A Scaled Linear Retrieval for IASI

Anu Dudhia & Robert Hargreaves, NCEO & AOPP, Dept. Physics, University of Oxford

ABSTRACT

The IASI (Infrared Atmospheric Sounding Interferometer) instruments are nadir-viewing fourier-transform spectrometers flown on the MetOp satellites (currently two in orbit, with a third due for launch in November 2018). Each instrument routinely acquires over a million spectra a day. The challenge of processing such a large quantity of data has stimulated the development of a variety of fast 'linear' methods, essentially just taking the dot product of the IASI spectrum with a pre-defined 'gain' vector to extract a number representing the concentration of a target molecule.

Linear methods work well enough for the qualitative detection of sporadically occurring species such as volcanic SO2, or NH3, but a fundamental limitation in deriving more quantitative estimates is that the gain vector itself depends on the atmospheric state: the size of a molecular absorption peak depends not only on the concentration of the molecule, but also on other factors such as the temperature contrast between the earth's surface and the atmosphere.

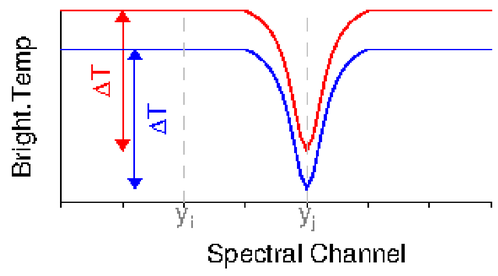
A modified linear retrieval scheme has been developed where, in addition to the standard linear retrieval, a second parameter is retrieved which represents the scaling factor for the gain vector. As well as providing a more accurate estimate of the concentration, this also generates a meaningful error estimate.

LINEAR RETRIEVALS

If the IASI spectrum is represented as a vector **y** and the retrieved parameter (eg column amount of NH3) as a scalar *x*, then a linear retrieval relates the two using a pre-defined gain vector **g**:

$$x = \mathbf{g} \cdot \mathbf{y}$$

A simple example of this is the Brightness Temperature Difference (BTD) method. This derives the strength of an absorption feature relative to a variable background (e.g. surface temperature) by taking the difference in radiance, expressed as brightness temperature, between points outside (y_i) and inside (y_j) an absorption line.



In this case the column amount $x = g(y_i - y_j)$, where *g* is a constant scale factor relating ΔT to column amount, so here **g** = *g*(1,-1) for the two spectral points (*i,j*), and zero elsewhere.

The BTD method is limited in using just a single spectral line and a single error term (ie the background radiance). However, it can be generalised if we reformulate this as a least squares fit (LSF) problem

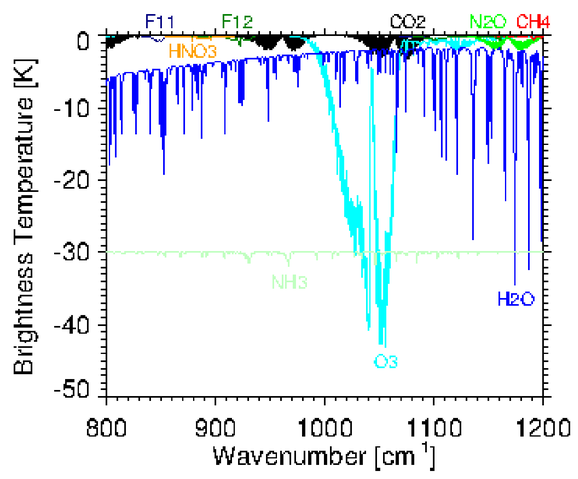
$$\mathbf{x} = (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{y} \equiv \mathbf{G}\,\mathbf{y}$$

where **x** is now a vector of jointly-retrieved parameters, **K** the Jacobian matrix d**y**/d**x**, and $\mathbf{S}_y$ the measurement error covariance matrix.

Various linear retrieval methods have been developed where error terms are either represented as jointly-retrieved parameters in **x** and/or additional components of $\mathbf{S}_y$. However these all suffer from the common problem of assuming a fixed relationship (*g,* **g** or **G**) between absorption depth and concentration.

