An algorithm for hyperspectral remote sensing of aerosols: 3. Application to the GEO-TASO data in KORUS-AQ field campaign

Weizhen Hou [a], Jun Wang [a,*], Xiaoguang Xu [a,e], Jeffrey S. Reid [b], Scott J. Janz [c], James W. Leitch [d]

[a] Department of Chemical and Biochemical Engineering, Center for Global & Regional Environmental Research, and Iowa Technology Institute, The University of Iowa, Iowa City, IA 52242, United States
[b] Marine Meteorology Division, Naval Research Laboratory, 7 Grace Hopper Ave, Stop 2, Monterey, CA 93943, United States
[c] Atmospheric Chemistry and Dynamics, NASA Goddard Space Flight Center, Code 614, 8800 Greenbelt Rd., Greenbelt, MD 20771, United States
[d] Ball Aerospace & Technologies, 10 Longs Peak Drive, Broomfield, CO 80021, United States
[e] Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, Maryland 21250, United States

ARTICLE INFO

ABSTRACT

This paper describes the third part of a series of investigations to develop algorithms for simultaneous retrieval of aerosol parameters and surface spectral reflectance from GEOstationary Trace gas and Aerosol Sensor Optimization (GEO-TASO) instrument. Since the algorithm is designed for future hyperspectral and geostationary satellite sensors, such as Tropospheric Emissions: Monitoring of Pollution (TEMPO), it is applied to GEO-TASO data measured over the same area by different flights as part of the Korea-United Stated Air Quality Study (KORUS-AQ) field campaign in 2016. While GEO-TASO has a spectral sampling interval of ~0.28 nm in the visible, its data is thinned through a band selection approach with consideration of atmospheric transmittance and different surface types, which yields 20 common spectral bands to be used by the algorithm. The algorithm starts with 4 common principal components (PCs) for surface spectral reflectance extracted from various spectral libraries; constraints of surface reflectance and aerosol model parameters are obtained respectively from k-means clustering analysis of the Rayleigh-corrected GEO-TASO spectra and AERONET data. The algorithm then proceeds iteratively with an optimal estimation approach to update PCs and retrieve aerosol optical depth (AOD) from GEO-TASO measured spectra until state vector converges. The comparison of AODs between GEO-TASO retrievals (y) and 7 AERONET (x) sites reveals that the iterative updates of surface PCs (and so surface reflectance) improve the inversions of fine-mode AOD, fine-mode fraction of AOD, Ångström exponent, and AOD at all (440, 550, 550, 675 nm) wavelengths. At 440 nm, the linear fitting equation, the Pearson correlation coefficient (R^2), and mean absolute error are improved respectively from $y = 0.72x + 0.11$, 0.53, and 0.05 (without update of PCs) to $y = 1.055x + 0.01$, 0.76, and 0.033. Future work is to prepare the algorithm for TEMPO that carries an enhanced version of GEO-TASO instrument.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Aerosols are an important component of the global atmosphere. They are composed of liquid and solid particles suspended in the air and can originate from either natural or manmade sources [1,2]. By scattering and absorbing solar radiation and further modifying the properties of clouds, aerosols have a significant impact on climate change, air pollution, visibility and the ecological environment [3-5]. In the past decades, polar-orbiting satellites have

demonstrated the ability to observe aerosols and other pollutants (such as SO_2 and NO_2) affecting air quality [6-14]. However, the impact of polar-orbiting observations has been limited by their coarse temporal resolution (usually once per day) with respect to change of aerosol source distributions that have distinct diurnal cycle (such as urban emissions) or are sporadic (such as fires and dust). This limitation is insufficient to observe the details of the diurnal nature of air pollution events that can develop over timescales of an hour to within a day [7,15,16]. Conversely, observations of geostationary (GEO) satellites can overcome these problems and provide observations many times throughout the day [17-23]. Recently, an international effort has been developed to

Table 1

List of the instruments and their associated models for OE-based inversion of aerosols.

Acronym	Full name	Surface model	Forward/OE model*	Reference
AATSR	Advanced Along Track Scanning Radiometer	MODIS surface albedo	ORAC	[45,46]
AirMSPI	Airborne Multi-angle spectroPolarimeter Imager	RPV, Modified Fresnel model	MarCh-AD	[62-64]
APEX/ESA	Airborne Prism Experiment/European Space Agency	ASTER spectral library	libRadtran	[67]
AVIRIS-NG	Next Generation Visible/Near Infrared Imaging Spectrometer	–	MODTRAN	[68,69]
CAPI	Cloud and Aerosol Polarimetric Imager	Linear mixture of actual surface spectra	UNL-VRTM, VLIDORT	[52,53]
DPC	Direction Polarimetric Camera	Improved BRDF, Maignan BPDF model	UNL-VRTM	[54]
EPIC/DSCOVR	Earth Polychromatic Imaging Camera / Deep Space Ovservatory	–	ARTDECO	[56]
GEMS	Geostationary Environment Monitoring Spectrometer	OMI surface reflectance dataset	VLIDORT	[24]
GEO-TASO	Geostationary Trace gas and Aerosol Sensor Optimization	PCA	UNL-VRTM	[34,35]
IASI	Atmospheric Sounding Interferometer	–	Atmosphit	[44]
MISR	Multi-angle Imaging SpectroRadiometer	–	SCIATRAN	[47]
OMI	Ozone Monitoring Instrument	Lambertian, TOMS climatology database	VLIDORT	[42]
POLDER/PARASOL	POLarization and Directionality of the Earth's Reflectances / Polarization and Anisotropy of Reflectances for Atmoshperic Science coupled with observation from a Lidar	RPV, Maignan BPDF model	GRASP	[48,49]
		Ocean model	SRON	[50,51]
PSAC	Chinese Polarized Scanning Atmospheric Corrector	Maignan BPDF model	UNL-VRTM	[55]
RSP	Research Scanning Polarimeter	Bréon model	VSOS	[57]
		RPV, Modified Fresnel model	SRON	[59]
		Ross-Li BRDF, Maignan BPDF model	SRON	[58]
		Ocean model	MAPP, VRT-SOS	[60,61]
SAGE II	Stratospheric Aerosol and Gas Experiment II	–	Mie/OE	[43]
SeaWiFS	Sea-Viewing Wide Field-of-View Sensor	Ocean model	CAO-LDISORT	[37]
SEVIRI	Spinning Enhanced Visible and Infrared Imager	MODIS BRDF	ORAC	[39]
		RPV model	SVATs/LDA	[38,40]
		RPV model	CISAR	[41]
SPEX airborne/ACEPOL	Spectropolarimeter for Planetary Exploration / Aerosol Characterization from Polarimeter and Lidar campaign	RPV, Maignan BPDF model	SRON	[66]
		Ocean model	SRON	[65]

* The definition of the acronyms in this column are as follows. ORAC: Oxford-RAL aerosol and cloud, MarCh-AD: Markov chain/adding-doubling, libRadtran: library for radiative transfer, MODTRAN: Moderate spectral resolution atmospheric transmittance algorithm and computer model, UNL-VRTM: Unified linearized vector radiative transfer model, VLIDORT: Linearized pseudo-spherical vector discrete ordinate radiative transfer, ARTDECO: Atmospheric radiative transfer database for earth and climate observation, Atmosphit: Atmosphere radiative transfer model, SCIATRAN: radiative transfer model for SCIAMACHY, GRASP: Generalized retrieval of aerosol and surface properties, SRON: Netherlands Institute for Space Research model, VSOS: Successive order of scattering model for solving vector radiative transfer, MAPP: Microphysical aerosol properties from polarimetry algorithm, Mie: Mie scattering code, OE: optimal estimation, CAO-LDISORT: Coupled atmosphere–ocean linearized discrete ordinate radiative transfer, SVATs: Soil-vegetation atmosphere transfer schemes, LDA: Land daily aerosol algorithm, CISAR: Combined inversion of surface and aerosol.

launch a constellation of GEO satellite instruments focused on air quality over Asia, North America and Europe, including the Geostationary Environment Monitoring Spectrometer (GEMS) [24,25], Tropospheric Emissions: Monitoring of Pollution (TEMPO) [26,27] and Sentinel-4 [28] missions, respectively. However, while GEO sensors can have an advantage in temporal sampling, they do suffer from reduced spatial resolution and must cope with increased challenges associated with viewing geometry and scattering angles.

This paper presents the third part of a series of studies that aim to develop hyperspectral remote sensing techniques for aerosol retrievals from a newly developed instrument called Geostationary Trace gas and Aerosol Sensor Optimization (GEO-TASO) [29,30]. As the airborne version of the upcoming air quality satellite instruments onboard GEMS and TEMPO, GEO-TASO measures hyperspectral backscattered ultraviolet (UV), visible (VIS) and near-infrared (NIR) radiation. TEMPO has been selected as the first Earth Venture Instrument by the National Aeronautics and Space Administration (NASA), and will be launched in 2022 to measure atmospheric pollution over North America from space, by using the hyperspectral UV and VIS spectroscopy hourly and with the spatial resolution about 4×4 km^2 [26]. Meanwhile, to improve the use of satellites to monitor air quality for public health and benefit, several field campaigns have been carried out by NASA, such as the Deriving Information on Surface Conditions form Column and Vertically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) in Houston and Denver during 2013–2014 [29,31], and the Korea-United Sated Air Quality Study (KORUS-AQ) in Korea during 2016 [32,33]. From the DISCOVER-AQ and KORUS-AQ campaigns, a series of airborne hyperspectral observation of GEO-TASO have been obtained and a three-part of study was carried out to explore and develop algorithms to retrieve aerosols from the real hyperspectral data in support of missions such as TEMPO.

