Aerosol retrieval over snow using RemoTAP

Zihan Zhang¹², Guangliang Fu¹, Otto Hasekamp¹

¹SRON Netherlands Institute for Space Research, Leiden, 2333CA, the Netherlands
²School of Earth and Space Science, Peking University, Beijing, 100871, China

Correspondence to: Otto Hasekamp (o.p.hasekamp@sron.nl)

Abstract. In order to conduct accurate aerosol retrieval over snow, the Remote Sensing of Trace Gases and Aerosol Products (RemoTAP) algorithm developed by SRON Netherlands Institute for Space Research is extended with a Bi-directional Reflection Distribution Function (BRDF) for snow surfaces. The capability of the extended algorithm is validated with both synthetic measurements and real satellite measurements from PARASOL and a comparison has been made to retrievals with the baseline RemoTAP (without snow kernel). For retrievals on real PARASOL observations, we use pixels over AERONET 10 stations for validation and we use the MODIS snow cover products to identify pixels over snow. We evaluate the retrieved aerosol optical thickness (AOT) at 550 nm ($\tau_{550}$), single-scattering albedo (SSA) at 550 nm ($\omega_{550}$), and Angstrom exponent (AE) for 440 nm – 870 nm ($AE_{440−870}$). Both the experiments with synthetic- and real data show that the extended RemoTAP maintains capability on snow-free pixels and has obvious advantages on accuracy and fraction of successful retrievals for retrieval over snow, especially over surfaces with snow cover > 75%. According to the real data experiment, we find that the retrieval algorithm has difficulty in fitting the PARASOL 1020 nm band, where snow reflectance is significantly lower than that for the visible bands. When we perform a 4-band retrieval (490 nm, 565 nm, 670 nm, 865 nm) with the extended RemoTAP, we obtain a good retrieval result for $\tau_{550}$, $\omega_{550}$ and $AE_{440−870}$. Therefore, the 4-band retrieval with the extended RemoTAP is recommended for aerosol retrieval over snow.

1 Introduction

Global climate change is greatly influenced by aerosol through aerosol cloud interaction (e.g., Twomey, 1974; Li et al., 2011; Rosenfeld et al., 2014; Hasekamp et al., 2019a; Gryspeerdt et al., 2020; Quaas et al., 2020) and aerosol radiation interaction (e.g., Koren et al., 2004; Yu et al., 2006; Myhre, 2009; Guo et al., 2017; Lacagnina et al., 2017; Witthuhn et al., 2021). These effects cause a significant radiative forcing of climate (Bellouin et al., 2020; Haywood et al., 2021). According to the sixth Assessment Report (AR6) of Intergovernmental Panel on Climate Change (IPCC), aerosols still represent the largest uncertainty in our quantification of global climate change (IPCC, 2023). Therefore, in order to better study climate change, it is essential to give an accurate estimate of global aerosol properties. Multi-Angle Polarimetric (MAP) satellite measurements provide the richest set of information on aerosol properties from a passive remote sensing point of view. So far, the only MAP instrument that has provided a multi-year dataset has been POLDER-3/PARASOL (here-after referred to as PARASOL), which was active from 2005-2013. In the near future a number of new MAP instruments will be launched, including SPEXone
(Hasekamp et al., 2019b) and HARP-2 (Mcbride et al., 2020) on the NASA PACE mission (Werdell et al., 2019) and the 3MI instrument (Fougnie et al., 2018) on METOP-SG.

After years of development, there exist a number of aerosol retrieval algorithms which are available for multi-angle and multi-spectral polarization sensors. The algorithms can be classified into two groups: algorithms based on lookup-tables (LUTs) and full inversion algorithms. LUT-based algorithms (Herman et al., 1997; Deuzé et al., 2000; Waquet et al., 2016) provide faster calculation, but are less accurate than full inversion algorithms. Among all the full inversion algorithms are the Microphysical Aerosol Properties from Polarimetry (MAPP) algorithm (Stamnes et al., 2018), the algorithm developed at the Jet Propulsion Laboratory (JPL) (Xu et al., 2016; Xu et al., 2017; Xu et al., 2018), the Generalized Retrieval of Aerosol and Surface Properties (GRASP) algorithm (Dubovik et al., 2011; Dubovik et al., 2014; Chen et al., 2020; Dubovik et al., 2021) and the Remote-sensing of Trace-gas and Aerosol Product (RemoTAP) algorithm developed by SRON (Hasekamp et al., 2011; Fu and Hasekamp, 2018; Fu et al., 2020; Lu et al., 2022). Till now, only the RemoTAP algorithm and the GRASP algorithm have demonstrated capability at the global scale.

In order to generate global aerosol products, aerosol retrieval over snow remains an important challenge for the above-mentioned algorithms. Until now, studies on aerosol retrieval over snow have focused mainly on intensity sensors (Istomina et al., 2011; Mei et al., 2013), focusing on AOT retrievals in the arctic region.

