

# Water vapour profiles by ground-based FTIR spectroscopy: study for an optimised retrieval and its validation

M. Schneider, F. Hase, and T. Blumenstock

IMK-ASF, Forschungszentrum Karlsruhe and Universität Karlsruhe, Germany

Received: 7 July 2005 – Published in Atmos. Chem. Phys. Discuss.: 29 September 2005

Revised: 19 January 2006 – Accepted: 25 January 2006 – Published: 14 March 2006

**Abstract.** The sensitivity of ground-based instruments measuring in the infrared with respect to tropospheric water vapour content is generally limited to the lower and middle troposphere. The large vertical gradients and variabilities avoid a better sensitivity for the upper troposphere/lower stratosphere (UT/LS) region. In this work an optimised retrieval is presented and it is demonstrated that compared to a commonly applied method, it improves the performance of the FTIR technique. The reasons for this improvement and the possible deficiencies of the method are discussed. Only by applying the method proposed here and using measurements performed at mountain observatories can water vapour variabilities in the UT/LS be detected in a self-consistent manner. The precision, expressed as noise to signal ratio, is estimated at 45%. In the middle and lower troposphere, precisions of 22% are achieved. These estimations are confirmed by a comparison of retrieval results based on real FTIR measurements with coinciding measurements of synoptical meteorological radiosondes.

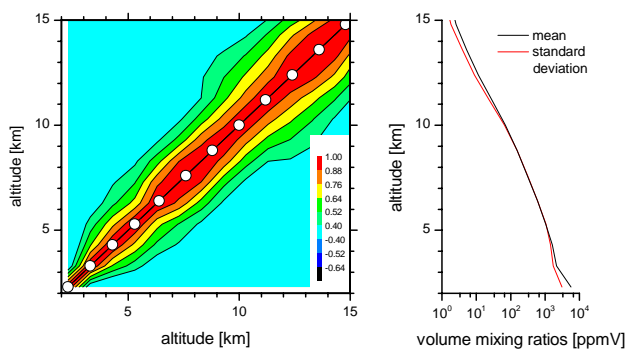
## 1 Introduction

The composition of the Earth's atmosphere has been profoundly modified throughout the last decades mainly by human activities. Prominent examples are the stratospheric ozone depletion and the upward trend in the concentration of greenhouse gases. While studies about the stratospheric composition have progressed rather well, there still exists a considerable deficiency for data from the free troposphere. Knowing the composition and evolution of these altitude regions is essential for the scientific verification of the Kyoto and Montreal Protocols and Amendments and for global climate modelling. Water vapour is the dominant greenhouse

gas in the atmosphere, and in particular its concentration and evolution in the upper troposphere and lower stratosphere (UT/LS) are of great scientific interest for climate modelling (Harries, 1997; Spencer and Braswell, 1997). Currently there is no outstanding routine technique for measuring water vapour in the UT/LS. The quick changes of atmospheric water vapour concentrations with time, their large horizontal gradients, and their decrease of several orders of magnitude with height makes their accurate detection a challenging task for any measurement technique. Traditionally tropospheric water vapour profiles are measured by synoptical meteorological radiosondes. However, this method has some deficiencies at altitudes above 6–8 km, which are mainly due to uncertainties in the pre-flight calibration and temperature dependence (Miloshevich, 2001; Leiterer et al., 2004). Other applied techniques are remote sensing from the ground by Lidar or Microwave instruments. Both are limited in their sensitivity: the Lidar generally to below 8–10 km, and the microwave measurements to above 15 km (SPARC, 2000). Satellite instruments also struggle to reach below this altitude. In this context the suggested formalism of retrieving upper tropospheric water vapour amounts from ground-based FTIR measurements aims to support efforts to obtain quality UT/LS water vapour data for research. To our knowledge, it is the first time that water vapour profiles measured by this technique are presented. A great advantage is that high quality ground-based FTIR measurements have already been performed during the last 10–15 years within the Network for Detection of Stratospheric Change (Kurylo, 1991, 2000; NDSC, web site). Therefore a long-term record of water vapour could be made available, with both temporal and to some extent, spatial coverage.

The structure of the article is as follows: first it is argued how the suggested optimisation acts in the context of inversion theory. Its advantages and deficiencies compared to a method, commonly used for trace gas retrievals, are discussed. In the third section an error assessment adds precise

Correspondence to: M. Schneider  
(matthias.schneider@imk.fzk.de)



**Fig. 1.** Description of a-priori state. Left panel: correlation matrix. Right panel: black line: mean state; red line: standard deviation of mean state.

quantitative estimations about the expected improvements to these qualitative considerations. It is also shown how possible deficiencies of the optimised method can be eliminated. Finally, these estimations are validated by a comparison of retrieval results based on real measurements with coinciding in-situ measurements.

## 2 Optimised water vapour retrieval

An inversion problem is generally under-determined. Many state vectors ( $\mathbf{x}$ ) are consistent with the measurement vector ( $\mathbf{y}$ ). If one also considers measurement noise ( $\epsilon_y$ ), there is an even wider range of possible solutions within  $\epsilon_y$ , in accordance to the measurement vector: in the equation,

$$\hat{\mathbf{y}} = \mathbf{y} + \epsilon_y = \mathbf{K}\mathbf{x} \quad (1)$$

the matrix  $\mathbf{K}$  is ill-conditioned. Its effective rank is smaller than the dimension of state space, i.e. it is singular and cannot be simply inverted. To come to an unique solution of  $\mathbf{x}$ , the state space is constrained by requiring:

$$\mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{x}_a \quad (2)$$

where  $\mathbf{x}_a$  is a “typical” or a-priori state and the matrix  $\mathbf{B}$  determines the kind of required similarity of  $\mathbf{x}$  with  $\mathbf{x}_a$ . This equation constrains the solution independently from the measurement, i.e. before the measurement is made. Therefore  $\mathbf{B}$  and  $\mathbf{x}_a$  contain the kind of information known about the state prior to the measurement. Subsequently, assuming Gaussian statistics for the error term in Eq. (1) and the a-priori distribution in Eq. (2) leads to the cost function:

$$\sigma^{-2}(\mathbf{y} - \mathbf{K}\mathbf{x})^T(\mathbf{y} - \mathbf{K}\mathbf{x}) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{B}^T \mathbf{B}(\mathbf{x} - \mathbf{x}_a) \quad (3)$$

The most probable state is the one which minimises Eq. (3). Here  $(\epsilon_y^T \epsilon_y)^{-1}$  was identified by  $\sigma^{-2}$ . It is obvious that the applied a-priori information ( $\mathbf{B}$  and  $\mathbf{x}_a$ ) influences the solution. For water vapour the large amount of synoptical meteorological sonde (ptu-sonde) data allows a detailed

study of the a-priori state. In the following it is discussed whether the extensive a-priori information can be used to optimise the performance of the retrieval. The study of a-priori data is done for the island of Tenerife, where ptu-sondes are launched twice daily (at 00:00 and 12:00 UT) within the global radiosonde network and where an FTIR instrument has been operating since 1999 at a mountain observatory (Izaña Observatory, Schneider et al., 2005).

### 2.1 Characterisation of a-priori data

The study is based on the daily 12:00 UT soundings performed from 1999 to 2003. It has been observed that an in-situ instrument – located at the mountain observatory – and the sonde, when measuring at the observatory’s altitude, detect quite different humidities because of their different locations, i.e. on the surface and in the free troposphere (see Sect. 4). For this reason the analysed profiles are built up by a combination of the in-situ measurements at the instrument’s site (for the lowest grid point; applied sensor: Rotronic MP100H), and sonde measurements (for all other grid points below 16 km). For higher altitudes a mean mixing ratio of 2.5 ppmv and covariances like those at 16 km are applied. The left panel of Fig. 1 shows the correlation matrix  $\Gamma_a$  determined from these a-priori profiles. Here correlation matrices are presented instead of the commonly shown covariance matrixes. The reason is that they can be more easily presented. Their elements are all of the same order of magnitude (between  $-1$  and  $1$ ), whereas in the case of water vapor the elements of the covariance matrices extend over 8 orders of magnitude. Figure 1 demonstrates how variabilities at different altitudes typically correlate with each other. In the real atmosphere the mixing ratios for different altitudes show correlation coefficients of at least 0.5 within a layer of around 2.5 km. The a-priori covariance matrix  $\mathbf{S}_a$  is calculated from  $\Gamma_a$  by  $\mathbf{S}_a = \Sigma_a \Gamma_a \Sigma_a^T$ , where  $\Sigma_a$  is a diagonal matrix containing the a-priori variabilities at a certain altitude. These variabilities are depicted as a red line in the right panel of Fig. 1. The black line shows the mean mixing ratios. The determined mean and covariances only describe the whole ensemble completely if mixing ratios are normally distributed. This is generally assumed and often justified by the fact that entropy is then maximised: if only the mean and the covariance are known a supposed normal distribution is thus the least restricting assumption about the a-priori state (Sect. 10.3.3.2 in Rodgers, 2000). However, this does not necessarily reflect the real situation!

A further examination of the sonde data reveals that the mixing ratios at a certain altitude are not normally but log-normally distributed. Their pdf is:

$$P_x = \frac{1}{x\sigma\sqrt{2\pi}} \exp - \frac{(\ln x - \ln x_a)^2}{2\sigma^2} \quad (4)$$

with a shape parameter  $\sigma$  ranging from 1.15 ppmv in the middle troposphere to 0.55 ppmv above 10 km, and a median

$x_a$  between 5000 ppmv close to the surface and 1.5 ppmv in the stratosphere. The only exception of this distribution is the first  $\approx 100$  m above the surface, where the mixing ratios are more normally distributed. It is possible to sample all this additional information in a simple mean state vector and a covariance matrix. This is achieved by transforming the state on a logarithmic scale, which transforms the log-normal pdf to a normal pdf. A normal pdf can be completely described by its covariance and its mean. A  $\chi^2$ -test reveals how the description of the a-priori state is improved by this transformation. This test determines the probability of a particular random vector of belonging to an assumed normal distribution. If a vector  $\mathbf{x}$  is supposed to be a member of a Gaussian ensemble with the mean  $\mathbf{x}_a$  and covariance  $\mathbf{S}$  the quantity considered is:

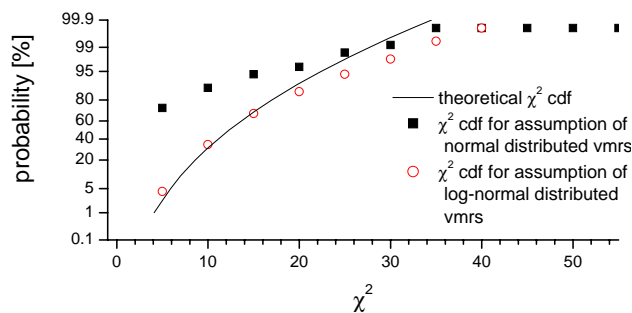
$$\chi^2 = (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{x}_a) \quad (5)$$

The  $\chi^2$  test clearly rejects a normal distribution of the mixing ratios. This can be seen by comparing the theoretical cumulative distribution function (cdf) of  $\chi^2$  with the one determined by Eq. (5). Figure 2 demonstrates that the theoretical  $\chi^2$  cdf differs clearly from the cdf obtained from the ensemble's state vectors if they are assumed to be normally distributed (difference between black line and black squares). More than 95% of the ensemble's state vectors are not consistent with this assumption. On the other hand, a prior log-normal pdf is well confirmed. If the mixing ratios and the covariances are transformed to a logarithmic scale, only approximately 10% of the ensemble's states fail the test (compare black line and red circles).

## 2.2 Discussion of two retrieval methods

This section discusses the differences between an inversion performed on a linear scale, which is the method commonly used for trace gas retrievals, and one performed on a logarithmic scale. The logarithmic retrieval is occasionally applied as a positivity constraint, since it avoids negative components in the solution vector. In the case of water vapour it has a further advantage. It converts the state for which Eq. (3) minimises in a statistically optimal solution: on a logarithmic scale the a-priori state can be described correctly in the form of a mean and covariance. Under these circumstances, substituting  $\mathbf{B}^T \mathbf{B}$  and  $\mathbf{x}_a$  in Eq. (3) by the inverse of the logarithmic a-priori covariance ( $\mathbf{S}_a^{-1}$ ) and the median state vector, leads to a cost function, which is directly proportional to the negative logarithm of the a-posteriori probability density function (pdf) of the Bayesian approach. This posterior pdf is the conditional pdf of the state given the measurement, or in other words, the a-priori pdf of the state updated by the information given in the measurement. The minimisation of Eq. (3) thus yields the maximum a-posteriori solution, i.e. it is the most probable state given the measurement.