In the first two parts of this study [34,35], we have developed a theoretical framework to retrieve aerosol and surface properties simultaneously from hyperspectral measurements in the visible spectrum. Here, we apply that theoretical algorithm to the observation data collected by GEO-TASO during KORUS-AQ. Our retrieval is based on an optimal-estimation (OE) approach to conduct the spectral fitting with inline radiative transfer, because the high dimensionality of state vector with multiple retrieved parameters may led to the much too large lookup-tables (LUTs). For computational reasons, the retrievals use the inline forward model instead of the LUTs approach [9,36] in this study. This strategy is consistent with the past studies that have applied OE method to invert aerosol properties from a wide range of optical instruments onboard either space-borne or airborne platforms, including the SeaWiFS [37], SEVIRI [38-41], OMI [42], SAG II [43], IASI [44], AASTR [45,46], MISR [47], POLDER/PARASOL [48-51], CAPI [52,53], DPC [54], PSAC [55], GEMS [24], EPIC/DSCOVR [56], RSP [57-61], AirMSPI [62-64], SPEX airborne/ACEPOL [65,66], APEX/ESA [67], AVRIS-NG [68,69] and GEO-TASO [34,35]. The definition of these sensors' acronyms is listed in Table 1, so are the names of the surface models, the forward models, and the OE models used by these past studies to implement OE-based inversion. The Levenberg-Marquardt method and Quasi-Newton methods are usually used for optimization in the framework of the inversion theory [70].

For the hyperspectral remote sensing, the OE approaches have been already used in the retrieval of gas concentration [71,72] and atmospheric correction of hyperspectral imaging [68,69] for AVRIS-NG. However, challenges in these past studies remain regarding the retrieval of aerosol parameters, especially in the treatment of surface reflectance. In the inversion, since the accuracy of surface reflectance estimation can directly affect the separation of the atmospheric contribution of radiances (or path radiance) from the surface contribution to the hyperspectral measurements at altitude of satellite or aircraft, one major source of uncertainty in the algorithm of aerosol retrievals arise from the approach to characterize the hyperspectral surface reflectance. To tackle this issue, our algorithm framework as shown in the first two-parts of this study [34,35], is to simultaneously retrieve the aerosol and surface properties. The unique part of this framework is to decompose the surface reflectance spectra into different principal components (PCs). Thus we only need to retrieve several weighting coefficients of PCs instead of the reflectance band-by-band for full spectra [34]. Indeed, for the hyperspectral instruments like GEO-TASO, they have tens to hundreds of bands in the ultraviolet (UV) and visible (VIS) spectrum and a PC-based characterization of surface reflectance in these bands are not only computationally appealing but also physically sound because the reflectance in these bands are often highly correlated. Hou et al. [35] also showed that the PCs of surface spectra in the atmospheric window channel could be approximately derived from the top-of the atmosphere reflectance in the cases of the low AOD. Here, to investigate the feasibility of the OE-based retrieval framework with principal component analysis (PCA), we apply its algorithm theoretical basis (as described in the part 1 and part 2 of this study [34,35]) to conduct the retrievals of aerosol and surface properties from the real data, i.e., the hyperspectral measurements of GEO-TASO during the KORUS-AQ campaign. For completeness, we briefly describe the methodology of our OE-based inversion in Section 2, and then present the GEO-TASO data processing and preparation in Section 3. Afterward, the retrieval results are investigated and validated in Section 4. Finally, the summary and collusion are provided in Section 5.

2. Methodology

Fig. 1 shows the flowchart of our algorithm. The algorithm is based on the OE inversion framework described in Section 2.1. Because of the fine spectral resolution (~0.28 nm) of the GEO-TASO, to expedite the inversion, GEO-TASO data thinning is conducted through band selection method in Section 2.2. To reconstruct the spectral surface reflectance, as an initial guess, the common PCs for various surface reflectance spectra are first extracted from the spectral reflectance libraries for different surface types in Section 2.3. In addition, the Rayleigh correction method from GEO-TASO measurements is described in Section 2.4. After that, the approach to update the surface PCs is discussed based on the real hyperspectral measurements in Section 2.5, and the detailed steps of inversion are summarized in Section 2.6.

2.1. OE-based inversion method

Based on optimal estimation inversion framework [48,70], the cost function used in this study can be written as

$$J(\mathbf{x}) = \frac{1}{2}[\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b})]$$
$$+ \frac{1}{2}\gamma (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a), \tag{1}$$

where the superscript "T" and "−1" represents the transpose and inverse operation, respectively, γ is the regularization parameter to adjust the cost function's contribution of measurement and a priori, $\mathbf{y}$ represents the selected hyperspectral measurement vector of

GEO-TASO, $\mathbf{x}$ and $\mathbf{b}$ represent the state vector of the retrieval and the vector of other variables not retrieved but important for the forward model calculation (hereafter model error), $\mathbf{F}$ is a forward model and corresponds to the Unified Linearized Vector Radiative Transfer Model (UNL-VRTM) used in this study [19,73–76], $\mathbf{S}_a$ refers the error covariance matrix of the a priori estimate $\mathbf{x}_a$. Besides, the covariance matrix $\mathbf{S}_\epsilon$ can be written as

$$\mathbf{S}_\epsilon = \mathbf{S}_y + \mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T, \tag{2}$$

here $\mathbf{S}_y$ means the measurement error covariance matrix, the expression of $\mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T$ represents the error propagation in the forward model, in which $\mathbf{S}_b$ is the error covariance matrix for the vector $\mathbf{b}$, and $\mathbf{K}_b$ is the Jacobians matrix of measurements $\mathbf{y}$ with respect to $\mathbf{b}$.

Correspondingly, the OE-based inversion model can be seen as a nonlinear optimization problem subject to (s.t.) some constraint conditions [34,77]:

$$\begin{aligned} &\min J(\mathbf{x}) \\ &\text{s.t. } \mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \end{aligned} \tag{3}$$

where $\mathbf{l}$ and $\mathbf{u}$ are the per-variable constants representing the lower and upper bounds for the retrieval parameter vector, $\mathbf{x}$, respectively.

To solve the optimization problem in Eq. (3), the gradient vector $\nabla_\mathbf{x} J(\mathbf{x})$ is usually used in the optimized iterations, which can be represented in the form of

$$\nabla_\mathbf{x} J(\mathbf{x}) = -\mathbf{K}^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b})] + \gamma \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a), \tag{4}$$

here, $\mathbf{K}$ means the Jacobian matrix of $\mathbf{F}(\mathbf{x})$ with respect to $\mathbf{x}$. In this study, the code of Limited-memory Broyden-Flectcher-Goldfarb-Shanno algorithm for Bound constrained optimization (L-BFGS-B) has been used for the optimized iteration, which is an extension of the L-BFGS algorithm to handle the problems with simple bounds [78,79], such as the one in this study where the only constraint is of the form $\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$ in Eq. (3).

In the inversion with real GEO-TASO data, we set the state vector $\mathbf{x}$ as

$$\mathbf{x} = \left[V_0^f, V_0^c, w_1, w_2, w_3, w_4\right]^T, \tag{5}$$

where V_0 means aerosol volume concentration, superscripts "f" and "c" mean the fine-mode and coarse-mode parameter, respectively, $w_1 - w_4$ stands for the weighting coefficients of predefined PCs. Here, only 4 PCs are used to reconstruct the hyperspectral surface reflectance [34,35]. For simplicity, we assume that each parameter in the state vector is independent of each other, and the corresponding error covariance matrix of state vector can be written as

$$\mathbf{S}_a = \text{diag}\left(\left[\sigma_{V_0^f}^2, \sigma_{V_0^c}^2, \sigma_{w_1}^2, \sigma_{w_2}^2, \sigma_{w_3}^2, \sigma_{w_4}^2\right]^T\right). \tag{6}$$

Here, the symbol diag means the assignment of vector along the main diagonal of a matrix.