In our paper, we extend the RemoTAP algorithm to carry out aerosol retrieval over snow from MAP measurements of PARASOL. We evaluate the capability of the extended RemoTAP algorithm using synthetic observations as well as real PARASOL retrievals which are validated by AERONET. The paper is organized as follows: Section 2 introduces the methodology of RemoTAP for aerosol retrieval over snow; Section 3 describes the satellite data and ancillary data used in our real data retrieval and the validation data; Section 4 shows results for synthetic retrievals; Section 5 shows the results for real PARASOL retrievals and provides a recommended routine for aerosol retrieval over snow from PARASOL; Section 6 discusses the results of the paper and future research, and concludes the paper.

2 Methodology

2.1 Forward model

For aerosol retrieval over snow, the detailed information for the extended RemoTAP algorithm is described below.

The aim of the RemoTAP algorithm is to retrieve a state vector \( \mathbf{x} \) from measurement vector \( \mathbf{y} \) by inverting the forward model \( \mathbf{F} \):

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

where \( \mathbf{e}_y \) refers to the measurement error vector.

The measurement by polarization sensors is described by the intensity vector \([I, Q, U, V]\) (Stokes, 1851). For Earth observation, \(V\) parameter can be ignored in most cases. In this study, the Stokes parameters are calculated by the SRON
radiative transfer model LINTRAN v2.0 (Schepers et al., 2014). The measurement vector $\mathbf{y}$ contains parameters including the Top-of-Atmosphere (ToA) apparent reflectance $\rho$ and Degree of Linear Polarization (DoLP) at different spectral bands and different observation geometries:

$$\rho = \frac{i}{E_0},$$  \hspace{1cm} (2)

$$\text{DoLP} = \frac{\sqrt{Q^2 + U^2}}{I},$$  \hspace{1cm} (3)

where, $E_0$ is the ToA solar irradiance perpendicular to the solar beam.

The state vector $\mathbf{x}$ contains parameters related to aerosol and surface characteristics. To describe aerosol properties in the state vector, following Lu et al., (2022), 3 log-normal modes are applied: (1) Mode 1 is a fine mode for which the state vector includes aerosol column number ($N_{aer}$), effective radius ($r_{eff}$), effective variance ($v_{eff}$), spherical fraction ($f_{sph}$), aerosol layer height ($z_{aer}$) and the refractive index coefficients corresponding to the standard spectra (D'almeida et al., 1991; Kirchstetter et al., 2004) of Inorganic aerosol, Black Carbon (imaginary part) and Organic Carbon (imaginary part). (2) Mode 2 is an insoluble coarse mode, which consists of non-spherical dust. The state vector includes $N_{aer}$, $r_{eff}$, $v_{eff}$, $z_{aer}$ and refractive index coefficients corresponding to the standard spectra of Dust (imaginary part). The fixed parameter is $f_{sph} = 0$ and the coefficient for the real part of Dust refractive index is fixed to 1. $z_{aer}$ is assumed to be the same for mode 1 and mode 2. (3) Mode 3 is a soluble coarse mode. The state vector includes $N_{aer}$, $r_{eff}$, $v_{eff}$ and the refractive index coefficient corresponds to the standard spectra of Inorganic aerosol. The fixed parameters are $f_{sph} = 1$ and $z_{aer} = 500$ m.

To describe the surface, the surface reflection matrix $\mathbf{R}_{surf}(\lambda, \theta_s, \theta_v, \varphi)$ is given by:

$$\mathbf{R}_{surf}(\lambda, \theta_s, \theta_v, \varphi) = r_{11} (\lambda, \theta_s, \theta_v, \varphi) \mathbf{D} + \mathbf{R}_{pol}(\theta_s, \theta_v, \varphi),$$  \hspace{1cm} (4)

where $\lambda$ is the wavelength, $\theta_s$ and $\theta_v$ are the sun zenith angle and view zenith angle respectively, $\varphi$ is the relative azimuth angle, $\mathbf{D}$ is a null matrix except $\mathbf{D}_{11} = 1$, $r_{11}(\lambda, \theta_s, \theta_v, \varphi)$ is described by the Ross-Li BRDF model, extended by a snow kernel:

$$r_{11}(\lambda, \theta_s, \theta_v, \varphi) = A(\lambda) \left[ 1 + k_{geo} f_{geo}(\theta_s, \theta_v, \varphi) + k_{vol} f_{vol}(\theta_s, \theta_v, \varphi) + k_{snow} f_{snow}(\theta_s, \theta_v, \varphi) \right],$$  \hspace{1cm} (5)

where $A(\lambda)$ is the isotropic reflectance, $f_{geo}(\theta_s, \theta_v, \varphi)$ and $f_{vol}(\theta_s, \theta_v, \varphi)$ are respectively the geometric (Li-Sparse) kernel and volumetric (Ross-Thick) kernel function (Wanner et al., 1995), $f_{snow}(\theta_s, \theta_v, \varphi)$ is the snow kernel function (Jiao et al., 2019), $k_{geo}, k_{vol}, k_{snow}$ are the coefficients for the Li-Sparse, Ross-Thick and snow kernel respectively. It is important to note that the inclusion of $k_{snow} f_{snow}(\theta_s, \theta_v, \varphi)$ is the major difference between the baseline RemoTAP and the extended RemoTAP of the present work. The Li-Sparse kernel function $f_{geo}(\theta_s, \theta_v, \varphi)$, Ross-Thick kernel function $f_{vol}(\theta_s, \theta_v, \varphi)$ and snow kernel function $f_{snow}(\theta_s, \theta_v, \varphi)$ are given in Eq. (6) and Eq. (7):