To the contrary, on a linear scale setting  $\mathbf{B}^T \mathbf{B}$  as  $\mathbf{S}_a^{-1}$  and  $\mathbf{x}_a$  as mean state in Eq. (3) does not lead to a statistically



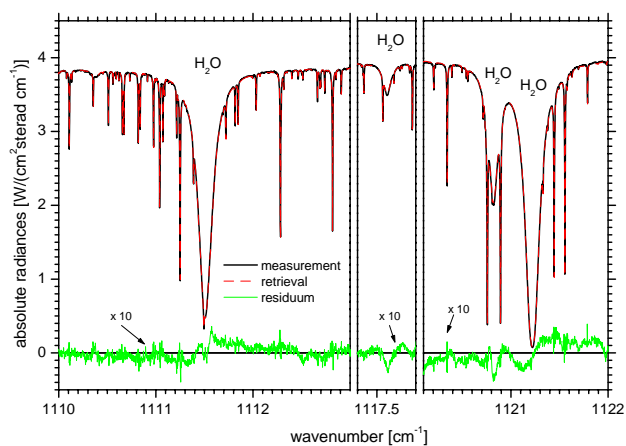
**Fig. 2.**  $\chi^2$  test for different of a-priori assumptions. Black line: theoretical  $\chi^2$  cumulative distribution function (cdf); black filled squares:  $\chi^2$  cdf of ensemble for assumed normal pdf on a linear scale; red circles:  $\chi^2$  cdf of ensemble for assumed normal pdf on a logarithmic scale.

optimal solution. It is not related to the a-posteriori pdf in the Bayesian sense. On a linear scale the a-priori state is log-normally distributed. Therefore, seen from a statistical point of view, the second term of the cost function over-constrains states above the mean and under-constrains states below the median. As a consequence, the probability of states above the mean is underestimated and below the median overestimated – the overestimation is greater the further away it is from the centre of the a-priori distribution. Thus, if compared to a correct maximum a-posteriori solution, the retrieval tends to underestimate the values of the real state both far above and far below the mean state.

However, the transformation on a logarithmic scale introduces some other problems: it significantly increases the non-linearity of the forward model, which requires decreasing the differences between each iteration step, thus lowering the speed of convergence. This difficulty is overcome within the inversion code PROFFIT by using a refined minimisation scheme. A further drawback is that, in the retransformed linear scale the constraints now depend on the solution, which may cause misinterpretations of the spectra. To assess whether the linear or logarithmic retrieval performs better both retrieval approaches are extensively examined first by a theoretical (Sect. 3) and second by an empirical validation (Sect. 4).

## 2.3 Applied inversion code and spectral region

PROFFIT (Hase et al., 2004) is the inversion code used. It applies the Karlsruhe Optimised and Precise Radiative Transfer Algorithm (KOPRA, Höpfner et al., 1998; Kuntz et al., 1998; Stiller et al., 1998) as the forward model, which was developed for the analysis of MIPAS-Envisat limb sounder spectra. PROFFIT enables the inversion on a linear and logarithmic scale. Hence, in the case of water vapour, it enables the correct application of prior information to obtain a statistically optimal solution. PROFFIT does not employ a fixed



**Fig. 3.** Spectral regions applied for retrieval. Plotted is the situation for a real measurement taken on 10 March of 2003 (solar elevation angle  $50^\circ$ ). Black line: measured spectrum; red line: simulated spectrum; green line: difference between simulation and measurement.

a-priori value for the measurement noise ( $\sigma$  of Eq. 3). This value is taken from the residuals of the fit itself, performing an automatic quality control of the measured spectra. Furthermore, if the observed absorptions depend on temperature, PROFFIT allows the retrieval of temperature profiles. For both the linear and logarithmic retrieval, the same retrieval setup is applied: three microwindows between 1110 and 1122  $\text{cm}^{-1}$  are fitted. Figure 3 shows a typical situation for an evaluation of a real measurement. The black line represents the measurement, the red dotted line the simulated spectrum and the green line the difference between both. The  $\text{H}_2\text{O}$  signatures are marked in the Figure. One can observe that two stronger lines (at 1111.5 and 1121.2  $\text{cm}^{-1}$ ) and two relatively weak lines (at 1117.6 and 1120.8  $\text{cm}^{-1}$ ) lie within these spectral regions, where additionally  $\text{O}_3$  is an important absorber (numerous thin strong signatures). The profile of this species is thus simultaneously retrieved. Other interfering gases are  $\text{CO}_2$ ,  $\text{N}_2\text{O}$ , and  $\text{CH}_4$ , whereby the latter two are also simultaneously retrieved by scaling their respective climatological profiles, the former is kept fixed to a climatological profile. Spectroscopic line parameters are taken from the HITRAN 2000 database Rothman et al. (2003), except for  $\text{O}_3$ , where parameters from Wagner et al. (2002) are applied.

### 3 Error analysis and sensitivity assessment

Assuming linearity for the forward model  $F$  and the inverse model  $I$  within the uncertainties of the retrieved state and the model parameters it is (Rodgers, 2000):

$$\hat{x} - x = \left( \frac{\partial I[F(\hat{x}, \hat{p}), \hat{p}]}{\partial y} \frac{\partial F(\hat{x}, \hat{p})}{\partial x} - \mathbf{I} \right) (x - x_a)$$

$$\begin{aligned} & + \frac{\partial I[F(\hat{x}, \hat{p}), \hat{p}]}{\partial y} \frac{\partial F(\hat{x}, \hat{p})}{\partial p} (p - \hat{p}) \\ & + \frac{\partial I[F(\hat{x}, \hat{p}), \hat{p}]}{\partial y} (y - \hat{y}) \\ & = (\hat{\mathbf{A}} - \mathbf{I})(x - x_a) + \hat{\mathbf{G}}\hat{\mathbf{K}}_p(p - \hat{p}) + \hat{\mathbf{G}}(y - \hat{y}) \end{aligned} \quad (6)$$

i.e. the difference between the retrieved and the real state ( $\hat{x} - x$ ) – the error – can be linearised about a mean profile  $x_a$ , the estimated model parameters  $\hat{p}$ , and the measured spectrum  $\hat{y}$ . Here  $\mathbf{I}$  is the identity matrix,  $\hat{\mathbf{A}}$  the averaging kernel matrix,  $\hat{\mathbf{G}}$  the gain matrix, and  $\hat{\mathbf{K}}_p$  a sensitivity matrix to model parameters:

$$\begin{aligned} \hat{\mathbf{A}} &= \hat{\mathbf{G}}\hat{\mathbf{K}} \\ \hat{\mathbf{G}} &= \frac{\partial I[F(\hat{x}, \hat{p}), \hat{p}]}{\partial y} \\ \hat{\mathbf{K}} &= \frac{\partial F(\hat{x}, \hat{p})}{\partial x} \\ \hat{\mathbf{K}}_p &= \frac{\partial F(\hat{x}, \hat{p})}{\partial p} \end{aligned} \quad (7)$$

whereby  $\hat{\mathbf{K}}$  is the Jacobian. Equation (6) identifies three principle error sources. These are the inherent finite vertical resolution, the input parameters applied in the inversion procedure, and the measurement noise. This analytic error estimation may be applied if the inversion is performed on a linear scale. In this case, the constraints and consequently  $\hat{\mathbf{G}}$  are constant within the uncertainty of  $\hat{x}$ . However, if the inversion is performed on a logarithmic scale the constraints are constant on this scale, but variable on the retransformed linear scale. Changes of the state vector towards values above the a-priori value are only weakly constrained, while changes towards smaller values are more strongly constrained. As a consequence  $\hat{\mathbf{G}}$  cannot necessarily be considered constant within the uncertainty of the retrieved state and some model parameters. The latter is particularly problematic for water vapour. The phase error of the instrumental line shape and the temperature profile have a large impact on the spectra. This is due to the broad and strong absorption signatures of water vapor. Consequently, all these errors can only be estimated by a full treatment. Two forward calculations are performed for each error estimation and for all profiles of the large ensemble of the a-priori profiles: a first calculation with correct parameters and a second with erroneous parameters. Subsequently both spectra are retrieved with the correct parameter as input data. The parameter error is then given by the difference of the two retrievals. The smoothing error is the difference between the correct parameter retrieval and the a-priori profile. In this work, all errors are estimated by this full treatment for consistency reasons for both the linear and the logarithmic retrieval.

Together with the error estimation a sensitivity assessment is performed. Generally the averaging kernels (columns of

$\hat{A}$ ) are used to estimate the sensitivity of the retrieval at certain altitudes. They document by how much ppmv the retrieved solution will change due to a variability of 1 ppmv in the real atmosphere. They may inform that 1 ppmv more at 5 km is reflected in the retrieval by an extra of 0.1 ppmv at 8 km. However, the typical real atmospheric variabilities at different altitudes are not considered and hence to what extent the typical variability as retrieved at 8 km is disturbed by typical variabilities at 5 km. This is a minor problem if the mixing ratio variabilities have the same magnitude throughout the atmosphere. The variabilities of water vapour decrease by 3–4 orders of magnitude from the surface to the tropopause (see Fig. 1), thus the interpretation of the averaging kernels is quite limited. Alternatively, one may produce adequately normed kernels to address this deficiency. Furthermore the averaging kernels depend strongly on the actual water vapor content, i.e. there is no typical kernel and non-linearities play an important role. For all these reasons here a full treatment, consisting of forward calculation of assumed real states and subsequent inversion, is used to estimate the response of the retrieval on real atmospheric variabilities. Therefore, the real state vectors are correlated linearly to their corresponding retrieved vectors. The correlation coefficient ( $\rho$ ) considers the different magnitudes of the variabilities. For instance,  $\rho$  between the real state at 5 km and the retrieved state at 8 km gives the typical fraction of the retrieved variabilities at 8 km due to disturbances from 5 km. These correlation matrices give a good overview of the relation between real atmospheric variabilities and the retrieved variabilities.

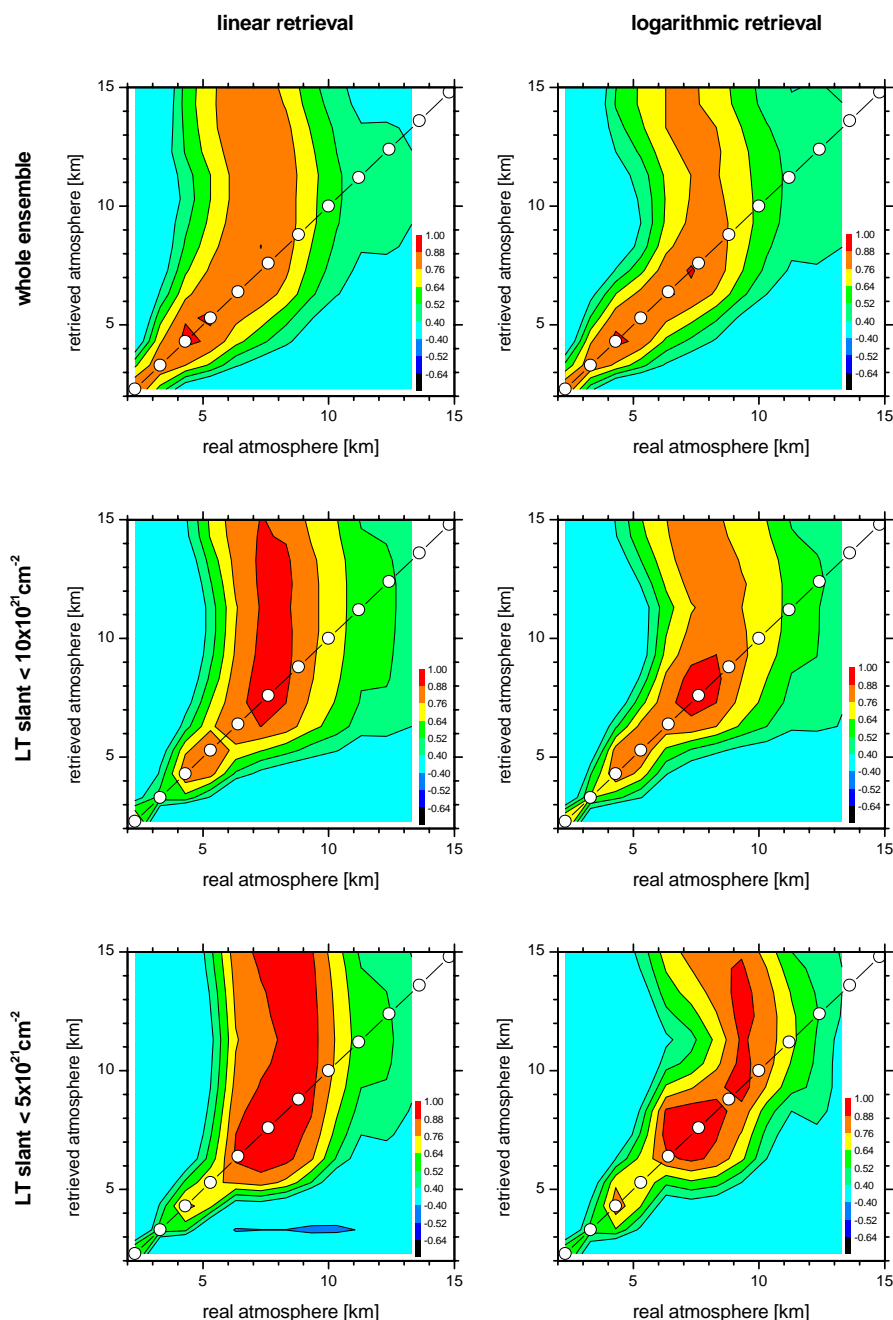
Error estimation and sensitivity assessment are performed for the whole ensemble (the ensemble used for calculating the a-priori mean and covariances), and for two sub-ensemble of selected conditions, when especially good upper tropospheric sensitivity and even sensitivity in the tropopause region are expected. Sensitivity in the UT and tropopause region requires the strong absorption lines to be unsaturated. Furthermore, the signal to noise ratio, which at Izaña is occasionally decreased by high aerosol loading owing to Saharan dust intrusion events, should be acceptable (above 200 at  $1100\text{ cm}^{-1}$ ). In 30% of all measurement days the lower tropospheric water vapour slant column amounts (slant column amounts between surface and 4.3 km) are below  $10 \times 10^{21}\text{ cm}^{-2}$  (LT slant  $< 10 \times 10^{21}\text{ cm}^{-2}$  criterion), which means that the strong absorption lines are unsaturated. On these days, good sensitivity for the UT can be expected. The observing system should perform even better if the lower tropospheric slant column amounts are below  $5 \times 10^{21}\text{ cm}^{-2}$  (LT slant  $< 5 \times 10^{21}\text{ cm}^{-2}$  criterion). This is however only the case for 10% of all possible observations.

