Uncertainty Quantification for the SPEXone/PACE Aerosol Product

Otto Hasekamp and Guangliang Fu

July 12, 2023

Abstract

This technote describes the inversion procedure, including uncertainty quantification, of the Remote sensing of Trace gas and Aerosol Products (RemoTAP) algorithm, to be used for operation aerosol Level-2 processing for SPEXone on PACE. RemoTAP uses an analytical inversion approach, allowing also an analytical error propagation. The error covariance matrix of the retrieved state vector contains contributions from the regularization error covariance matrix and the retrieval error covariance matrix. Evaluation with synthetic observations shows shows the RemoTAP uncertainty quantifcation represents the distribution true errors reasonably well.

1 RemoTAP Inversion procedure and Error Covariance Matrix

Any retrieval algorithm aims at inferring an atmospheric state vector $\mathbf{x}$ from a measurement vector $\mathbf{y}$. The state vector is linked to the measurement vector through a forward model $\mathbf{F}(\mathbf{x}, \mathbf{b})$ that depends on the state vector $\mathbf{x}$ and the vector $\mathbf{b}$ containing ancillary parameters that are not retrieved,

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \mathbf{e}_y \tag{1}$$

where $\mathbf{e}_y$ represents the measurement error vector. In our case the measurement vector consists of multi-spectral, multi-angle measurements of intensity and state of polarization performed by the SPEXone Multi-Angle-Polarimeter (MAP). In the following, we will omit the dependence of $\mathbf{F}$ on $\mathbf{b}$.

For the retrieval procedure it is needed that the non-linear forward model is linearized so that the retrieval problem can be solved iteratively. For iteration step n the forward model is approximated by

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_\mathrm{n}) + \mathbf{K}\,[\mathbf{x} - \mathbf{x}_\mathrm{n}] \tag{2}$$

where $\mathbf{x}_\mathrm{n}$ is the state vector for the current iteration step, and $\mathbf{K}$ is the Jacobian matrix with elements

$$K_{ij} = \frac{\partial F_i}{\partial x_j}(\mathbf{x}_\mathrm{n}). \tag{3}$$

In the inversion procedure, we invert the linearized forward model of Eq. (2) for iteration step n to find the state vector $\mathbf{x}_{\mathrm{n+1}}$ for iteration step $n+1$. Hereto, we minimize the following cost function (Tikhonov, 1963):

$$\mathbf{x}_{\mathrm{n+1}} = \min_{\mathbf{x}} \left([\mathbf{K}\,\mathbf{x} - \mathbf{y}]^T\ \mathbf{S}_y^{-1}\ [\mathbf{K}\,\mathbf{x} - \mathbf{y}] \right) + \left([\mathbf{x} - \mathbf{x}_a]^T\ \gamma^2 \mathbf{H}^{-1}\ [\mathbf{x} - \mathbf{x}_a] \right), \quad (4)$$

which we transform to

$$\tilde{\mathbf{x}}_{\mathrm{n+1}} = \min_{\tilde{\mathbf{x}}} \left([\tilde{\mathbf{K}}\,\tilde{\mathbf{x}} - \tilde{\mathbf{y}}]^T\ [\tilde{\mathbf{K}}\,\tilde{\mathbf{x}} - \tilde{\mathbf{y}}] \right) + \gamma^2\ \left([\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_a]^T\ [\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_a] \right), \quad (5)$$

where $\tilde{\mathbf{K}} = \mathbf{S}_y^{-\frac{1}{2}} \mathbf{K} \mathbf{H}^{\frac{1}{2}}$, $\tilde{\mathbf{x}} = \mathbf{H}^{-\frac{1}{2}} \mathbf{x}$ and $\tilde{\mathbf{y}} = \mathbf{S}_y^{-\frac{1}{2}} (\mathbf{y} - \mathbf{F}(\mathbf{x_n}))$. $\mathbf{x}_{\mathrm{a}}$ is the a priori state vector, $\mathbf{S}_y$ is the measurement error covariance matrix, γ is a regularization parameters, and $\mathbf{H}$ is a regularization matrix that ensures that all state vector parameters range within the same order of magnitude and determine the relative weight of parameters in the side constraint (Hasekamp, Litvinov, & Butz, 2011). Note that if $\gamma = 1$ and $\mathbf{H}$ is the prior error covariance matrix, Eq. (4) (and hence 5) reduce to the cost function of the optimal estimation method (Rodgers, 2000). We use a diagnonal matrix for $\mathbf{H}$ with diagonal elements $h_{ii} = w_i^2$.