We set aerosol model vector $\mathbf{b}$ as

$$\mathbf{b} = \left[r_{eff}^f, \ v_{eff}^f, r_{eff}^c, \ v_{eff}^c, m_r^f(\lambda_0), m_i^f(\lambda_0), b_r^f, \ b_i^f, m_r^c(\lambda_0),\right.$$
$$\left. m_i^c(\lambda_0), b_r^c, \ b_i^c\right]^T, \tag{7}$$

where r_{eff} and v_{eff} are the effective radius and effective variance to represent the particle size distribution, $m_r(\lambda_0)$ and b_r represent the coefficients of the real part of complex refractive index, $m_i(\lambda_0)$ and b_i represent the coefficients of the imaginary part of complex refractive index, λ_0 is the reference wavelength and corresponds to 550 nm in this study. Thus, the spectral complex refractive index can be further calculated by

$$\begin{cases} m_r(\lambda) = m_r(\lambda_0)\left(\frac{\lambda}{\lambda_0}\right)^{-b_r} \\ m_i(\lambda) = m_i(\lambda_0)\left(\frac{\lambda}{\lambda_0}\right)^{-b_i} \end{cases}, \tag{8}$$
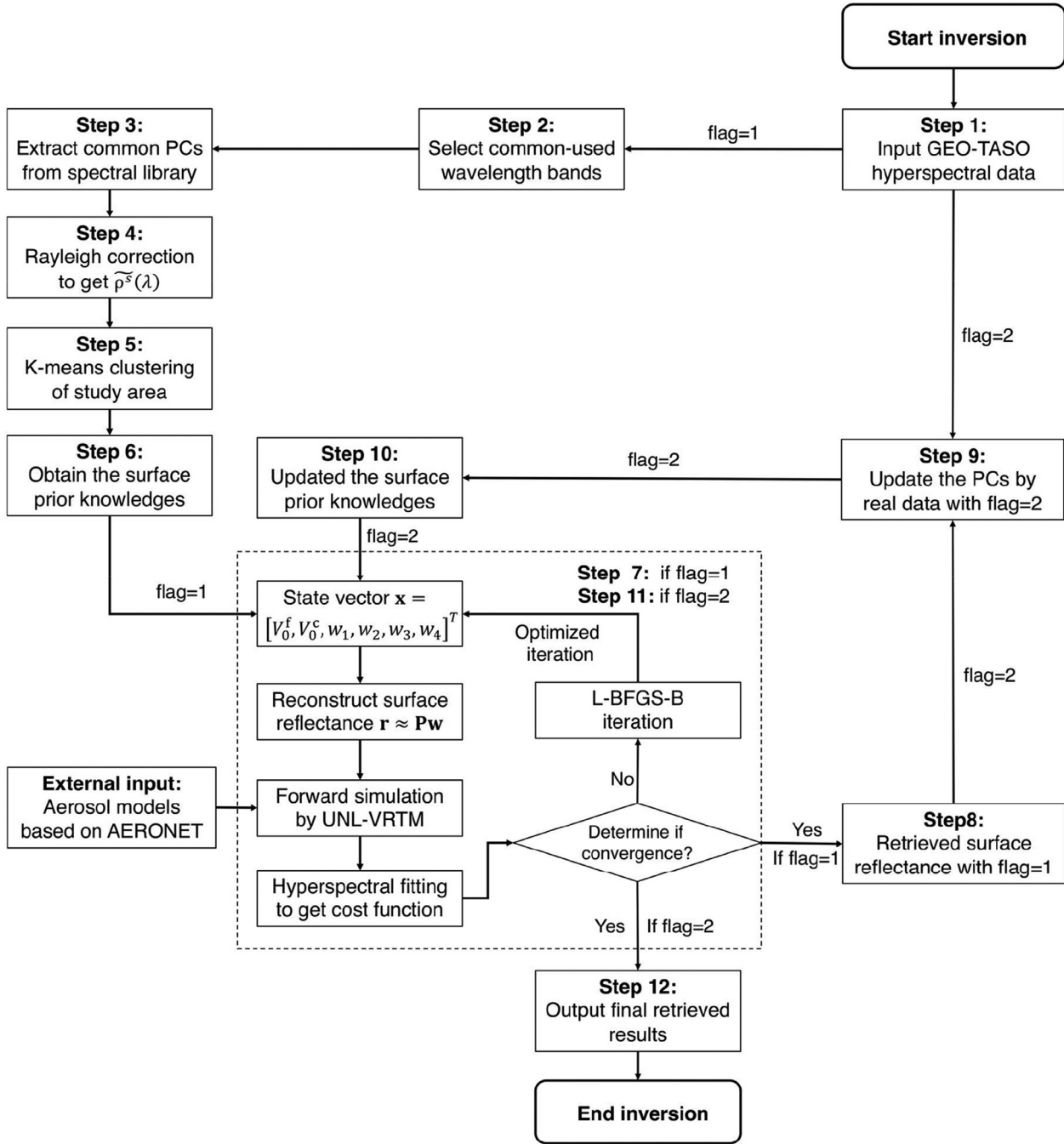
Fig. 1. Flowchart of PCA-based OE inversion of aerosol and surface reflectance from GEO-TASO hyperspectral measurement. Aerosol model parameters based on AERONET are the external inputs for the step 7 and step 11 in which the computational flows for the iterative retrievals (as enclosed in the box with dash lines) are the same regardless of flag values. If flag = 1, the step 7 is performed after step 6 in the flowchart; and if flag = 2, the step 11 will be performed after step 10 in the flowchart. Please see details in the main text.

where $m_r(\lambda)$ and $m_i(\lambda)$ are the complex refractive index dependent on the wavelength λ, respectively [34]. The aerosol model has been extracted from the Aerosol Robotic Network (AERONET) [80], which is discussed in next section.

Correspondingly, the spectral aerosol optical depth (AOD) can be further derived by

$$\tau_a(\lambda) = \tau_a^f(\lambda) + \tau_a^c(\lambda) = \frac{3V_0^f}{4r_{eff}^f}Q_{ext}^f(\lambda) + \frac{3V_0^c}{4r_{eff}^c}Q_{ext}^c(\lambda), \qquad (9)$$

where τ_a^f and τ_a^c mean the fine-mode AOD (AOD_f) and coarse-mode AOD (AOD_c) respectively, Q_{ext} represents the extinction efficiency factor and is usually calculated by the ratio of extinction cross section to geometric cross section. Consequently, the fine-mode fraction (FMF) of AOD and Ångström exponent (AE) can be written as

$$FMF(\lambda) = \tau_a^f(\lambda) / \tau_a(\lambda), \qquad (10)$$

and

$$AE = \ln\frac{\tau_a(\lambda_i)}{\tau_a(\lambda_j)} / \ln\frac{\lambda_j}{\lambda_i}, \qquad (11)$$

where the spectral wavelength pairs, λ_i and λ_j, correspond to 440 and 675 nm for GEO-TASO in this study.

For the uncertainty estimates provided by OE inversion, a posteriori error covariance matrix $\hat{\mathbf{S}}$ can be written as

$$\hat{\mathbf{S}} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1}, \tag{12}$$

and the square root of diagonal elements are the posterior errors (absolute errors). Because the retrieval uncertainties of hyperspectral surface reflectance have been investigated in our part 1 work [34], we only discuss the posterior errors of aerosol parameters in this paper. Here, the symbols $\epsilon_{V_0^f}$, $\epsilon_{V_0^c}$, $\epsilon_{\tau_a^f}$, $\epsilon_{\tau_a^c}$ and ϵ_{τ_a} are used to represent the posterior errors of V_0^f, V_0^c, AOD_f, AOD_c and AOD, respectively. Based on the posteriori error results of $\epsilon_{V_0^f}$ and $\epsilon_{V_0^c}$ calculated by Eq. (12), and we can further have

$$\begin{cases} \epsilon_{\tau_a^f} = \tau_a^f \epsilon_{V_0^f}/V_0^f \\ \epsilon_{\tau_a^c} = \tau_a^c \epsilon_{V_0^c}/V_0^c \\ \epsilon_{\tau_a} = \text{FMF } \epsilon_{\tau_a^f} + (1 - \text{FMF}) \ \epsilon_{\tau_a^c} \end{cases} . \tag{13}$$

In this way, the uncertainty estimates of retrieved AOD_f, AOD_c and AOD could be obtained.

2.2. GEO-TASO data thinning and error characterization

Owning to its dense spatial measurements in one thousand spectral channels with the spectral sampling interval of 0.28 nm in the visible wavelengths from 415 nm to 695 nm, GEO-TASO contains more aerosol and surface information than traditional multi-spectral measurements as shown in our previous work [34,35]. Although the spectral full width at half maxima (FWHM) is usually 2.5−3.6 times of the band or spectral sampling interval, considerable redundant information inevitably exist in hyperspectral measurements because of the high correlation of the measurement (or radiance intensity) between different bands [81]. To improve the efficiency of aerosol retrieval, the best bands containing most of information content and less noise are needed to reduce the dimensionality of hyperspectral measurements in the inversion. By combining the simulations and real data, the 20 spectral bands of GEO-TASO are selected for retrieval over different surface types [34,35]. We can set the vector of hyperspectral measurement in the form of

$$\mathbf{y} = \left[I_{\lambda_1}, I_{\lambda_2}, \cdots, I_{\lambda_{20}} \right]^{\mathrm{T}}, \tag{14}$$

where I_{λ_i} means the normalized GEO-TASO radiance (with respect to the incoming solar radiance at the top of the atmosphere at λ_i) after calibrations in the common wavelength bands λ_i with $i = 1$, $\cdots$, 20.