$$f_{geo}(\theta_s, \theta_v, \varphi) = O(\theta_s', \theta_v', \varphi') - \left[ \sec \theta_s' + \sec \theta_v' - \frac{1}{2} (1 - \cos \theta') \sec \theta_s' \sec \theta_v' \right],$$  \hspace{1cm} (6)

$$f_{vol}(\theta_s, \theta_v, \varphi) = \frac{(\theta - \pi/2) \cos(\pi - \theta) + \sin(\pi - \theta)}{\cos \theta_s + \cos \theta_v} \frac{\pi}{4},$$  \hspace{1cm} (7)
where $\Theta$ is the scattering angle ($\cos \Theta = -\cos \theta_s \cos \theta_v - \sin \theta_s \sin \theta_v \cos \phi$), $\theta_s', \theta_v'$ are the equivalent zenith angles (the transformation is $\theta' = \tan^{-1}(b/r \cdot \tan \theta)$ for sun zenith angle and view zenith angle respectively, $b$ and $r$ are vertical and horizontal crown radius respectively), $O(\theta_s', \theta_v', \phi)$ is the overlap function given by Li and Strahler (1992).

The snow kernel function $f_{snow}(\theta_s, \theta_v, \phi)$ is given by Jiao et al. (2019) in Eq. (8):

$$f_{snow}(\theta_s, \theta_v, \phi) = R_0(\theta_s, \theta_v, \phi)[1 - \alpha \cdot \cos(\pi - \Theta) \cdot e^{-\cos(\pi - \Theta)}] + 0.4076 \cdot \alpha - 1.1081,$$

where $R_0(\theta_s, \theta_v, \phi)$ is the reflectance for a semi-infinite, non-absorbing snow layer at zero absorption and $\alpha$ is an empirical parameter used to correct $R_0(\theta_s, \theta_v, \phi)$ for the underestimation of the reflectance in the forward-scattering direction. $R_0(\theta_s, \theta_v, \phi)$ is given by Kokhanovsky et al. (2005):  

$$R_0(\theta_s, \theta_v, \phi) = K_1 + K_2 \cdot (\cos \theta_s + \cos \theta_v) + K_3 \cdot \cos \theta_s \cos \theta_v + P(\Theta),$$

where $K_1$, $K_2$ and $K_3$ are three constants. In our algorithm, $\alpha = 0.3$, $K_1 = 1.247$, $K_2 = 1.186$ and $K_3 = 5.157$ are fixed when calculating snow kernel, as suggested by Jiao et al. (2019) and Kokhanovsky et al. (2005).

$R_{pol}(\theta_s, \theta_v, \phi)$ in Eq. (4) is given by:

$$R_{pol}(\theta_s, \theta_v, \phi) = B_{pol} \frac{\exp\left(-\tan\left(\frac{\pi}{2} - \theta\right)\right) \exp(-v) \cdot F_p(m, \Theta)}{4(\cos \theta_s + \cos \theta_v)},$$

where $B_{pol}$ is the free linear parameter, $F_p(m, \Theta)$ is the Fresnel scattering matrix and $m$ is the refractive index (Maignan et al., 2009). In our experiment, $m = 1.5$ and $v = 0.1$ are fixed when calculating $R_{pol}$.

To characterize the surface properties, we include $A(\lambda)$, $k_{geo}$, $k_{vol}$, $k_{snow}$ and $B_{pol}$ as the surface parameters in the state vector. Theoretically, the surface model still maintains the ability of depicting snow-free surfaces, because $k_{snow}$ is fitted in the retrieval (for snow-free surfaces, it should retrieve $k_{snow} = 0$).

The aerosol and surface parameters in the state vector $\mathbf{x}$ are shown in Table 1.
Table 1: Parameters in the state vector $\mathbf{x}$ utilized in the retrieval. $c_1$,… correspond to coefficients for standard refractive index of Inorganic aerosol (INORG), Black Carbon (BC), Organic Carbon (OC), Dust (DU).

