



PERGAMON

Journal of Quantitative Spectroscopy &  
Radiative Transfer 73 (2002) 529–543

Journal of  
Quantitative  
Spectroscopy &  
Radiative  
Transfer

www.elsevier.com/locate/jqsrt

# A Hotelling transformation approach for rapid inversion of atmospheric spectra

Patrick Eriksson<sup>a,\*</sup>, Carlos Jiménez<sup>a</sup>, Stefan Bühler<sup>b</sup>, Donal Murtagh<sup>a</sup>

<sup>a</sup>*Department of Radio and Space Science, Chalmers University of Technology, SE-412 96 Göteborg, Sweden*

<sup>b</sup>*Institute of Environmental Physics, University of Bremen, P.O. Box 330440, D-28334 Bremen, Germany*

Received 20 February 2001; accepted 24 May 2001

---

## Abstract

Atmospheric observations from space often result in spectral data of large dimensions. To allow an optimal inversion of the observed spectra it can be necessary to map the data into a space of smaller dimension. Here several data reduction techniques based on eigenvector expansions of the spectral space are compared. The comparison is done by inverting simulated observations from a microwave limb sounder, the Odin-SMR. For the examples tested, reductions exceeding two orders of magnitude with no negative influence on the retrieval performance are demonstrated. The techniques compared include a novel method developed especially for atmospheric inversions, based on the weighting functions of the variables to be retrieved. The new method shows an excellent performance in practical tests and is both computationally more effective and more flexible than the standard Hotelling transformation. © 2002 Elsevier Science Ltd. All rights reserved.

*Keywords:* Hotelling transformation; SVD; Inversion problem; Passive observations

---

## 1. Introduction

Observations from space are becoming more and more important for investigations of our atmosphere. The sensors are normally placed on-board satellites [1], but the space shuttle has also been used [2] and the international space station can become another relevant platform in the future [3]. The atmospheric sensors are becoming more numerous, using both new

---

\* Corresponding author. Fax: +46-31-7721884.

*E-mail address:* patrick@rss.chalmers.se (P. Eriksson).

and broader wavelength regions, and measuring with better frequency resolution and lower noise. As the data analysis methods mature at the same time, there is a parallel shift toward more complex calculation approaches for the inversion, the process of extracting the information from the observed spectra. All this results in much larger data amounts that must be handled during the inversions and effective means of reducing the data sizes would be advantageous.

Especially occultation and emission measurements performed in limb sounding mode give rise to large scale inversion problems. Particularly as it is normal practice to treat all spectra from a scan sequence as a single observation and to invert them simultaneously. Accordingly, the length of the measurement vector is  $n = mp$ , where  $m$  is the number of spectrometer channels and  $p$  the number of spectra in a scan. In addition, for some sensors it is planned to invert jointly several scans, or even the scans from a total orbit [4,5], increasing the data size even further with about 1–2 magnitudes. The normal strategies to handle these inversions are to invert parts of the total spectra separately and to handle the spectral uncertainties in a simplified manner, what can result in that the full capability of the measurements may not be reached.

For most traditional inversion methods, such as the optimal estimation method [6], the critical step is to calculate the inverse of the covariance matrix describing the uncertainties of observed spectra,  $S_e$ . The general cost for inverting this matrix is proportional to  $n^3$ , which puts severe limitations on  $n$ . To avoid this problem the matrix  $S_e$  is in general set to be diagonal, modeling only errors uncorrelated between the spectrometer channels, that is, thermal noise. Inclusions of other error sources in  $S_e$  can improve the inversion accuracy and helps to stabilize the inversion problem [7]. To allow  $S_e$  to have off-diagonal elements is especially important for cases where the magnitude of the thermal noise is low [8], a situation becoming more common with newer, less noisy, receivers.

Inversion approaches exist that do not require the inverse of  $S_e$ , one example is the use of neural networks (e.g. [9]). Reducing the dimension of the input data to the neural network is a common procedure in order to reduce the computational burden and enhance the generalization properties of the neural network, so a reduction technique is normally an essential component of the required pre-processing of the spectral data.

Accordingly, methods to reduce the size of the measured data without information loss could enable more detailed treatments of observation uncertainties, more large scale inversions and the application of novel inversion approaches, that should result in improved accuracy of the retrieved atmospheric data. Several data reduction methods based on eigenvector expansions, here called Hotelling transformations, are compared. This includes a novel approach, designed especially for the purpose of atmospheric inversions, that has advantages regarding performance, calculation efficiency and flexibility.

The next section gives a more detailed discussion about the limitations set by the inversion methods on the data size. Section 3 describes the Hotelling transformations that have been considered. The data reduction approaches have been compared by simulating observations of the Odin sub-mm radiometer, and the simulation conditions are given in Section 4. The results are found in Section 5 and finally Section 6 gives the conclusions.

## 2. Retrieval theory and methods

Following Rodgers [10], a forward model,  $F$ , is introduced to describe the relationship between the various influential variables and the observed spectra,  $\mathbf{y}$ :

$$\mathbf{y} = F(\mathbf{x}, \mathbf{b}) + \boldsymbol{\varepsilon}, \quad (1)$$

where  $\mathbf{x}$  is the state vector representing the variables to be retrieved from the observation,  $\mathbf{b}$  contains additional atmospheric and sensor variables, and  $\boldsymbol{\varepsilon}$  represents measurement errors such as thermal noise. It will be assumed here that the forward model is sufficiently linear to allow a linearization around an a priori state  $(\mathbf{x}_a, \mathbf{b}_a)$ :

$$\mathbf{y} = F(\mathbf{x}_a, \mathbf{b}_a) + \mathbf{K}_x(\mathbf{x} - \mathbf{x}_a) + \mathbf{K}_b(\mathbf{b} - \mathbf{b}_a) + \boldsymbol{\varepsilon}, \quad (2)$$

where  $\mathbf{K}_x = \partial F / \partial \mathbf{x}$  and  $\mathbf{K}_b = \partial F / \partial \mathbf{b}$  are the state and model parameter weighting function matrix, respectively. The a priori state is our best beforehand estimate of  $(\mathbf{x}, \mathbf{b})$ . Two inversion methods to retrieve the wanted information ( $\mathbf{x}$ ) from the observed spectra ( $\mathbf{y}$ ) are presented below, with focus on the importance of the length of  $\mathbf{y}$ .

