HYPERSPECTRAL RETRIEVALS AND SUBSPACES

Tim.Hultberg@EUMETSAT.INT and Thomas.August@EUMETSAT.INT : EUMETSAT-Allee 1, 64295 Darmstadt, Germany

The high compression rate of IASI principal component (PC) compression, relies on the ability to suppress most of the random instrument noise from the measurements. This is achieved by removing the component of the measurements, which is orthogonal to the signal space defined by a truncated set of PCs. Due to the high spectral correlation of IASI radiances, the number of retained PCs (the dimension of the signal space) can be chosen to be much smaller than the number of channels, without introducing any significant atmospheric loss when suppressing the orthogonal complement of the signal space. Similarly, it is possible to determine a small dimensional subspace within which simulated spectra originating from a given forward model are contained. In the case of a PC forward model the vectors spanning this subspace are readily available. Following the characterization of the signal and forward model subspaces, any spectrum can be uniquely written as a sum of four components: 1) a component which belongs to both the signal and forward model space, 2) a component which belongs to signal space but is orthogonal to the forward model space, 3) a component which belongs to the forward model space but is orthogonal to the signal space and 4) the rest component, orthogonal to both signal and forward model space. For a simulated spectrum, the third component represents features, which are not observed in any measurements and must therefore be classified as forward model error. The underlying causes might, for example, be spectroscopy errors, instrument and processing artefacts or unrealistic shape of the surface emissivity spectrum in the input state vector. This component can be suppressed from simulated spectra simply by projecting them onto the signal space. In this case we show that the optimal estimation cost function using a subset of reconstructed radiances equals the one where the full spectra of radiances are represented by PC scores. The size of the subset of reconstructed radiances is equal to the number of PC scores and must be selected such that the sub-matrix corresponding to the selected rows of the eigenvectors has a small condition number. A simple channel selection method for reconstructed radiances is presented and retrievals using the resulting subset are performed and analysed in order to characterize the effect of the suppression of the part of the forward model error mentioned above.

On the equivalence of doing retrievals in PC space of the radiances and of using a small subset of reconstructed radiances.

Recently much work has been done to investigate efficient retrieval (or assimilation) using information from the full IASI spectrum by representing the measurements as PC scores. Another, perhaps technically easier, way to use information from the full spectrum in an efficient manner, is to use a subset of reconstructed radiances. This turns out to be exactly equivalent to the representation of the measurements as PC scores under two conditions:

1. The forward model space is a subspace of the signal space.
2. The sub-matrix of the eigenvector matrix underlying the reconstruction corresponding to the selected sub-set of reconstructed channels in non-singular.

To check the first condition, we projected synthetic spectra using RTTOV 10 as forward model onto the signal space and looked at the orthogonal complement. It was seen that the condition is not fulfilled for RTTOV 10 and in fact for some channels the mean or standard deviation of the orthogonal complement (computed over a big sample of spectra based on the 'Chevallier profile-set') exceeds the instrument noise. This means that the synthetic spectra have features which are never observed in any real IASI measurements and as such must be classified as observation errors. In order not to upset the manufacturers of forward models, it is important to emphasize that these errors can also be caused by unrealistic inputs (like emissivity spectra never occurring in real world) or instrument artefacts not reproduced by the forward model. Nevertheless it is possible to avoid these errors by projecting the output of the forward model onto the signal space (obviously the Jacobian must also be filtered accordingly). For a traditional forward model this implies that all channels must be included in the forward computation, but when using a PC based forward model it would be possible to remove these observation errors efficiently by prior projection of the forward model PCs onto the signal space. Here we have used the traditional RTTOV 10 combined with projection onto signal space.

The second condition can always be fulfilled by a suitable selection of a sub-set of channel with the same number of elements as the number of eigenvectors, s, used in the reconstruction (since the matrix, E, composed of the s leading eigenvectors is of rank s, it is possible to select s linearly independent rows of E). Numerically it is also important that the condition number of the s times s sub-matrix, Es, of E is kept as low as possible. This observation leads to a simple heuristic for selecting a suitable sub-set of channel, which will be presented below.

When these two conditions are satisfied, the measurement term of the cost function is seen to agree for the two different representations of the measurement.

$$y^T E E_s^T (E_s E^T R E E_s^T)^{-1} E_s E^T y \;=\; y^T E E_s^T E_s^{-T} (E^T R E)^{-1} E_s^{-1} E_s E^T y \;=\; y^T E (E^T R E)^{-1} E^T y$$

IASI channel selection for reconstructed radiances.