<table>
<thead>
<tr>
<th>Property</th>
<th>Full name</th>
<th>A-priori</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine mode (mode 1)</td>
<td>$N_{\text{aer}}$</td>
<td>Aerosol column number (mode 1)</td>
</tr>
<tr>
<td></td>
<td>$r_{\text{eff}}$</td>
<td>Effective radius (mode 1)</td>
</tr>
<tr>
<td></td>
<td>$v_{\text{eff}}$</td>
<td>Effective variance (mode 1)</td>
</tr>
<tr>
<td></td>
<td>$f_{\text{sph}}$</td>
<td>Spherical fraction (mode 1)</td>
</tr>
<tr>
<td></td>
<td>$z_{\text{aer}}$</td>
<td>Aerosol layer height (mode 1)</td>
</tr>
<tr>
<td></td>
<td>$c_1$ (INORG real)</td>
<td>Refractive index coefficient (INORG real)</td>
</tr>
<tr>
<td></td>
<td>$c_2$ (BC imaginary)</td>
<td>Refractive index coefficient (BC imaginary)</td>
</tr>
<tr>
<td></td>
<td>$c_3$ (OC imaginary)</td>
<td>Refractive index coefficient (OC imaginary)</td>
</tr>
<tr>
<td>Coarse insoluble mode (mode 2)</td>
<td>$N_{\text{aer}}$</td>
<td>Aerosol column number (mode 2)</td>
</tr>
<tr>
<td></td>
<td>$r_{\text{eff}}$</td>
<td>Effective radius (mode 2)</td>
</tr>
<tr>
<td></td>
<td>$v_{\text{eff}}$</td>
<td>Effective variance (mode 2)</td>
</tr>
<tr>
<td></td>
<td>$z_{\text{aer}}$</td>
<td>Aerosol layer height (mode 2)</td>
</tr>
<tr>
<td></td>
<td>$c_1$ (DU imaginary)</td>
<td>Refractive index coefficient (DU imaginary)</td>
</tr>
<tr>
<td>Coarse soluble mode (mode 3)</td>
<td>$N_{\text{aer}}$</td>
<td>Aerosol column number (mode 3)</td>
</tr>
<tr>
<td></td>
<td>$r_{\text{eff}}$</td>
<td>Effective radius (mode 3)</td>
</tr>
<tr>
<td></td>
<td>$v_{\text{eff}}$</td>
<td>Effective variance (mode 3)</td>
</tr>
<tr>
<td></td>
<td>$c_1$ (INORG)</td>
<td>Refractive index coefficient (INORG)</td>
</tr>
</tbody>
</table>

2.2 Inversion algorithm

To retrieve the state vector $\mathbf{x}$ from the measurement $\mathbf{y}$, a damped Gauss-Newton iteration method with Phillips-Tikhonov regularization is employed (Hasekamp et al., 2011; Fu and Hasekamp, 2018). We shortly summarize the method here. The aim of the inversion algorithm is to find the solution $\hat{\mathbf{x}}$ of the minimization-optimization problem:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \left[ \left\| \mathbf{S}_y^{-1}(\mathbf{F}(\mathbf{x}) - \mathbf{y}) \right\|^2 + \gamma^2 \left\| \mathbf{W}^{-1}(\mathbf{z}(\mathbf{x}) - \mathbf{z}_a) \right\|^2 \right],$$  \hspace{1cm} (12)
where $\mathbf{x}_a$ is the a-priori state vector, $\mathbf{W}$ is the diagonal weight matrix in order to remove the order-of-magnitude difference of each state parameters ($W_{ii} = 1/x_{a,i}$), $\mathbf{S}_y$ is the diagonal measurement error covariance matrix which is related to the sensor, $\gamma$ is the regularization parameter (Hasekamp et al., 2011).

Since the forward model $\mathbf{F}$ is nonlinear, an iterative strategy is utilized to conduct the inversion. For each iteration $n$, the linear approximation of the forward model is given in Eq. (13):

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_n) + \mathbf{K}(\mathbf{x} - \mathbf{x}_n),$$

where $\mathbf{K}$ is the Jacobian matrix containing the derivatives of the forward model $\mathbf{F}$ with respect to the parameters in state vector $\mathbf{x}$, and $K_{ij} = \frac{\partial F_i}{\partial x_j}(\mathbf{x}_n)$.

With linear approximation, Eq. (12) can be simplified:

$$\tilde{\mathbf{x}}_{n+1} = \min_{\tilde{x}} \left[ \|\mathbf{R}(\tilde{x} - \tilde{\mathbf{x}}_n) - \tilde{\mathbf{y}}\|^2 + \gamma^2 \|\tilde{x} - \tilde{\mathbf{x}}_a\|^2 \right],$$

(14)

where $\tilde{\mathbf{x}} = \mathbf{W}^{-\frac{1}{2}}\mathbf{x}$, $\tilde{\mathbf{y}} = \mathbf{S}_y^{-\frac{1}{2}}(\mathbf{y} - \mathbf{F}(\mathbf{x}_n))$ and $\tilde{\mathbf{K}} = \mathbf{S}_y^{-\frac{1}{2}}\mathbf{KW}^{\frac{1}{2}}$.

The solution to Eq. (14) for iteration step $n$ is given by:

$$\tilde{x}_{n+1} = \tilde{x}_n + \Lambda(\tilde{\mathbf{R}}^T \tilde{\mathbf{K}} + \gamma^2 \mathbf{I})^{-1}\left[ \tilde{\mathbf{R}}^T \tilde{\mathbf{y}} - \gamma^2 (\tilde{\mathbf{x}}_n - \tilde{\mathbf{x}}_a) \right],$$

(15)

where $\mathbf{I}$ is the identity matrix, $\Lambda$ is the filter factor ($0 \leq \Lambda \leq 1$), which is utilized to control the step size per iteration. For each iteration, the optimal $\Lambda$ and $\gamma$ are chosen via goodness-of-fit assessment by comparing $\chi^2$ given in Eq. (16):

$$\chi^2 = \frac{1}{N_{\text{meas}}} \sum_{i=1}^{N_{\text{meas}}} \left( \frac{y_i - F_i}{e_i} \right)^2,$$

(16)

where $N_{\text{meas}}$ is the length of measurement vector $\mathbf{y}$, $y_i$ and $F_i$ refer to the measurements and the results of forward model respectively, $e_i$ is the measurement uncertainties determined by the instrument. After the final iteration step, we use $\chi^2 < 5$ as the threshold to determine whether the retrieval was successful or not. For a set of retrievals, the fraction of successful retrievals (FoSR) is determined based on this filter.