The solution of Eq. (4) is given by:

$$\tilde{\mathbf{x}}_{n+1} = \tilde{\mathbf{x}}_n + \Lambda(\tilde{\mathbf{K}}^T \tilde{\mathbf{K}} + \gamma^2 \mathbf{I})^{-1} (\tilde{\mathbf{K}}^T \tilde{\mathbf{y}} - \gamma^2 (\tilde{\mathbf{x}}_n - \tilde{\mathbf{x}}_a)). \quad (6)$$

Λ is a filter/damping factor between 0 and 1, which limits the step size for each iteration of the state vector. In this way, we use a Gauss-Newton scheme with reduced step size to avoid diverging retrievals.

The error covariance matrix $\mathbf{S}_x$ of the retrieved state vector is given by

$$\mathbf{S}_x = \mathbf{S}_r + \mathbf{S}_e, \quad (7)$$

where $\mathbf{S}_r$ is the regularization error covariance matrix which describes the effect of the a priori error covariance matrix $\mathbf{S}_a$ on $\mathbf{x}$,

$$\mathbf{S}_r = (\mathbf{I} - \mathbf{A})\ \mathbf{S_a}\ (\mathbf{I} - \mathbf{A})^{\mathbf{T}}, \quad (8)$$

and $\mathbf{S}_e$ is the retrieval error covariance matrix that describes the effect of measurement- and forward model errors on $\mathbf{x}$,

$$\mathbf{S}_e = \mathbf{D}\ \mathbf{S}_{\mathrm{y,true}}\ \mathbf{D}^T, \quad (9)$$

where $\mathbf{D}$ is the contribution- or gain matrix

$$\mathbf{D} = \left(\mathbf{K}^T\ \mathbf{S}_y^{-1}\ \mathbf{K} + \gamma^2 \mathbf{H}^{-1} \right)^{-1}\ \mathbf{K}^T\ \mathbf{S}_y^{-1}, \quad (10)$$

and $\mathbf{A}$ is the averaging kernel

$$\mathbf{A} = \mathbf{D}\ \mathbf{K} \quad (11)$$

Ideally, $\mathbf{S}_{\mathrm{y,true}}$ and $\mathbf{S}_y$ are the same matrix. There may be reasons to choose $\mathbf{S}_y$ different in the inversion equations to give some spectral bands and/or angles more weights than others.

The standard deviation σ_τ on the Aerosol Optical Depth (AOD) can be obtained from the retrieval error covariance matrix $\mathbf{S}_x$ via

$$\sigma_\tau = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} S_{i,j}\,\frac{\partial\tau}{\partial x_i}\,\frac{\partial\tau}{\partial x_j}} \tag{12}$$

where $S_{i,j}$ denotes element (i,j) of $\mathbf{S}_x$. A similar expression holds for the other optical properties.

2 Performance Analysis and Verification of Uncertainties

We investigated the performance of MAP retrievals based on synthetic measurements created for the SPEXone instrument (Hasekamp et al., 2019). Here, we use a simplified error model that assumes the total error on reflectance and Degree of Linear Polarization (DoLP) can be described as a random error with a magnitude that corresponds to the instrument accuracy requirements: 2% on reflectance I/F_0 and 0.003 on DoLP. We simulate 4 days of synthetic measurements, each day consisting of 14 orbits of which the location (latitude, longitude) and geometry (solar zenith angle, viewing zenith angles, relative azimuth angles) are obtained from the SPEXone orbit simulator. We use MODIS cloud data of the year 2006 to simulate a realistic cloud mask and only retrievals are being performed for cloud free pixels. The atmospheric and surface properties needed to simulate the measurements are described below.

2.1 Ensemble Description

2.1.1 Aerosol Properties

We take the microphysical aerosol properties of our synthetic ensemble from the ECHAM-HAM model (Stier et al., 2005). ECHAM-HAM provides mass-mixing ratio in different vertical layers of the atmosphere of different aerosol species (Sulfate, Organic Carbon, Black Carbon, Dust, Sea Salt) in seven different size modes: Nucleation Soluble (NS), Aitken Soluble (KS), Accumulation Soluble (AS), Coarse Soluble (CS), Aitken Insoluble (KI), Accumulation Insoluble (AI), Coarse Insoluble (CI). Using the airmass in each model layer, we compute total mass in each layer per species per mode, which is translated into total volume in each layer per species per mode using the specific density per species. Also, for each layer the sub-column number of aerosol particles and the volume of aerosol water is provided. We sum up the different layers to obtain per mode the total column volume per species (including water), as well as the column number per mode. From the total volume V (all species together) and column number per mode N, we compute the mode radius under the assumption of a log-normal mode description:

$$r_g = \left(\frac{3V}{4\pi N} \, e^{-\frac{9 \ln^2 \sigma_g}{2}} \right)^{\frac{1}{3}}, \tag{13}$$

with $\sigma_g = 1.59$ for modes NS, KS, AS, KI, AI, and $\sigma_g = 2.0$ for modes CS and CI. The refractive index for each mode is obtained using a volume weighted mean of the refractive index of each species. We take the altitude of the layer with maximum sub-column number as the aerosol Layer Height for that mode and further assume a Gaussian altitude distribution. The Gaussian altitude distribution is chosen because it is needed as input of out forward model. ECHAM-HAM does not provide information on the shape of particles. Here, we assume that Dust is purely non-spherical and hence take the fraction of non-spherical particles as the volume fraction of Dust.

The AOD of the ECHAM-HAM ensemble is severely underestimated compared to POLDER-3 satellite retrievals. Therefore, we scaled the column number of all fine and coarse modes such that they agree with the fine- and coarse mode AOD as retrieved from POLDER-3 (Lacagnina, Hasekamp, & Torres, 2017).

The aerosol description in the synthetic ensemble differs substantially from the aerosol description used in the retrieval (which is based on 3 size modes: a fine mode, a coarse insoluble mode (dust), and a coarse insoluble mode (hydrated sea salt)). Namely, the number of modes as well as the effective radius/variance differs from what is assumed in the retrieval and in the synthetic ensemble each mode hase it's own refractive index. Further, in the synthetic ensemble, each mode has its own vertical profile whereas in the retrieval only distinction is made between vertical profiles in the fine and coarse mode, respectively. This makes the ensemble suitable to test performance and robustness against assumptions in the state vector definition.

2.1.2 Surface Properties

We use the directional parameters ($k_{\text{geo}}, k_{\text{vol}}$ of the Ross-Li model from MODIS and the spectral dependent scaling parameter A from GOME-2. Further, the surface polarization parameter is taken from POLDER-3 retrievals. For synthetic measurements over ocean, we use the chlorophyll-a concentration x_{chl} from MODIS and the wind-speed from NCEP meteorological data.

2.2 Results

Figure 1 shows the performance of the MAP-only retrieval algorithm for AOD, SSA, r_{eff}^f, and m_r. for retrievals over ocean and Fig. 2 for retrievals over land. Here, the retrieved r_{eff}^f are compared against the volume-weighted average of modes NS, KS, AS, KI, AI from ECHAM-HAM and for and m_r the volume-weighted retrieved values for the 3 modes are compared against the volume weigthed value of all modes from ECHAM-HAM.

We show both scatter-plots of retrieved versus true values as well as for the given state vector parameters the distrubution of $(x_{\mathrm{i,ret}} - x_{\mathrm{i,true}})/\sigma_i$, where σ_i is the standard deviation that follows for the state vector error covariance matrix of Eq. (7). If Eq. (7) would correctly describe the 'true' uncertainties in the retrieved aerosol parameters, then the distribution of $(x_{\mathrm{i,ret}} - x_{\mathrm{i,true}})/\sigma_i$ is given by a Gaussian with a mean value of 0 and a Full Width at Half Maximum (FWHM) of 2. Here, $x_{\mathrm{i,ret}}$ is the i-th element of the retrieved state vector, $x_{\mathrm{i,true}}$ is the corresponding true value. For retrievals over ocean, the Mean Absolute Error (MAE) is 0.007 for AOD, 0.011 for SSA, 0.016 for r_{eff}^f, and 0.017 for m_r. For retrievals over land, the MAE is 0.032 for AOD, 0.024 for SSA, 0.02μm for r_{eff}^f, and 0.028 for m_r. From the histogram plots, we conclude that Eq. (7) provides a reasonable estimate of the retrieval uncertainty, although for r_{eff} and m_r the uncertainty seems somewhat underestimated. This might also be caused by the fact that these properties are not defined in exactly the same manner for SPEXone retrievals and ECHAM-HAM. Further, the estimated uncertainties per definition do not describe a retrieval bias.

We propose to use the same method for validation of uncertainties for real retrievals, where also uncertainties in the reference measurements (e.g. AERONET) are taken into account.