As discussed above we want to select a channel sub-set, which minimizes the condition number of the corresponding sub-matrix of E, in order to preserve the full information content. This leads to a simple heuristic method, in which the first channel chosen corresponds to the row of the eigenvector matrix, E, with the largest norm. After a row has been chosen we subtract the part which lies in the subspace spanned by the chosen row from each of the rows in E (including the chosen row itself, which then becomes zero) and repeat the process until all s channels have been selected.

```
X = E(1:m,1:s)
Cs = {}
For j=1 to s
      i = argmax{ norm(Xi) : i=1,...,m}
      Cs += {i}
      For k=1 to m
            Xk -= (Xk·Xi / Xi·Xi) Xi
      End
End
```

2	9	15	21	26	33	72	86
90	92	94	96	101	154	186	218
243	259	292	297	302	307	314	332
339	344	359	386	388	395	399	427
437	448	526	546	558	579	589	596
607	616	643	654	698	722	763	814
832	847	892	912	936	941	954	969
977	1004	1058	1073	1109	1188	1219	1242
1290	1373	1436	1479	1521	1539	1555	1560
1586	1597	1646	1675	1686	1696	1706	1714
1786	1791	1825	1842	1857	1937	1960	1963
1966	1994						

Several variations of this basic channel selection scheme are possible, for example by considering only the first j columns when computing the norm for selecting the j'th channel. I.e.

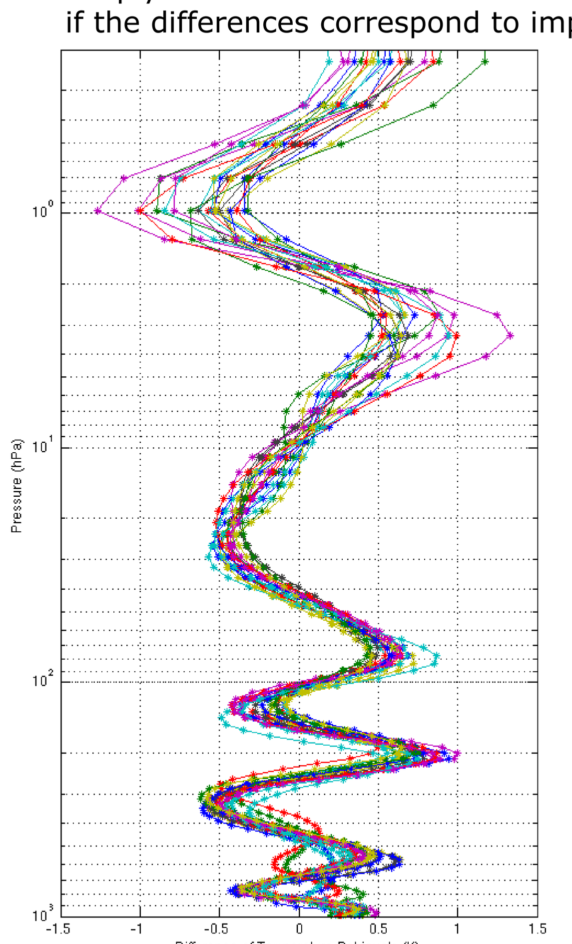
```
i = argmax{ norm(X(i,1:j)) : i=1,...,m}
```

The 90 channels selected from Band 1, following this procedure are shown to the left. These 90 channels were supplemented with 120 channels from Band 2, selected with the same procedure, for the retrievals below.