3 Data

3.1 PARASOL data

The microsatellite Polarization & Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar (PARASOL), equipped with POLarization and Directionality of the Earth’s Reflectances-3 instrument, was launched on 18th December 2004 (Fougnie et al., 2007; Lier and Bach, 2008). PARASOL includes 9 spectral bands and 3 bands contain polarization information (the details of the bands are shown in Table 2).
Table 2: Spectral band information for PARASOL

<table>
<thead>
<tr>
<th>PARASOL band (nm)</th>
<th>Central Wavelength (nm)</th>
<th>Band Width (nm)</th>
<th>Polarization</th>
</tr>
</thead>
<tbody>
<tr>
<td>443</td>
<td>443.9</td>
<td>13.5</td>
<td>×</td>
</tr>
<tr>
<td>490</td>
<td>491.5</td>
<td>16.5</td>
<td>✓</td>
</tr>
<tr>
<td>565</td>
<td>563.9</td>
<td>15.5</td>
<td>×</td>
</tr>
<tr>
<td>670</td>
<td>669.9</td>
<td>15.0</td>
<td>✓</td>
</tr>
<tr>
<td>763</td>
<td>762.8</td>
<td>11.0</td>
<td>×</td>
</tr>
<tr>
<td>765</td>
<td>762.5</td>
<td>38.0</td>
<td>×</td>
</tr>
<tr>
<td>865</td>
<td>863.4</td>
<td>33.5</td>
<td>✓</td>
</tr>
<tr>
<td>910</td>
<td>906.9</td>
<td>21.0</td>
<td>×</td>
</tr>
<tr>
<td>1020</td>
<td>1019.4</td>
<td>17.0</td>
<td>×</td>
</tr>
</tbody>
</table>

PARASOL provides multidirectional and multispectral data from December 2004 to December 2013, which has been used for aerosol retrieval (e.g., Tanré et al., 2011; Dubovik et al., 2011; Fu and Hasekamp, 2018; Chen et al., 2020), and the resolution is approximately 6 km in the nadir. In our experiments, we consider 5 bands (490 nm, 565 nm, 670 nm, 865 nm, 1020 nm), because 443 nm band suffers from stray light which may be in particular problematic over bright snow surfaces (Fougnie et al., 2007), 763 nm and 765 nm are mainly used to retrieve cloud oxygen pressure and 910 nm is usually used to retrieve the water vapor (Leroy et al., 1997).

3.2 Ancillary data

As input to the retrieval, ancillary data are needed including surface pressure, and profiles of air temperature, relative humidity after moist, ozone mass mixing ratio. These data are obtained from the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) (Gelaro et al., 2017). Additionally, the cloud fraction is from the MODerate-resolution Imaging Spectroradiometer (MODIS). In order to make a reliable cloud screening, only pixels with cloud fraction under 0.2 are used in our experiments.

3.3 Snow cover data

During the process of retrieval and validation, pixels with snow cover are selected based on the MODIS/Aqua Snow Cover Daily L3 Global 500m SIN Grid (MYD10A1) product (Hall et al., 2019). Each tile is generated from MODIS/Aqua Snow Cover 5-Min L2 Swath 500m (MYD10_L2) product. Given the row and column in a MODIS tile, the 500 m MODIS sinusoidal grid is converted to PARASOL sinusoidal grid. It is important to note that the MODIS snow cover product has been pre-processed with a cloud filtering procedure.
3.4 AERONET data

In this paper, the main concern is aerosol retrieval over snow. The RemoTAP-retrieved aerosol properties are validated with AERONET data (Holben et al., 1998). AERONET provides 3 levels of data quality: level 1.0, level 1.5 and level 2.0. Level 1.0 provides unscreened data which have been rarely used for validation. Level 1.5 provides near-real-time, cloud-screened data with instrument quality control. Level 2.0 provides quality-assured data on the basis of level 1.5 by applying pre-field and post-field calibrations. The retrieved AOT and AE are validated with AERONET direct sun level 2.0 AOT data (Giles et al., 2019). The retrieved SSA is validated with AERONET-Inversion level 2.0 SSA data (Sinyuk et al., 2020).

3.5 Data pre-processing

In the real data experiments, the PARASOL data are pre-processed with in steps. The first step is to match global PARASOL L1 data with global AERONET data. For each successfully-matched group of pixels (here-after referred to as a colocation), the difference between the measurement time of AERONET data and PARASOL pixels is within 1 hour and the distance between AERONET data and PARASOL pixels is within 20 km. The second step is to match MODIS snow cover data with AERONET-colocated PARASOL data, thus divide the colocated PARASOL data into snow pixels and snow-free pixels.