### 2.1. The optimal estimation method, OEM

The most commonly applied retrieval method for atmospheric passive observations is the optimal estimation method [6]. Other plausible names for this method are the minimum variance method (e.g. [11]) and statistical regularization (e.g. [7]). If the forward model is linear with respect to  $\mathbf{x}$ , the OEM solution can be written as

$$\hat{\mathbf{x}} = \mathbf{x}_a + (\mathbf{K}_x^T \mathbf{S}_e^{-1} \mathbf{K}_x + \mathbf{S}_x^{-1})^{-1} \mathbf{K}_x^T \mathbf{S}_e^{-1} [\mathbf{y} - F(\mathbf{x}_a, \mathbf{b}_a)], \quad (3)$$

where  $\mathbf{S}_x$  is a covariance matrix reflecting our knowledge of  $\mathbf{x}$  and  $\mathbf{S}_e$  is the covariance matrix for the observation uncertainties:

$$\mathbf{S}_e = \mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T + \mathbf{S}_\varepsilon, \quad (4)$$

where  $\mathbf{S}_b$  and  $\mathbf{S}_\varepsilon$  are the covariance matrix for  $\mathbf{b}$  and  $\boldsymbol{\varepsilon}$ , respectively. The expected error can be estimated as

$$\mathbf{S}_\delta = (\mathbf{K}_x^T \mathbf{S}_e^{-1} \mathbf{K}_x + \mathbf{S}_x^{-1})^{-1}. \quad (5)$$

Other forms of the OEM solution exist but when the length of  $\mathbf{y}$  exceeds that of  $\mathbf{x}$ , which should normally be the case, Eq. (3) is the most computationally effective expression. The computational cost of Eq. (3) depends greatly on the size and structure of  $\mathbf{S}_e$ . This matrix is often set to be diagonal and models only thermal noise ( $\mathbf{S}_\varepsilon$ ), then the calculation of  $\mathbf{S}_e^{-1}$  is a simple operation. However, to improve the retrieval accuracy and to decrease the sensitivity to the assumptions connected to  $\mathbf{S}_x$ , forward model uncertainties ( $\mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T$ ) should also be included in  $\mathbf{S}_e$  [7] and the inversion of  $\mathbf{S}_e$  becomes a crucial step. The computational cost to invert  $\mathbf{S}_e$  is proportional to  $n^3$ , where  $n$  is the length of  $\mathbf{y}$ . In other words, if  $\mathbf{S}_e$  is set to be diagonal, OEM allows basically any length of  $\mathbf{y}$ , but if  $\mathbf{S}_e$  has off-diagonal elements the data size is a crucial factor.

## 2.2. Neural networks, NNs

A more novel approach for inversion of passive atmospheric observations is the application of neural networks (NNs) (e.g. [12–14,9]). NNs perform non-linear mappings between sets of variables. They can be regarded as functions with a set of adjustable parameters that are determined during a training phase. In this case, the training set consists of pairs of spectra and the corresponding atmospheric variables. After training, inverting an atmospheric observation is reduced to the computation of a function, typically involving a few simple matrix operations. Traditional approaches, such as OEM, require much more lengthy iterative calculations for non-linear inversions, involving for example the recalculation of  $\mathbf{K}_x$ , so the main advantage of NNs is that they allow comparably fast calculations for non-linear cases.

If there is no loss of essential information, reducing the dimension of the input space to a NN affects positively its mapping ability and computational efficiency. The larger the dimension of the input data, the larger the number of adjustable parameters, and the larger the training set needed to properly constrain them, phenomenon known as the “curse of dimensionality” [15]. A smaller number of parameters also contributes to enhancing the generalization properties of the NN, as NNs with a small number of parameters provide smoother mappings. A lower dimensional space is also beneficial for the representation of the underlying function generating the mapping, because it helps to discard the undesired features of the mapping introduced by the limitations of the training set. All these benefits make the reduction in the dimension of the input data a normal procedure when designing NNs, even for a relatively low dimensional input space.

## 3. Hotelling transformations

### 3.1. Standard formulation

A very common approach for compression of geophysical data is the Hotelling transformation. The method is also known under the names: principal component analysis, Karhunen–Loeve transformation and empirical orthogonal function analysis. The Hotelling transformation is based on a decomposition of the covariance matrix describing the variability of the observed quantity:

$$\mathbf{S}_y = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T, \quad (6)$$

where  $\mathbf{E}$  is an orthogonal unitary matrix ( $\mathbf{E}^T\mathbf{E} = \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix) and  $\mathbf{\Lambda}$  is a diagonal matrix with non-negative values. The columns of  $\mathbf{E}$  are denoted as eigenvectors and the diagonal elements of  $\mathbf{\Lambda}$ , the eigenvalues, are ordered in non-increasing order. The eigenvalues express the fraction of the total variance that is associated with the corresponding eigenvector. With these definitions, the transformation, for a given dimension  $k$ , minimizing the mean-square error between original ( $\mathbf{y}$ ) and transformed ( $\bar{\mathbf{y}}$ ) vectors is

$$\bar{\mathbf{y}} = \mathbf{E}_k^T \mathbf{y} \quad (7)$$

where  $\bar{\mathbf{y}}$  has length  $k$  and  $\mathbf{E}_k$  is the part of  $\mathbf{E}$  holding the first  $k$  eigenvectors.

