Uncertain Model Parameters for SCIAMACHY Limb Retrieval

Adrian Doicu(1), **Siegfried Hilgers**(1), **Albrecht von Bargen**(1) **and Bernd Aberle**(1)

German Aerospace Center, Oberpfaffenhofen, 82234 Wessling, Germany, Email: adrian@doicu.dlr.de

1. INTRODUCTION

Atmospheric profiles are retrieved by the inverse solution of the radiative transfer equation. The solution is afected by uncertain model parameters, which add an error term to the solution. The uncertain parameters can be estimated simultaneously with the atmospheric profiles to be retrieved [1], or equivalently a generalized covariance matrix approach can be applied [2]. In both situations, the determination of the weighting factor giving the contribution of each regularization matrix into the global regularization matrix plays an important role. In this paper we present an a priori parameter choice method for computing the optimal value of the regularization parameter and the weighting factor for the SCIAMACHY limb retrieval processor.

2. THEORY

Let us assume that the forward model $\mathbf{F}(\mathbf{x}, \mathbf{b})$ describing the complete physics of the measurement (including for example the radiative transfer theory required to relate the state to the measured signal) depends on the state vector $\mathbf{x}$ and a set of auxiliary parameters $\mathbf{b}$. The vector of auxiliary parameters $\mathbf{b}$ comprises those quantities which influence the measurement but are not intended to be retrieved with an acceptable accuracy. The auxiliary parameters can be classified as deterministic or random according to whether they are constant between consecutive measurements or vary randomly. In order to simplify our analysis we will consider a linearization of the forward model $\mathbf{F}$ around the a priori $(\mathbf{x}_{\mathtt{a}}, \mathbf{b}_{\mathtt{a}})$,

$$\mathbf{F}(\mathbf{x}, \mathbf{b}) \approx \mathbf{F}(\mathbf{x}_{\mathtt{a}}, \mathbf{b}_{\mathtt{a}}) + \mathbf{K}_{\mathtt{x}}(\mathbf{x} - \mathbf{x}_{\mathtt{a}}) + \mathbf{K}_{\mathtt{b}}(\mathbf{b} - \mathbf{b}_{\mathtt{a}}).$$

Denoting by $\mathbf{F}^{\delta}$ the measurement vector and making the change of variables $\mathbf{F}^{\delta} - \mathbf{F}(\mathbf{x}_{\mathtt{a}}, \mathbf{b}_{\mathtt{a}}) \to \mathbf{y}^{\delta}$, $\mathbf{x} - \mathbf{x}_{\mathtt{a}} \to \mathbf{x}$ and $\mathbf{b} - \mathbf{b}_{\mathtt{a}} \to \mathbf{b}$ we are led to the general linear data model

$$\mathbf{y}^{\delta} = \mathbf{K}_{\mathtt{x}}\mathbf{x} + \mathbf{K}_{\mathtt{b}}\mathbf{b} + \delta, \tag{1}$$

where δ is the experimental error term. The random vector $\delta = [\delta_1, ..., \delta_m]^T$ is assumed to have zero mean and the covariance matrix $\mathbf{C}_{\delta}$, i.e., $\mathcal{E}\{\delta\} = 0$ and $\mathbf{C}_{\delta} = \mathcal{E}\{\delta\delta^T\}$. Explicitly we have $\mathcal{E}\{\delta_i\} = 0$ and $[\mathbf{C}_{\delta}]_{ij} = \mathcal{E}\{\delta_i\delta_j\}$ for $i, j = 1, ..., m$. The random vector δ is called a discrete white noise vector if $\mathbf{C}_{\delta} = \sigma^2\mathbf{I}_m$, with σ^2 being the noise variance. Note that each data model with an arbitrary noise covariance matrix can be transformed into a data model with white noise by using the prewhitening technique. For a symmetric and positive definite covariance matrix with the singular value decomposition $\mathbf{C}_{\delta} = \mathbf{U}_{\delta}\Sigma_{\delta}\mathbf{U}_{\delta}^T$, where $\Sigma_{\delta} = \left[\mathrm{diag}\left(\eta_k\right)_{m\times m}\right]$, we define $\sigma^2 = (1/m)\sum_{k=1}^{m}\eta_k^2$ and $\widehat{\mathbf{C}}_{\delta} = \left(1/\sigma^2\right)\mathbf{C}_{\delta}$. Multiplying equation (1) by $\widehat{\mathbf{C}}_{\delta}^{-1/2}$, where $\widehat{\mathbf{C}}_{\delta}^{-1/2} = \mathbf{U}_{\delta}\widehat{\Sigma}_{\delta}^{-1/2}\mathbf{U}_{\delta}^T$ and $\widehat{\Sigma}_{\delta} = \left(1/\sigma^2\right)\Sigma_{\delta}$, and considering the change of variables $\widehat{\mathbf{C}}_{\delta}^{-1/2}\mathbf{y}^{\delta} \to \mathbf{y}^{\delta}$, $\widehat{\mathbf{C}}_{\delta}^{-1/2}\mathbf{K}_{\mathtt{x}} \to \mathbf{K}_{\mathtt{x}}$, $\widehat{\mathbf{C}}_{\delta}^{-1/2}\mathbf{K}_{\mathtt{b}} \to \mathbf{K}_{\mathtt{b}}$, $\widehat{\mathbf{C}}_{\delta}^{-1/2}\delta_{\mathtt{m}} \to \delta_{\mathtt{m}}$, and $\widehat{\mathbf{C}}_{\delta}^{-1/2}\delta \to \delta$, we obtain the data model (1) with $\mathcal{E}\{\delta\} = 0$ and $\mathbf{C}_{\delta} = \mathcal{E}\{\delta\delta^T\} = \sigma^2\mathbf{I}_m$.

