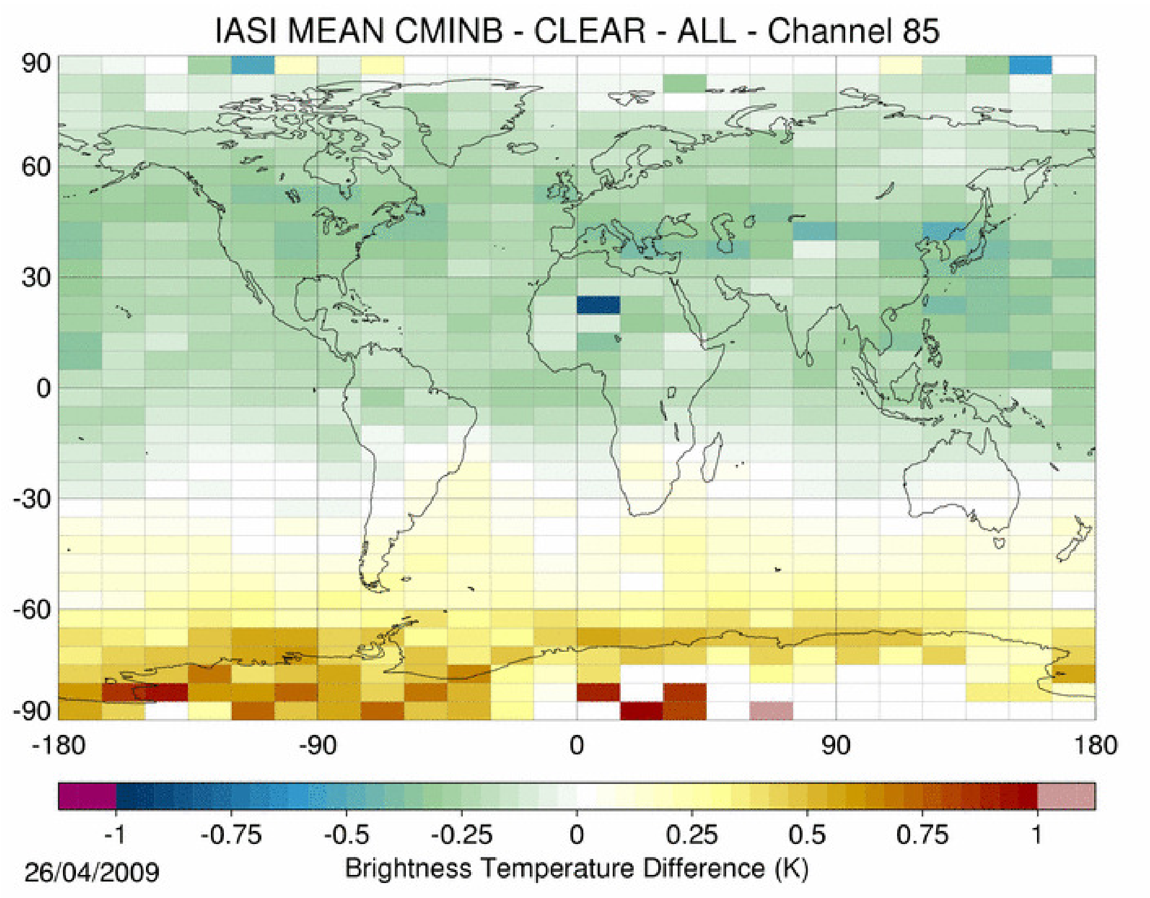

Figure 7: Residual bias after bias correction for IASI channel 85 out of 314, which corresponds to instrument channel 280 at 714.75cm^{-1}—a temperature sounding channel peaking at approximately 400hPa. One week of thinned data is shown from April 2009.