References

Hasekamp, O. P., Fu, G., Rusli, S. P., Wu, L., Noia, A. D., aan de Brugh, J., … van Amerongen, A. (2019). Aerosol measurements by spexone on the nasa pace mission: expected retrieval capabilities. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *227*, 170 - 184. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0022407318308653` doi: https://doi.org/10.1016/j.jqsrt.2019.02.006

Hasekamp, O. P., Litvinov, P., & Butz, A. (2011, July 27). Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements. *J. Geophys. Res.*, *116*(D14), D14204+. doi: 10.1029/2010jd015469

Lacagnina, C., Hasekamp, O. P., & Torres, O. (2017, February). Direct radiative effect of aerosols based on PARASOL and OMI satellite observations. *Journal of Geophysical Research (Atmospheres)*, *122*, 2366–2388. doi: 10.1002/2016jd025706

Rodgers, C. (2000). *Inverse methods for atmospheric sounding: Theory and practice*. World Sc.,River Edge, N. J.

Stier, P., Feichter, J., Kinne, S., Kloster, S., Vignati, E., Wilson, J., … Petzold, A. (2005, March). The aerosol-climate model ECHAM5-HAM. *Atmospheric Chemistry & Physics*, *5*, 1125-1156.

Tikhonov, A. (1963). On the solution of incorrectly stated problems and a method of regularization. *Dokl. Akad. Nauk SSSR*, *151*, 501-504.

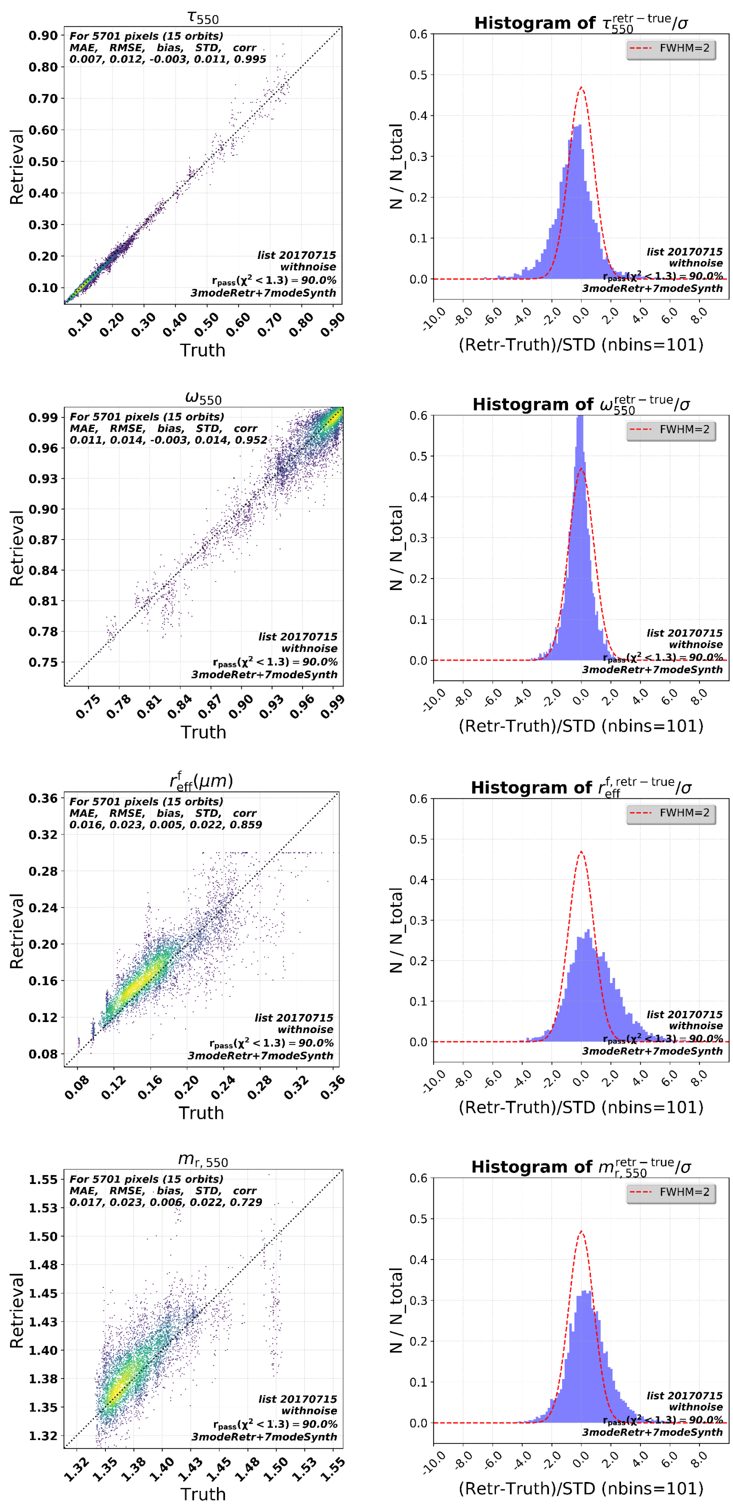

Figure 1: Scatterplot of retrieved versus truth (left plots) and histograms of differences, normalized by estimated uncertainty (right plots) for the following aerosol properties (from top to bottom): AOD, SSA, fine-mode-effective-radius, fine-mode-refractive index. The dotted line is a Gaussian function with mean=0 and FWHM=2. Retrievals over ocean.