In the framework of the Bayesian estimation, the computation of the maximum a posteriori estimator (MAP) depends on the construction of the a priori terms for $\mathbf{b}$ and $\mathbf{x}$. Let $\mathbf{C}_{\mathtt{x}}$ and $\mathbf{C}_{\mathtt{b}}$ denote the a priori covariance matrices of the state vector and auxiliary parameters, respectively. Then, assuming the representations $\mathbf{C}_{\mathtt{x}} = \sigma_{\mathtt{x}}^2\widehat{\mathbf{C}}_{\mathtt{x}}$ and $\mathbf{C}_{\mathtt{b}} = \sigma_{\mathtt{b}}^2\widehat{\mathbf{C}}_{\mathtt{b}}$, and the Choleski factorizations $\widehat{\mathbf{C}}_{\mathtt{x}}^{-1} = \mathbf{L}_{\mathtt{x}}^T\mathbf{L}_{\mathtt{x}}$ and $\widehat{\mathbf{C}}_{\mathtt{b}}^{-1} = \mathbf{L}_{\mathtt{b}}^T\mathbf{L}_{\mathtt{b}}$, the negative of the a posteriori log likelihood function read as

$$-2\sigma^2\mathcal{L}(\mathbf{x}) = \left\|\mathbf{y}^{\delta} - [\mathbf{K}_{\mathtt{x}}, \mathbf{K}_{\mathtt{b}}]\begin{bmatrix}\mathbf{x}\\ \mathbf{b}\end{bmatrix}\right\|^2 + \sigma^2\left\|\begin{bmatrix}\frac{1}{\sigma_{\mathtt{x}}}\mathbf{L}_{\mathtt{x}} & \mathbf{0}\\ \mathbf{0} & \frac{1}{\sigma_{\mathtt{b}}}\mathbf{L}_{\mathtt{b}}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{b}\end{bmatrix}\right\|^2. \tag{2}$$

Since both $\sigma_{\mathtt{b}}$ and $\sigma_{\mathtt{x}}$ are unknown, we consider the multi-parameter problem

$$\mathcal{F}_{\alpha\omega}(\mathbf{x}) = \left\|\mathbf{y}^{\delta} - [\mathbf{K}_{\mathtt{x}}, \mathbf{K}_{\mathtt{b}}]\begin{bmatrix}\mathbf{x}\\ \mathbf{b}\end{bmatrix}\right\|^2 + \alpha\left\|\mathbf{L}_{\omega}\begin{bmatrix}\mathbf{x}\\ \mathbf{b}\end{bmatrix}\right\|^2,$$

with

$$\mathbf{L}_\omega = \begin{bmatrix} \sqrt{1-\omega}\mathbf{L}_{\mathtt{x}} & \mathbf{0} \\ \mathbf{0} & \sqrt{\omega}\mathbf{L}_{\mathtt{b}} \end{bmatrix},$$

and

$$\frac{1}{\sigma_{\mathtt{x}}^2} = \frac{\alpha(1-\omega)}{\sigma^2}, \ \frac{1}{\sigma_{\mathtt{b}}^2} = \frac{\alpha\omega}{\sigma^2}.$$

The optimal values of the regularization parameters $\alpha_{\mathtt{opt}}$ and $\omega_{\mathtt{opt}}$ can be obtained by using an a priori parameter choice method. The a priori parameter choice method is a two-dimensional optimization problem and consists in the minimization of the expected value of the error $\left\|\mathbf{x}_{\alpha\omega}^{\delta} - \mathbf{x}^{\dagger}\right\|^2$ with respect to α and ω for a set of true solutions $\mathbf{x}^{\dagger}$, i.e., $(\alpha_{\mathtt{opt}}, \omega_{\mathtt{opt}}) = \arg\min_{\alpha,\omega} \mathcal{E}\left\{\left\|\mathbf{x}_{\alpha\omega}^{\delta} - \mathbf{x}^{\dagger}\right\|^2\right\}$. Solving the minimization problem for different noise variances yields the a priori selection criterion $\alpha_{\mathtt{opt}} = \alpha_{\mathtt{opt}}(\sigma^2)$ and $\omega_{\mathtt{opt}} = \omega_{\mathtt{opt}}(\sigma^2)$.