For the design of GEO-TASO instrument, the calibration error is about 2.8% in VIS [29,30]. Because the measurement errors are not totally independent among the common-used wavelength bands, the measurement error covariance matrix $\mathbf{S}_y$ is a non-diagonal matrix [9], and in the past we have used Monte Carlo methods to estimate the its off-diagonal values [82]. But, this is shown to be challenging because it requires knowledge of instrument and forward model error ($\mathbf{K}_b \mathbf{S}_b \mathbf{K}_b^T$) covariances (both of which are not diagonal). To facilitate the analytic expressions of cost function and gradient vector, following our past work [75,76], we simplify $\mathbf{S}_\epsilon$ as a diagonal matrix and its diagonal element $(\mathbf{S}_\epsilon)_{i,i}$ follows

$$(\mathbf{S}_\epsilon)_{i,\ i} = \sigma_{I_{\lambda_i}}^2, \ (i = 1, \cdots, \ 20), \tag{15}$$

with

$$\sigma_{I_{\lambda_i}} = e_i I_{\lambda_i}, \ (i = 1, \cdots, \ 20), \tag{16}$$

here $e_i = 5\%$ represents the total relative error of measurement and forward model at each common band for simplicity, which is calculated by the combination of 2.8% calibration error of GEO-TASO and about 2.2% mean error propagation of forward model to the airborne measurement based on the information content analysis in our part 2 work [35].
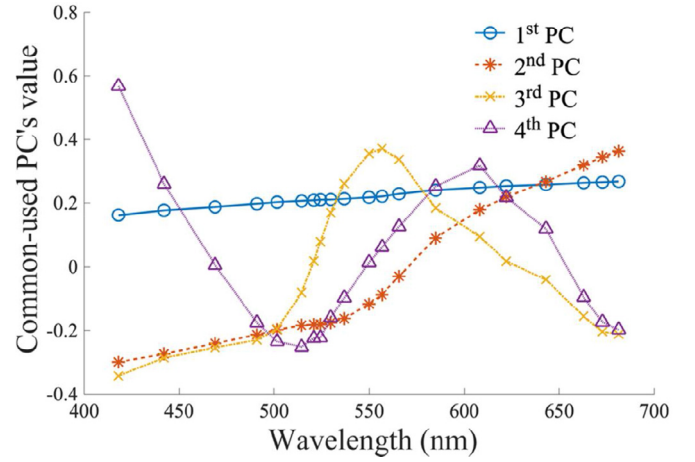
Fig. 2. Four PCs extracted from USGS and ASTER spectral libraries at 20 selected bands for different surface types. Those PCs are referred as common PCs in the text and are used as the first guess in the retrieval algorithm. The specific values for the 20 bands can be found at Section 3.2.

2.3. Extract the common PCs of surface spectra from spectral libraries

Following the theoretical framework for hyperspectral remote sensing [34], a Lambertian surface has been assumed in the forward model, and surface reflectance spectrum for a given sample can be approximately reconstructed by the PCs as

$$\mathbf{r} \approx \mathbf{Pw}, \tag{17}$$

where $\mathbf{r}$ is the vector of surface reflectance spectrum in 20 common used bands, $\mathbf{P}$ is the matrix that constituted by the first 4 PCs as the column vectors, and weighting coefficients vector $\mathbf{w} = [w_1, w_2, w_3, w_4]^T$. Since $\mathbf{P}$ is an orthogonal matrix by the column vector and can satisfy the condition that $\mathbf{P}^T \mathbf{P} = \mathbf{E}$ with $\mathbf{E}$ as the identity matrix (also called unit matrix), $\mathbf{w}$ can be calculated by

$$\mathbf{w} \approx \mathbf{P}^{\mathrm{T}} \mathbf{r}. \tag{18}$$

Fig. 2 shows the extracted common-used PCs from USGS and ASTER spectral library [83,84], in which the various datasets of the green vegetation, yellow vegetation, bare soil, rangeland, rock, concrete, water and man-made martials (such as roofing martial) have been considered. With the 4 PCs (e.g., common PCs for different surface types), the spectral reflectance datasets in these spectral libraries can be reconstructed with the relative error of ~2%–3%.

In this study, the common PCs are only used as the *initial guess* to start the retrieval, and then the PCs will be updated interactively by the retrieved surfaces from real GEO-TASO data to conduct the inversion again, which is further described in Section 2.5. In addition, because these 4 PC vectors are orthogonal to each other, they unavoidably have the negative elements in the PC vectors, and thus the negative values cannot be interpreted from a physics point of view. However, in relatively sense, the linear increase in the first PC may reflect the spectral variation of soil, while other PCs having peaks in the 500–600 nm may reflect the spectral variation of greenness for various canopies.

For the purpose of optimization, the Jacobian results of hyperspectral measurements with respect to weighting coefficients vector $\mathbf{w}$ has been implemented as part of the new development of UNL-VRTM [34]:

$$\frac{\partial I_{\lambda_i}}{\partial \mathbf{w}} = \frac{\partial I_{\lambda_i}}{\partial r_{\lambda_i}} [P_{i,1}, P_{i,2}, P_{i,3}, P_{i,4}]^T, (i = 1, \cdots, d), \tag{19}$$

where I is the normalized GEO-TASO radiance, $P_{i,1}$, $P_{i,2}$, $P_{i,3}$ and $P_{i,4}$ are the elements in matrix $\mathbf{P}$, d is the number of spectral bands used, and is equal to 20 here.
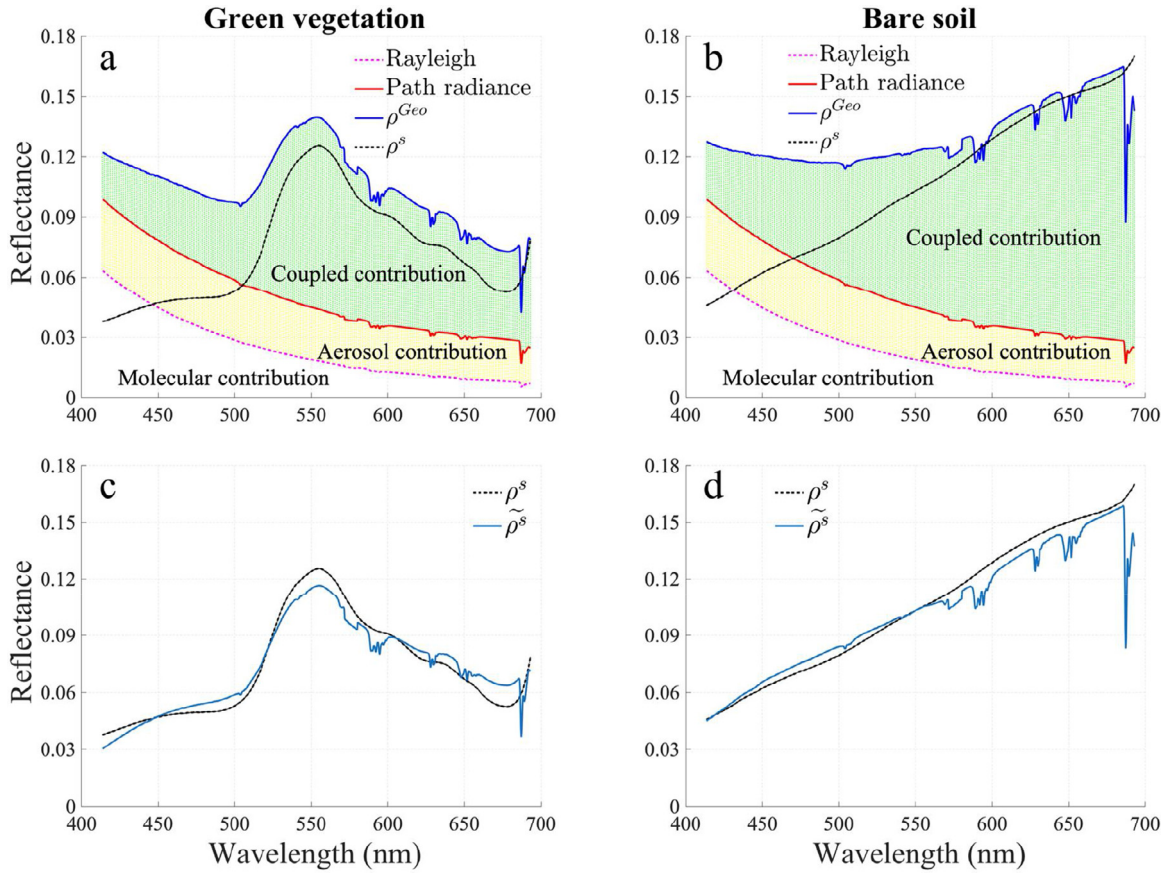
Fig. 3. The simulated coupling contribution and Rayleigh correction to GEO-TASO hyperspectral measurements. Panel ($a−b$): contribution of Rayleigh scattering, path radiance and surface reflectance (ρ^s) to the GEO-TASO reflectance (ρ^{Geo}) at about 8.4 km simulated by UNL-VRTM with consideration of gas absorption and AOD $\tau_a = 0.5$ at 550 nm for green vegetation and bare soil surface types, respectively. Panel ($c−d$): comparisons of the real surface reflectance (known from the spectral libraries and used in the forward calculation) and the approximated surface reflectance ($\widetilde{\rho^s}$) by the Rayleigh correction of TOA measurements for vegetation and soil surfaces, respectively. See the text in Section 2.4 for details.

2.4. Rayleigh correction of hyperspectral measurements

UNL-VRTM has been used as the forward model to simulate hyperspectral measurements and corresponding Jacobian results, which comprises a linearized pseudo-spherical vector discrete ordinate radiative code (VLIDORT) for radiative transfer [85], a linearized Mie/T-Matrix code for aerosol single scattering, a Rayleigh scattering module, and line-by-line gas absorption calculation with HITRAN database, as well as liberature-based modules for characterizing the surface properties [19,73].