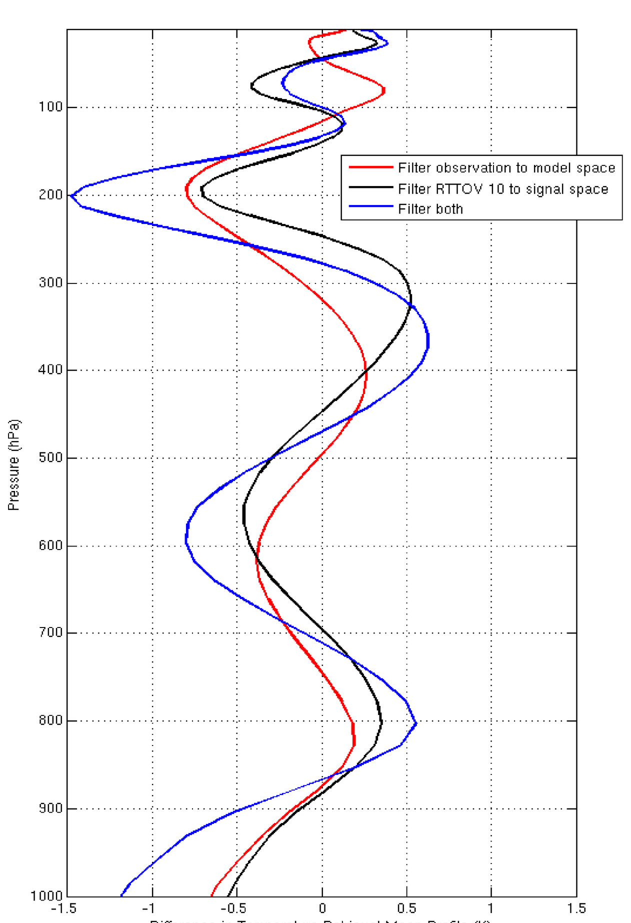
It would also be straight-forward to adapt the procedure by excluding specific eigenvectors dominated by instrument artefacts or banning certain channels, for example within the band overlap regions, where the noise characteristics is very different within the 4 IASI FOVs, or channels highly sensitive to model parameters not being actively retrieved.

Retrieval Experiments.

A retrieval experiment was performed using the 210 IASI channels in Band 1 and 2 selected above. The main purpose was to compare retrievals using standard RTTOV 10 as forward model with retrievals obtained with a forward model based on RTTOV 10, adapted by projecting the simulated spectra onto the signal space. The retrievals were done with the 1DVar software used in the EUMETSAT IASI L2 operational product processing with an active state-vector consisting of surface skin temperature as well as temperature, water vapour and ozone profiles represented as PC scores (30, 30 and 15 respectively). The observation error covariance matrix (full matrix) was based on obs. minus calc. statistics and the background error covariance matrix was based on global climatology. The retrievals were performed on 34 clear sky near nadir fields of view in the Tasman Sea measured by IASI the 15th February 2012 local night time. The left figure shows the difference in the retrieved temperature profiles with and without filtering of the forward model radiances for 20 of the 34 cases. At this early stage we simply note that the differences are big enough to be potentially important, but have not yet attempted to assess if the differences correspond to improvements as would be expected.

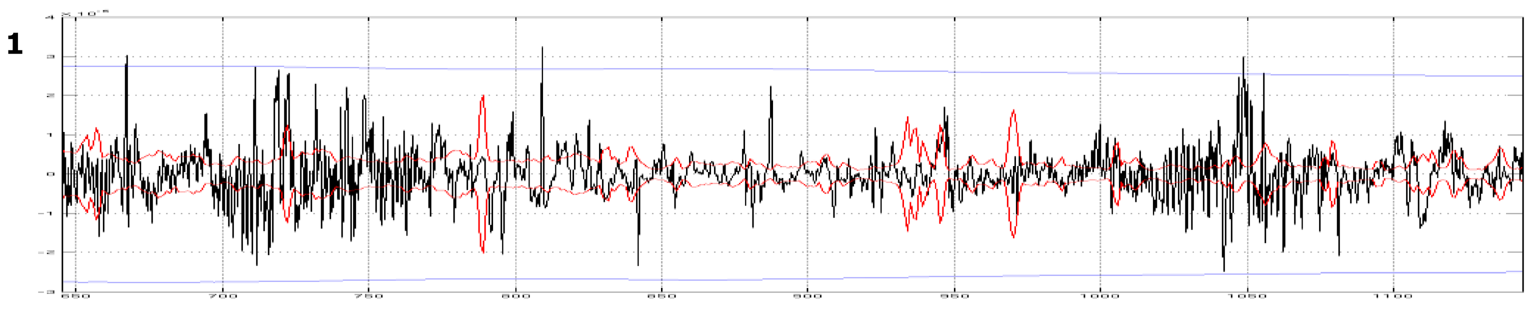


We also tried to filter the measurements by projecting them on the model space. In this context, the model space was computed from a training set of about 4000 near nadir ocean synthetic spectra and 90 (band1) and 120 (band 2) eigenvectors were used. That the signal space is not contained in the forward model space is expected. The component of the signal space which is orthogonal to the forward model space could for example originate from residual cloud effects and trace-gas amounts being kept constant in the model. A major part of this component lies in the null space of the retrieval gain matrix, G, i.e. does not affect the retrievals. Nevertheless our limited experiments shows that its removal can affect the retrievals considerably (red curve in the figure to the right). For these cases the impact on the retrieved temperature profile seems to be of the same order (and mostly in the same direction) as the effect of filtering the forward model radiances.