The standard Hotelling transformation has some drawbacks that should be considered. The matrix  $\mathbf{S}_y$  is determined empirically from a set of measurements and during the preparatory phase of a project  $\mathbf{S}_y$  cannot be calculated as relevant data do not exist. The transformation only considers the variability in the data, independently of whether it is noise or interesting information, and there is no mechanism for putting emphasis on special features in the data. If the length of  $\mathbf{y}$  is large it can even be impossible to handle  $\mathbf{S}_y$  practically due to limited computer memory. However, it should be noted that it is not needed to calculate  $\mathbf{S}_y$  explicitly to determine the eigenvectors. If  $\mathbf{Y}$  is an ensemble with  $n$  measurements where the mean values have been subtracted, the covariance matrix can be estimated as

$$\mathbf{S}_y = \frac{1}{n-1} \mathbf{Y}\mathbf{Y}^T. \quad (8)$$

The eigenvectors are then obtained by a singular value decomposition (SVD) of  $\mathbf{Y}$ :

$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{Y}, \quad (9)$$

as the properties of SVD give that ( $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ )

$$\frac{1}{n-1} \mathbf{Y}\mathbf{Y}^T = \mathbf{U} \frac{\mathbf{\Sigma}^2}{n-1} \mathbf{U}^T. \quad (10)$$

By comparing Eqs. (6) and (10) we see that

$$\mathbf{E} = \mathbf{U}. \quad (11)$$

### 3.2. Correlation formulation

Another common procedure is to base the Hotelling transformation on the correlation coefficient matrix,  $\mathbf{R}_y$ , instead of  $\mathbf{S}_y$ :

$$\mathbf{R}_y = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T. \quad (12)$$

The correlation coefficient matrix is the covariance matrix obtained when the data are normalized with the standard deviations. If the covariance matrix is decomposed, most emphasis will be put on the data points with the largest variance, while if the correlation matrix is used all the data will be weighted equally. This fact can be of importance if the species of interest give emission of very different magnitudes, as investigated below.

### 3.3. Formulation based on $\mathbf{K}_x$

Passive atmospheric observations are normally very well characterized as this is necessary for the retrieval process. If Eq. (2) is valid, the matrix  $\mathbf{S}_y$  can be expressed analytically as

$$\mathbf{S}_y = \mathbf{K}_x \mathbf{S}_x \mathbf{K}_x^T + \mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T + \mathbf{S}_\varepsilon. \quad (13)$$

The Hotelling transformation aims at maintaining a maximum fraction of the variability of  $\mathbf{y}$ , but Eq. (13) shows that this variability consists partly of noise ( $\mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T + \mathbf{S}_\varepsilon$ ), that is only a

disturbing factor for the retrieval. To optimize the data reduction for the retrieval, the reduction should only be based on the part of  $\mathbf{S}_y$  containing the interesting information:

$$\mathbf{E}\mathbf{A}\mathbf{E}^T = \mathbf{K}_x\mathbf{S}_x\mathbf{K}_x^T. \quad (14)$$

If the Cholesky decomposition of  $\mathbf{S}_x$  is calculated,

$$\mathbf{L}\mathbf{L}^T = \mathbf{S}_x, \quad (15)$$

we have that  $\mathbf{K}_x\mathbf{S}_x\mathbf{K}_x^T = \mathbf{K}_x\mathbf{L}(\mathbf{K}_x\mathbf{L})^T$  and the matrix  $\mathbf{E}$  can be determined by a SVD of the product  $\mathbf{K}_x\mathbf{L}$  (cf. Eqs. (10) and (11)). For further details of the Cholesky decomposition and SVD, see e.g. [16].

These calculations are simplified if the matrix  $\mathbf{S}_x$  is set to be diagonal. The Cholesky decomposition is then simply obtained by taking the square root of the diagonal elements,  $d^i = \sqrt{\mathbf{S}_x^{ii}}$ . In general, it is not possible to set  $\mathbf{S}_x = \mathbf{I}$  as the weighting functions can correspond to different units, etc. The scaling by  $d^i$  is needed to put all weighting functions on the same scale. Even if only species are retrieved and the same unit, for example the volume mixing ratio (VMR), is used throughout, a scaling is needed as the species abundance is not constant, neither between the species or as a function of altitude. However, as only species are retrieved here and the retrievals are performed in relative units (with respect to the a priori profiles), all diagonal elements of  $\mathbf{S}_x$  will be of the same order (see Eq. (18)) and it is possible to set  $\mathbf{S}_x$  to equal  $\mathbf{I}$ . The implications of using a diagonal  $\mathbf{S}_x$  are further discussed in Section 5.1.

The scheme based on  $\mathbf{K}_x$  requires in most cases less computations and computer storage than the standard Hotelling transformation. The calculations are especially simple when  $\mathbf{S}_x$  is set to be diagonal. The  $\mathbf{K}_x$  scheme has also the potential of being more effective as it is based directly on the weighting function matrix of the variables to be retrieved. As shown below, it gives a further possibility to control the data reduction to ensure that the information needed to retrieve variables causing only a low variability in the observations is not lost. However, Eq. (14) assumes linearity (for  $\mathbf{x}$ ) and the performance of the scheme in non-linear cases must be tested for each individual retrieval situation.