### 3.1 Smoothing error

Figure 4 shows correlation matrices in the absence of parameter errors. They document the sensitivity of the retrieval

if the smoothing error alone is taken into account. The left panels show the linear retrieval, the right panels the logarithmic retrieval, the upper panels the whole ensemble, and the middle and lower panels the LT slant  $< 10 \times 10^{21}\text{ cm}^{-2}$  and LT slant  $< 5 \times 10^{21}\text{ cm}^{-2}$  sub-ensembles. Considering the whole ensemble the sensitivity is limited to altitudes below 8–9 km. Furthermore, the upper tropospheric mixing ratios of the linear retrieval tend to depend more on variabilities at lower altitudes. For example, the value retrieved at 9 km is mainly influenced by the real atmospheric situation at 7 km. This incorrect altitude attribution is less pronounced in the logarithmic retrieval. For the LT slant  $< 10 \times 10^{21}\text{ cm}^{-2}$  sub-ensemble the sensitivity is extended by 1–2 km towards higher altitudes. In this case, the observing system provides good information about the atmospheric water vapour variabilities up to 10 km ( $\rho$  at the diagonal above 0.7). As before, for the linear retrieval, the amounts at higher altitudes are strongly disturbed by the real states at lower altitudes, while, for the logarithmic retrieval, high correlation coefficients are more concentrated around the diagonal of the matrix. If the LT slant is smaller than  $5 \times 10^{21}\text{ cm}^{-2}$ , the logarithmic retrieval's  $\rho$  values at the diagonal are still 0.7 at 11 km. The  $\rho$  values of the linear retrieval are slightly lower (0.61 at 11 km). But the most pronounced difference between both methods is the incorrect altitude attribution in case of the linear retrieval. For example, the mixing ratio retrieved by the linear method at 11 km is strongly correlated to real values at 8 km ( $\rho$  of 0.9). These disturbances are significantly reduced in case of the logarithmic retrieval ( $\rho$  of 0.63). Thus the error of the state retrieved with the logarithmic method at 11 km can already be sufficiently reduced by considering the disturbances originating from altitudes down to about 8 km only. The linear method, on the other hand, should very likely take into account values from further down in order to reach a similar error level. This means that the correlation length of the smoothing error is larger for the linear retrieval. To determine the amount of a layer with a certain uncertainty the layer must be broader for the linear retrieval if compared to the logarithmic retrieval.

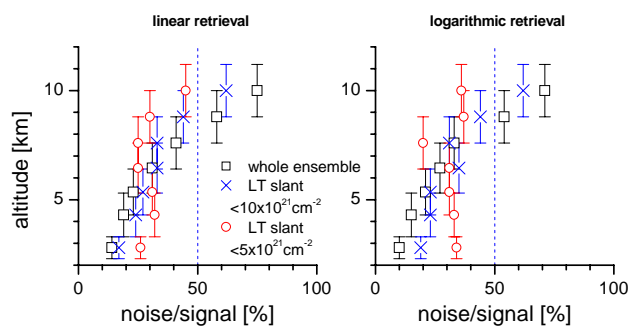
The smoothing error is commonly presented as a mean error and a covariance matrix. However, for a non-Gaussian distribution this kind of error presentation is inappropriate. For a pure log-normal distribution, errors might be presented on a logarithmic scale as mean and covariance. This is not possible for the linear retrieval, since it is expected to have normal as well as log-normal characteristics. To overcome this problem the errors are firstly assessed for layers and not for a single altitude. Bearing in mind the modest vertical resolution of trace gas profiles determined by ground-based FTIR spectroscopy, the objective of this technique should consist of retrieving the amount of a certain layer rather than a concentration at a single altitude. Secondly, least squares fits are applied to estimate the errors. This enables all systematic errors to be separated from the random errors. The regression curves of the least squares fits demonstrate how



**Fig. 4.** Sensitivity of observing system in the absence of parameter error. Depicted are correlation matrices between assumed real profiles and retrieved profiles. Left panels for retrieval on a linear scale, right panels for retrieval on a logarithmic scale. Upper panels for the whole ensemble, middle panels for the LT slant  $< 10 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble, and lower panels for the LT slant  $< 5 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble. Colors mark the values of the correlation coefficients ( $\rho$ ) as given in legend.

the real atmosphere – as a mean – is mapped by the retrieval: their difference from the diagonal describes the systematic smoothing error. The scattering around the regression curve describes its pure random error. For a linear least squares fit the correlation coefficient ( $\rho$ ) can be used to estimate this pure random error.  $\rho^2$  is the ratio of the variance of the re-

gression line ( $\sigma_{reg}^2$ ) to the variance of the retrieved amount ( $\sigma_{\hat{x}}^2$ ):  $\rho^2 = \sigma_{reg}^2 / \sigma_{\hat{x}}^2$ . It gives the proportion of the variance of the retrieved amount that is systematically linked to the real atmospheric variance. The remaining variance of  $\hat{x}$  is the scattering around the regression line (the random error:



**Fig. 5.** Smoothing errors in the retrieved profiles. Left panel: linear retrieval. Right panel: logarithmic retrieval. Colors as described in legend.

$\sigma_{\epsilon_{reg}}^2$ ). It can be calculated from (e.g. Wilks, 1995):

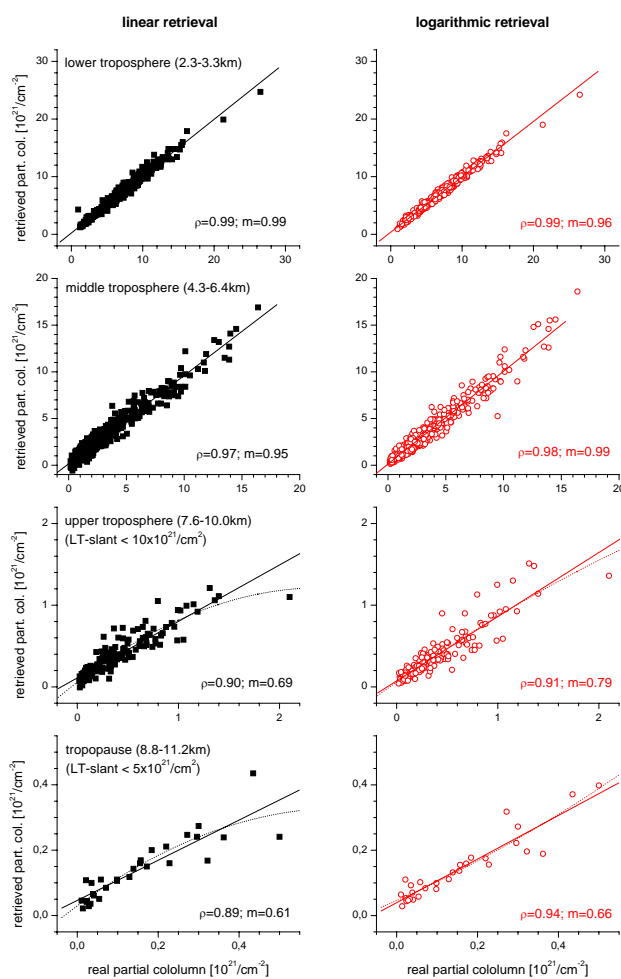
$$\sigma_{reg}^2 + \sigma_{\epsilon_{reg}}^2 = \sigma_{\hat{x}}^2 \quad (8)$$

or

$$\rho^2 + \frac{\sigma_{\epsilon_{reg}}^2}{\sigma_{\hat{x}}^2} = 1 \quad (9)$$

Figure 5 depicts the random smoothing errors relative to the variability of the retrieved value (noise to signal error:  $\sigma_{\epsilon_{reg}}/\sigma_{\hat{x}}$ ) of several layers throughout the troposphere. The altitude region of each layer is indicated by the error bars. The left panel shows estimations for the linear retrieval and the right panel for the logarithmic retrieval. The errors for both retrieval methods are quite similar. The black squares represent the error for the whole ensemble. It confirms the observation made in Fig. 4 that above 8 km the retrieval generally contains limited information about the real atmosphere: the signal/noise ratio lies above 50%. The blue crosses show the same but for the LT slant  $< 10 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble. Here the smoothing errors above 6 km are reduced, e.g. from 54% to 44% for the 7.6–10 km layer and the logarithmic retrieval. The red circles show the situation for the LT slant  $< 5 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble. Under these conditions, the random smoothing error of the logarithmic retrieval for the 8.8–11.2 km layer is as small as 36%. The random errors calculated with Eq. (9) are similar for the linear and logarithmic retrieval.

The better performance of the logarithmic approach becomes visible in Fig. 6, which shows the real characteristics of the correlations for four different layers representing the lower troposphere (LT, 2.3–3.3 km), the middle troposphere (MT, 4.3–6.4 km), the upper troposphere (UT, 7.6–10 km), and the tropopause region (8.8–11.2 km). Depicted are all single ensemble members and curves of linear least squares fits (solid lines) and second order polynomial least squares fits (dotted lines). The correlation coefficient ( $\rho$ ) and the slope ( $m$ ) of the regression line are given in the panels. The left panels show the linear and the right panels the logarithmic retrievals. Above  $\approx 6$  km the linear retrieval regression



**Fig. 6.** Correlations between assumed real partial column amounts and their corresponding retrieved amounts in the absence of parameter errors. From the top to the bottom: lower troposphere, middle troposphere, upper troposphere (for the LT slant  $< 10 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble), and tropopause region (for the LT slant  $< 5 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble). Left panels, black squares and black lines: retrieval on a linear scale and corresponding least squares fits. Right panels, red circles: retrieval on a logarithmic scale and corresponding regression line. Solid lines: linear least squares fit. Dotted lines: second order polynomial least squares fit.

line slope always lies below the relevant logarithmic retrieval slope: the linear method has a larger systematic error. Furthermore, the amounts retrieved by the logarithmic method are always linearly correlated to the real amounts. For the linear retrieval this is only valid at low altitudes. At altitudes above 6 km the linear retrieval is more sensitive at small amounts than at large amounts. In these cases the systematic behavior of the linear retrieval is insufficiently described by a linear regression line. A linear regression line then systematically overestimates very low and very large amounts and underestimates amounts between the median and the mean

**Table 1.** Assumed uncertainties.

error source	uncertainty
measurement noise	S/N of 500
phase error	0.02 rad
modulation eff.	2%
T profile <sup>a</sup>	up to 2.5 K at surface 1 K rest of troposphere
solar angle	0.1°
line intensity	5%
pres. broad. coef.	1%