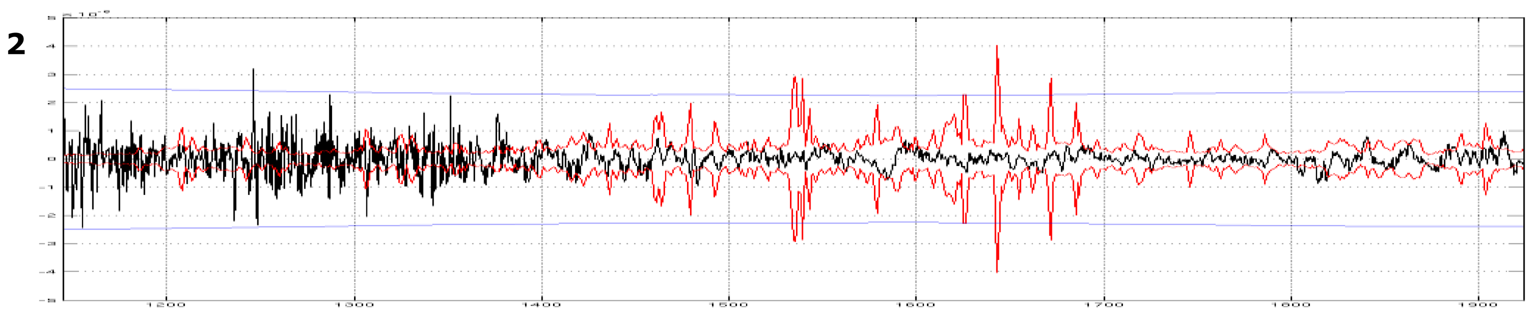


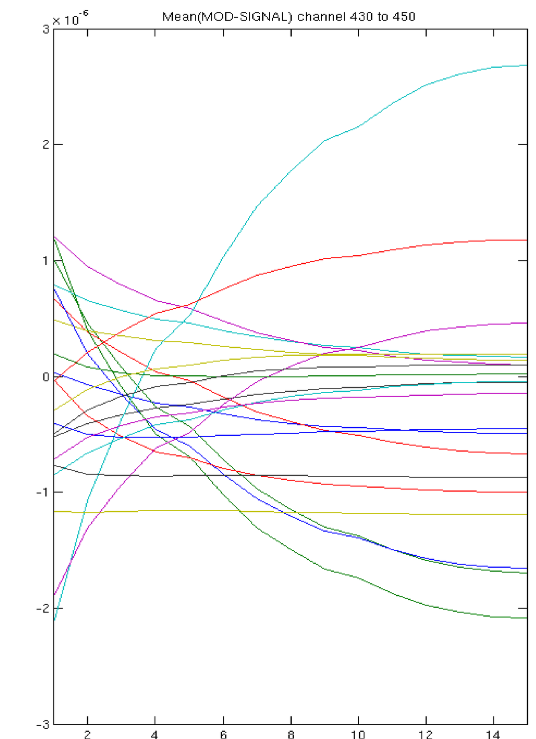
Bias and standard deviation of residual after projecting RTTOV 10 spectra onto signal space.