To investigate the coupling contribution of aerosol and surface to the GEO-TASO hyperspectral measurement for aerosol retrieval, the forward simulations are also considered along the visible bands of GEO-TASO. Fig. 3(a−b) present the separate contributions of Rayleigh scattering, aerosol contribution, surface reflectance, and gas absorption to the GEO-TASO measurements for green vegetation and bare soil surface type, respectively, in which solar zenith angle $\theta_0 = 40°$, viewing zenith angle $\theta_v = 20°$, relative azimuth $\phi = 20°$, the AOD is assumed to 0.5 at 550 nm and flight altitude is ~8.4 km. The results in Fig. 3 is obtained from several realization of UNL-VRTM with different input options. For the separate contributions of Rayleigh scattering, aerosol scattering/absorption and surface-atmosphere coupled result, different input options are set in the UNL-VRTM. With the combination of gas absorption in forward simulation, the Rayleigh contribution is obtained by $\rho^s = 0$ and $\tau_a = 0$. The aerosol contribution is approximately calculated by subtracting Rayleigh contribution from path radiance with $\rho^s = 0$. The coupled contribution is approximately obtained

by subtracting path radiance from simulated GEO-TASO measurements with hyperspectral surface reflectance.

In the visible spectral range, as expected, contribution of Rayleigh scattering decreases sharply (e.g., λ^{-4}) with the increase of wavelength, while the variety of aerosol's contribution with the wavelength is relatively flat in atmospheric window channels. Deviations include the oxygen B bands in the range of 685–695 nm and other gas absorption bands. The smoothed spectral curve shape of GEO-TASO reflectance after the correction of Rayleigh scattering is similar as the spectral shape of surface reflectance, which is illustrated in Fig. 3(c−d). Thus, it is foreseeable that the a priori spectral curve shape can be approximately obtained after the Rayleigh correction from the GEO-TASO reflectance.

The wavelength dependent correction of Rayleigh scattering is made by using a scaling approach, and therefore the approximated surface reflectance $\widetilde{\rho^s}(\lambda)$ before the aerosol retrieval can be calculated by

$$\widetilde{\rho^s}(\lambda) = \rho_{meas}^{Geo}(\lambda) - k(\lambda)\ \rho_{simu}^{Ray}(\lambda), \qquad (20)$$

where ρ_{meas}^{Geo} means the real measured reflectance of GEO-TASO, and ρ_{simu}^{Ray} means the simulated Rayleigh reflectance, k is an empirical wavelength-dependent scaling coefficient, usually in range of 0.9–1.4 empirically. For the scaled Rayleigh correction of hyperspectral measurements, we use a lookup table that has been calculated by UNL-VRTM with discrete values of Sun zenith angle (θ_0), viewing zenith angle (θ_v), relative azimuth angle (ϕ), flight attitude and ground surface pressure. Then, the corresponding reflectance of Rayleigh scattering could be obtained by interpolation

from the lookup table. With the defined vector of spectral approximated surface reflectance $\widetilde{\boldsymbol{\rho}^s} = [\widetilde{\rho^s}(\lambda_1), \widetilde{\rho^s}(\lambda_2), \cdots, \widetilde{\rho^s}(\lambda_d)]^T$, the vector of approximated weighting coefficient $\tilde{\mathbf{w}}$ could be calculated by

$$\tilde{\mathbf{w}} \approx \mathbf{P}^T \widetilde{\boldsymbol{\rho}^s}, \tag{21}$$

here $\mathbf{P}$ corresponds to the common PCs and $d = 20$.

Based on the Rayleigh correction results of GEO-TASO in the common wavelength bands and Eq. (21), the prior knowledge could be extracted to constrain the retrieval of surface parameters, which include the lower bound $\mathbf{l} = [l_1, l_2, l_3, l_4]^T$, the upper bound $\mathbf{u} = [u_1, u_2, u_3, u_4]^T$, the mean value $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2, \bar{w}_3, \bar{w}_4]^T$, and the standard deviation $\boldsymbol{\sigma}_w = [\sigma_{w_1}, \sigma_{w_2}, \sigma_{w_3}, \sigma_{w_4}]^T$ of weighting coefficients. Correspondingly, the elements of $\mathbf{l}$, $\mathbf{u}$, $\bar{\mathbf{w}}$ and $\boldsymbol{\sigma}_w$ could be calculated by

$$\begin{cases} l_i = \min\left(\tilde{w}_i^j\right) - \sigma_{w_i} \\ u_i = \max\left(\tilde{w}_i^j\right) + \sigma_{w_i} \\ \overline{w}_i = \sum_j \tilde{w}_i^j / N_{\text{pixel}} \\ \sigma_{w_i} = \text{std}\left(\tilde{w}_i^j\right) \end{cases}, \ \left(i = 1, \cdots, 4; j = 1, \cdots, N_{pixel}\right), \tag{22}$$

where min, max and std stand for the minimum, maximum and standard deviation operator of the approximated weighting coefficients, the superscript j corresponds to the serial number of pixels in the same surface type, and N_{pixel} represents the total number of pixels used for each surface type classified by the k-mean clustering method [86].

2.5. Iterative update PCs based on real data

For the hyperspectral airborne and satellite remote sensing, a pixel usually corresponds to a large area on the ground, making it a mixture of reflectance spectra of multiple materials [87,88]. Because the common PCs are extracted from the spectral datasets for various pure materials and typical surface scences, obviously these datasets are limited and not able to cover all of the surface materials in reality. Once the training data used to define PCs are missing some components of real-world variability, or if the surface types are not a mixture of those typical pure materials (used for constructing the spectral libraries), the real surface spectra will not be reconstructed reliably by the common PCs. Therefore, to better reconstruct the reflectance of complicated surfaces or to account of local features that are not described by the training datasets for deriving initial PCs, we update the PCs iteratively in the retrieval based on the real data of GEO-TASO and the surface reflectance derived at the end of each iteration. In other words, the PCs shown in Fig. 2 are used as the initial guess to start the iterative retrieval.

Specifically, to update the surface PCs, we define the hyperspectral apparent reflectance vector of GEO-TASO as

$$\boldsymbol{\rho}_{\text{meas}}^{\text{Geo}} = \left[\rho_{\text{meas}}^{\text{Geo}}(\lambda_1), \rho_{\text{meas}}^{\text{Geo}}(\lambda_2), \cdots, \rho_{\text{meas}}^{\text{Geo}}(\lambda_d)\right]^T, \tag{23}$$

and define the simulated apparent reflectance with the common-used PCs as

$$\boldsymbol{\rho}_{\text{simu}}^{\text{Geo}} = \left[\rho_{\text{simu}}^{\text{Geo}}(\lambda_1), \rho_{\text{simu}}^{\text{Geo}}(\lambda_2), \cdots, \rho_{\text{simu}}^{\text{Geo}}(\lambda_d)\right]^T, \tag{24}$$

where the superscript "Geo" represents the GEO-TASO, the subscripts "meas" and "simu" represent the measurements and the convergence simulation results, respectively. Meanwhile, the vector of retrieved surface reflectance by the common PCs (at the initial step of retrieval) can be represented by

$$\boldsymbol{\rho}_{\text{simu}}^{\text{s}} = \left[\rho_{\text{simu}}^{\text{s}}(\lambda_1), \rho_{\text{simu}}^{\text{s}}(\lambda_2), \cdots, \rho_{\text{simu}}^{\text{s}}(\lambda_d)\right]^T. \tag{25}$$

Correspondingly, we have the retrieval residual of reflectance as

$$\Delta\boldsymbol{\rho} = \boldsymbol{\rho}_{\text{meas}}^{\text{Geo}} - \boldsymbol{\rho}_{\text{simu}}^{\text{Geo}}, \tag{26}$$

which can be used to the adjust surface reflectance vector for the next iteration of retrieval

$$\boldsymbol{\rho}_{\text{adju}}^{\text{s}} = \boldsymbol{\rho}_{\text{simu}}^{\text{s}} + \Delta\boldsymbol{\rho}. \tag{27}$$

With the adjusted surface reflectance vectors for pixels with the same surface type (which is obtained from the k-means clustering), the PCs are updated by performing PCA for that surface type, which will be discussed in detail in next section with real hyperspectral data. Correspondingly, the updated weighting coefficient vector $\mathbf{w}$ of the pixel with $\boldsymbol{\rho}_{\text{adju}}^{\text{s}}$ can be calculated by

$$\mathbf{w} \approx \mathbf{P}_{\text{adju}}^{\text{T}}\boldsymbol{\rho}_{\text{adju}}^{\text{s}}, \tag{28}$$

and the elements of $\mathbf{l}$, $\mathbf{u}$, $\bar{\mathbf{w}}$, and $\boldsymbol{\sigma}_w$ of weighting coefficients are updated by

$$\begin{cases} l_i = \min\left(w_i^j\right) \\ u_i = \max\left(w_i^j\right) \\ \overline{w}_i = \sum_j w_i^j / N_{\text{pixel}} \\ \sigma_{w_i} = \text{std}\left(w_i^j\right) \end{cases}, \ \left(i = 1, \cdots, 4; j = 1, \cdots, N_{\text{pixel}}\right), \tag{29}$$

where $\mathbf{P}_{\text{adju}}$ is the updated PCs, and the definition of operator symbols are same as Eq. (22). Note, $\mathbf{P}_{\text{adju}}$ varies with surface type, and is the same for those pixels with the same surface type. In contrast, $\mathbf{w}$ in Eq. (28) can vary with different pixels even for the same surface type because of different $\boldsymbol{\rho}_{\text{adju}}^{\text{s}}$ values; this is understood because the pixels of the same surface type in optical sense have the similarity in spectral variation of reflectance, but not in absolutely value of the reflectance value itself. Therefore, the retrieved values of surface reflectance are pixel dependent and for the same pixel, their values can vary during different times of the day or different days, which effectively allow the simultaneous retrieval of aerosol and surface properties at the same time.