After the optimal values $\alpha_{\mathtt{opt}}$ and $\omega_{\mathtt{opt}}$ have been determined, the dimension of the minimization problem (2) can be reduced as follows. Assuming that $\mathbf{b}$ is stochastic with zero mean and covariance matrix $\mathbf{C}_{\mathtt{b}} = \sigma_{\mathtt{b}}^2\widehat{\mathbf{C}}_{\mathtt{b}} = \left(\sigma^2/\alpha_{\mathtt{opt}}\omega_{\mathtt{opt}}\right)\widehat{\mathbf{C}}_{\mathtt{b}}$, we define the total error as $\delta_{\mathtt{t}} = \mathbf{K}_{\mathtt{b}}\mathbf{b} + \delta$, and we have $\mathcal{E}\{\delta_{\mathtt{t}}\} = 0$ and $\mathbf{C}_{\delta_{\mathtt{t}}} = \mathcal{E}\left\{\delta_{\mathtt{t}}\delta_{\mathtt{t}}^T\right\} = \mathbf{K}_{\mathtt{b}}\mathbf{C}_{\mathtt{b}}\mathbf{K}_{\mathtt{b}}^T + \mathbf{C}_{\delta}$. Using the prewhitening technique with $\mathbf{C}_{\delta_{\mathtt{t}}}$ for the equation $\mathbf{y}^{\delta} = \mathbf{K}_{\mathtt{x}}\mathbf{x} + \delta_{\mathtt{t}}$, we transform the problem in the standard form $\mathbf{y}^{\delta} = \mathbf{K}_{\mathtt{x}}\mathbf{x} + \delta_{\mathtt{t}}$, where now $\mathcal{E}\{\delta_{\mathtt{t}}\} = 0$ and $\mathbf{C}_{\delta_{\mathtt{t}}} = \sigma_{\mathtt{t}}^2\mathbf{I}_m$. The MAP estimator is then the minimizer of the one-dimensional functional

$$-2\sigma_{\mathtt{t}}^2\mathcal{L}(\mathbf{x}) = \left\|\mathbf{y}^{\delta} - \mathbf{K}_{\mathtt{x}}\mathbf{x}\right\|^2 + \frac{\sigma_{\mathtt{t}}^2}{\sigma_{\mathtt{x}}^2}\left\|\mathbf{L}_{\mathtt{x}}\mathbf{x}\right\|^2,$$

with $1/\sigma_{\mathtt{x}}^2 = \alpha_{\mathtt{opt}}\left(1-\omega_{\mathtt{opt}}\right)/\sigma^2$.

3. NUMERICAL SIMULATION

In our numerical simulations we consider the retrieval of O_3 profile from SCIAMACHY limb scatter measurements. The forward model for SCIAMACHY limb radiance simulation is a single–scattering model, while the multiple scattering effect is taken into account by using look-up table corrections. As uncertain parameter we consider the tangent height variation. The regularization matrix for ozone is the Choleski factor of an Markov a priori constraint, while the regularization matrix for the tangent height variation is a diagonal matrix. The atmospheric profile is retrieved from simulated data in the spectral domain 520-590 nm. The a priori and initial gas profile were assumed to be identical and were choosen from the U.S. standard atmosphere. The true profile was obtained by scaling and shifting the a priori profile. The scale and shifting factors were 1.2 and 2 km, respectively. An altitude retrieval grid with 24 grid points between 13 and 100 km is considered. The number of limb scans is 15 and the corresponding tangent altitudes vary between 13 and 60 km. The entire sequence of limb scans has been shifted with 1 km. The noisy data vector (or the contaminated spectrum) is generated by adding a white noise to the exact data vector. The variance of the white noise was chosen such that measurement error vector δ does not dominate the term $\mathbf{K}_{\mathtt{b}}\mathbf{b}$. For the considered tangent height variation, four noise variances were assumed: 0.005, 0.001, 0.002 and 0.003.

The minimizer of $\mathcal{E}\left\{\left\|\mathbf{x}_{\alpha\omega}^{\delta} - \mathbf{x}^{\dagger}\right\|^2\right\}$ has been computed with the BFGS routine with line search. For $\alpha_{\mathtt{opt}}$ we found an optimal domain of variation of $\left[0.95\sigma^2, \sigma^2\right]$, while the optimal domain of variation for $\omega_{\mathtt{opt}}$ is $\left[10^{-5}, 10^{-3}\right]$. For $\alpha_{\mathtt{opt}} = \sigma^2$, we plot in Figure 1 the relative errors $\left\|\mathbf{x}_{\alpha\omega}^{\delta} - \mathbf{x}^{\dagger}\right\|^2$ as functions of ω. The plots show that the optimal value of the weighting factor $\omega_{\mathtt{opt}}$ does not depend significantly on the noise variance σ^2.