4 Synthetic data experiments

The forward model of RemoTAP is used to generate the synthetic PARASOL measurement and noise is subsequently added according to a Gaussian distribution. For ToA reflectance, the simulated noise (1 standard deviation) is 1%, and for DoLP it is 0.007 (absolute). The set of synthetic measurements contains 1000 pixels with randomly-generated input land properties, aerosol properties, auxiliary properties and geometry properties. The settings for the properties are shown in Table 3.

Table 3: Observation geometry, aerosol properties and surface properties used to create synthetic PARASOL observations.

<table>
<thead>
<tr>
<th>Property</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_s$</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>$\theta_v$</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>$k_{geo}$</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>$k_{vol}$</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>$k_{snow}$</td>
<td>-0.2</td>
<td>1.5</td>
</tr>
<tr>
<td>$A(\lambda)$</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>$\tau_{550}$ (mode 1)</td>
<td>0.005</td>
<td>1.0</td>
</tr>
<tr>
<td>$\tau_{550}$ (mode 2)</td>
<td>0.0025</td>
<td>0.25</td>
</tr>
<tr>
<td>$\tau_{550}$ (mode 3)</td>
<td>0.0025</td>
<td>0.25</td>
</tr>
</tbody>
</table>
The surface properties in the synthetic data set are created by mixing the contribution of the surface reflection by vegetation, soil, and snow. By controlling the fraction of vegetation, soil and snow, 4 sets of synthetic measurements are created, where the detailed information for these 4 synthetic measurements are listed in Table 4. The isotropic part of the BRDF for snow, soil, and vegetation, as used in this study, is shown in Fig. 1. For the directional BRDF parameters for soil and vegetation, we use the values found by Litvinov et al. (2011).
Table 4: Descriptions for different synthetic measurements

<table>
<thead>
<tr>
<th>Synthetic measurement</th>
<th>Description</th>
<th>Snow cover fraction ($c_{\text{snow}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>snow_free</td>
<td>Ground surfaces without snow, randomly mixed with vegetation and soil</td>
<td>$c_{\text{snow}} = 0%$</td>
</tr>
<tr>
<td>snow_pure</td>
<td>Ground surfaces completely covered by snow</td>
<td>$c_{\text{snow}} = 100%$</td>
</tr>
<tr>
<td>snow_domi</td>
<td>Ground surfaces randomly mixed with vegetation, soil and snow, but snow is dominant in landcover</td>
<td>$c_{\text{snow}} &gt; 75%$</td>
</tr>
<tr>
<td>snow_rand</td>
<td>Ground surfaces randomly mixed with vegetation, soil and snow, without limitation for snow cover fraction</td>
<td>$0% \leq c_{\text{snow}} \leq 100%$</td>
</tr>
</tbody>
</table>

These 4 sets of synthetic measurements are taken as the input for RemoTAP to conduct the retrieval with the baseline RemoTAP setup (without snow BRDF) and the extended RemoTAP setup (with snow BRDF). During validation, $\tau_{550}$, $\omega_{550}$, and $\text{AE}_{440-870}$ are chosen as the main performance indicators. In our experiment, a 5-band (490 nm, 565 nm, 670 nm, 865 nm, 1020 nm) retrieval is conducted and the validation results for different synthetic situations are shown in Figure 2.
Comparing the RMSE, bias and fraction of successful retrievals between the baseline- and extended RemoTAP in Figure 2, we can conclude that the baseline RemoTAP, which utilizes the Ross-Li model to characterize the ground surfaces, has poor capability to retrieve aerosol properties over snow. For synthetic measurements of pure snow (“snow_pure”) and dominated by snow (“snow_domi”), the fraction of successful retrieval is low (34.8% and 51.6%, respectively) and the retrieval accuracy for the successfully-retrieved pixels is low, with an RMSE of 0.268 and 0.197 respectively for $\tau_{550}$. The extended version of RemoTAP with the Ross-Li-Snow model on the other hand has much better performance with high fraction of successful retrievals (89.1% to 97.4%), with an RMSE on $\tau_{550}$ of 0.084 and 0.097 respectively. As expected, the performance over snow-covered pixels is still worse than for snow free pixels, because the high signal from the bright snow surface overwhelms the aerosol signal, leading to a reduced aerosol information content. For the synthetic measurements in “snow_rand”, in which...
snow is not dominant, the baseline RemoTAP gets acceptable performance, but the performance is still worse than the extended RemoTAP both in accuracy and fraction of successful retrievals. For $\omega_{550}$ retrieval, the performance of the baseline is slightly better than extended RemoTAP but it should be noted that only few pixels are left after the $\chi^2$ filtering. Overall, the results demonstrate the importance of extending RemoTAP with the snow kernel in order to get the capability for aerosol retrieval over snow.

Looking at the validation of the synthetic measurement “snow_free” in which there is no snow present, the extended RemoTAP maintains a similar capability as the baseline RemoTAP in terms of RMSE, bias and fraction of successful retrievals. For synthetic measurement with snow (“snow_pure”, “snow_domi” and “snow_rand”), especially when snow is dominant among the landcover (“snow_pure” and “snow_domi”), the extended RemoTAP has good performance for aerosol retrieval over snow. The scatter-plot for synthetic measurement “snow_pure” retrieved with extended RemoTAP is shown in Figure 3. The accuracy for $\tau_{550}$, $\omega_{550}$ and $AE_{440-870}$ retrievals is good, with an RMSE of 0.084 for $\tau_{550}$ retrieval, an RMSE of 0.021 for $\omega_{550}$ retrieval and an RMSE of 0.293 for $AE_{440-870}$ retrieval.