2.6. Inversion steps

Fig. 1 illustrate the flowchart of OE-based inversion step by step, in which "flag = 1″ represents the initial inversion with the common PCs, while "flag = 2″ represents another inversion with the updated PCs. Besides, based on the aerosol products of AERONET, the parameters of aerosol model have been counted and extracted in advance for constraining the inversion. The corresponding inversion has the following 12 steps as described below:

Step 1: Read formated GEO-TASO data files to obtain the necessary measurement information.

Step 2: Select the GEO-TASO bands for the retrieval (hereafter common bands for retrieval, which will be discussed in next section).

Step 3: PCs for surface reflectance spectral reconstruction are extracted from the spectral library for the common wavelengths obtained in Step 2.

Step 4: The scaled Rayleigh correction is applied to each hyperspectral pixel to obtain the approximated surface reflectance $\widetilde{\rho^s}(\lambda)$.

Step 5: Use the unsupervised k-means clustering method to classify the surface types in the selected region.

Step 6: Based on all of the approximated weighting coefficient vector $\tilde{\mathbf{w}}$ by Eq. (21) for the pixels with each surface type, the corresponding prior constraints for the retrieval of surface reflectance are calculated by Eq. (22) for each surface type.

Step 7: By setting the parameters of aerosol model, state vector, measurement vector, aerosol models and corresponding prior knowledge, UNL-VRTM and L-BFGS-B are integrated to carry out the forward simulation and optimized iteration for

retrievals, from which the multiple parameters for aerosol and surface reflectance (e.g., state vector in eq. (5)) for each pixel are simultaneously retrieved.

Step 8: If the optimization gets convergence with flag = 1, the spectral surface reflectance can be reconstructed by the common PCs and retrieved weighting coefficients; consequently, the value of flag changes to 2 and the algorithm moves to Step 9. If the optimization gets convergence with flag = 2, the algorithm moves to Step 12. Here, the convergence is defined as the change of the state vector between the two iterations is less than 0.1% (for all retrieved parameters).

Step 9: Obtain the adjusted spectral surface reflectance dataset for each pixel, and the PCs are updated by PCA (flag = 2) for each surface type.

Step 10: Derive the prior knowledge of corresponding weighting coefficient for each surface type based on the updated PCs and analysis in Step 9 Eqs. (28–29).

Step 11: Repeat the inversion following Steps 7 – 10 until the optimization is converged with flag = 2.

Step 12: Output retrieved state vector and reconstruct surface reflected with the retrieved weights and the PCs used in the last optimization and the algorithm stops.

3. Data processing and preparation

GEO-TASO data in KORUS-AQ field campaign are processed to suit two purposes. The first is to imitate the geostationary spatial distribution of TEMPO by focusing on the observations of different flights in the same study area, while the second is to analyze observations over 7 AERONET sites to evaluate the retrieved properties of aerosols. Meanwhile, a prior aerosol model has been extracted based on the AERONET to constrain the inversion in KORUS-AQ field campaign.

3.1. GEO-TASO aircraft instrument

Under the NASA Earth Science Technology Office Instrument Incubator Program, the GEO-TASO aircraft instrument has been developed by Ball Aerospace in order to support geostationary satellite measurements of the trace gas abundance of a variety of molecules important to air quality, and was originally conceived as a test bed instrument for GEOstationary Coastal and Air Pollution Event (GEO−CAPE) [29,89]. Operating as a hyperspectral push broom scanner, the GEO-TASO instrument measures solar backscattered light in the nadir direction. The incoming radiation is split into two broad spectral bands by utilizing the 1st and 2nd order diffraction products in combination with a dichroic filter. The separated orders are then imaged onto 2 detectors, one for the 290−400 nm (UV) range and the other for the 415−695 nm (VIS) range, in which the spectral sampling is 0.14 nm and 0.28 nm for the UV and VIS, and the corresponding spectral resolution in FWHMs are 0.34−0.49 nm and 0.70−1.00 nm, respectively. For the main characteristics and detailed optical design of the instruments, readers can refer to the work of Leitch et al. [30] and Nowlan et al. [29]. By boarding on the NASA UC-12B King Air with the typical flight altitude of 8.7 km, the $45°$ field of view (FOV) of GEO-TASO results in a cross-track FOV of about 7.2 km for each cross-track swath in KORUS-AQ field campaign. The algorithm based on spatial covariance for cloud mask is first applied to GEO-TASO level 1b product files in VIS and cloud-free pixels are subsequently used for the OE-based inversion in this study. The nominal resolution of GEO-TASO data at the ground is 250 m × 250 m but can vary due to aircraft altitude, latitude, speed and telescope viewing angle. Because the level 1b data already has the cloud mask processing, the

Table 2
Date and time of 6 flights selected for this study.

Flight No.	Day	Time (UTC hour)
Flight 1	May 17, 2016	0.00–0.06, 0.50–0.55
Flight 2	May 17, 2016	2.28–2.35
Flight 3	May 17, 2016	5.51–5.58, 6.02–6.18
Flight 4	May 17, 2016	7.77–7.84
Flight 5	May 18, 2016	4.14–4.21
Flight 6	May 18, 2016	7.23–7.30

Table 3
List of AERONET sites where validation of retrieval is performed.

AERONET site	Latitude (Deg)	Longitude (Deg)
Hankuk_UFS	37.34	127.27
KORUS_Baeksa	37.41	127.57
KORUS_NIER	37.57	126.64
KORUS_Olympic_Park	37.52	127.12
KORUS_Taehwa	37.31	127.31
KORUS_Songchon	37.34	127.49
Yonsei_University	37.56	126.93

description below focuses only the processing of cloudless pixels only for the retrieval.

3.2. GEO-TASO data processing and thinning

In order to imitate the geostationary hyperspectral measurements of TEMPO, we select the cloudless GEO-TASO data measured over the same area during different flights in the adjacent days of the KORUS-AQ campaign (https://espo.nasa.gov/korus-aq), and then resample and match the cloudless pixels to the same latitude and longitude coordinate grid. Here, the grid step of latitude and longitude is constructed such that the spatial resolution is still corresponding to original GEO-TASO pixel spatial resolution of about 250 m × 250 m. Fig. 4(a−f) show the true color images of selected 6 flights over the same coordinate grid, and Fig. 5 illustrates the sun zenith angle, viewing zenith angle and relative azimuth angle distribution, and flight direction of these 6 flights correspondingly. Following the definition of observation geometry in UNL-VRTM [73], the relative azimuth angle (ϕ) is calculated by

$$\begin{cases} \phi = 180° - |\phi_v - \phi_0|, if\ |\phi_v - \phi_0| < 180° \\ \phi = -180° + |\phi_v - \phi_0|, if\ |\phi_v - \phi_0| > 180° \end{cases}, \quad (30)$$

where ϕ_0 and ϕ_v mean the solar azimuth and viewing azimuth respectively. Since both flight 1 and flight 3 had data collected within half an hour, these corresponding data are merged as one data granule for each of these two flights and mapped over the same study domain that other 4 flights also cover (Table 2). To validate the retrieved results by AERONET, we select the GEO-TASO data within the spatial range of 1 km × 1 km over 7 AERONET sites whose latitude, longitude and site names are listed in Table 3.

To extract common spectral bands from GEO-TASO for the retrieval, both the real and simulated data have been considered, and the following steps are taken:

Step 1: Select the measurements of 6 flights over the same pixel. An example is shown in Fig. 6(a) for a selected pixel.

Step 2: Transform the radiance to the reflectance, as illustrated in Fig. 6(b).

Step 3: Smooth the spectral reflectance and remove the noise, as plotted in Fig. 6(c). Here, we use the "wden" function code from the wavelet toolbox of Matlab to smooth the spectral noise.

