Thank-you for the careful and thorough reading of my manuscript and for all of the helpful comments. I very much appreciate the time and effort you've both taken!

Responses to Referee #2

Specific Comments

p1186, l2-5: Point taken – I've moved these two sentences from the abstract to the intro. I've also added some quantifications of the method's strengths.

p1186, 112: The 30% noted comes from investigations done with the ensemble of RFM spectra to see what filling of the MIPAS FOV was required for the current operational (CI) method to detect cloud, as parameterised by the EF (which isn't introduced until later in the paper, which is why I've stuck to giving the value for thick-clouds only, for which the geometric filling is approximately equal to the EF), which combines information about both the geometrical proportion of the FOV filled as well as the optical thickness of the filling cloud. Hence, as shown in Fig. 2, in order for the CI method (with its current thresholds/formulations) to flag cloud, EF>~30%. I've changed this phrase so that it is clear that it's not just a geometric filling that is being referred to, but optical 'filling' as well.

p1186, l17: Fig. 3 shows that even if the settings of the CI method (at least in terms of threshold choice) are adjusted to include thinner cloud, that it misdetects many clear spectra as cloudy – so it appears that it is not increasing detection as much as it is loosening reliability. I've readjusted this sentence to clarify this.

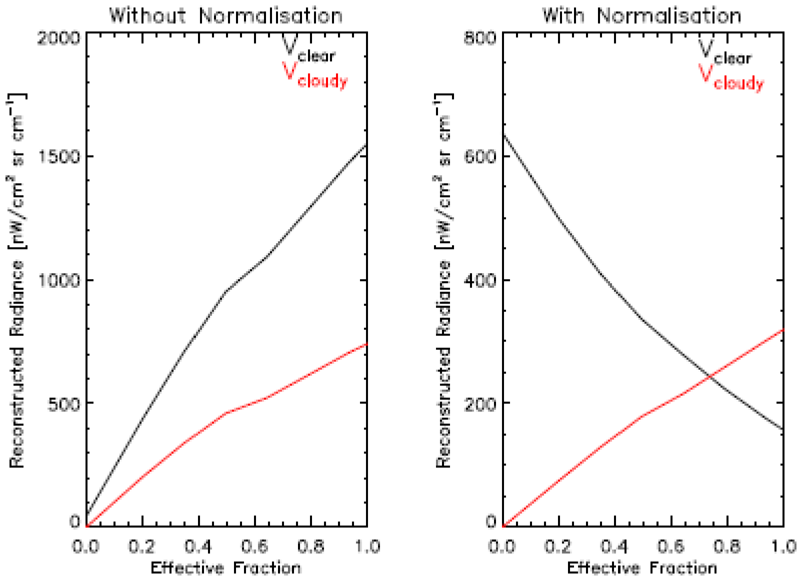
p1186, l26: Definitely true. Have amended this.

p1187, l18: Have adopted this phrasing.

p1187, l25: have changed phrasing so that frequency -> tangent height spacing, and have mentioned the lower stratosphere.

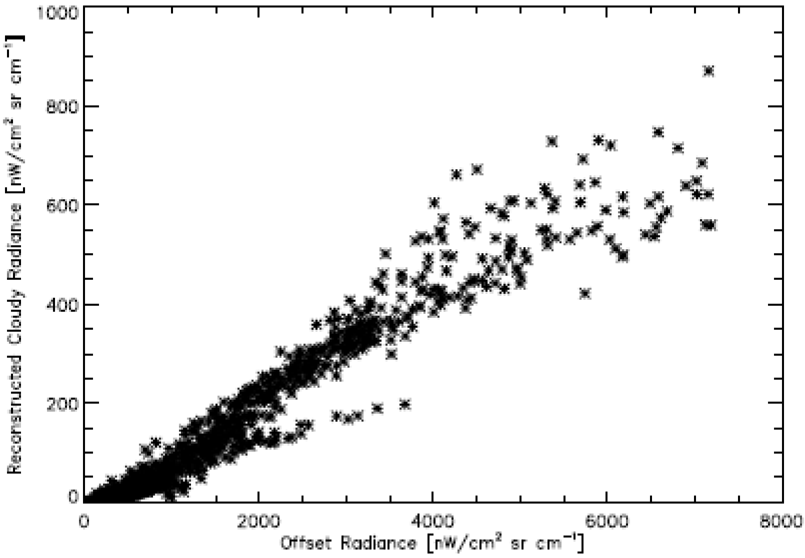
Section 4: Referee 1 also mentioned this point, and I've put in more of the specifics of the simulations/assumptions – horizontal homogeneity, non-scattering … I've also included a paragraph detailing how the assumption of non-scattering doesn't greatly affect the results, as least for clouds thicker than 10^{-4} km^{-1}. I would think that use of clear singular vectors (in conjunction with the cloudy singular vectors) when fitting real cloud signals would reduce the effect on detection that inhomogeneities in the cloud field would introduce.

