

IASI Principal Component Compression (IASI PCC) FAQ

Tim Hultberg (tim.hultberg@eumetsat.int), March 2009

1. How does it work?

Principal component compression works by representing multidimensional data, like IASI spectra, in a lower dimensional space, which accounts for most of the variance seen in the data. This space is spanned by a truncated set of the eigenvectors of the data covariance matrix. By noise-normalising the spectra prior to the application of the compression technique, the ability to fit the data is enhanced by avoiding giving too much weight to variance caused by noise.

The coefficients used for the compression are determined by 1) the training set of IASI L1C spectra, $X \in \mathbb{R}^{m \times n}$, where m is the number of channels and n is the number of spectra in the training set, 2) the noise normalisation matrix, $N \in \mathbb{R}^{m \times m}$ and 3) the number, s , of eigenvectors to retain.

Let $N^{-1}\bar{x} \in \mathbb{R}^m$ and $C \in \mathbb{R}^{m \times m}$ be the mean and covariance of the normalised training dataset $N^{-1}X$ and let $E \in \mathbb{R}^{m \times s}$ be the s most significant eigenvectors of C . The compressed representation (the PC scores), $p \in \mathbb{R}^s$, of a IASI spectrum, $x \in \mathbb{R}^m$, can now be computed as

$$p = E^T N^{-1}(x - \bar{x})$$

from which a noise-reduced approximation, $\tilde{x} = NEp + \bar{x} \in \mathbb{R}^m$ of x can be reconstructed. For this noise-reduced approximation of the radiances we use the terms “reconstructed” and “noise filtered” radiances synonymously.

2. How can IASI PCC discriminate between atmospheric signal and random noise?

The radiances measured by IASI are spectrally highly correlated and the atmospheric information they contain can therefore be explained by a small subset of the leading eigenvectors. On the contrary, this is not the case for the noise. In fact, if the random noise is Gaussian, uniform and uncorrelated, all possible directions (including all eigenvectors) explain exactly the same amount of noise.

3. How big should the training set be?

The training set should include spectra observed over different types of atmospheric/surface conditions at different scan angles and for different pixel numbers to ensure that a truncated set of eigenvectors can be adequately used to represent any observed spectrum (see “What about rare events?”).

Additionally, if the training set is too small, the specific outcome of the random noise will not be sufficiently uncorrelated and uniform and will therefore have an influence on the computed eigenvectors (and eigenvalues). This effect can be observed by plotting the eigenvalues obtained from training sets of different sizes. Experiments show that 70,000 spectra are enough.

4. What is the residual?

We use the term “residual” for the difference between the original and reconstructed spectra, i.e. $x - \tilde{x}$. Sometimes, if it is clear from the context, we also use the term “residual” as short hand notation for the noise normalised residual, i.e. $N^{-1}(x - \tilde{x})$.

5. What is the reconstruction score?

The reconstruction score of a spectrum is defined as the root mean square of the noise

normalised residual, i.e. $\sqrt{\frac{1}{m} \sum_{i=1}^m r_i^2}$, where $r = N^{-1}(x - \tilde{x})$. If the noise normalisation

corresponds well to the actual noise seen in the data, the average reconstruction score is close to one with a relatively small standard deviation. If the reconstruction score of a particular spectrum is significantly higher than the average, this is a sign that the residual does not consist of residual noise only, but is affected by non-negligible reconstruction error (see “What is reconstruction error?”). This can happen if the spectrum contains features which are not well represented in the training set (see “What about rare events?”) and is easily detected by the user since the reconstruction score is disseminated along with the PC scores.

6. I have been told, the compression is “lossy”, is that true?

When spectra are reconstructed from a truncated set of PC scores, most of the noise contained in the original spectra is filtered away. The compression is therefore “lossy” in the sense that reconstructed and original spectra are different. However it is important to remember that this difference is almost exclusively a result of reduced noise in the reconstructed spectra.

7. What is reconstruction error?

In the following we assume that each measured spectrum, x , can be written as a sum, $x = x_0 + \varepsilon_x$, of true radiances (the atmospheric signal), x_0 , and random measurement error (the instrument noise), ε_x .