Step 4: Scatter plot the real reflectance versus the smooth reflectance, just as shown in Fig. 6(d), in which the root mean
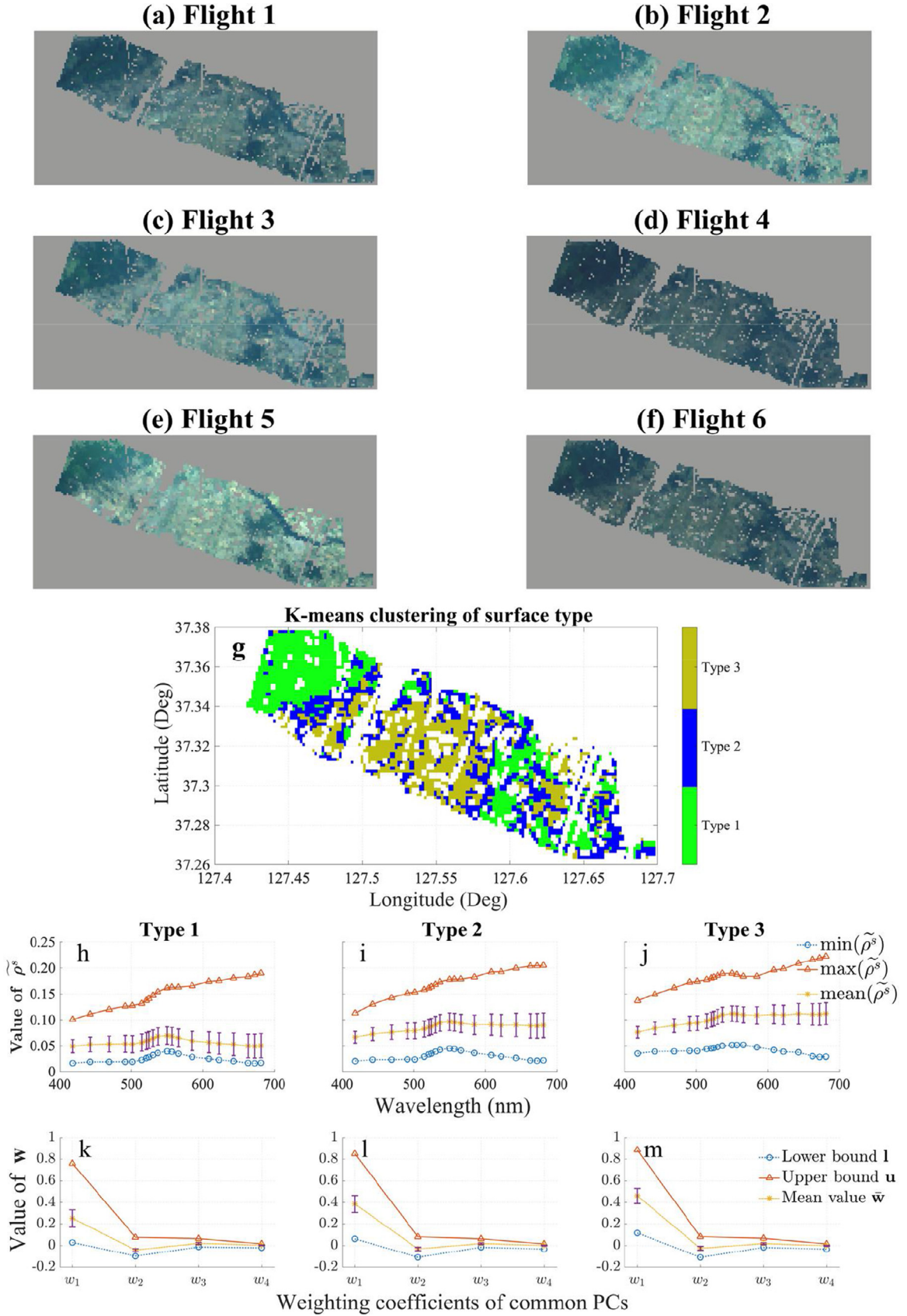
Fig. 4. Panel (a−f): GEO-TASO three-band (440, 550, 670 nm) true color images resampled over the study area with the same coordinate grids for 6 flights (Table 2) in KORUS-AQ, respectively. Panel (g): the k-means clustering results for 3 different surface types in the study area. Panel (h−j): the minimum, maximum and mean of approximated surface reflectance with standard deviation error bar in 20 common wavelengths for three surface types, respectively. Panel (k−m): the extracted prior knowledge to constrain the retrieval of surface parameters, which contains the lower bound **l**, the upper bound **u** and the mean value $\bar{\mathbf{w}}$ of weighting coefficients with the standard deviation $\sigma_{\boldsymbol{w}}$ error bar for 3 different surface types, respectively.
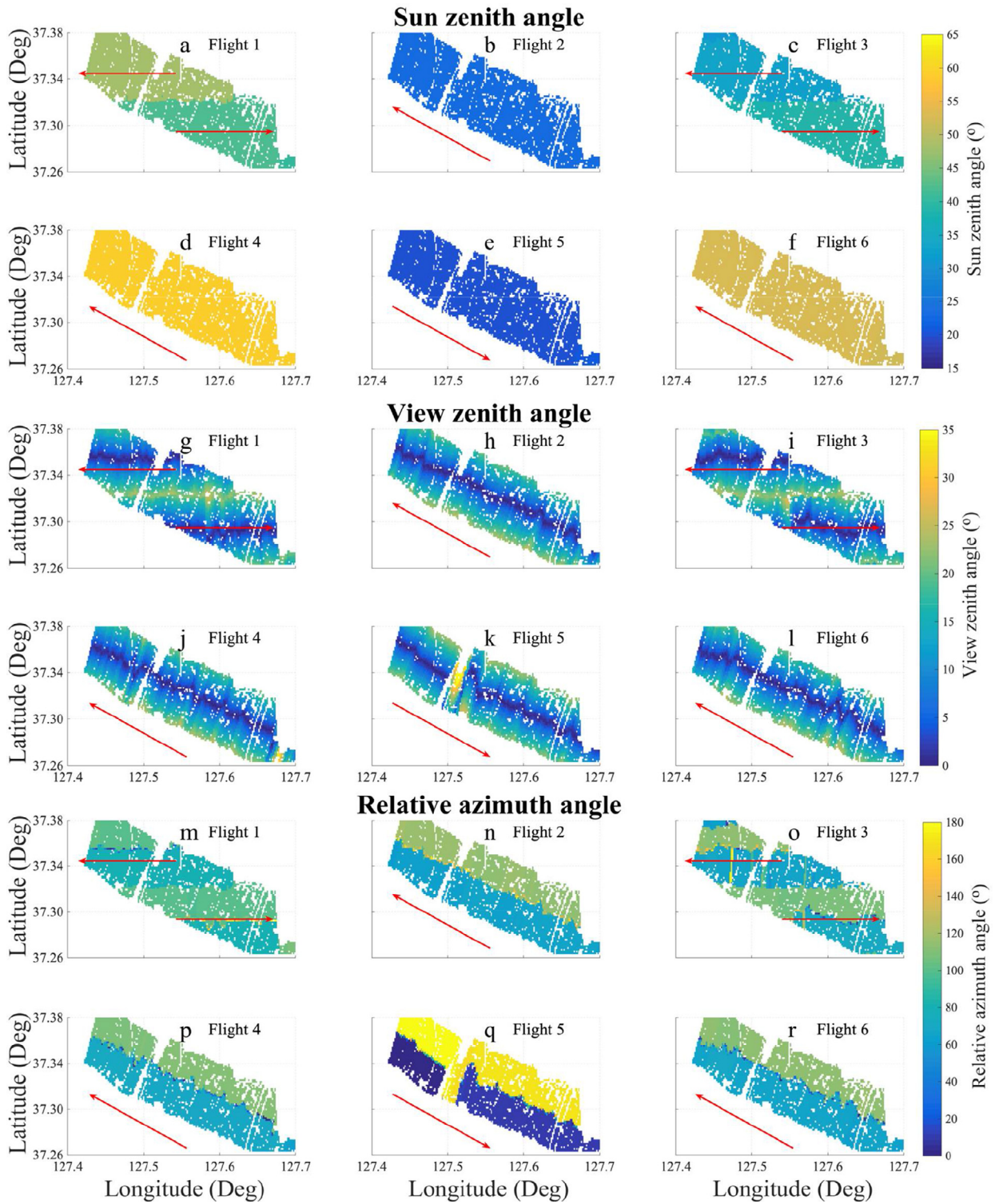
Fig. 5. Spatial distributions of the solar zenith angle (a−f), viewing zenith angle (g−l), and relative azimuth angle (m−r) for the GEO-TASO data collected during the 6 flights (Table 2 and Fig. 4a−f); the red arrow represents the direction for each flight. Because the images are constructed by merging the discontinuous measurements within half an hour for flight 1 and flight 3, there are 2 opposite flight directions and 2 different values of sun zenith angle in these 2 flights. In addition, relative azimuth angle (ϕ) is calculated from the solar relative azimuth and viewing azimuth by Eq. (30), and $\phi = 180°$ means that the GEO-TASO instrument and the Sun face the same direction and are located on the same side of the main plane.
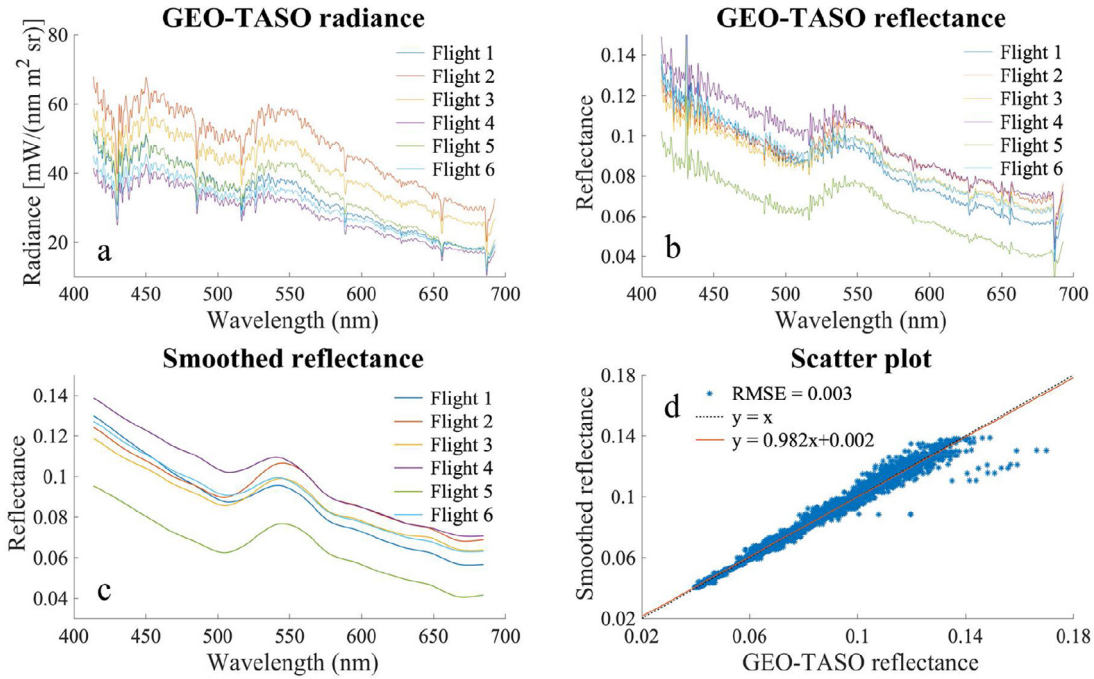
Fig. 6. (a) the radiances measured by GEO-TASO for one selected pixel from 6 flights. (b) and (c) are same as (a) but for calculated reflectance and smoothed reflectance, respectively. (d) the scatterplot of GEO-TASO reflectance versus smoothed reflectance, in which the root mean square error (RMSE) is about 0.003 and the linear fit line is y = 0.982x + 0.002. See Section 3.2 for details.