The synthetic results suggest that PARASOL measurements have enough information to allow the inclusion of the snow kernel in the retrieval state vector. Also, they show that by inclusion of the snow kernel, the extended RemoTAP is capable of performing aerosol retrievals over snow in a consistent setup. The next section will show the performance of real measurements.

5 Real data experiments

In our experiments, 5-band (490 nm, 565 nm, 670 nm, 865 nm, 1020 nm), 4-band (490 nm, 565 nm, 670 nm, 865 nm) and 3-band (490 nm, 565 nm, 670 nm) retrievals are conducted with the extended RemoTAP, as well as with the baseline RemoTAP. The validation of $\tau_{550}$ for different retrieval setups over snow-dominant surfaces ($c_{\text{snow}} > 75\%$) is shown in Figure 4.
Figure 4: Real data retrievals of $\tau_{550}$ among 5-band (panels (a1, a2)), 4-band (panels (b1, b2)) and 3-band (panels (c1, c2)) RemoTAP versus AERONET over snow-dominant surfaces ($c_{snow} > 75\%$). Panels (a1, b1, c1) show the scatter-plot of baseline RemoTAP and the panels (a2, b2, c2) show the scatter-plot of extended RemoTAP. The statistics with threshold required by Global Climate Observing System (GCOS) are also shown on the right bottom of each panel.
According to the validation with AERONET for $\tau_{550}$ retrieval over snow-dominant surfaces shown in Figure 4, we find that:

1. For the 5-band retrieval, the extended RemoTAP offers a better performance than the baseline RemoTAP, but the accuracy is still far from acceptable. This is caused by the fact that the retrieval algorithm has difficulty in fitting the 1020 nm band, where snow reflectance decreases significantly, making the reflectance much lower than that of the visible bands (see Figure 1). In order to get a better result than the 5-band retrieval, we investigate retrievals with a reduced number of spectral bands. First of all, the 1020 nm band is excluded because the algorithm has difficulties in fitting the strong spectral change between the visible bands and the 1020 nm band (synthetic experiments indicate that this only works with a very accurate first guess value for the BRDF in the 1020 nm band). We also investigate an even more reduced set of wavelength bands including only the 490 nm, 565 nm, and 670 nm bands (3-band retrieval). The performance of the 3-band and 4-band retrieval is much better than that of the 5-band retrieval, e.g., for the 4-band retrieval the RMSE is reduced from 0.307 to 0.133. Interestingly, for the 3-band retrieval the RMSE is further reduced to 0.101. We do not show the scatter-plot validation figures for SSA and AE retrievals over snow-dominant surfaces ($c_{\text{snow}} > 75\%$) because not enough AERONET data are available for validation for snow-dominated pixels.

2. The extended RemoTAP has a significantly better agreement with AERONET than the baseline RemoTAP for the 3-, 4-, and 5-band retrievals. We would also like to emphasize that the fraction of pixels which meet the GCOS accuracy requirement also has a huge increase compared to the baseline RemoTAP of up to 45 percent-points for 3- and 4-band retrievals. This demonstrates the importance of adding the snow kernel to our BRDF model for retrievals over snow.

Figure 6 summarizes the RMSE and bias on $\tau_{550}$, $\omega_{550}$ and $\text{AE}_{440-870}$ for different retrieval setups for both snow-free ($c_{\text{snow}} = 0\%$) and snow-covered pixels ($0\% < c_{\text{snow}} \leq 100\%$). For SSA, we would like to note again that the statistics over snow have limited value and no strong conclusions should be drawn from these numbers. For AE, although there is not enough statistics over snow-dominant surfaces ($c_{\text{snow}} > 75\%$), there are enough pixels for validation when $0\% < c_{\text{snow}} \leq 100\%$. 

https://doi.org/10.5194/amt-2023-127
Preprint. Discussion started: 26 July 2023
© Author(s) 2023. CC BY 4.0 License.
According to the validation results for retrievals over snow-free surfaces ($c_{\text{snow}} = 0\%$) and snow surfaces ($0\% < c_{\text{snow}} \leq 100\%$) shown in Figure 5, we can conclude that:

1. For snow-free surfaces ($c_{\text{snow}} = 0\%$), the extended RemoTAP offers similar performance as the baseline RemoTAP for 3-band and 4-band retrievals, respectively. The only exception is the bias of $\omega_{550}$ retrieval, where extended RemoTAP retrievals have larger bias than baseline RemoTAP, but the difference is small ($< 0.01$), especially given also the AERONET uncertainty on SSA and the limited number of validation points for SSA.

2. For snow-covered surfaces ($0\% < c_{\text{snow}} \leq 100\%$), the baseline RemoTAP 4-band and 3-band retrievals fail to provide useful results with good accuracy, while both 3-band and 4-band extended RemoTAP retrievals are able to have good performance.