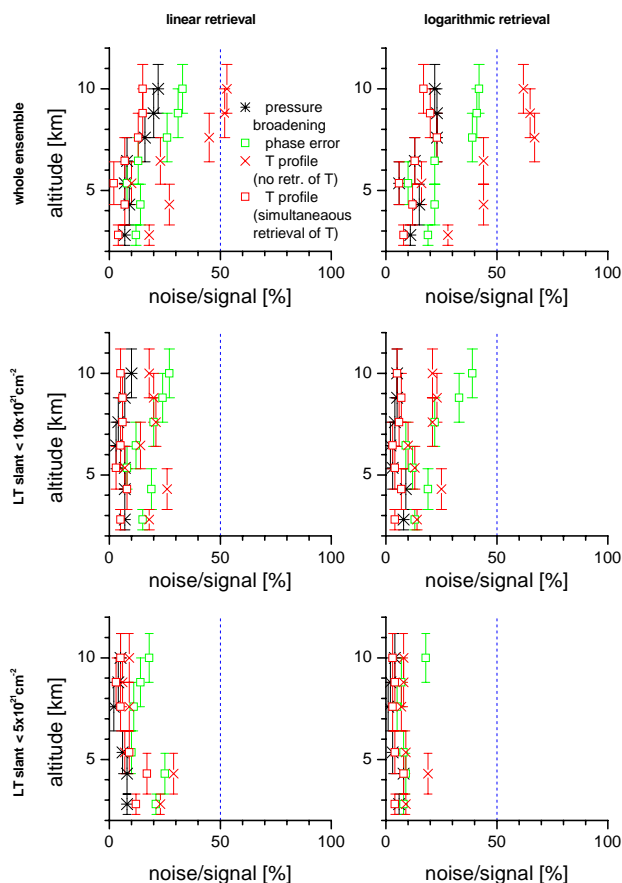
<sup>a</sup> detailed description see text

value. For a better description of this systematic behavior a second order polynomial would be needed. This additional characteristic of the linear retrieval's smoothing error has important consequences: it limits the linear retrieval in correctly detecting variabilities present in time series. It underestimates alterations towards large amounts and overestimate alterations towards small amounts. For an analysis of water vapour time series above 6 km the logarithmic retrieval is the better choice!

### 3.2 Model parameter error

As for the smoothing error, the random and systematic errors caused by parameter uncertainties are separated by means of least squares fits. Therefore, the retrievals of spectra simulated with correct parameters are correlated to the retrievals of spectra simulated with erroneous parameters. In this subsection, errors due to measurement noise, uncertainties in solar angle, instrumental line shape (ILS: modulation efficiency and phase error Hase et al., 1999), temperature profile, and spectroscopic parameters (line intensity and pressure broadening coefficient) are estimated. The assumed parameter uncertainties are listed in Table 1. Two sources are considered as errors in the temperature profile: first, the measurement uncertainty of the sonde, which is assumed to be 0.5 K throughout the whole troposphere and to have no interlevel correlations. Second, the temporal differences between the FTIR and the sonde's temperature measurements, which are estimated to be 1.5 K at the surface and 0.5 K in the rest of the troposphere, with 5 km correlation length for the interlevel correlations.

Random errors due to measurement noise, uncertainties in the modulation efficiencies, the solar angle and the line intensity are situated below or around 5%. They may be neglected if compared to the errors caused by phase error, temperature profile, or pressure broadening coefficient uncertainties. Figure 7 shows the latter errors for the whole ensemble (upper panels) and for the sub-ensembles with low LT slant columns (middle and lower panels). It should be remarked that, owing to the aforementioned nonlinearity of  $\hat{G}$ , it is impossible to



**Fig. 7.** Parameter errors in the retrieved profiles. Upper panels: for the whole ensemble. Middle panels: for the LT slant  $< 10 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble. Bottom panels: for LT slant  $< 5 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble. Left panels: linear retrieval. Right panels: logarithmic retrieval. Symbols as described in the legend.

separate the parameter errors completely from the smoothing errors. As a consequence, even systematic error sources may produce random errors (line intensity and pressure broadening parameter). Furthermore, the correlation plots are expected to show some of the characteristics of the smoothing error: e.g. above 6 km the linear retrieval's sensitivity towards parameter uncertainties is expected to be smaller at large amounts than at small amounts. This is the main reason for the linear retrieval's high random errors above 6 km for the LT slant  $< 5 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble caused by the uncertainties in the phase error parameter. The corresponding errors of the logarithmic retrieval are smaller (at least for the MT and UT).

Considering the whole ensemble, the temperature uncertainty provides the largest errors (red crosses). The errors are generally larger for the logarithmic retrieval, in particular the temperature error. Here 23% at 6.5 km for the linear retrieval is much lower than 44% for the logarithmic retrieval.



This is due to the retrieval's misinterpretation of spectral signatures arising from errors in the temperature profile. Since  $\hat{\mathbf{G}}\hat{\mathbf{K}}_p$  from Eq. (6) is generally not equal to zero, the parameter error in the measurement space may be transformed into the state space. This is a minor problem when the minimisation of the cost function (Eq. 3) is performed on a linear scale. Then changes of the state vector with respect to its a-priori state and the magnitude of the constraining term are linearly correlated. A misinterpretation would thus mean a large value of the constraining term and consequently Eq. (3) would never be minimised. On a logarithmic scale, however, a linear increase of the constraining term is related to an exponential increase of the retransformed state vector. Hence, a significant change of the state vector is not avoided by the constraining term. The problem can be reduced by a simultaneous retrieval of the temperature profile, which adds two terms to the cost function:

$$\begin{aligned} & \sigma^{-2}(\mathbf{y} - \mathbf{K}\mathbf{x})^T(\mathbf{y} - \mathbf{K}\mathbf{x}) + (\mathbf{x} - \mathbf{x}_a)^T\mathbf{S}_a^{-1}(\mathbf{x} - \mathbf{x}_a) \\ & + \sigma^{-2}(\mathbf{y} - \mathbf{K}_t\mathbf{t})^T(\mathbf{y} - \mathbf{K}_t\mathbf{t}) + (\mathbf{t} - \mathbf{t}_a)^T\mathbf{S}_{et}^{-1}(\mathbf{t} - \mathbf{t}_a) \end{aligned} \quad (10)$$

Here  $\mathbf{t}$  and  $\mathbf{t}_a$  are the real and the assumed temperature state vector,  $\mathbf{K}_t$  the sensitivity (or Jacobian) matrix for the temperature, and  $\mathbf{S}_{et}$  the error covariance matrix for the temperature. Thus a temperature error does not lead to an adjustment of the first term – a misinterpretation of spectral information –, but to an adjustment of the third term in Eq. (10). This reduces the probability of misinterpreting the temperature error. At 6.5 km, for example, the simultaneous fitting of the temperature profile reduces the error from 44% to 13%. This is seen by comparing the red crosses with the red squares in Fig. 7. This strategy leaves the uncertainty in phase error and pressure broadening coefficient as the most important error sources.

For the LT slant  $<10 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble (middle panels of Fig. 7), the errors are much smaller (generally below 30%). Except for the phase error, the errors for the linear and logarithmic retrieval are now similar. A misinterpretation of spectral signatures is less probable for this ensemble. Apparently, the condition of unsaturated absorption lines simultaneously eliminates days predestined for misinterpretation. However, a simultaneous retrieval of the temperature further improves the retrievals by reducing the temperature error to below 10% at all altitudes. The most important errors are due to uncertainties in the phase error.

The lower panel of Fig. 7 depicts the errors for the LT slant  $<5 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble. This condition further reduces all errors at altitudes above 5 km. A simultaneous fit of the temperature limits all errors for the logarithmic retrieval to below 10%. The only exception is the error owing to phase error uncertainties. It still reaches 18% around 10 km.

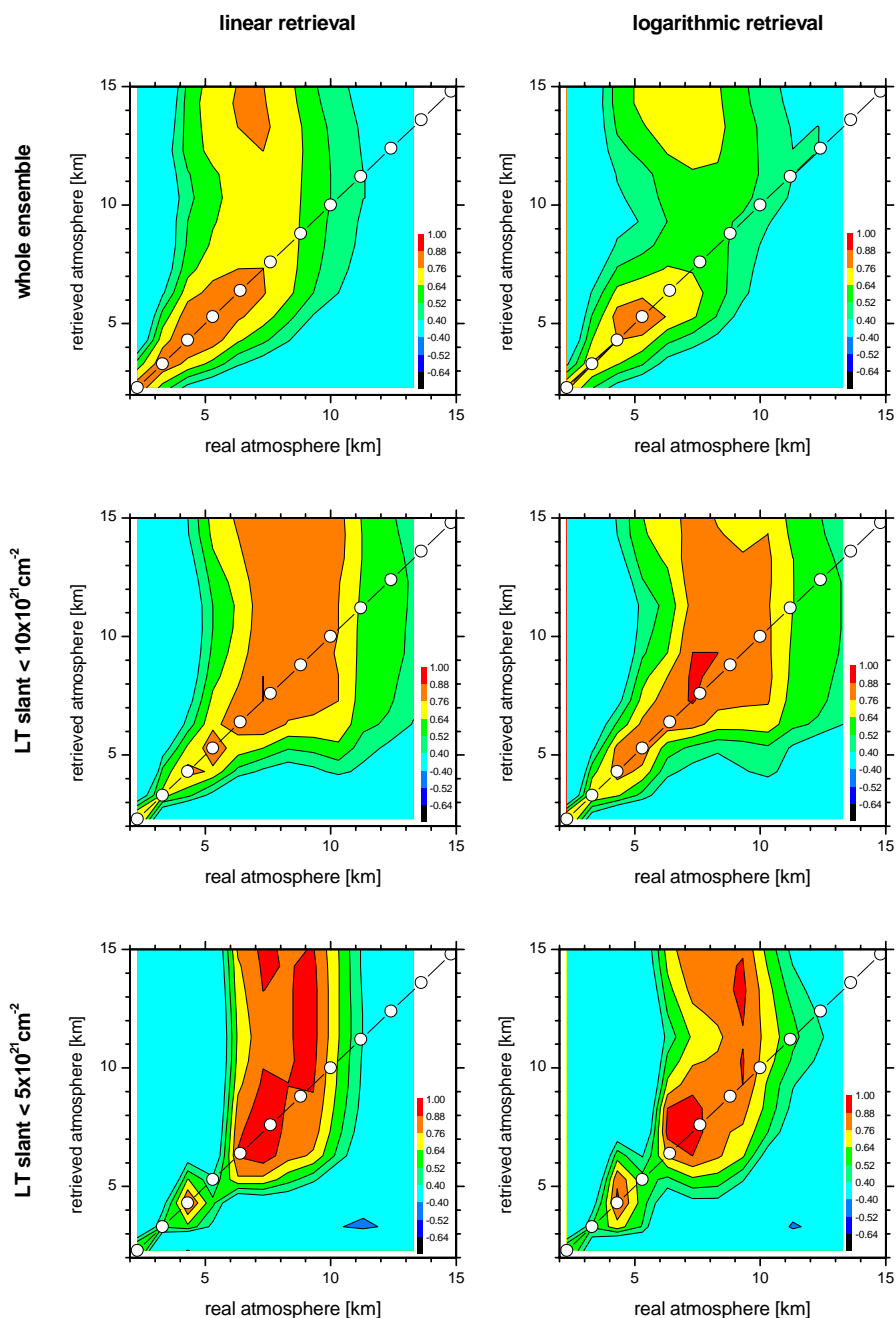
### 3.3 Total random errors

Due to the strong non-linearity of  $\hat{\mathbf{G}}$ , the total error cannot be deduced from the smoothing and parameter errors presented above. It has to be simulated separately by a full treatment. Figure 8 shows the correlation matrices for consideration of parameter errors according to Table 1 and for retrievals without simultaneous fitting of the temperature profile. It is the same as Fig. 4 but in the presence of parameter errors. The matrices for the whole ensemble (upper panels) show that the parameter errors reduce the sensitivity of both retrievals in the middle and upper troposphere. Additionally, the logarithmic retrieval performs poorly in the lower troposphere. For the low LT slant sub-ensembles (middle and lower panels), the differences to Fig. 4 are much smaller: The parameter errors are much more important for saturated than for unsaturated absorption lines, which was already observed in Fig. 7.