It is worth noting that this  $\mathbf{K}_x$  reduction approach has similarities with the truncated SVD (TSVD) retrieval method [17,18]. In fact, the TSVD solution is obtained by OEM if the  $\mathbf{K}_x$  data reduction is applied and  $\mathbf{S}_x^{-1}$  is set to zero, but there is a crucial difference. The eigenvector expansion is here used to find an effective data reduction, that is, no information shall be lost, and there is no risk connected with using a too high value for  $k$ . The situation is different for TSVD as the variable  $k$  is used to regularize the solution, thus replaces  $\mathbf{S}_x$  in OEM, and a too high value for  $k$  is disastrous. With other words, in TSVD the eigenvector expansion is used to filter the data and make the inversion problem well-posed.

A final remark is that the  $\mathbf{K}_x$  from Eqs. (3) and (14) do not have to be the same. For instance, if the species retrieval is performed on a coarse grid it can be justified to make a special calculation of  $\mathbf{K}_x$ , using a finer grid, when determining  $\mathbf{E}$ .

### 3.4. Truncation issues

There is no general criteria for deciding how many eigenvectors should be retained in  $\mathbf{E}_k$ . The simplest tests are based on dominant variance approaches, for instance, the scree test [19], that

consists of plot eigenvalues versus the corresponding eigenvector numbers to find the location of a breaking point, or the Kaiser criterion [20], that consists in retaining only eigenvectors with eigenvalues larger than the average amount of variance.

The criteria suggested here is based on the information content of the retrieval. The retrieval error  $\mathbf{S}_\delta$  (Eq. (5)) can be estimated for different  $\mathbf{E}_k$  and expressed in terms of information content as [21]

$$H = \frac{1}{2} \log_2 |\mathbf{S}_\delta^{-1}|, \quad (16)$$

where  $|\mathbf{S}_\delta^{-1}|$  is the matrix determinant. The information content gained by the retrieval is

$$\Delta H = \frac{1}{2} \log_2 |\mathbf{S}_\delta^{-1}| - \frac{1}{2} \log_2 |\mathbf{S}_x^{-1}|. \quad (17)$$

The information content increases with the number of eigenvectors retained until it reaches a maximum level that shall coincide with the information content obtained when no data reduction is applied.

## 4. Simulation conditions

### 4.1. Odin-SMR

Simulations of observations by the Odin sub-mm radiometer (Odin-SMR) were selected to exemplify the different data reduction approaches. Odin-SMR observations will result in large, high-resolution, data sets, so there is a clear need for a data reduction technique.

Odin is a small, low cost, satellite, build as a collaboration between Sweden, Canada, Finland and France, that will perform the first space based sub-mm observations of the atmosphere. The launch of Odin is scheduled for the year 2001. Technical details and the expected retrieval performance of Odin-SMR are presented by [22] and the Odin-SMR forward model, that is also used here, is described in [23]. The most important technical characteristics (in this context) are described below.

Odin-SMR has four sub-mm front-ends, tunable inside the frequency ranges 486.1–503.9 and 541.0–581.4 GHz. Three spectrometers can be connected to any of the front-ends. The four parts of the two hybrid autocorrelators (ACs) have around 420 channels each. The third spectrometer (an AOS with 1720 channel) is not considered here as it is only used during parts of the mission due to power limitations. Spectra will be recorded every 2 s, separated 1.5 km in tangent altitude.

During a scan through the atmosphere, Odin-SMR will record about 35–45 spectra depending on which of the four observation modes is used. The normal approach for limb sounding is to append all spectra from a scan and perform a joint inversion. The narrowest frequency range that will be considered for inversions corresponds to one AC part and accordingly 14 700 is the minimum length of  $\mathbf{y}$  for Odin-SMR retrievals. When the frequency bands of an AC are adjacent, it is justified to invert simultaneously all data from an AC, corresponding to a length of  $\mathbf{y}$  of 37,800 for the largest scan range.

Fig. 1 shows examples of noise free spectra from the three frequency bands considered in this study. All bands have two dominating features caused by a single or a cluster of transitions. Band 576.4 GHz covers also interesting transitions of BrO, NO<sub>2</sub> and ClO with weak emission.

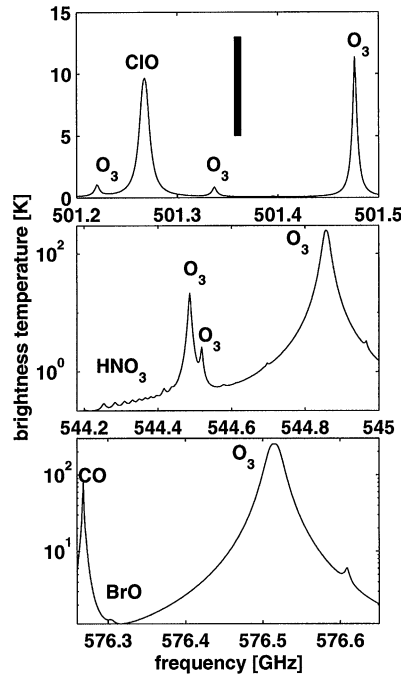


Fig. 1. Simulated noise-free spectra for the 501.4, 544.6 and 576.4 GHz Odin-SMR bands. The spectra correspond to a tangent altitude of 35 km. Unlabelled transitions correspond to  $O_3$  isotopes. The vertical mark indicates the magnitude of the thermal noise ( $\pm 1\sigma$ ).

The magnitude of the thermal noise is also indicated in the figure, about 8 K ( $\pm 1\sigma$ ) for an integration time of 0.875 s and a channel bandwidth of 1.16 MHz.