3. Comparing the extended RemoTAP for 3- and 4-band retrievals, the results of $\tau_{550}$ retrievals are similar, but the 4-band extended RemoTAP retrieval has slightly better performance for $\omega_{550}$ and $AE_{440-870}$ retrievals (comparing the blue and brown bars). Therefore, the 4-band extended RemoTAP retrieval is recommended for aerosol retrieval over snow from PARASOL.
In addition to the advantage in accuracy, taking 4-band RemoTAP as an example, there is also an obvious increase in the fraction of successful retrievals \((\text{chi}^2 < 5)\), as shown in Fig. 6 for different snow fractions bins. On average, inclusion of the snow kernel leads to an increase in successful retrievals by 5.8 percent-points and when the snow cover is large this increases to 18.9 percent-points, compared to the baseline, resulting in a factor \(~2\) more successful retrievals. This indicates that the extended RemoTAP provides better performance on both accuracy and goodness-of-fit for aerosol retrieval over snow.

![Figure 6: Fraction of successful retrievals (left y-axis) for real data retrievals among baseline 4-band RemoTAP (red bar) and extended 4-band RemoTAP (blue bar) over different snow cover intervals. The number of the matched snow pixels (right y-axis) over different snow cover intervals is plotted in green bar.](https://doi.org/10.5194/amt-2023-127)

Figure 7 shows an example of the difference between PARASOL measurements and the RemoTAP forward model simulations (after convergence) for ToA reflectance and DoLP for aerosol retrieval over a snow-dominated pixel. Comparing the ToA reflectance for the baseline- and extended RemoTAP in Figure 7 (a1) and (b1), the difference between PARASOL measurements and the results of extended RemoTAP forward model is much smaller than that of baseline RemoTAP. Comparing the ToA DoLP in Figure 7 (a2) and (b2), the performance is quite comparable, and there is slight advantage at the 865 nm band for the extended RemoTAP. It is important to note that there are some unphysical oscillations in the PARASOL measurements which are probably caused by interpolation error in the PARASOL level 1C processing for inhomogeneous scenes.
Figure 7: Example of PARASOL measurements and the RemoTAP results for ToA reflectance (panel (a1, b1)) and ToA DoLP (panel (a2, b2)). Solid lines in panels (a1, a2, b1, b2) refer to PARASOL measurements, dotted lines in panel (a1, a2) refer to baseline RemoTAP result and dotted lines in panel (b1, b2) refer to extended RemoTAP results. The red lines refer to 490 nm band, the blue lines refer to 670 nm band and the green lines refer to 865 nm band. The pixel is located at (43.75° N, 96.65° W), and the snow cover of this pixel is 68.41%.

6 Conclusions

We extended the RemoTAP algorithm with a snow kernel in the BRDF model to carry out aerosol retrieval over snow from PARASOL MAP measurements. We conducted synthetic retrievals to investigate the necessity and advantage of extending RemoTAP, and conducted retrievals on real measurements. For PARASOL retrievals with 4-bands (490 nm, 565 nm, 670 nm, 865 nm), the extended RemoTAP retrieval results agree well with AERONET of optical properties, both for retrievals over snow as well as for snow-free areas. The RMSE, bias, and fraction of retrievals within the GCOS requirements are 0.118, 0.023, and 57.2% respectively for \( \tau_{550} \) retrieval over partly snow-covered surfaces. This is much improved compared to the baseline RemoTAP (without snow kernel) that yields an RMSE, bias, and fraction of retrievals within the GCOS requirements of 0.221, 0.071, and 48.2% respectively. Furthermore, the fraction of successful retrievals also improves by up to 18.9 percent-
points compared with the baseline. The improvement is most striking for surfaces that have snow cover > 75%, where the number of successful retrievals more than doubles and the fraction of retrievals that are within the GCOS requirement increases from 3.4% to 48.1%. The performance of the extended RemoTAP on accuracy and goodness-of-fit is in good agreement with the expectation from the synthetic data experiments. A limitation of the extended RemoTAP is that it is not able to fit PARASOL measurements at 1020 nm, where the snow albedo is substantially lower than at lower wavelengths. Therefore, the 4-band extended RemoTAP is recommended as the best choice for aerosol retrieval over snow surfaces considering the performance for AOT, SSA and AE retrieval, both from synthetic- and real retrievals.

Data availability. The PARASOL level-1 data can be downloaded from the website http://www.icare.univ-lille1.fr/parasol/products (last access: 8 June 2023). The AERONET data can be downloaded from the website https://aeronet.gsfc.nasa.gov/ (last access: 8 June 2023). The MERRA-2 meteorological data can be accessed through the website https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/ (last access: 8 June 2023). The MODIS/Aqua snow cover data can be accessed through the website https://nsidc.org/data/ MODIS/Aqua snow cover data can be accessed through the website https://nsidc.org/data/myd10a1/versions/61 (last access: 8 June 2023).

Author contributions. ZZ, GF and OH designed the experiments. ZZ performed the experiments and data analysis and GF and OH contributed to the interpretation. GF and OH implemented the snow kernel in RemoTAP. ZZ wrote the first paper draft which was finalized by contributions from GF and OH.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Atmospheric Measurement Techniques.

Acknowledgements. Zihan Zhang is supported by a scholarship from China Scholarship Council (CSC).

References