Now we can write the residual as

$$x - \tilde{x} = x_0 + \varepsilon_x - (NEE^T N^{-1}(x_0 + \varepsilon_x - \bar{x}) + \bar{x}) = x_0 - \tilde{x}_0 + \varepsilon_x (I - NEE^T N^{-1}),$$

where \tilde{x}_0 denotes the reconstructed true radiances, i.e. $\tilde{x}_0 = NEE^T N^{-1}(x_0 - \bar{x}) - \bar{x}$.

The first term in the expression for the residual, $x_0 - \tilde{x}_0$, corresponds to the loss of atmospheric signal and is known as the reconstruction error. The second term, $\varepsilon_x (I - NEE^T N^{-1})$, is the part of the random measurement error which is removed by the reconstruction process.

8. How do you know that the number of PC scores is high enough to make the reconstruction error negligible?

Two methods are used to choose the number of PC scores to retain:

- 1) By looking at the second derivative of the average reconstruction score for an ensemble of spectra as a function of the number of PC scores, it is possible to determine at what number of PC scores the rate of decrease of the average reconstruction score stabilises.
- 2) If an eigenvector carries some non-negligible amount of atmospheric information, this is expected to result in some spatial correlation of the corresponding observed PC score. By plotting the PC score spatial correlation as a function of the eigenvector rank it is possible to select the number of PC scores to be high enough such that any significant spatial correlation gets included.

A posterior confirmation that the reconstruction error is negligible can be achieved by statistical analysis of the residual.

9. What about rare events?

By outlier spectra (or simply outliers) we understand spectra which are poorly represented in the training set and result in big residuals. Such outliers can be caused by rare events such as major volcanic eruptions. Outliers are easily detected by comparing the RMS of the noise-

normalised residuals (i.e. the reconstruction score) with a threshold value and are captured in an auxiliary product so that they can be added to the training set and a new set of eigenvectors can be generated, if needed. It is nevertheless our hope that we are able to include enough “rare event spectra” in the initial training set, by prior examination of at least one year of past data, such that a regeneration of the eigenvectors might never become necessary.

10. What about lossless compression alternatives?

If the quantisation of the residuals is chosen to correspond to the quantisation used in the L1C product (implicitly determined by the scale factors given in the product), the addition of the residual will result in exactly the original spectrum. Since the quantisation used in the L1C product is finer than half of the NEdN currently proposed as quantisation step for the residuals, the resulting numerical range and thereby the storage consumption (whether Huffman compressed or not) of the residuals would increase. Nevertheless, such a scheme would still represent an efficient lossless alternative.

There are however, in my opinion, no good reasons to prefer a quantisation based on the L1C scale factors to a quantisation based on the actual instrument noise, nor to desire the ability to reconstruct the random noise.

11. Can a subset of reconstructed radiances substitute the PC scores?

Yes. If the number of channels is at least as big as the number of scores, the sub-matrix of E consisting of the rows corresponding to the selected channels will normally be of rank s , in which case the subset of the reconstructed radiances will contain all the information present in the PC scores. This means that it would be possible to disseminate a subset of reconstructed radiances as an alternative to the PC scores and the users would still be able to reconstruct the full spectra just as well as they would be able to do from the PC scores.

12. What is the error characteristic of reconstructed radiances?

As already discussed, the noise remaining in the reconstructed radiances is much lower than in the original radiances. However, it should be noted that the reconstruction process can introduce spectral correlation of the observation errors not seen in the original data. Under the assumption that the reconstruction error is negligible, the observation error covariance matrix of the reconstructed radiances can be computed as $NEE^T N^{-1} R N^{-1} E^T E N$ where R is the error (noise) covariance matrix of the IASI Level 1C radiances.

13. What is the error characteristic of reconstructed radiances + the residual?

If the residuals are added to the reconstructed radiances we are able to reconstruct the original instrument noise. With the proposed quantisation step, radiances are rounded to the nearest multiple of 0.5 times the NEdN – the difference between the reconstructed (+residual) radiances and the original radiances is never higher than 0.25 times the NEdN. This rounding will sometimes decrease the error and sometimes increase it. It can be shown that the net effect is an increase of 1% of the standard deviation of the error.

14. Why is IASI PCC applied individually to each band?

The main reason is to be able to compress and disseminate radiances from one or two of the bands in the presence of IASI Level 1C bad quality radiances affecting only one (or two) of the bands (later in 2009 the IASI L1C product will be changed to include a quality flag for each individual band). Since single band failures are more frequent for band 3 than band 2 (which again are more frequent than for band 1), we choose the band limits at the lower end of the band overlap regions. This avoids a failure of band 3 affecting any of the other bands (the L1C radiances in the band overlap spectral regions are obtained by merging of radiances from two neighbouring bands).