The total errors for this kind of retrieval are depicted in Fig. 9. If the whole ensemble is considered (black squares) even the retrieval of the 6.4–8.8 km layer becomes uncertain (noise/signal of 62% and 75% for the linear and logarithmic retrieval). For the logarithmic retrieval the large error in the lower troposphere also stands out. For the LT slant  $<10 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble, the error in the 6.4–8.8 km layer is reduced to 45% and the retrieval of the 7.6–10 km layer is possible with an uncertainty of 53%. The condition of LT slant  $<5 \times 10^{21} \text{ cm}^{-2}$  further reduces the errors: the logarithmic retrieval enables the 8.8–11.2 km layer to be retrieved with an error of only 43%. This realistic error scenario suggests that, considering the whole ensemble, the linear retrieval performs better.

The reason for the poorer performance of the logarithmic retrieval is due to the misinterpretations of spectral signatures as discussed above. There it was shown that the misinterpretation of a temperature error is strongly reduced by simultaneously fitting this parameter. Figures 10 and 11 show that this strategy is also successful concerning the total error. For the logarithmic retrieval the respective correlation matrices (Fig. 10) are very similar to those without additional parameter errors (Fig. 4). It should now be possible to retrieve water vapour amounts up to at least 7–8 km under all conditions. Figure 11 demonstrates that, for a realistic error scenario and a simultaneous fit of temperature, both linear and logarithmic retrieval yield similar random errors. Considering the whole ensemble, LT and MT amounts can be determined with an acceptable noise to signal ratio of around 22%.

Tables 2 and 3 summarize random errors for the total column amount and for partial column amounts of the LT, MT, UT, and the tropopause region. Figure 12 depicts the correlations between real amount and retrieved amount of the 4 representative layers. The left panels represent the linear and the right panels the logarithmic retrieval. While correlation coefficients are quite similar, the better performance of the logarithmic retrieval manifests itself by the good linear cor-



**Fig. 8.** Same as Fig. 4 but in the presence of parameter error as listed in Table 1.

relation with the real amounts at all altitudes, which finally results in higher sensitivity (higher values of slopes), for altitudes above 6 km, if compared to the linear retrieval.

### 3.4 Systematic errors

Already in Sect. 3.1, compared to the logarithmic method, the systematic smoothing error of the linear method is shown to be larger. At the same time the linear retrieval's error de-

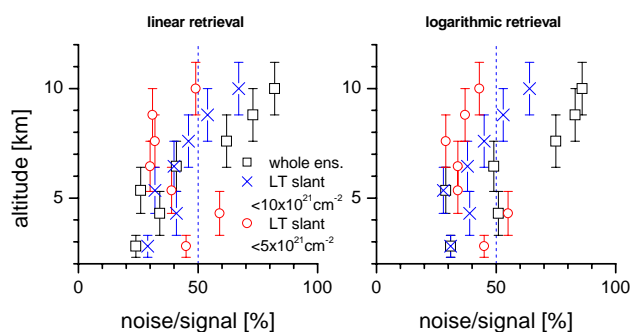
pends on the retrieved amounts: at higher altitudes it is more sensitive at small amounts than at large amounts. This inconsistency of the linear method complicates the interpretation of the amounts obtained with the linear retrieval. Neglecting this inconsistency and assuming linear correlation for both the linear and logarithmic method gives only a mean situation of their systematic smoothing error. Compared to the logarithmic retrieval the slope of the linear retrieval's regression line is further from unity, i.e. as a mean the absolute

**Table 2.** Estimated noise/signal of linear retrieval with simultaneous fitting of temperature [%]. The values for the 7.6–10.0 km and 8.8–11.2 km layers are for the LT slant  $<10 \times 10^{21} \text{ cm}^{-2}$  and LT slant  $<5 \times 10^{21} \text{ cm}^{-2}$  sub-ensembles, respectively.

error source	total	2.3–3.3 km	4.3–6.4 km	7.6–10.0 km	8.8–11.2 km
smoothing	3	14	23	44	45
meas. noise	<1	3	2	8	9
pha. err.	2	12	8	24	18
mod eff.	<1	1	<1	<1	2
T. profile	1	4	2	6	5
solar angle	<1	1	<1	<1	1
line int.	<1	<1	<1	2	1
pres. coef.	1	7	7	7	5
<b>total</b>	<b>4</b>	<b>21</b>	<b>24</b>	<b>50</b>	<b>47</b>

**Table 3.** Same as Table 2, but for logarithmic retrieval.

error source	total	2.3–3.3 km	4.3–6.4 km	7.6–10.0 km	8.8–11.2 km
smoothing	2	10	21	44	36
meas. noise	1	4	2	7	8
pha. err.	2	19	10	33	18
mod eff.	<1	1	<1	<1	<1
T. profile	1	8	6	7	3
solar angle	1	<1	<1	<1	<1
line int.	<1	1	1	1	1
pres. coef.	1	11	6	5	4
<b>total</b>	<b>4</b>	<b>22</b>	<b>24</b>	<b>49</b>	<b>42</b>

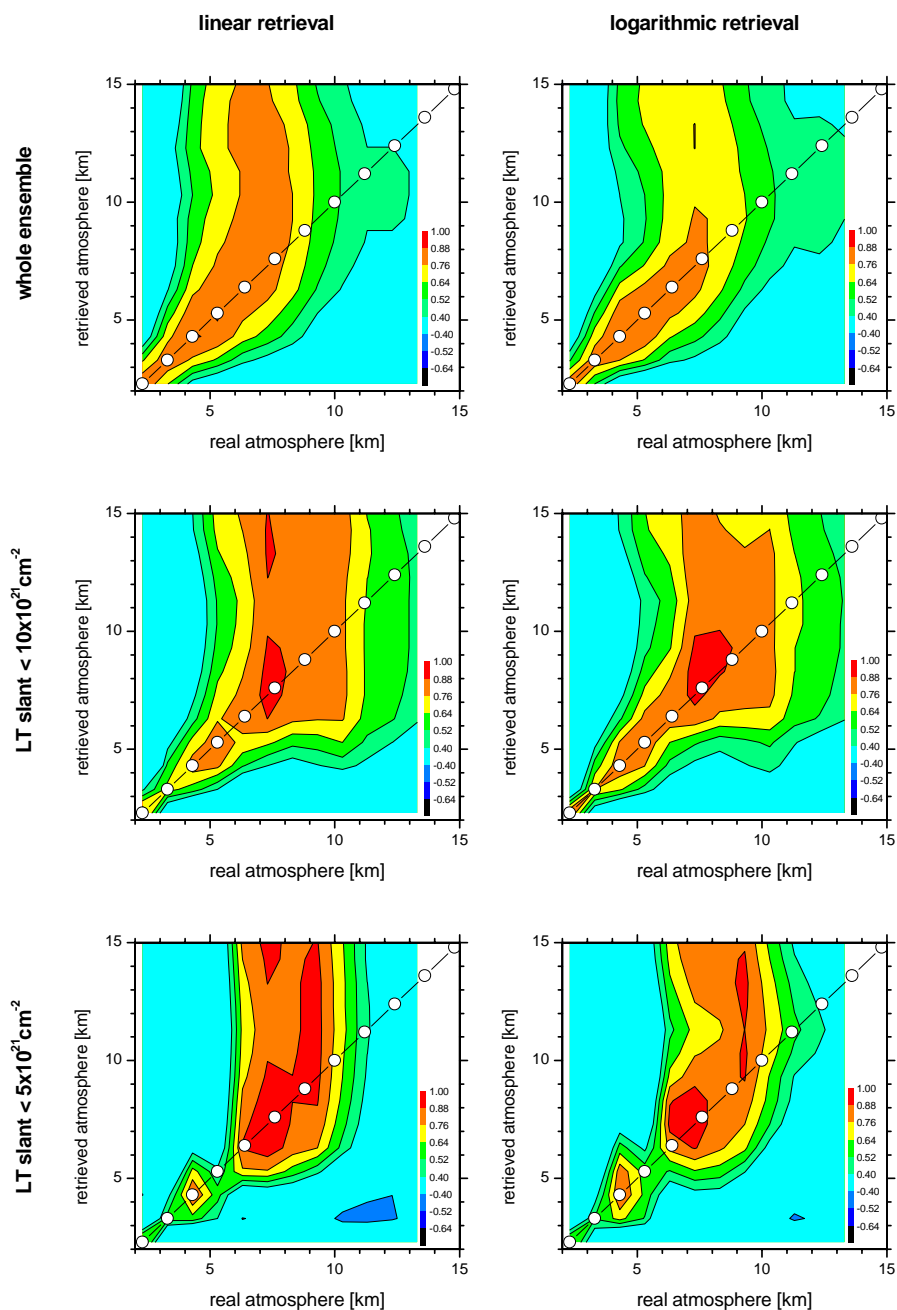
**Fig. 9.** Same as Fig. 5 but in the presence of parameter error as listed in Table 1.

variance of the linear retrieval's regression line ( $\sigma_{reg}^2$ ) agrees less well with the real variance. According to Eq. (9) and since  $\rho^2 = \sigma_{reg}^2 / \sigma_x^2$  is similar for both retrieval methods, the absolute variance of the scattering around the regression line (the random error) is then larger for the logarithmic retrieval. However, this is a mean value for the whole ensemble. A detailed analysis would reveal that the linear retrieval has increased absolute systematic errors and reduced absolute random errors only for large amounts. It is vice versa for small

amounts: the absolute systematic errors are increased and the random errors reduced. This once again manifests the dependency of the linear retrieval's errors on the retrieved amounts. On the other hand, the absolute errors of the logarithmic retrieval are practically independent from the retrieved amounts.

Additionally, systematic uncertainties of the spectroscopic line parameters may cause systematic errors. To estimate them, the retrievals of spectra simulated with correct parameters are linearly correlated to the retrievals of spectra simulated with erroneous parameters. The systematic errors are given as the difference of the regression line slope to unity.

Table 4 lists the systematic errors for the linear retrieval for the four partial column amounts representing the LT, MT, and UT, the tropopause region and for the total column amount. Table 5 lists these estimations for the logarithmic retrieval. Below 6 km they are very similar to the linear retrieval. At higher altitudes the linear retrieval's errors are larger. At these altitudes and as discussed above, it would be more correct to give two values for the linear retrieval's smoothing error: a first for low amounts, which would be reduced, and a second for large amounts, which would be increased if compared to the mean value presented. At higher altitudes the logarithmic retrieval is slightly more sensitive to the system-



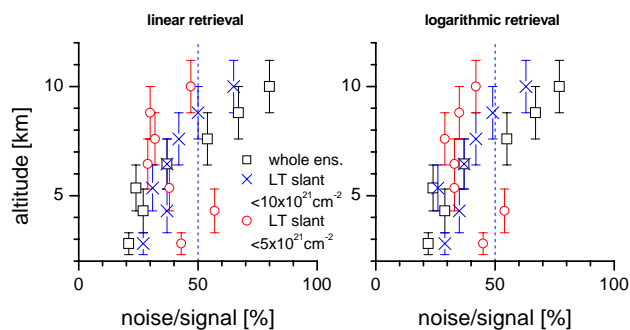
**Fig. 10.** Same as Fig. 8 but with simultaneous retrieval of temperature profile.