3.2 The assimilation of AIRS reconstructed radiances at ECMWF

Experiments at ECMWF using the NOAA/NESDIS AIRS reconstructed radiance dataset of the same channels as used operationally were conducted from 1 March to 30 April 2005. The observation errors were also the same as for the operational assimilation (0.6K standard deviation for short wave temperature sounding channels, 2K for water vapour and window channels, 0.4K for long wave tropospheric temperature channels and 1K for stratospheric temperature channels).

The use of reconstructed radiances had an effect on the McNally and Watts (2003) cloud detection scheme because the smoothing of the instrument noise between channels means that the algorithm picks a different cloud top in some cases. Overall it was found that typically less than 2% of channels were flagged as cloudy which were clear for raw radiances and less than 1% were clear for RRs and cloudy for raw observations. The overall results of the trial showed that the fit to other observations was unaffected by the change to RRs. The effect on the 500hPa geopotential height anomaly correlation was found to be neutral in both hemispheres.

Further experiments with reduced observation errors gave neutral to negative impact. Also tested was the use of a full covariance matrix for which the variance was taken from the operational diagonal matrix and the correlation structure from the matrix transformation in equation (3). Results of this trial were also neutral.

3.3 Assimilation trial conclusions

The assimilation trials of reconstructed radiances for both IASI and AIRS at the Met Office and ECMWF give confidence that PC-compressed hyperspectral sounding data can be used in place of raw observations with no detriment to the performance of the forecast model. No forecast benefit should be expected from a simple change from raw radiances to reconstructed radiances within the same assimilation scheme. Whilst we may expect to derive some benefit from a reduction in instrument noise, this benefit is hard to realise without correct treatment of inter-channel correlations and improvements in model-observation biases. It is also possible that more benefit could be achieved if the channel selection were modified to optimise the use of the reconstructed spectrum.

4 Conclusions

The use of principal component compression for hyperspectral sounding data has been investigated. At the present time, the direct assimilation of PC scores is not recommended for operational use. Significant questions remain as to how the data can be effectively quality controlled and more investigations are required into the impact of the highly non-localised jacobians of PC scores. The use of PC scores in retrieval schemes is under active investigation at the moment, and it is likely that significant progress will be made in these regards in the near future.

Reconstructing the radiances from the PC scores yields a dataset with reduced instrument noise, but with increased inter-channel correlations. These data can be assimilated into NWP models as a direct replacement for raw radiances with little effect on the forecast performance of the model. Although a new channel selection aimed at minimising error correlations may yield some improvements, the full benefit of the reduced instrument noise is only likely to be seen once the observation error covariance matrix is correctly specified and sources of model-observation biases are understood and reduced where possible.

References

Antonelli P., H. E. Revercomb, L. A. Sromovsky, W. L. Smith, R. O. Knuteson, D. C. Tobin, R. K. Garcia, H. B. Howell, H.-L. Huang, F. A. Best, 2004. A principal component noise filter for high spectral resolution infrared measurements. *J. Geophys. Res.*, **109**, D23102, DOI:10.1029/2004JD004862

Atkinson, N.C., P. Brunel, P. Marguinaud, T. Labrot, 2006. AAPP status report and preparations for processing MetOp data. *Proc. International TOVS Study Conference XV, Maratea, Italy, 4-10 October 2006. Available from http://cimss.ssec.wisc.edu/itwg/itsc/itsc15/proceedings/ (accessed on 20/10/2009).*

Chevallier, F. 2001. Sampled databases of 60-level atmospheric profiles from the ECMWF analyses. *EUMETSAT/ECMWF SAF Programme, Research Report No. 4. Available from: NWP SAF Helpdesk, Met Office, FitzRoy Road, Exeter, EX1 3PB, UK.*

Collard, A.D., 2007. Selection of IASI channels for use in Numerical Weather Prediction. *Q J R Meteorol. Soc.*, **133**, 1977-1991, DOI: 10.1002/qj.178

Collard, A.D. and A.P. McNally, 2009. The assimilation of Infrared Atmospheric Sounding Interferometer radiances at ECMWF. *Q J R Meteorol. Soc.*, **135**, 1044-1058, DOI: 10.1002/qj.410