#### 4.2. The atmosphere

The mean atmospheric conditions were taken from the data set REFMOD92, based on various measurements and models. For example, the pressure and temperature profiles are taken from the 1976 US Standard Atmosphere and the ozone profile from [24]. Atmospheric profiles were generated randomly by applying the Cholesky decomposition method [25]. Gaussian statistics and hydrostatic equilibrium were assumed throughout. Following [7,26], the variability of all species was modeled as

$$S(i, j) = \sigma^2 e^{-|z(i) - z(j)|/l_c}, \quad (18)$$

where  $z$  is the vertical altitude,  $i$  and  $j$  are altitude indices,  $\sigma$  is the standard deviation, set to 30%, and  $l_c$  is the correlation length, set to 4 km. The temperature variability was modeled following Eq. (18) with  $\sigma = 4$  K and  $l_c = 6$  km, but with a correlation linearly decreasing down to zero.



### 4.3. Retrieval approach

The inversions were performed by OEM where no other variables beside species profiles were retrieved. A perfect knowledge of the species variability was assumed. In other words, the  $\mathbf{S}_x$  matrix is given by Eq. (18). The matrix  $\mathbf{S}_e$  was set to only include thermal noise, this to enable retrievals without any data reduction. The retrievals are thus not “optimal” as no information on the temperature variability is provided to OEM.

## 5. Results

### 5.1. A linear retrieval case

The retrieval of  $\text{O}_3$  and  $\text{ClO}$  from Odin-SMR observations around 501.4 GHz was chosen as a first test. The emissions of  $\text{O}_3$  and  $\text{ClO}$  in this band have similar magnitudes (Fig. 1), and the retrieval problem can be handled by assuming linearity [27]. Information content as a function of  $k$  is plotted in Fig. 2, while Fig. 3 shows the practical retrieval error for 500 inversions where  $k$  was set to 50 for the reduction techniques considered. The most prominent result of these figures is the poor performance of the standard Hotelling transformation.

If the noise has zero mean, is uncorrelated and has the same magnitude for all measurement channels ( $\mathbf{S}_e = \sigma^2 \mathbf{I}$ ), it can be shown that the same eigenvectors are obtained with and without noise (e.g. [28]). Although these criteria are fulfilled during the simulations, close inspection of the  $\mathbf{S}_e$  derived from simulated noise shows a covariance matrix with off-diagonal elements having random numbers around zero, resulting from having a limited data set. This deviation from perfect statistics causes the noise to have a practical influence on the data reduction, despite the high number of 1000 spectra used to calculate the eigenvectors of  $\mathbf{S}_y$ . A decrease of the number of spectra used gives a further deterioration of the retrieval performance. To illustrate this negative effect of the thermal noise,  $\mathbf{E}_k$  was calculated from noise-free spectra. These results are indicated as  $\mathbf{S}_{y-\varepsilon}$  in Figs. 2 and 3, and a clear improvement compared to the noisy case is seen. Notice that this latter approach is included mainly as a reference as it cannot be based on, always noisy, measurement data. Note also that the thermal noise caused by the sub-mm receiver considered in this study has a high magnitude ( $\sigma = 4$  K) compared to, for example, infrared instruments, and the negative practical influence on the data reduction seen here can be considerably smaller for other sensors.

The results for  $\mathbf{K}_x \mathbf{K}_x^T$  and  $\mathbf{K}_x \mathbf{S}_x \mathbf{K}_x^T$  are very similar and just marginally poorer than when using noise-free spectra ( $\mathbf{S}_{y-\varepsilon}$ ). For the reductions based on  $\mathbf{K}_x$  it is important to notice that the exact variability of the retrieved variables was not needed to obtain good performance, setting  $\mathbf{S}_x$  to be diagonal did not deteriorate the results. This fact has positive practical consequences, such as a decrease in the computational burden. More important, the matrix  $\mathbf{S}_x$  is usually very difficult to estimate and, when lacking relevant data, the safest approach is to neglect all possible correlations for  $\mathbf{x}$ , thus giving a diagonal  $\mathbf{S}_x$ . This is the case as correlation for  $\mathbf{x}$  results in a higher correlation in the spectral data. Hence, overestimating the correlation for  $\mathbf{x}$  can result in that not all features of measured spectra can be captured by the data reduction, while with a diagonal  $\mathbf{S}_x$  all possible spectral patterns can be resolved, but at a plausible cost of needing

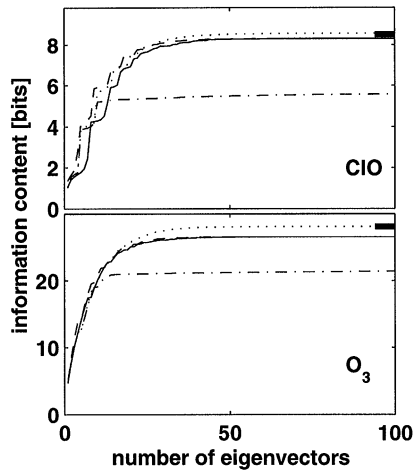


Fig. 2. Information content versus number of eigenvectors  $k$  for the retrieval of ClO and O<sub>3</sub> in the 501.4 GHz band. The results for  $\mathbf{E}_k$  derived from  $\mathbf{K}_x\mathbf{K}_x^T$  are plotted as —, from  $\mathbf{K}_x\mathbf{S}_x\mathbf{K}_x^T$  as ---, from  $\mathbf{S}_y$  as -·-, and from  $\mathbf{S}_{y-\varepsilon}$  as ···. The horizontal marks show the information content when no reduction is applied.