**Table 4.** Estimated systematic errors of linear retrieval [%]. The values for the 7.6–10.0 km and 8.8–11.2 km layers are for the LT slant  $< 10 \times 10^{21} \text{ cm}^{-2}$  and LT slant  $< 5 \times 10^{21} \text{ cm}^{-2}$  sub-ensembles, respectively.

error source	total	2.3–3.3 km	4.3–6.4 km	7.6–10.0 km	8.8–11.2 km
smoothing	0	–3	–6	–31	–38
line int.	–5	–5	–3	–3	–4
pres. coef.	0	+10	–8	–1	–4

**Table 5.** Same as Table 4, but for logarithmic retrieval.

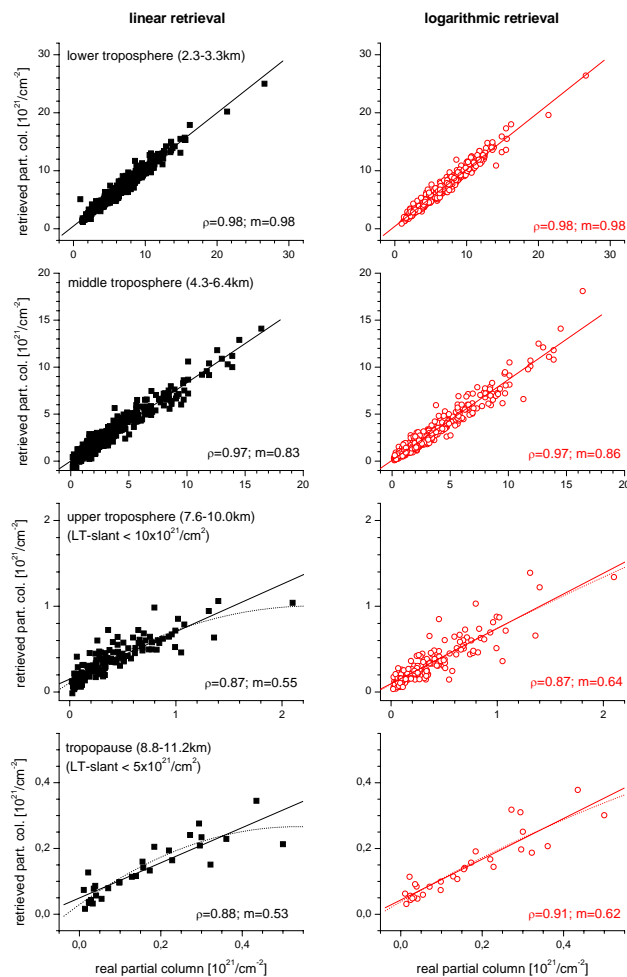
error source	total	2.3–3.3 km	4.3–6.4 km	7.6–10.0 km	8.8–11.2 km
smoothing	−1	−4	−1	−23	−33
line int.	−5	−5	−2	−4	−5
pres. coef.	+2	+19	−15	+4	−1

**Fig. 11.** Same as Fig. 9 but with simultaneous retrieval of temperature profile.

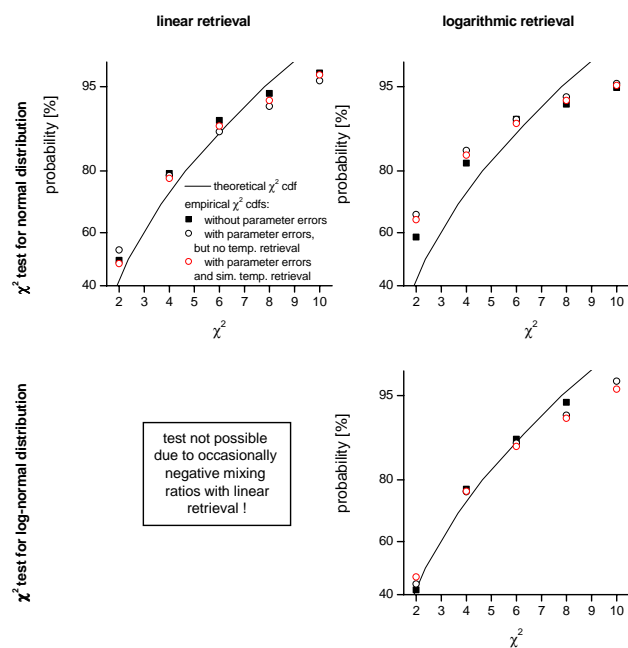
atic line intensity error, since the increasing sensitivity of the linear retrieval for high amounts reduces the slope of the regression line. The same can be observed for the pressure coefficient error: at high altitudes the linear retrieval's error always lies below the logarithmic retrieval's error.

### 3.5 Characterisation of posterior ensembles

On a logarithmic scale all involved pdfs are Gaussian distributions. A correctly working retrieval should therefore produce a normal pdf for the posterior ensemble, or if referred to the retransformed linear scale, a log-normal pdf. It should not change the principle distribution characteristics of the a-priori ensemble. The situation of the linear retrieval is different because it involves normal and log-normal pdfs. Consequently the posterior pdf may be something between a log-normal and normal pdf. A  $\chi^2$  test can check this issue. The posterior covariance matrix is  $\mathbf{S}_{\hat{\mathbf{x}}} = \epsilon \{ \hat{\mathbf{x}} \hat{\mathbf{x}}^T \}$ . In contrast to the a-priori covariance matrix  $\mathbf{S}_{\mathbf{a}}$ , the matrix  $\mathbf{S}_{\hat{\mathbf{x}}}$  is singular, since the solution space has fewer dimensions than the a-priori space. The calculation of the  $\chi^2$  values according to Eq. (5) is thus not straightforward. However, since the covariance matrix is symmetric its singular value decomposition leads to  $\mathbf{L}\mathbf{\Lambda}\mathbf{L}^T$ , with the columns of  $\mathbf{L}$  containing its eigenvectors and the diagonal matrix  $\mathbf{\Lambda}$  its corresponding eigenvalues. As  $\mathbf{S}^{-1}$  in Eq. (5) a pseudo inverse is applied, which only considers the 3 largest eigenvalues. The  $\chi^2$  calculated with this inverse would thus have 3 degrees of freedom. The test is performed for all aforementioned retrievals: with/without parameter errors and with/without simultaneous fitting of tem-

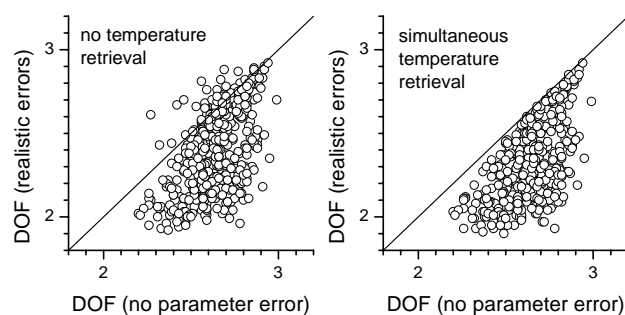
**Fig. 12.** Same as Fig. 6 but in the presence of parameter errors and with simultaneous retrieval of temperature profile.

perature. The calculations have to be performed on a logarithmic scale to check for a log-normal distribution and on a linear scale to check for a normal distribution. In Fig. 13 the theoretical  $\chi^2$  cumulative distribution function (cdf) for 3 degrees of freedom (black line) is compared to the  $\chi^2$  cdf derived from the different tests. The upper panels show the test assuming a normal distribution. The left panels show the linear retrieval and the right panel the logarithmic retrieval. The black squares (in the graph partially hidden by the red



**Fig. 13.**  $\chi^2$  test for posterior ensembles. Upper panels:  $\chi^2$  test assuming normal distribution. Lower panels:  $\chi^2$  test assuming log-normal distribution. Left panels: linear retrieval. Right panels: logarithmic retrieval. Black line: theoretical  $\chi^2$  cumulative distribution function (cdf) for 3 degrees of freedom; black filled squares: empirical  $\chi^2$  cdf of ensemble in absence of parameter errors; black circles: empirical  $\chi^2$  cdf of ensemble in the presence of parameter errors and without retrieval of temperature profile; red circles: empirical  $\chi^2$  cdf of ensemble in the presence of parameter errors and simultaneous retrieval of temperature profile.

circles) represent the posterior ensemble when no parameter errors are assumed. The linear posterior ensemble is quite consistent with a normal distribution. This means that the linear retrieval forces the originally log-normally distributed ensemble into a Gaussian ensemble. Additional errors push the solutions slightly away from a normal distribution. A simultaneous retrieval of the temperature enables a better exploitation of the information present in the spectra and leads nearly to the same distribution characteristic as if no errors were present. The logarithmic posterior ensemble has fewer characteristics of a normal distribution. Its empirical  $\chi^2$  cdfs differ considerably from the theoretical  $\chi^2$  cdf. The lower panel checks for a log-normal distribution. This test cannot be performed for the linear retrieval since it yields occasionally to negative retrieved values. In the absence of parameter errors, the logarithmic retrieval does not change the characteristics of the a-priori distribution. It is still a log-normal distribution (black squares). The presence of parameter errors pushes the posterior ensemble slightly away from a pure log-normal distribution (black and red circles).



**Fig. 14.** DOF values for logarithmic retrievals with realistic error assumptions compared to DOF values of logarithmic retrieval in the absence of errors. Left panel: no retrieval of temperature profile. Right panel: simultaneous retrieval of temperature profile.

In the case of misinterpretation of spectral signatures the logarithmic retrieval over-interprets spectral signatures. This can be demonstrated by analysing the trace of the averaging kernel matrix ( $\text{tr}(\hat{\mathbf{A}})$ ). It determines the amount of information present in the spectra used by the retrieval for updating the a-priori state. It is commonly called the degree of freedom of the measurement (DOF). Figure 14 compares the DOF values for the logarithmic retrievals with and without additional errors. If the retrieval is working correctly adding further errors should reduce the DOF value, since the information in the spectra is more uncertain. However, on a logarithmic scale occasionally the contrary is observed. If the temperature profile is not simultaneously fitted (left panel of Fig. 14) occasionally more information is retrieved from the erroneous spectra than from the spectra with only white noise, which means that errors in the spectra are misinterpreted as information. This problem disappears by fitting the temperature profile simultaneously (right panel).

## 4 Comparison of retrieval results to ptu-sonde measurements

### 4.1 The FTIR measurements

Since March 1999 measurements of highly-resolved infrared solar absorption spectra are routinely performed at the Izaña Observatory, situated on the Canary Island of Tenerife (28°18' N, 16°29' W) at 2370 m a.s.l. Its position in the Atlantic Ocean and above a stable inversion layer, typical for subtropical regions, provides clean air and clear sky conditions most of the year. This offers good conditions for atmospheric observations by remote sensing techniques. The spectra are obtained by a Bruker IFS 120M applying a resolution of 0.0036 to 0.005  $\text{cm}^{-1}$  and no numerical apodisation. The spectral intensities are determined by a liquid-nitrogen cooled HgCdTe detector, which, in order to ensure linearity, is operated in a photovoltaic mode. During short periods in 1999 and 2001 a photoconductive detector was

applied whose nonlinearities were corrected. The spectra are typically constructed by co-adding up to 8 scans recorded in about 10 or 13 min, depending on their resolution. Analysing the shape of the absorption lines (lines are widened by pressure broadening) and their different temperature sensitivities enables the retrieval of the absorbers' vertical distribution. Since the instrumental line shape (ILS) also affects the shape of the measured absorption lines, this instrumental characteristic should be determined independently from the atmospheric measurements. This is done on average every two months using cell measurements and LINEFIT software as described in Hase et al. (1999). The temperature and pressure profiles, necessary for the inversion, are taken from the synoptical meteorological 12:00 UT sondes. Above 30 km data from the Goddard Space Flight Center's automailer system are applied. Some results of these measurements are presented in Schneider et al. (2005) and references therein.

#### 4.2 The radiosonde measurements

Until September 2002 the meteorological soundings were launched from Santa Cruz de Tenerife, 35 km northeast of the observatory, and since October 2002 in an automatised mode from Güimar, 15 km southeast of the observatory. The sondes are equipped with a Vaisala RS80-A thin-film capacitive sensor which determines relative humidity. The sonde data are corrected by a method suggested by Leiterer et al. (2004), who reported a remaining random error of less than 5% throughout the troposphere. Other authors report correction methods with a remaining uncertainty of over 10% (Miloshevich, 2001). Furthermore, the precision of the water vapour measured by the RS80-A sensor may be degraded due to chemical contamination during storage. To avoid sondes with iced detectors, sondes that passed through clouds are not taken into account. Therefore sondes which detect a vapour pressure close to the liquid or ice saturation pressure are disregarded. Furthermore, sondes with unrealistic high humidities above 10 km, which may indicate an iced detector, are excluded. The corrected sonde mixing ratios are finally sampled on the altitude grid of the retrieval by requiring that linear interpolation of the mixing ratios between two grid levels yield the same partial columns as the original highly-resolved data.