Collard, A.D., 2009. The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Proc. ECMWF/EUMETSAT NWP-SAF Workshop on the assimilation of IASI in NWP, ECMWF, Reading, UK, 6th-8th May 2009. Available from http://www.ecmwf.int/publications/library/do/references/list/17784 (accessed on 16/10/2009)*

EUMETSAT, 2009. EARS Operational Service Specification, v4. *Document number EUM/OPS/SPE/01/0839 available from http://www.eumetsat.int/Home/Main/Documentation/Technical_and_Scientific_Documentation/OperationalServicesSpecifications/?l=en (accessed on 16/10/2009)*

Goldberg, M.D., Y. Qu, L.M. McMillin, W. Wolf, L. Zhou and M. Divakarla, 2003. AIRS Near-Real-Time Products and Algorithms in Support of Operational Numerical Prediction. *IEEE Trans. on Geoscience and Remote Sensing*, **41**, 379-389

Havemann, S., J.-C. Thelen, J.P. Taylor, A. Keil, 2009. The Havemann-Taylor Fast Radiative Transfer Code: Exact fast radiative transfer for scattering atmospheres using Principal Components (PCs). *Current Problems in Atmospheric Radiation (IRS 2008): Proc. of the International Radiation Symposium (IRC/IAMAS), AIP Conference Proceedings*, **1100**, 38-40

Hilton, F., N.C. Atkinson, S.J. English and J.R. Eyre, 2009a. Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments. *Q J R Meteorol. Soc.*, **135**, 495-505, DOI: 10.1002/qj.379

Hilton, F., A. Collard, V. Guidard, R. Randriamampianina and M. Schwaerz, 2009b. Assimilation of IASI radiances at European NWP Centres. *Proc. ECMWF/EUMETSAT NWP-SAF Workshop on the assimilation of IASI in NWP, ECMWF, Reading, UK, 6th-8th May 2009. Available from http://www.ecmwf.int/publications/library/do/references/list/17784 (accessed on 16/10/2009)*

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

Liu, X., W.L. Smith, D.K. Zhou, and A.M. Larar, 2005. A principal component-based radiative transfer forward model (PCRTM) for hyperspectral instruments. *In: Multipsectral and hyperspectral remote sensing instruments and applications II. Ed. Allen M. Larar, Makoto Suzuki, Qingxi Tong. Proceedings of SPIE*, **5655**, 96-105

Liu, X., D.K. Zhou, A. Larar, W.L. Smith, and S.A. Mango, 2007. Case-study of a principal-component-based radiative transfer forward model and retrieval algorithm using EAQUATE data. *Q J R Meteorol. Soc.*, **133**, 243-256, DOI: 10.1002/qj.156

McNally, A.P. and P.D. Watts, 2003. A cloud detection algorithm for high-spectral-resolution infrared sounders. *Q J R Meteorol. Soc.*, **129**, 2411-2323, Saunders, R., S. English, P. Francis, P. Rayer, P. Brunel, G. Kelly, P. Bauer, D. Salmond and D. Dent, 2005. RTTOV-8—the latest update to the RTTOV model. *Proc. International TOVS Study Conference XIV, Beijing, China, 25-31 May 2005. Available from http://cimss.ssec.wisc.edu/itwg/itsc/itsc14/proceedings/ (accessed on 16/10/2009).*

Turner, D.D., R.O. Knuteson, H.E. Revercomb, C. Lo and R.G. Dedecker, 2006. Noise reduction of Atmospheric Emitted Radiance Interferometer (AERI) observations using principal component analysis. *J Atmos. Oceanic Technol.*, **23**, 1123-1238

Met Office
FitzRoy Road, Exeter
Devon EX1 3PB
United Kingdom

Tel: 0870 900 0100
Fax: 0870 900 5050
enquiries@metoffice.gov.uk
www.metoffice.gov.uk