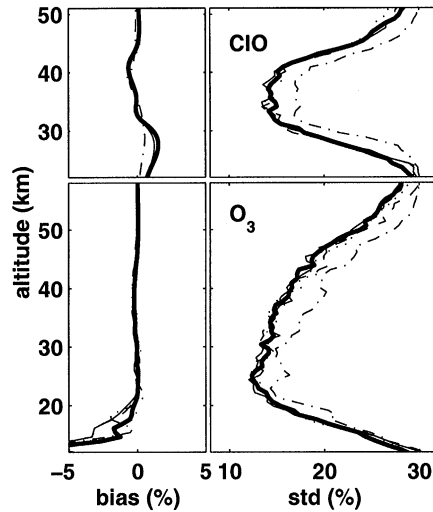


Fig. 3. Retrieval error for ClO and O<sub>3</sub> in the 501.4 GHz band. The mean (bias) and standard deviation (std) of the retrieval error for the different  $\mathbf{E}_k$ ,  $k=50$ , are plotted. The results without any reduction are plotted as a solid thick line, the remaining curves are labelled as in Fig. 2.

a slightly higher  $k$ , as can be observed in Fig. 2. Following this discussion and the results found in this section, of the two approaches for  $\mathbf{K}_x$  only the one with a diagonal  $\mathbf{S}_x$  matrix is considered below.

The information content has basically reached the maximum level already at  $k=50$  for all methods apart from when using  $\mathbf{S}_y$ . The worst retrieval performance in Fig. 3 corresponds to  $\mathbf{S}_y$ , while all the other reduction strategies give a similar performance to the case without reduction. These facts indicate that the information content works very well as a practical criterion to select the value of  $k$ .

## 5.2. A non-linear retrieval case

The derivation of  $\mathbf{E}_k$  based on  $\mathbf{K}_x$  assumes a linear inversion problem ( $\mathbf{K}_x$  is valid for all  $\mathbf{x}$ ), so it is interesting to see the performance of this reduction technique for non-linear cases. As an example, information content and inversions of 100 spectra in the 544.6 GHz band were calculated (Figs. 4 and 5). The 544.6 GHz band presents an inversion problem that can be described as moderately non-linear [27], and the inversions were done by using the Marquardt–Levenberg technique [8,21].

In contrast to Fig. 2, the approach based on  $\mathbf{K}_x$  shows for band 544.6 GHz a somehow faster increase, as a function of  $k$ , in information content than when using  $\mathbf{S}_{y-\varepsilon}$ . According to Fig. 4 the maximum level of information content is reached for a value of  $k$  around 50 for the  $\mathbf{K}_x$

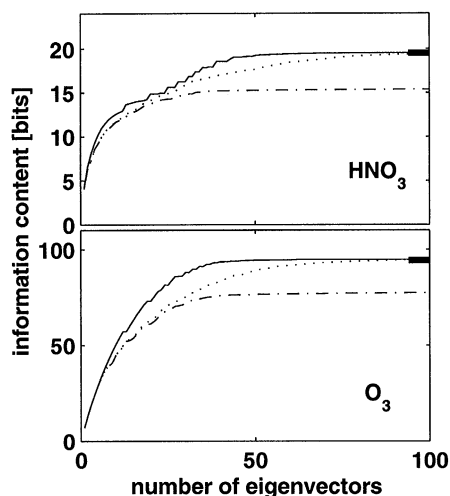


Fig. 4. Information content versus number of eigenvectors  $k$  for the retrieval of  $\text{HNO}_3$  and  $\text{O}_3$  in the 544.6 GHz band. The results for  $\mathbf{E}_k$  derived from  $\mathbf{K}_x\mathbf{K}_x^T$  are plotted as —, from  $\mathbf{S}_y$  as ---, and from  $\mathbf{S}_{y-\varepsilon}$  as ···. The horizontal marks show the information content when no reduction is applied.

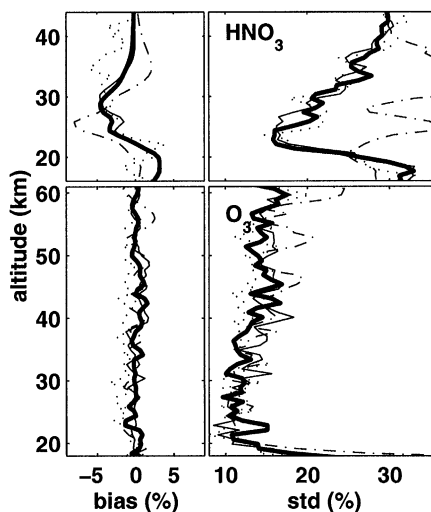


Fig. 5. Retrieval error for  $\text{HNO}_3$  and  $\text{O}_3$  in the 544.6 GHz band. The mean (bias) and standard deviation (std) of the retrieval error for the different  $\mathbf{E}_k$ ,  $k=50$  for  $\mathbf{K}_x$  and  $k=100$  for  $\mathbf{S}_y$  and  $\mathbf{S}_{y-\varepsilon}$ , are plotted. The results without any reduction are plotted as a solid thick line, the remaining curves are labelled as in Fig. 4. The curves are not as smooth as in Fig. 3 due to the smaller number of inversions.

reduction, while  $k$  around 100 is needed for the  $\mathbf{S}_{y-\varepsilon}$  reduction. Fig. 5 shows the retrieval error for these reductions and  $k$  values. For  $\mathbf{S}_y$   $k$  was set to 100 also, with the reduction performing poorly, for similar reasons as in the band 501.4 GHz.

Fig. 4 indicates the same relative performance between  $\mathbf{S}_y$  and the other two methods for  $\text{O}_3$  and  $\text{HNO}_3$  but the practical inversions using  $\mathbf{S}_y$  for  $\text{HNO}_3$  show notably large errors. The emission of  $\text{HNO}_3$  originates from a cluster transitions giving ragged spectra, while all other species have separated transitions, and a plausible cause to the poor  $\text{HNO}_3$  inversions is that the presence of noise in  $\mathbf{S}_y$  makes the non-linear inversions especially hard to handle for this particular case.