SLR METHOD

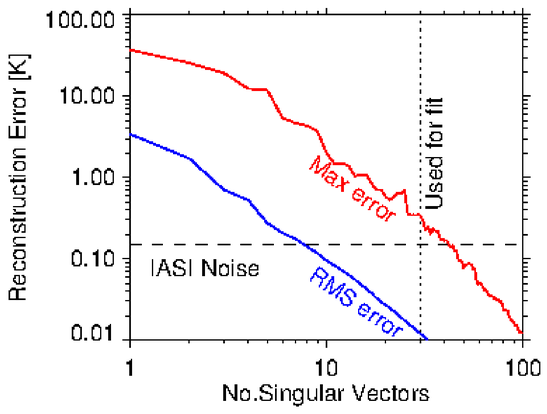
1 The first step is to identify a spectral region containing the signature from the target molecule.



Simulation of IASI measurements in the 800-1200 cm^{-1} (12.5-8.3 μm) spectral region showing the absorption features of various molecules, with NH3 offset and magnified for clarity.

2 Spectra for this region are simulated (using the RFM) for a set of 83 atmospheric profiles, chosen to represent the full range of global diversity, coupled with 4 different surface/cloud conditions and two viewing angles: a total of 664 spectra. These calculations exclude the target molecule.

3 Apply Singular Vector Decomposition (SVD) to extract the 30 leading spectral patterns.

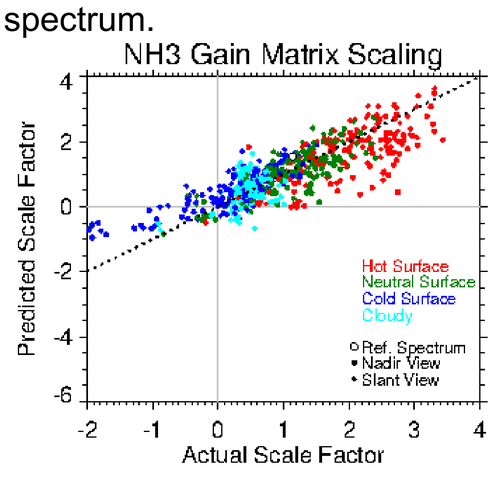


Error reconstructing simulated spectra from singular vectors. 10 SVs are sufficient to obtain a RMS error smaller than the IASI noise, while around 40 will fit every point within the noise.

4 Define a linear LSF retrieval with **x** representing a joint retrieval of the target molecule with the amplitudes of the 30 SVs. Since only the target molecule concentration is required (one row of the gain matrix **G**) this is still just a dot product: $x = \mathbf{g}.\mathbf{y}$. The SVs basically act an extension of the 'flat' background term in the BTD method to include other patterns of variability.

Up to this point, the technique suffers the same limitation as other linear methods: **G** is determined assuming a fixed **K**, ie assuming the depth of an absorption line depends only on concentration.

5 Recalculate the 664 spectra, this time with the target molecule incuded, and apply the linear LSF retrieval to these. By comparing the result with the known concentration used in each case we can determine a correction 'scale factor' for each spectrum.



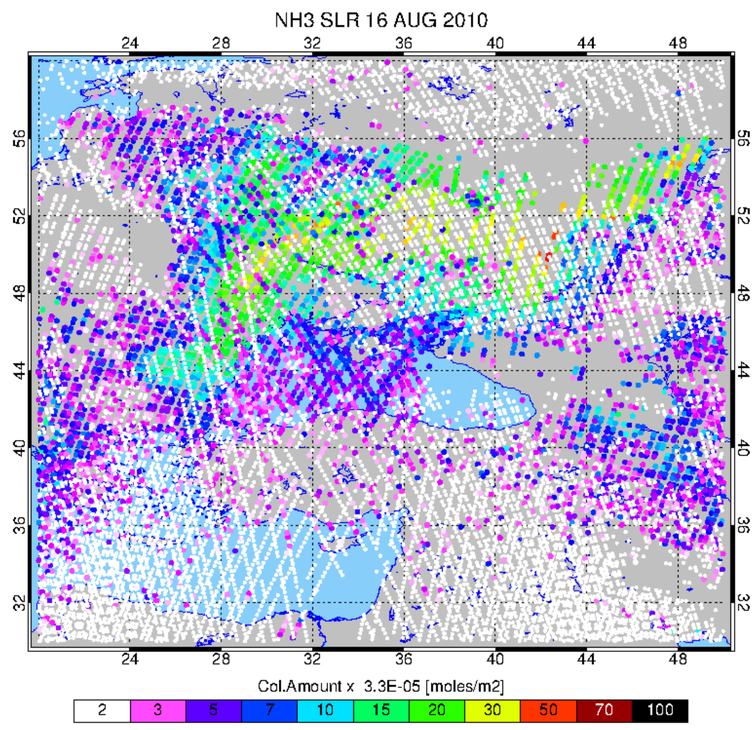
The x-axis shows the linear retrieval of NH3 expressed as a multiple of the 'true' NH3 for 664 spectra.