BAND 1



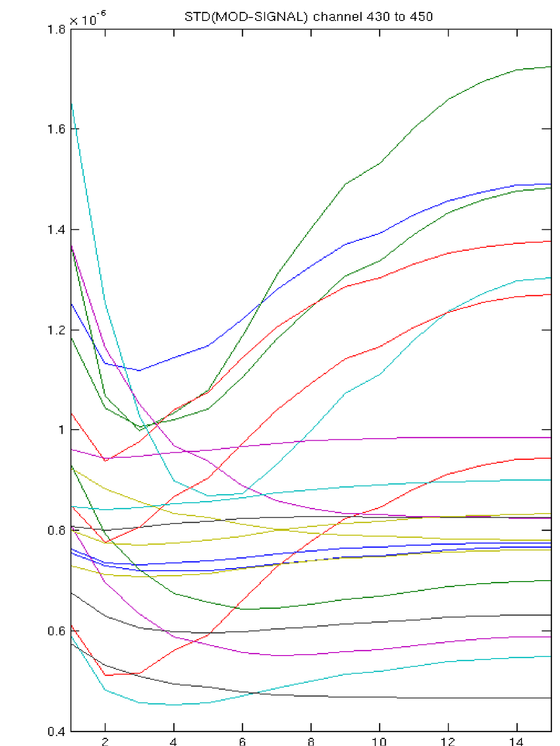
BAND 2

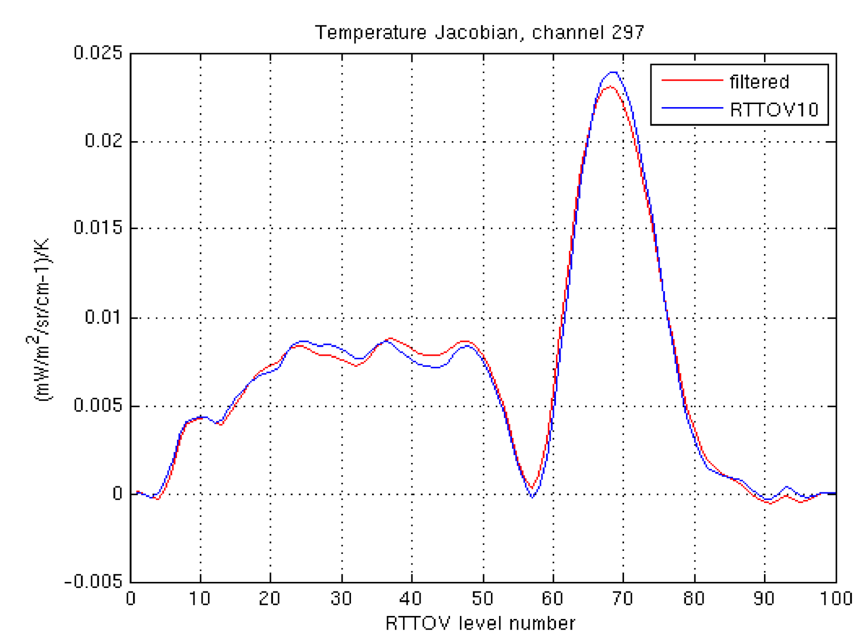




Mean and standard deviation of observation error removed by projecting synthetic RTTOV 10 spectra onto signal space, for selected channels, as function of scan position.

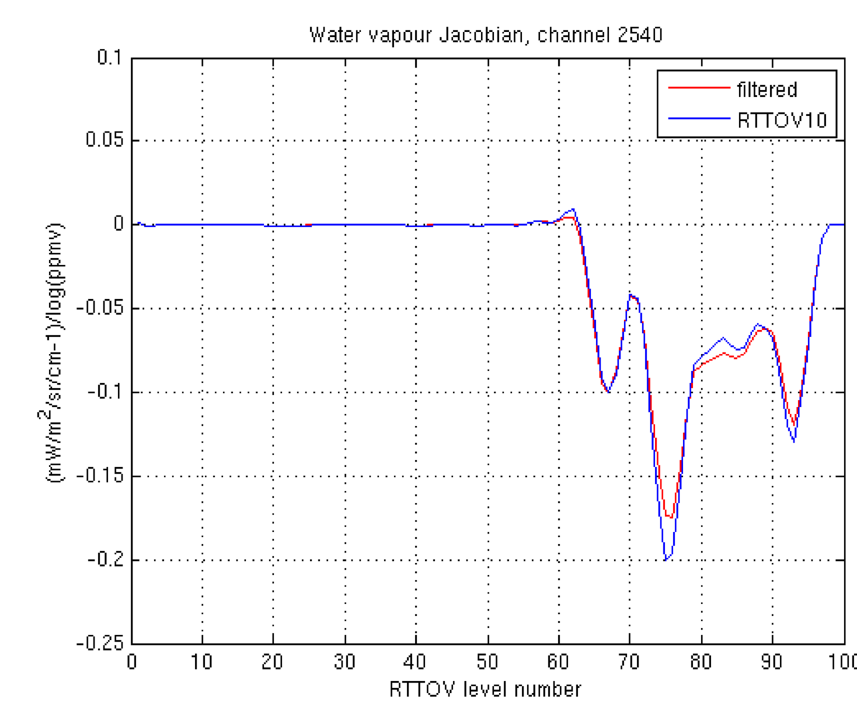
The component belonging to forward model space but being orthogonal to signal space shows a clear scan angle dependence for some, but not all, channels. This is consistent with the behaviour of the scan angle dependence observed in IASI obs minus calc statistics and provides a justification for using scan angle dependent bias correction.