The example here for 544.6 GHz shows that, for moderately non-linear situations, the reduction based on  $\mathbf{K}_x$  with a selection criteria for  $k$  based on the information content works well, but the more non-linear the problem, the more cautiously this approach has to be used. This is the case as a basic assumption for the  $\mathbf{K}_x$  method is that a linear inversion problem is at hand (Eq. (13)) and for more non-linear inversion problems the reduction based on  $\mathbf{K}_x$  is expected to be less optimal. The information content is also based on an assumption of linearity, that the retrieval error can be calculated by Eq. (5), and the more non-linear the inversion problem is, the more likely it is that the information content misrepresents the information gain, making the selection of  $k$  more difficult.

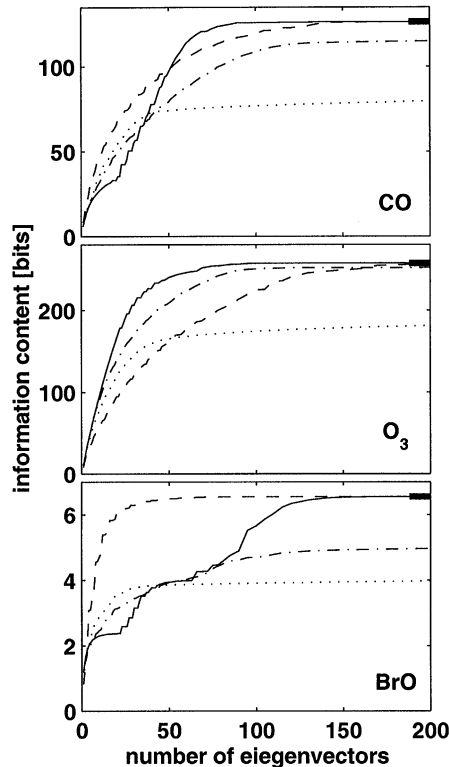


Fig. 6. Information content versus number of eigenvectors  $k$  for the retrieval of CO, O<sub>3</sub>, and BrO in the 576.4 GHz band. The results for  $\mathbf{E}_k$  derived from  $\mathbf{K}_x\mathbf{K}_x^T$  are plotted as —, from  $\mathbf{K}_x\mathbf{K}_x^T$  for each species individually as ---, from  $\mathbf{S}_y$  as -·-, and from  $\mathbf{R}_y$  as ···. The horizontal marks show the information content when no reduction is applied.

### 5.3. Handling weak emissions

A reduction technique has sometimes to deal with variability of very different ranges. The Odin-SMR band 576.4 GHz is an example of such a situation. The chemistry involving BrO is of high importance for ozone depletion [29], but no global vertically resolved observations of this species exist, and Odin will offer several ways to determine profiles of BrO. The 576.4 band offers one of the possibilities (but not the best) to measure BrO. The BrO emission is found on the wing of a CO line and is relatively close to a strong ozone transition (Fig. 1). The BrO emission is around two orders of magnitude smaller than the CO emission. As the emission of BrO is very weak, spectra have to be averaged in order to get useful results and here averages of 500 scans have been used.

Fig. 6 gives the information content for O<sub>3</sub>, CO and BrO when retrieved from the band 576.4 GHz by using different reduction schemes. The reductions based on  $\mathbf{K}_x$  and  $\mathbf{S}_y$  show clearly how the reduction put more weight in the information coming from the species with the strongest emission, O<sub>3</sub>, CO and BrO in decreasing order. The  $\mathbf{S}_y$  reduction performs more poorly, although it should be noted that, as the noise is reduced due to the averaging of scans,

the influence is smaller compared with the previous cases. For instance, for  $O_3$  the combination of a high emission and reduced noise makes the  $S_y$  reduction to work nearly as well as the  $K_x$  reduction. The results when the reduction is based on correlation ( $R_y$ ) instead of covariance ( $S_y$ ) does not work for this simulation. The problem is that a considerable number of spectral channels have very little information from the species, so their variability is mainly due to the thermal noise. As the correlation approach put the same weight on all the channels, there is an evident deterioration of the reduction performance as the noise features becomes more relevant.

The method based on  $K_x$  can be modified to find a better way to assure that information from species giving rise to only small variations in the spectra is not lost. The modified approach consists of deriving eigenvectors from the weighting functions of the different species individually, in contrast to using the whole of  $K_x$ , and appending them to construct  $E_k$  (first the most representative eigenvector for each species, then the second most, and so on). It can be seen in Fig. 6 that this approach indeed results in a more effective reduction for BrO, but the price to pay is the resulting lower efficiency for the other species. If the constrain to select  $k$  is that the information content shall have reached its maximum for all species, this approach does not result in an improvement, but for a dedicated retrieval of BrO (in this case) this method has some potential. It should be noted that  $E_k$  is not orthogonal when constructed in this way.

## 6. Conclusions

Several Hotelling transformation approaches for data reduction have been compared and tested practically by simulating the retrieval of atmospheric species from the sub-mm limb sounding observations of the Odin-SMR. The term Hotelling transformation covers here any method that uses some eigenvector expansion of observed spectra.

The simulated observations have a high magnitude of thermal noise and basing the reduction on the complete covariance matrix for the spectra ( $S_y$ ) gives, for the methods tested, the poorest performance. Theoretically the noise should not have this negative influence but the performance is deteriorated in practice as perfect statistics never are obtained from a limited data set.

An alternative approach was to base the data reduction on the correlation matrix ( $R_y$ ) instead of the covariance matrix. The information from all the channels is then weighted equally, independently of their variability. This is favorable when the species of interest give rise to emissions of very different magnitudes. The problem however is that the noise becomes more relevant as it can be the major source of variability in some channels, variability that is now highly weighted. This is the case for these simulations, so the approach worked very poorly. Another drawback, shared with the  $S_y$  reduction, is that the schemes are computational demanding due to the large size of the matrix in question or the large measurement ensemble needed to obtain good statistics.