p1194, l25: Basically, if no normalisation is carried out, the radiance contributing to the heightened baseline of cloudy signals is fitted by causing the clear singular vectors to have increasingly large peaks and not by increasing the contribution of the cloudy singular vectors. However including the normalisation step has the effect of "equalising the field" for both the clear and cloudy singular vectors to share the reconstructive responsibility of the raised cloud radiance baseline. Depending upon the choice of input spectra in the ensemble considered for SVD analysis and the way in which they are pre-processed, different sets of singular vectors will result — as will how these singular vectors fit an arbitrary signal. However, in order to be used in a meaningful way in cloud detection, fundamentally one requires that the fit of arbitrary spectra be dominated by clear singular vectors v_clear if the spectra are clear and by cloudy singular vectors v_cloudy if the spectra are fully cloud-filled. Basically as one moves from the totally clear state to the fully cloudy state, it is required that the dominance of the v_cloudy increases monotonically while that of the v_clear remains constant or decreases monotonically. Two different approaches can then be considered: to normalise or not to normalise.



The attached figure highlights the result of following each of these two slightly different methods, in terms of the radiance that can be reconstructed by integrating the contribution from each singular vector as scaled by the fit coefficient. When no normalisation is carried out, as the cloud amount increases, the v_clear have to account for proportionately more radiance than do the v_cloudy, as indicated by the larger slope of the radiance-EF line for v_clear. The v_clear are simply forced to accommodate more of the radiance coming as a result of the cloud presence than are the v_cloudy. However, when normalisation is used as indicated, as the cloud amount in the FOV increases, the radiance picked up by the v_clear decreases while that accounted for by the v_cloudy increases, as desired. Thus, it is clearly more desirable to carry out normalisation when determining the appropriate singular vectors.

A valid concern is that by normalising the input spectra, a significant piece of information is thrown away: the baseline radiance of the spectra. From a first order perspective, cloud detection can be carried out by visually noticing a raised spectral baseline. It has been shown in the previous analysis that normalisation is required in order to successfully fit an arbitrary signal — however does the fitting process using the clear and cloudy singular vectors capture the information lost by discarding the baseline radiance? Consider the radiance spectrum that is reconstructed by integrating the cloudy singular vectors weighted by their corresponding fit coefficients. If the mean of this reconstructed cloudy radiance spectrum is correlated with the mean offset radiance which is subtracted off of the input spectrum, then one can be assured that the SVD and linear least squares fit captures the information about the baseline radiance. For MIPAS data on 6 August 2003 having tangent height less than 22.5 km, these quantities were compared.



The above figure shows this comparison — and it is quite easy to see that indeed these two quantities are well correlated. In fact, the correlation between the mean offset radiance and reconstructed cloudy radiance is 97.7% — which is convincing proof that the SV analysis well captures the information coming from the raised baseline. So it appears that normalisation, then, is both a necessary step in the calculation of singular vectors and in the fitting of arbitrary spectra as well as a step which does not, in fact, cause loss of baseline radiance information.