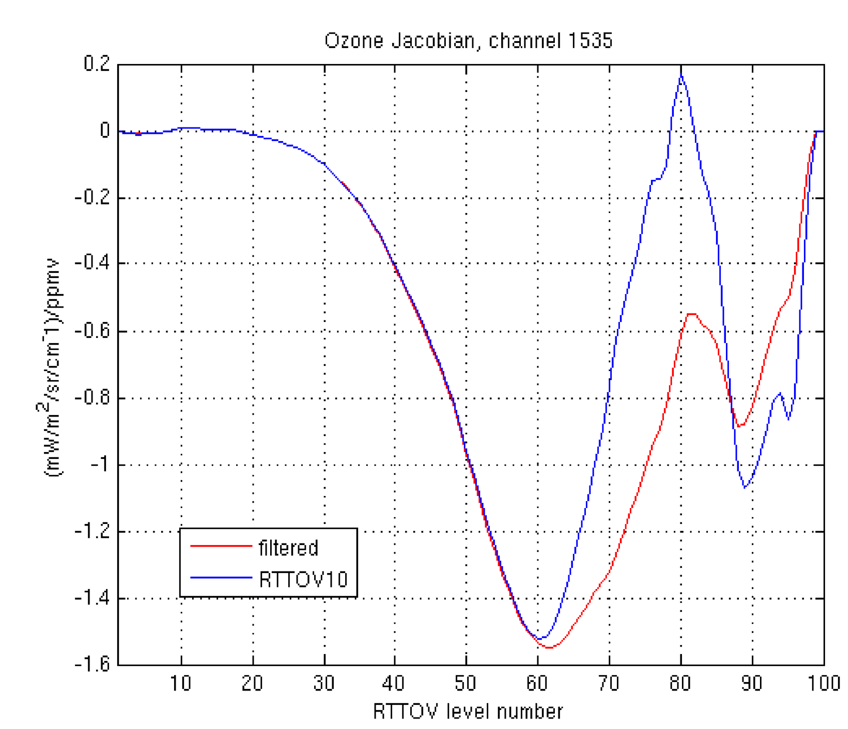




Comparison of Temperature, Water vapour and Ozone Jacobian in three selected channels obtained with RTTOV 10 and with RTTOV 10 filtered by projection onto signal space.

In most of the channels the effect on the filtering can hardly be noticed in the Jacobians, but here we plot Jacobians for some of the channels where the effect was found to be largest. Especially the effect on the tropospheric ozone Jacobian in channel 1535 is striking and suggests that the effect of the forward model filtering on tropospheric ozone retrievals could significant.





Conclusions.

In the classical 1D-Var, we iteratively adjust the atmospheric state vector in order to fit the measurements with synthetic spectra computed with a forward model - with the underlying assumption that the forward model space is a subspace of the signal space. We have computed eigenvectors generating the two subspaces and put in evidence some components of each subspace which are orthogonal to the other. We propose an approach where the forward model outputs are projected onto the signal space in order to perform the minimisation of the cost function in the space spanned by the observations. Since the purpose of synthetic spectra is to simulate measurements, it seems natural to expect that the forward model space is a subspace of the signal space and furthermore this condition must be fulfilled in order to ensure that the same results can be achieved with reconstructed radiances as with radiance PC scores. While this is not normally the case, we have shown how this can be achieved by projecting the forward model outputs onto the signal space, but to use this operationally a computationally efficient implementation within a PC forward model would have to be used. As long as this is not available, a temporary solution could be to correct for these observation errors by bias correcting using the mean of the residuals after projecting a representative sample of forward model spectra onto the signal space and adding the covariance to the observation error covariance used in the retrieval scheme.

Our initial experiments show that the effect of the proposed projection of the forward model outputs onto the signal space can be significant on the retrievals. But we are aware, that these results are sensitive to the retrieval set-up, in particular to the observation and background error covariance matrices. Further investigation is therefore required to determine if this proposal can improve retrievals from hyper-spectral instruments meaningfully. The same applies to the similar idea of projecting the measurements onto the forward model space.

The philosophy of EUMETSAT's operational IASI L2 1D-Var scheme - using a subset of reconstructed radiances combined with a full observation error covariance matrix - has been supported theoretically and is believed to be one of the reasons behind the very good quality of the temperature profile product. We plan to maintain this philosophy and improve it further by using a channel selection specifically tailored for reconstructed radiances as presented in this poster.