15. What are the band limits?

The band limits, chosen as described above, are:

Band 1 covers channel 0 to 1996 (wavenbrs. 645 – 1144 cm^{-1}) (1997 channels)

Band 2 covers channel 1997 to 5115 (wavenbrs. 1144.25 – 1923.75 cm^{-1}) (3119 channels)

Band 3 covers channel 5116 to 8460 (wavenbrs. 1924 – 2760 cm^{-1}) (3345 channels)

16. Does the band separation introduce “discontinuities” at band limits, in the reconstructed spectra?

If enough PC scores are used for all three bands, the reconstruction error is negligible for them all and no “discontinuity” of the signal is introduced. The resulting noise covariance matrix (NCM) of the reconstructed radiances will however be slightly different than if all channels were compressed simultaneously – we get a block diagonal NCM (with three blocks) instead of a full NCM (see “What is the error characteristic of reconstructed radiances?”). But that should be an advantage rather than a disadvantage.

17. What is the compression factor?

A total of 290 PC scores will be used for the three bands (90 for Band 1, 120 for Band 2 and 80 for Band 3). When all PC scores are quantised with the same step size, the scores corresponding to the lower ranking eigenvectors have a smaller dynamic range and can be encoded with fewer bits (for example one byte per score). If the residuals are not disseminated, we end up with a raw (i.e. before BUFR encoding) compression factor of about 50 without considering potential further compression of the scores (like Huffman encoding or exploitation of spatial correlation). The final compression factor when considering BUFR products is still to be determined.

18. How often will the eigenvectors be updated?

The eigenvectors will only be updated if a clear need is identified. It is not unlikely that this will never happen in the lifetime of the first Metop satellite.

19. Will a change of the IASI instrument noise trigger an update of the eigenvectors?

No, not necessarily. By monitoring the residuals we are able to detect changes in the LIC noise characteristics, for example as resulting from an instrument decontamination. But as long as the shape of the noise spectrum does not change dramatically within each band, there is no need to change the noise normalisation used for the eigenvector generation.

20. Can noise filtering give better retrievals?

The use of reconstructed radiances for the retrieval of atmospheric parameters is currently being investigated at EUMETSAT and elsewhere. Preliminary results show a neutral to positive effect on the retrievals even without changing the measurement error covariance matrix. To fully harvest the benefits of noise filtering, the measurement error covariance matrix should be adjusted accordingly, including the full use of its off-diagonal elements. Noise filtering is expected to be particularly beneficial for retrievals where only a small subset of all channels is used and where the spectral signals are small.

21. How does the omission of the off-diagonal elements of the noise covariance matrix used for constructing the noise normalisation matrix affect the compression scheme?

The noise filtering effect of the IASI PCC is most effective when applied to data with white noise. This can be achieved if the matrix square root of the measurement error covariance matrix is used as the noise normalisation matrix. However, as a result of apodisation, the noise covariance matrix of the IASI Level 1C radiances is band diagonal and if we disregard

its off-diagonal elements and use a diagonal noise normalisation matrix formed by the square roots of the noise variances, the noise does not get spectrally de-correlated by the noise normalisation. While this does not introduce any reconstruction error, it could make the noise filtering slightly less efficient.

22. Where can I read more?

There are numerous papers/documents concerned with PC compression in general and applied to hyper-spectral infrared observations in particular. The four I recommend to read are:

[R1] Antonelli P., Revercomb H., Stromovsky L., Smith W., Knuteson R., Tobin D., Garcia R., Howell H., Huang H., and F. Best, 2004: "A principal component noise filter for high spectral infrared measurements", JGR, 109.

[R2] Goldberg M., Zhou L., Wolf W., Barnet C. and M. Divakarla, 2005: "Applications of Principal Component Analysis (PCA) On AIRS Data", Proc of SPIE, 5655.

[R3] Lee A. C. L. and S. Bedford, 2004: "Support Study on IASI Level 1c Data Compression", Technical Report, EUMETSAT Contract EUM/CO/03/1155/PS.

[R4] A. Collard and Marco Matricardi, 2005: "Definition of an efficient interface to NWP for assimilating IASI radiances", Contract No. EUM WP 989-2 December 2005.

They are all excellent and complement each other well.