Does this clarify? I've added a few words to the phrase in the paper – but would prefer to reference the full explanation …

p1195, l24: I'm not sure I understand. If ones aims to reconstruct a spectrum from which the singular vectors have been calculated (ie. any of the RFM ensemble), if all of the singular vectors are used, there should be an exact fit, and no error – but if fewer than all the vectors are use there will of course be an error associated with reconstruction (but decreasing to zero as more are used). If spectra not included in the original ensemble from which the vectors are calculated are reconstructed using either the full set of vectors – or some subset thereof – there will be errors in fit from spectral details and variances not accounted for in the ensemble (which cannot be reduced regardless of how many of the vectors are used in the fitting process), but the fit should definitely improve (and the error decrease to some small but non-zero value) as more vectors are employed.

p1195, l24: Yes, non-linearity is definitely an issue. But for the purposes of detection, it is really important (at least from a first order) that it is sufficiently linear that the cloudy singular vectors are dominant for cloudy spectra and clear vectors for clear spectra (which, upon investigation appears to be the case). It turns out that this method is adept at differentiating between clear and cloudy spectra, but that it does not give very much information about 'how' cloudy a cloudy measurement is – certainly not in the thick limit (where linearity is going to be a big issue) – but as long as it gets the clear/cloud identification correct, then for the purposes of this work, surely it is a non-issue? Does this need to be addressed in the manuscript?

p1196, l18: Singular vectors are orthonormal by design. So, if one takes the ensemble of clear spectra and runs SVD on it, a set of orthonormal clear singular vectors are produced. Then, if these clear singular vectors are used to isolate the clear component of the combined clear+cloudy signal in the RFM cloud simulations, one is left with the component of signal due to the cloud itself (well … in any case, the component of the signal NOT attributable to the clear spectra) – and by the properties of least squares fitting, this will be orthogonal to the original fitting vectors, otherwise the information would have been accounted for in some way by the vectors. When this 'leftover' component is decomposed, we get another orthonormal set of singular vectors (the cloudy singular vectors), which are also orthogonal to the original clear singular vectors. I've done the checks to make sure this is indeed the case in practice – and it is. Does this explanation need to go in the manuscript?

p1198, l1-13 and Figure 7: Yes – the first cloudy singular vector should have the largest singular value for cloudy spectra, as it contains most of the variance expected in cloudy spectra – and in a sense the 'higher order' singular vectors simply increase the accuracy of the reconstructed radiance. That said, the singular values assigned to the first few cloudy singular vectors aren't negligible – although they are about 10% that of the first. Preliminary work used only the first cloudy singular value as the cloud flag, but it was not as reliable as when all the vectors were used to reconstruct the radiance. I've noted the relative magnitudes of the first few singular values in the manuscript.

Figure 8: Oops! Yes, what I've outputted is the root squared error – not the root MEAN squared error, as intended. So need to divide through by the sqrt(number of spectral points) = sqrt(5701)=75.5. Have updated the plot.

p.1200, l6: this is the unapodised value – I've a footnote in explaining this, but perhaps it needs to be a bit closer to the actual stated noise value – I've moved the footnote to clarify.

p.1202, l17: please see my response to Referee 1: "It isn't necessarily better to use a small spectral range as opposed to a large one. The reasoning was that one would expect that the 960-961 cm^{-1} range to have the highest chance of being well-fit, as it is the most atmospherically (gaseously) transparent part of the larger spectral region – and hence is most likely to have the gaseous component well-isolated. The choice of a smaller spectral region is also useful in practicalities of application, as loading in large portions of spectra can be time consuming."

I did, I preliminary investigations use the first singular value itself as a cloud flag, however it did not detect cloud as reliably as using the Integrated Radiance Ratio over the smaller spectral range.

p.1203: In Hurley (2008) there were extensive comparisons carried out between the operational cloud detection method, and the proposed SVD method – using both simulated and real MIPAS data for both clear and cloudy cases. However, these were considered too detailed for the purposes of this paper – is it sufficient to note the sensitivity of the operational method to threshold (although Fig. 3 shows that the CI method cannot be reliably used to detect thin cloud by simply loosening the threshold) and to direct the reader to the appropriate reference? I've made a note of this in the 'comparison' section.

Technical Comments

p.1186, l6: Have done.

p.1191, l6: Typo fixed.