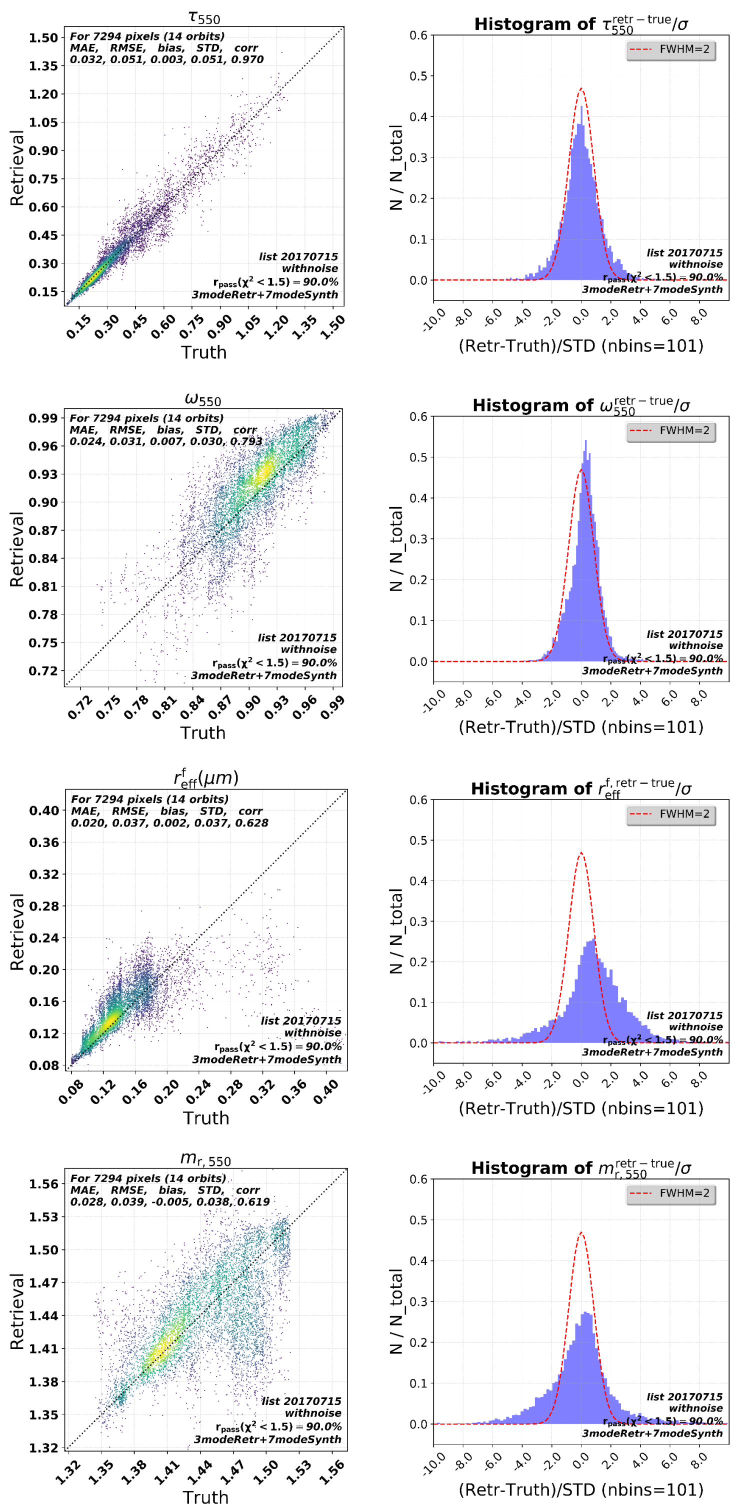

Figure 2: Scatterplot of retrieved versus truth (left plots) and histograms of differences, normalized by estimated uncertainty (right plots) for the following aerosol properties (from top to bottom): AOD, SSA, fine-mode-effective-radius, fine-mode-refractive index. Retrievals over ocean.