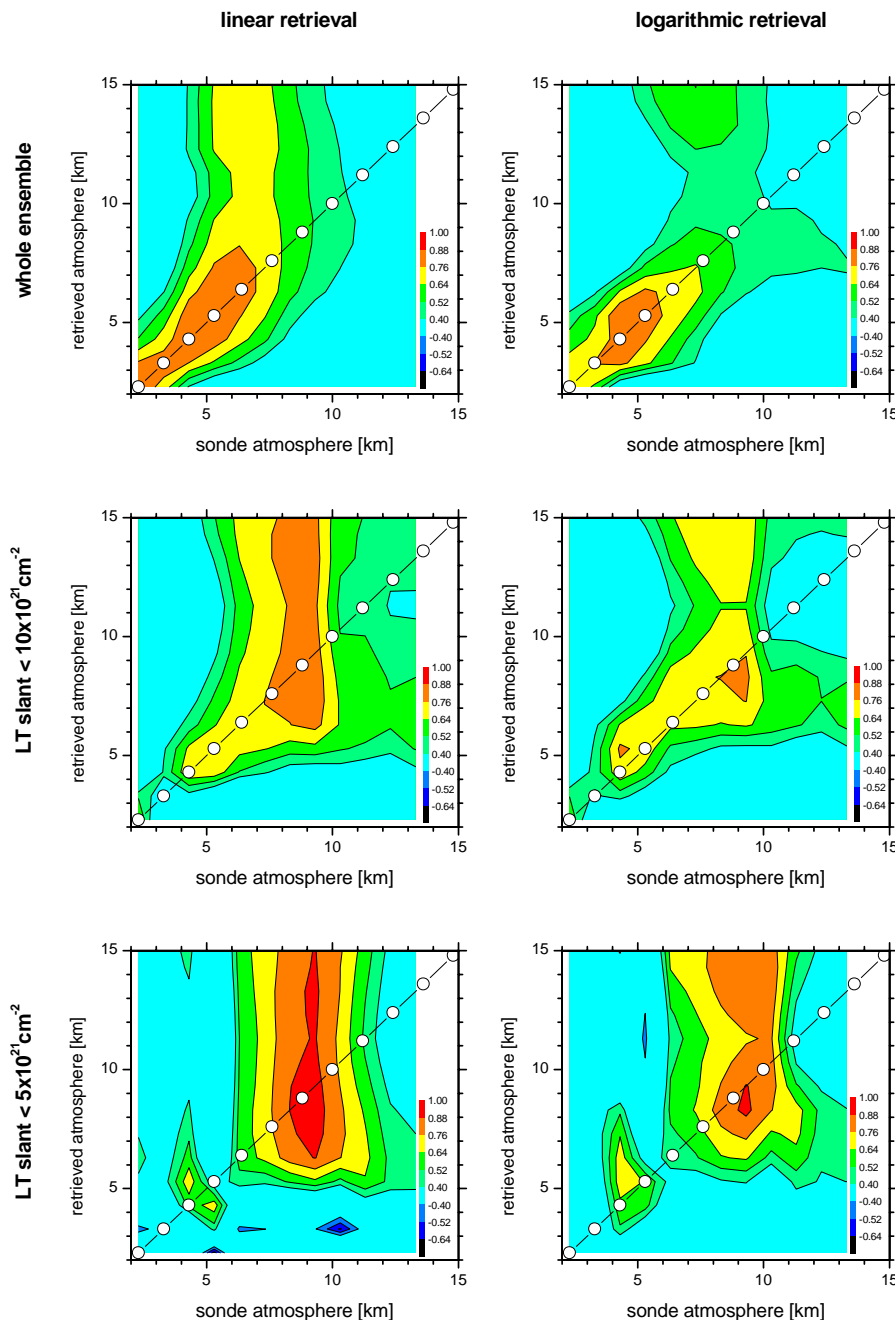
#### 4.3 Temporal and spatial variability

The large temporal and spatial variabilities of atmospheric water vapour are problematic when measurements conducted from different platforms are to be compared. Both experiments should be conducted at the same time and sound the same atmospheric location. For this reason only sonde measurements coinciding within 2 h of the FTIR measurements are used for the comparison. Spatial coincidence is difficult to achieve. The sonde measures in-situ and will always be situated at a certain distance from the imaginary line between

the FTIR instrument and the sun. This is particularly problematic for the lowest layer above the FTIR instrument as, while the FTIR instrument is located at the surface the sonde is typically floating around 30 km south of the observatory in the free troposphere. A comparison between the humidity measured in-situ at the observatory and the sonde's humidity demonstrated that the water vapour amounts close to the surface are more variable and on average 40% larger than those in the free troposphere.

#### 4.4 Comparison

Within the comparison period, from March 1999 to January 2004, the criteria for sonde quality (no clouds, realistic humidity above 10 km) and temporal coincidence with FTIR measurements are fulfilled in 157 occasions only. 59 of them also belong to the LT slant  $< 10 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble and 19 to the LT slant  $< 5 \times 10^{21} \text{ cm}^{-2}$  sub-ensemble. Figure 15 presents correlation matrices of FTIR and sonde profiles. They are the experimental analogue to the simulated correlations shown in Fig. 10. The upper panels show the situation for all coincidences and the lower panels for low LT slant column amounts. Keeping in mind the errors of the sonde data and temporal and spatial mismatching, the overall agreement of these correlation patterns with the simulated pattern is very good. Considering all situations the linear retrieval is apparently more consistent with the sonde measurements than the logarithmic retrieval, since it has slightly larger  $\rho$  values along the diagonal of the matrix. The degraded performance of the logarithmic retrieval may be due to a slight misinterpretation of an incorrect ILS characterisation. As seen in Fig. 7, the phase error is similar to the temperature error and may cause similar problems if the assumptions of Table 1 are too optimistic for the applied Bruker IFS 120M spectrometer. In this case, the logarithmic retrieval may be improved even further by a simultaneous retrieval of the ILS. However, it should be considered that the linear retrieval has large outer diagonal elements, in particular above 5 km. For the logarithmic retrieval, on the other hand, large correlation coefficients are well centred around the diagonal, which counterbalances the lower diagonal values, since it means that the correlation lengths towards sonde mixing ratios are smaller compared to those of the linear retrieval. This is a consequence of the poorer vertical resolution of the latter (see explanations about smoothing error in Sect. 3), and even more important considering the situation of the upper troposphere and tropopause region for days with low LT slant column amounts. Here the  $\rho$  values on the diagonal are quite similar for the linear and logarithmic retrieval. However, the logarithmic amounts above 9 km are less correlated with the sonde measurements around 8 km. For example, the state retrieved at 10 km has a  $\rho$  value with the real state at 8 km of 0.82 in the linear and 0.70 in the logarithmic case only. This results in a much smaller systematic difference between the sonde and FTIR for the logarithmic



**Fig. 15.** Same as Fig. 10 but for correlation matrices between measured sonde and FTIR profiles.

retrieval at high altitudes if compared to the linear retrieval. Tables 6 and 7 list these differences. They are calculated from least squares fits as described in section 3.1. The difference to unity of the slope gives the systematic deviation and the scattering around the regression line gives the random deviation. This scattering describes the level of consistency between the variabilities detected by the sonde and the FTIR measurements. It may also be seen as the overall precision of FTIR and sonde experiments together. For the UT and tropopause layer and considering the coincidences with

low LT slant column amounts only, it is situated around 55% for the linear as well as the logarithmic retrieval. These calculations even disregard temporal and spatial mismatching of both measurements. These values are therefore – at least qualitatively – well consistent with the simulations in Sect. 3, where the total random error of the FTIR measurements for these altitudes is estimated to be situated around 45% for the linear and logarithmic retrieval (see total error in Tables 2 and 3).



**Table 6.** Differences between sonde and FTIR column amounts as estimated from the correlation plots (Fig. 16). The values for the 7.6–10.0 km and 8.8–11.2 km layers are for the LT slant  $<10 \times 10^{21} \text{ cm}^{-2}$  and LT slant  $<5 \times 10^{21} \text{ cm}^{-2}$  sub-ensembles, respectively.

	total	2.3–3.3 km	4.3–6.4 km	7.6–10.0 km	8.8–11.2 km
random	25	40	32	54	56
systematic	+6	+3	−4	−40	−47

**Table 7.** Same as Table 6 but for logarithmic retrieval.

	total	2.3–3.3 km	4.3–6.4 km	7.6–10.0 km	8.8–11.2 km
random	25	47	33	58	51
systematic	+6	−4	+1	+2	−10

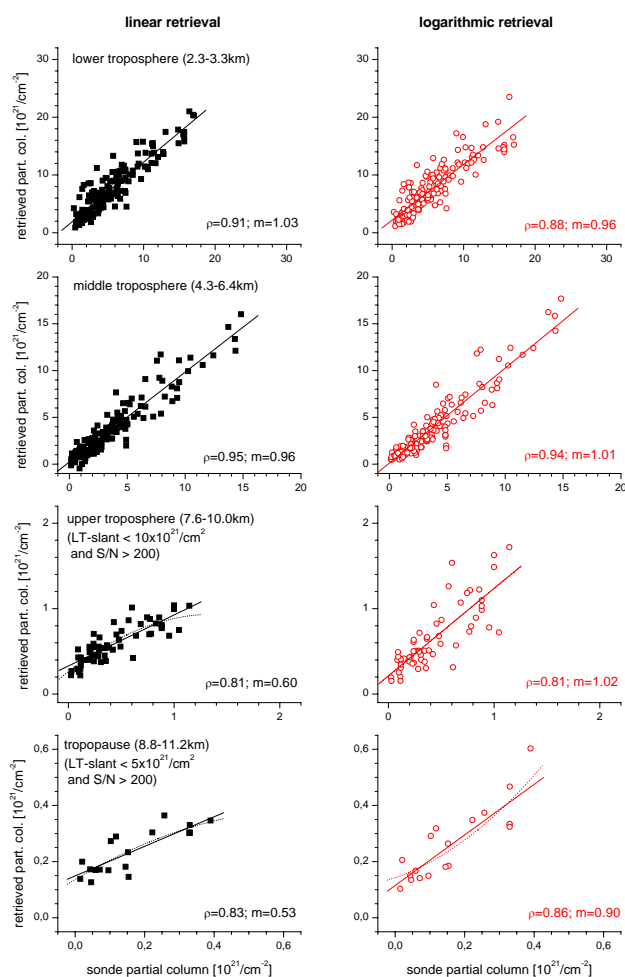
An outstanding difference to Tables 2 and 3 is the poorer consistency for the LT layer of FTIR when compared to sonde than when compared within the simulations: empirical standard deviation of  $\approx 45\%$  compared to the estimated values of below 22%. This is due to the aforementioned different conditions in the lowermost layer above the instrument (surface influences) and the corresponding layer at the sonde (free troposphere). Since the LT mainly determines the total column amount, the latter is also largely affected by these differences. The estimated and empirically observed precision for the MT are much more consistent: estimated noise to signal for the FTIR of 24% versus measured  $\approx 32\%$  for both experiments together.

Figure 16 shows the correlation between LT, MT, UT, and tropopause partial column amounts of FTIR and sonde measurements. The greatest differences with Fig. 12 are observed for the LT (as discussed above), where the regression line between sonde and FTIR data has an offset of  $\approx 2.5 \times 10^{21} \text{ cm}^{-2}$ : the LT at the site of the instrument is more humid than the free tropospheric LT. For the MT the consistency between the simulations and the empirical observations is excellent, even though the errors of the sonde measurements and temporal and spatial mismatching are still disregarded. The regression lines for the UT and the tropopause region show a small offset. Retrieved amounts are  $\approx 1 \times 10^{20} \text{ cm}^{-2}$  larger than the sonde amounts. This may confirm a dry bias of the sonde measurements (Turner et al., 2003). The slopes of the linear regression lines for the UT and tropopause region are 1.02 and 0.90 for the logarithmic retrieval. These values are much larger than the simulated slopes of 0.64 and 0.62. The reason may be that the humidity applied for the first layer in the simulations differs from the real humidity of this layer. Due to the aforementioned different condition at the sonde and at instrument altitude (2.3 km), a mixing ratio determined by an in-situ instrument was applied for the simulation. This relatively high value is then spread out up to the next grid point (3.3 km). However, the

enhanced humidity due to surface conditions is very likely limited to the lowest 100 m of the atmosphere. This overestimation of simulated LT amounts reduces the mean estimated sensitivity in the UT and tropopause region.

Figure 16 further demonstrates that the logarithmic retrieval is correlated linearly to the sonde data at all altitudes, whereby for high altitudes the linear retrieval's regression line underestimates both especially large and small amounts. This is consistent with the simulations (Fig. 12). The experiments confirm that at high altitudes the linear retrieval is less sensitive at large amounts if compared to small amounts. The empirical validation suggests that the differences between the linear and logarithmic retrievals' systematic errors are even more pronounced than proposed by the theoretical study performed in Sect. 3. This is reflected in the larger differences between the slopes of the regression lines for the linear and logarithmic retrieval. While at higher altitudes and for days with low LT slant column amounts, slopes of around 0.53 for the linear retrieval versus 0.63 for the logarithmic retrieval are simulated, the empirical validation yields 0.57 versus 0.96. An explication may be that the assumed measurement noise is underestimated in the simulations, since all spectra were calculated for no aerosol loading. More measurement noise would mean that the a-priori information is more important and, since the linear retrieval applies a wrong a-priori, the caused systematic error would increase.

The empirical validation not only confirms the increased systematic error of the linear retrieval, it even suggests that the improvements by the logarithmic method are more pronounced than indicated in the simulations. The empirical validation – in consistency with the simulations – shows that the linear retrieval of water vapour at high altitudes has important inconsistencies. Only the logarithmic method is a self-consistent retrieval method. Its retrieved amounts are linearly correlated with the real amounts and are well-suited for detection of water vapour variabilities in the UT and the tropopause region.