Beside these standard Hotelling transformations, a novel approach based on the weighting functions of the species to be retrieved ( $K_x$ ) was introduced. This new method gave in the test cases the same general performance as using noise-free simulated spectra ( $S_{y-\varepsilon}$ ), but at a much lower computational burden. It was further shown that the method does not require detailed information on the species variability, and that it can be applied for, at least, moderately non-linear inversion problems. The proposed technique also gives flexibility in controlling the

data reduction, for instance, it was demonstrated how the method can be used to ensure that the same emphasis is put on all the species to be retrieved. Summarizing, the proposed reduction technique has a high computational efficiency, is flexible and showed in the test simulations an excellent performance.

## References

- [1] Reber CA, Trevethan CE, McNeal RJ, Luther MJ. The Upper Atmosphere Research Satellite (UARS) mission. *J Geophys Res* 1993;98:10643–7.
- [2] Kaye JA, Miller TL. The ATLAS series of shuttle missions. *Geophys Res Lett* 1996;23:2285–8.
- [3] Masuko H, et al. Superconducting submillimeter-wave limb emission sounder (SMILES) onboard Japanese experimental module (JEM) of international space station. Proceedings of the Quadrennial Ozone Symposium, Sapporo, 3–8 July 2000. p. 535–6.
- [4] Reburn WJ, Siddans R, Kerridge BJ, Bühler S, von Engel A, Erikson P, Kuhn-Sander T, Verdes C, Künzi K. Critical assessments in millimetre-wave atmospheric limb sounding. Final report, ESTEC/contract no. 13348/98/NL/GD, 2000.
- [5] Ridolfi M. An optimized forward model and retrieval scheme for MIPAS near real time data processing. Proceedings of the 2000 ERS-ENVISAT Symposium, Göteborg, 16–20 October, 2000.
- [6] Rodgers CD. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev Geophys* 1976;14:609–24.
- [7] Eriksson P. Analysis and comparison of two linear regularization methods for passive atmospheric observations. *J Geophys Res* 2000;105:18157–67.
- [8] Marks C, Rodgers CD. A retrieval method for atmospheric composition from limb emission measurements. *J Geophys Res* 1993;98:14939–53.
- [9] Jimenez C, Eriksson P, Askne J. Non-linear inversion of Odin sub-mm observations in the lower stratosphere by neural networks. *Microwave radiometry and remote sensing of the Earth's surface and atmosphere*. Zeist, The Netherlands: VSP 2000;200:503–511.
- [10] Rodgers CD. Characterization and error analysis of profiles retrieved from remote sounding measurements. *J Geophys Res* 1990;95:5587–95.
- [11] Jackson DD. The use of a priori data to resolve non-uniqueness in linear inversion. *Geophys. J R Astron Soc* 1979;57:137–57.
- [12] Butler CT, Meredith RZ, Stogryn AP. Retrieving atmospheric temperature parameters from DMSP SSM/T-1 data with a neural network. *J Geophys Res* 1996;101:7075–83.
- [13] Del Frate F, Schiavon G. Neural networks for the retrieval of water vapor and liquid water from radiometric data. *Radio Sci* 1998;33:1373–86.
- [14] Hadji-Lazaro J, Clerbaux C, Thiria S. An inversion algorithm using neural networks to retrieve atmospheric CO total columns from high-resolution nadir radiances. *J Geophys Res* 1999;104:23841–54.
- [15] Bellman R. *Adaptive control process: a guided tour*. New Jersey: Princeton University Press, 1961.
- [16] Björk Å. *Numerical methods for least squares problems*. Philadelphia, PA: SIAM, 1996.
- [17] Hansen PC. The truncated SVD as a method for regularization. *BIT* 1987;27:534–53.
- [18] Hansen PC. Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank. *SIAM J Sci Stat Comput* 1990;11:503–18.
- [19] Cattell RB. The scree test for the number of factors. *Multivariate Behavioral Research* 1966;1:245–76.
- [20] Kaiser HF. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 1960;20:141–51.
- [21] Rodgers CD. *Inverse methods for atmospheric sounding: theory and practice*. Singapore: World Scientific, 2000.
- [22] Merino F, Murtagh D, Eriksson P, Baron P, Ricaud P, de la Nöe J. Studies for the Odin sub-millimetre radiometer: 3. Performance simulations. *Can J Phys* 2002, in press.

- [23] Eriksson P, Merino F, Murtagh D, Baron P, Ricaud P, de la Nöe J. Studies for the Odin sub-millimetre radiometer: 1. Radiative transfer and instrument simulation. *Can J Phys* 2002, in press.
- [24] Krueger AJ, Minzner RA. A mid-latitude ozone model for the 1976 US standard atmosphere. *J Geophys Res* 1976;81:4477–81.
- [25] Cressie N. *Statistics for spatial data*. New York: Wiley Interscience, 1993.
- [26] Hoogen R, Rozanov VV, Burrows JP. Ozone profiles from GOME satellite data: Algorithm and first validation. *J Geophys Res* 1999;104:8263–80.
- [27] Eriksson P, Jimenez C. Non-linear profile retrievals for observations in the lower stratosphere with the Odin sub-mm radiometer. *Proceedings of IGARSS'98, Seattle, USA, 1998*. p. 1420–23.
- [28] Ready PJ, Wintz PA. Information extraction, SNR improvement and data compression in multispectral imagery. *IEEE Trans Commun* 1973;21:10,1123–30.
- [29] Lary DJ, Chipperfield MP, Toumi R, Lenton T. Heterogeneous atmospheric bromine chemistry. *J Geophys Res* 1996;101:1489–504.