In Figures 2 and 3 we plot the retrieved O_3 profiles for synthetic and real data. The results correspond to the a priori choice $\alpha_{\mathtt{opt}} = \sigma^2$ and $\omega_{\mathtt{opt}} = 10^{-3}$. It is apparent that the retrieval which neglects the uncertainties due to the tangent height variations leads to relative errors of about 30%.

4. CONCLUSIONS

In this paper we present an a priori selection criterion for computing the optimal values of the regularization parameter and the weighting factor when the solution of the radiative transfer equation is affected by uncertain model parameters. The technique is based on the minimization of the expected value of the error for a set of true profiles. The model parameter uncertainties are then transformed into the measurement domain and included in the measurement covariance matrix. The method is applied for the uncertainties due to the tangent height variation in the case of the SCIAMACHY limb retrieval processor and yields a substantial improvement of the solution.

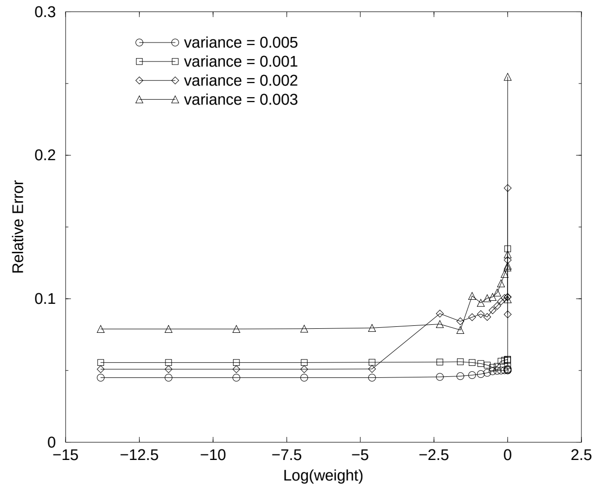

Figure 1: Relative error as function of the weighting factor ω.

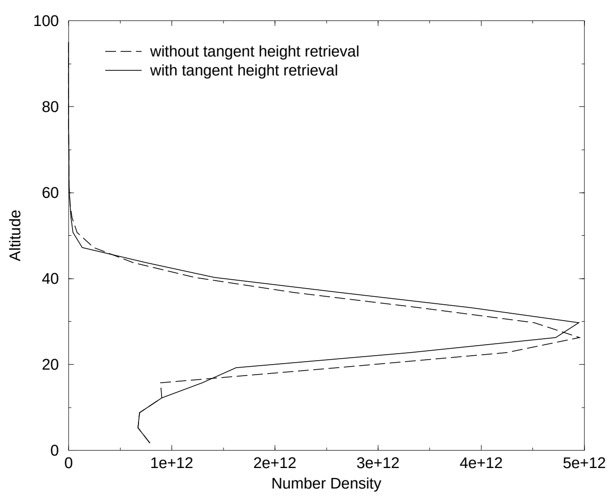

Figure 2: Retrieved ozone profiles for synthetic data.

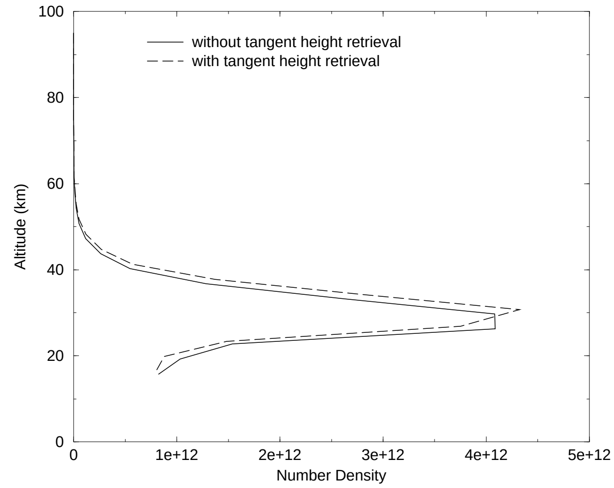

Figure 3: Retrieved ozone profiles for real data.

References

1. Rodgers C.D., Retrieval of atmospheric temperature and composition from remote measurementa of thermal radiation, Rev. Geophy. Space Phys. 14, 609-624, 1976.

2. von Clarmann T., Grabowski U. and Kiefer M., On the role of non-random errors in inverse problems in radiative transfer and other applications, J. Quant. Spectrosc. Radiat. Transfer 71, 39-46, 2001.