**Fig. 16.** Same as Fig. 12 but for measured sonde and FTIR partial column amounts.

## 5 Subtropical water vapour time series

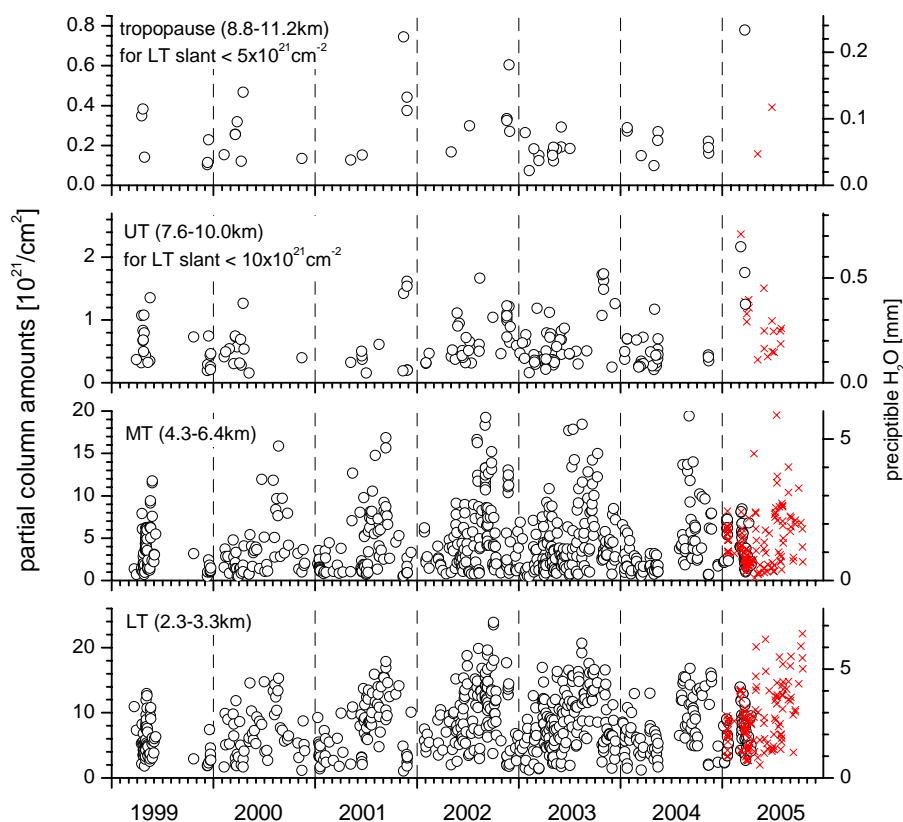
Figure 17 depicts a nearly 7 year record of tropospheric water vapour amounts as determined by the logarithmic retrieval with simultaneous fitting of the temperature. The black circles show data from the Bruker IFS 120M, which was operated until April 2005. The red crosses are results as obtained from a Bruker 125HR, which measures since January 2005. While for the lower and middle tropospheric values all measurement days are depicted, the upper tropospheric and tropopause values are presented only when the LT slant column amounts are lower than  $10 \times 10^{21} \text{ cm}^{-2}$  and  $5 \times 10^{21} \text{ cm}^{-2}$ , respectively. For the lower and middle troposphere a well pronounced seasonal cycle is observed. Values are highest at the end of summer and lowest in the winter months. A similar clear seasonal dependence is not observed for the upper tropospheric amounts and the amounts of the tropopause region. Values are sometimes even especially high in autumn/winter, which demonstrates their inde-

pendence from lower tropospheric levels. A quick view may give the impression of increasing water vapour contents in the upper troposphere; however, for a serious trend analysis a longer and more continuous time series would be needed.

## 6 Summary and conclusions

Compared to other atmospheric components, the retrieval of atmospheric water vapour from ground-based FTIR measurements has additional difficulties. Water vapour has very large vertical gradients and variabilities, which generally limit the sensitivity of the ground-based technique to the lower and middle troposphere. The spectral signatures originating from the upper troposphere are rather weak and thus their retrieved values depend to an important extent on a-priori assumptions. Water vapour mixing ratios are log-normally distributed and an inversion on a logarithmic scale enables the correct application of this a-prior knowledge and consequently leads to a statistically optimal retrieval. However, this method introduces the risk of misinterpreting spectral signatures produced by errors in assumed model parameters. It is shown that the misinterpretations can be controlled by simultaneously fitting the temperature profile. A logarithmic retrieval should therefore perform better than the commonly applied linear retrieval, in particular for high altitudes where the spectral signatures are similar to the measurement noise. It is found that the linear retrieval leads to large systematic errors, which are difficult to characterise. They can be observed in correlation plots between retrieved and real amounts. The complex character of the linear retrieval's systematic error has important consequences: it limits the linear retrieval in correctly detecting variabilities present in time series. It would underestimate alterations towards large amounts and overestimate alterations towards small amounts. The systematic error of the logarithmic retrieval is smaller. Its amounts are almost linearly correlated to the real amounts, i.e. its sensitivity is independent from the retrieved amount. For an analysis of water vapour time series of the upper troposphere and the tropopause region the logarithmic retrieval has to be applied. A realistic error scenario simulates random errors of 4% for the total column amounts and around 23% for amounts of the lower and middle troposphere. On days with low LT slant amounts, amounts of the upper tropospheric and the tropopause region can also be determined with an uncertainty of around 45%. Furthermore, it is found that, in addition to the limited vertical resolution, the uncertainties in the instrumental line shape (phase error) are responsible for the most important errors. All these estimations are confirmed by a comparison to sonde measurements.

The advantage of the FTIR technique compared to the meteorological sondes is that the errors are well understood and water isotope evaluation is possible. This may allow a study of hydrometeorological processes in the atmosphere.



**Fig. 17.** Time series of water vapour above Tenerife island determined from FTIR measurements. From the top to the bottom: tropopause column amounts (8.8–11.2 km), upper tropospheric column amounts (7.6–10.0 km), middle tropospheric column amounts (4.3–6.4 km), and lower tropospheric column amounts (2.3–3.3 km). Black circles: measurements of Bruker IFS 120M. Red crosses: Measurements of Bruker IFS 125HR.

The suggested method can be applied to other dataset of highly-resolved infrared spectra (e.g. to measurements made within the Network for Detection of Stratospheric Change). However, the capability of the method would have to be investigated for each measurement site individually. If the instrumental line shape (phase error) is well characterised and the temperature profile data are reliable even on days with saturated absorption lines, lower and middle tropospheric amounts can be detected. The upper tropospheric sensitivity is expected to be better the lower the water vapour content in the lowest layers and the stabler the instrumental line shape. In this context the subtropical site of Izaña, located on an island, and the application of a Bruker IFS 120M are not the best conditions. For measurements made by a Bruker IFS 125HR even lower uncertainties should be expected. At less humid or higher located measurement sites, the conditions necessary for the detection of water vapour in the tropopause region are more frequently fulfilled. Applying the proposed retrieval method to spectra measured during the last 20–25 years at FTIR sites such as Jungfraujoch or Kitt Peak could produce unique continuous long-term series of UT/LS water vapour amounts.

*Acknowledgements.* We would like to thank the Bundesministerium für Bildung und Forschung for funding via the DLR (contracts 50EE0008 and 50EE0203). Furthermore, we are grateful to the Izaña Observatory for facilitating the sonde data and for allowing us to use its infrastructure and to the Goddard Space Flight Center for providing the temperature and pressure profiles of the National Centers for Environmental Prediction via the automailer system.

Edited by: M. G. Lawrence

## References

- Harries, J. E.: Atmospheric radiation and atmospheric humidity, *Q. J. R. Meteorol. Soc.*, 123, 2173–2186, 1997.
- Hase, F., Blumenstock, T., and Paton-Walsh, C.: Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software, *Appl. Opt.*, 38, 3417–3422, 1999.
- Hase, F., Hannigan, J. W., Coffey, M. T., Goldman, A., Höpfner, M., Jones, N. B., Rinsland, C. P., and Wood, S. W.: Intercomparison of retrieval codes used for the analysis of high-resolution,

- ground-based FTIR measurements, *J. Quant. Spectrosc. Radiat. Transfer*, 87, 25–52, 2004.
- Höpfner, M., Stiller, G. P., Kuntz, M., Clarmann, T. v., Echle, G., Funke, B., Glatthor, N., Hase, F., Kemnitzer, H., and Zorn, S.: The Karlsruhe optimized and precise radiative transfer algorithm, Part II: Interface to retrieval applications, *SPIE Proceedings 1998*, 3501, 186–195, 1998.
- Kuntz, M., Höpfner, M., Stiller, G. P., Clarmann, T. v., Echle, G., Funke, B., Glatthor, N., Hase, F., Kemnitzer, H., and Zorn, S.: The Karlsruhe optimized and precise radiative transfer algorithm, Part III: ADDLIN and TRANSF algorithms for modeling spectral transmittance and radiance, *SPIE Proceedings 1998*, 3501, 247–256, 1998.
- Kurylo, M. J.: Network for the detection of stratospheric change (NDSC), *Proc. SPIE–Int. Co. Opt. Eng.* 1991, 1491, 168–174, 1991.
- Kurylo, M. J. and Zander, R.: The NDSC – Its status after 10 years of operation, *Proceedings of XIX Quadrennial Ozone Symposium*, Hokkaido University, Sapporo, Japan, 167–168, 2000.
- Leiterer, U., Dier, H., Nagel, D., Naebert, T., Althausen, D., Franke, K., Kats, A., and Wagner, F.: Correction Method for RS80-A Humicap Humidity Profiles and their Validation by Lidar Backscattering Profiles in Tropical Cirrus Clouds, *J. Atmos. Oceanic Technol.*, 22, 18–29, 2005.
- Miloshevich, L. M., Vömel, H., Paukkunen, A., Heymsfield, A. J., and Oltmans, S. J.: Characterization and correction of relative humidity measurements from Viasalla RS80-A radiosondes at cold temperatures, *J. Atmos. Oceanic Technol.*, 18, 135–155, 2001.
- NDSC: <http://www.ndsc.ws/>, 2005.
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Praxis*, World Scientific Publishing Co., Singapore, 2000.
- Rothman, L. S., Barbe, A., Benner, D. C., Brown, L. R., Camy-Peyret, C., Carleer, M. R., Chance, K. V., Clerbaux, C., Dana, V., Devi, V. M., Fayt, A., Fischer, J., Flaud, J.-M., Gamache, R. R., Goldman, A., Jacquemart, D., Jucks, K. W., Lafferty, W. J., Mandin, J.-Y., Massie, S. T., Newnham, D. A., Perrin, A., Rinsland, C. P., Schroeder, J., Smith, K. M., Smith, M. A. H., Tang, K., Toth, R. A., Vander Auwera, J., Varanasi, P., and Yoshino, K.: *The HITRAN Molecular Spectroscopic Database: Edition of 2000 Including Updates through 2001*, *J. Quant. Spectrosc. Radiat. Transfer*, 82, 5–44, 2003.
- Schneider, M., Blumenstock, T., Chipperfield, M., Hase, F., Kouker, W., Reddmann, T., Ruhnke, R., Cuevas, E., and Fischer, H.: Subtropical trace gas profiles determined by ground-based FTIR spectroscopy at Izaña (28°, 16°): Five year record, error analysis, and comparison with 3D-CTMs, *Atmos. Chem. Phys.*, 5, 153–167, 2005.
- SPARC: *Assessment of Upper Tropospheric and Stratospheric Water Vapour*, edited by: Kley, D., Russell III, J. M., and Phillips, C., WCRP-113, WMO/TD-No. 1043, SPARC report No. 2, December, 2000.
- Spencer, R. W. and Braswell, W. D.: How dry is the tropical free troposphere? Implications for global warming theory, *Bull. Am. Meteorol. Soc.*, 78, 1097–1106, 1997.
- Stiller, G. P., Höpfner, M., Kuntz, M., Clarmann, T. v., Echle, G., Fischer, H., Funke, B., Glatthor, N., Hase, F., Kemnitzer, H., and Zorn, S.: The Karlsruhe optimized and precise radiative transfer algorithm, Part I: Requirements, justification and model error estimation, *SPIE Proceedings 1998*, 3501, 257–268, 1998.
- Turner, D. D., Lesht, B. M., Clough, S. A., Liljegren, J. C., Revercomb, H. E., and Tobin, D. C.: Dry Bias and Variability in Vaisala RS80-H Radiosondes: The ARM Experience, *J. Atmos. Oceanic Technol.*, 20, 117–132, 2003.
- Wagner, G., Birk, M., Schreier, F., and Flaud, J.-M.: Spectroscopic database for Ozone in the fundamental spectral regions, *J. Geophys. Res.*, 107, 4626–4643, 2002.
- Wilks, D. S.: *Statistical methods in the atmospheric science*, Academic Press, ISBN 0-12-751965-3, 1995.