The y-axis shows the attempt to recover this scale factor from the co-retrieved SV amplitudes.

Note in particular the negative factors, where the atmosphere is warmer than the surface so the 'absorption' lines invert and appear as 'emission' lines. Use linear regression to find best fit **a** of scale factor *s* to retrieved SV amplitudes: $s = \mathbf{a}.\mathbf{x}$

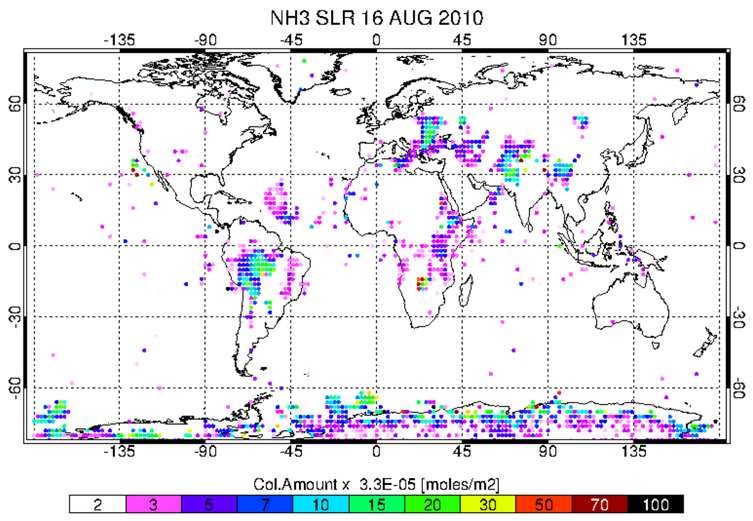
6 This scale factor *s* can be incorporated into the linear LSF retrieval as a second explicitly retrieved parameter. **x** is then retrieved as a 2-element vector whose components are finally multiplied to generate the SLR product.

RESULTS



The above plot shows results for individual IASI pixels on 16 August 2010 showing enhanced NH3 levels associated with forest fires over western Russia. This combines 2 day-time orbits (09:30 local time) with the satellite moving NE to SW, and 2 night-time (21:30), SE to NW. Gaps are where more the pixel is flagged as >50% cloudy and no retrieval is attempted.

The plot below shows the same day's data for the whole globe, gridded in 2x2° lat,lon bins. This takes of the order of 15 minutes to generate, using a single CPU on a linux desktop machine.



This also shows enhancements over other regions which are still to be investigated, although the Antarctic values are almost certainly an artefact associated with the cold surface temperatures.

ERROR ESTIMATION

Conventionally the retrieval error covariance $\mathbf{S}_x$ associated with a LSF gain matrix **G** is related to the measurement error covariance $\mathbf{S}_y$ by

$$\mathbf{S}_x = \mathbf{G}\,\mathbf{S}_y\,\mathbf{G}^T = (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K})^{-1}$$

($\mathbf{S}_x$ reduces to just the variance of *x* in the case of a single parameter retrieval.) Since $\mathbf{S}_y$, representing the instrument noise, is constant it can be seen that if **G** is also constant this will generate the same retrieval error for all cases, ie not useful for quality control.

However, for the SLR technique, **G** is scaled by a factor *s*: the larger the sensitivity of the absorption line to concentration (**K**↑), the smaller the scale factor *s*, hence the smaller the retrieval error ($\mathbf{S}_x$↓), as would be expected.

$$\mathbf{S}_x^{SLR} = s^2\,\mathbf{S}_x^{std}$$

Hence a second benefit of the SLR technique is that a meaningful error can be assigned.