square error (RMSE) is about 0.003 with the linear fit line y = 0.982x + 0.002. Afterward, we select these bands corresponding to the absolute error of reflectance smaller than 0.001 for the next band selection. Therefore, these gas absorption bands at which the outlier data in Fig. 6 were measured are indeed not used in the retrieval.

- Step 5: Based on the degree of freedom for signal (DFS) analysis and considering the spectral shape, the sequential forward selection (SFS) method is used to common bands selection [35].

Following the steps above, 20 spectral bands are selected as common bands for the retrieval in this study, including the wavelengths in 418.09, 442.11, 468.93, 491.01, 501.63, 514.50, 520.93, 524.29, 529.88, 536.87, 550.02, 556.74, 565.69, 585.01, 608.25, 622.25, 642.98, 662.88, 672.69 and 681.38 nm. Here, those 20 wavelength bands can cover more than 85% total information content of 1000 bands of GEO-TASO measurements for the retrieval of aerosol and surface parameters [35].

To obtain the prior knowledge of surface spectra from GEO-TASO measurements over land, the k-means clustering method based on the 20 common wavelengths is further used to classify the surface types in the selected region. Specifically, to do the clustering of the study area in Fig. 4(a−f), the Rayleigh correction has been firstly made, and then the mean approximated surface reflectance of 6 flights could be obtained for each pixel. With the mean approximated surface reflectance of each pixel in 6 flights, we carry out the k-means clustering and divide the study area into 3 surface types approximately, as shown in Fig. 4(g). Here, the number of clusters is mainly decided based on the visual examination of the images since the study area is rather small. The eventual application of this algorithm to TEMPO application may not require the image-based classification of surface types to obtain reprehensive PCs for each pixel, because for each pixel the TEMPO can provide multiple observations each day. A brief discussion of this issue is provided in Section 5.

Table 4
List of AERONET sites where the aerosol model is derived.

AERONET site	Years	Data level
Yonsei_University	2011–2015	Level 2.0
Seoul_SNU	2000–2003, 2012, 2013, 2015	Level 2.0

Fig. 4(h−j) further provides the minimum, maximum, mean and standard deviation results of all approximated surface reflectance in 20 common wavelength bands for 3 different surface types in 6 flights. Correspondingly, Fig. 4(k−m) illustrates the extracted prior knowledge to constrain the retrieval of surface parameters, which contain the lower/upper bounds, the mean value vector and the standard deviation of weighting coefficients.

3.3. AERONET data processing for aerosol model

The long-term aerosol inversion data from 2 AERONET Yonsei_University and Seoul_SNU sites (Table 4) are used to derive the suitable aerosol optical model for the retrieval algorithm. From a total of 12 accumulative years of observation at these two AEROENT sites (~ 6 yr for each site, Table 4), only ~2102 data points have the AOD value larger than 0.4 at 440 nm (the requirement for the valid AERONET level-2 inversion product), and they are subsequently analyzed. Results shown in Fig. 7(a) are the frequency distribution of the fine-mode fraction of volume (fmf$_V$). In Fig. 7(b−f), shown as a function fmf$_V$, are the mean and standard deviation values of Ångström exponent (AE), the real and imaginary part (m_r, m_i) of complex refractive index, and the effective radius (r_{eff}) and effective variance (v_{eff}), respectively. Our retrieval algorithm assumes a bi-lognormal size distribution for aerosols, with one mode representing fine and another representing coarse. Therefore, the volumetric fine-mode fraction (FMF$_V$) can be derived from the retrieved fine- and coarse-mode aerosol volume concentration (V_0^f, V_0^c) [75,76]. Though the parameters of fine- and coarse-mode aerosol size distribution are prescribed (and static) in our
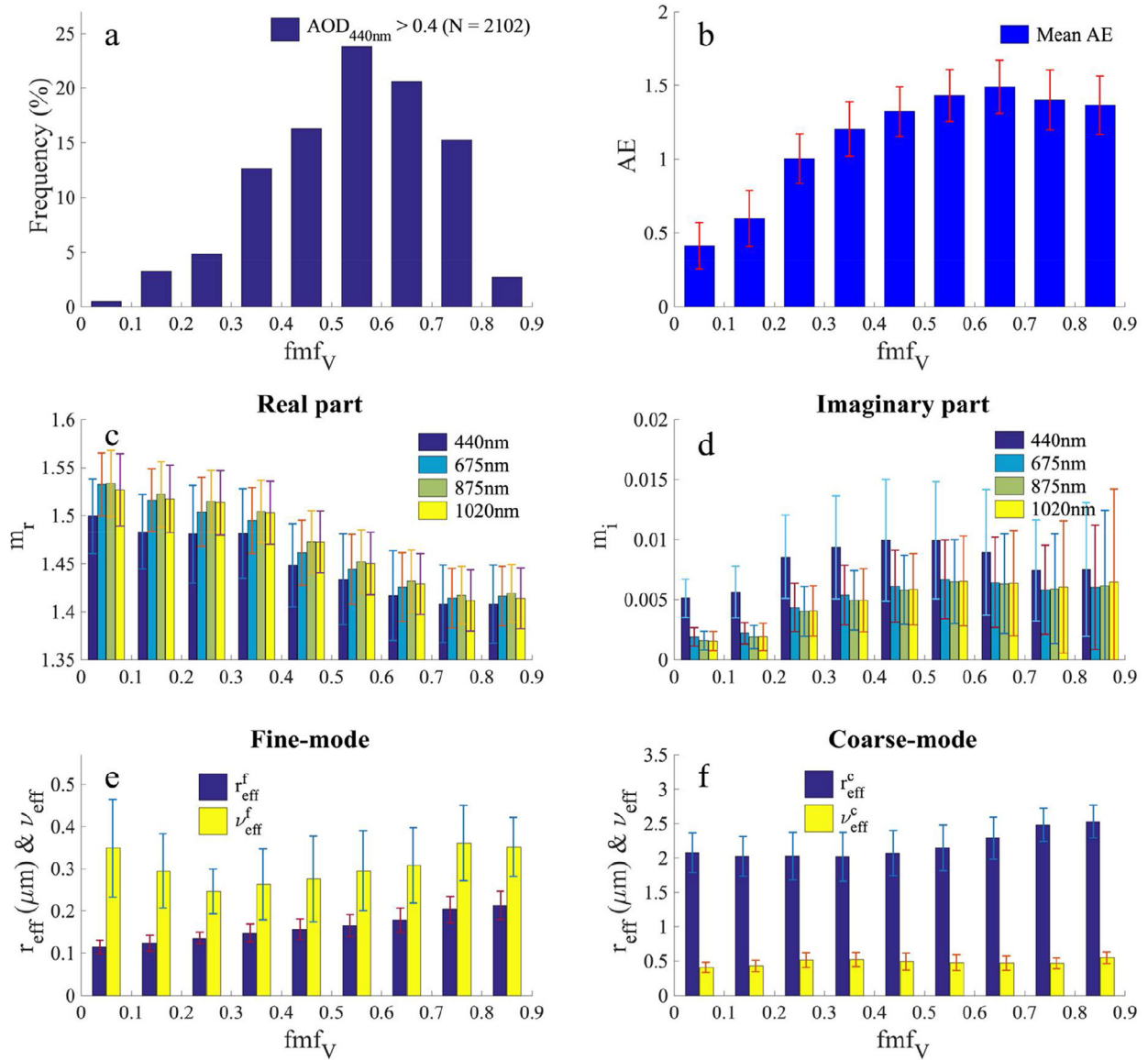
Fig. 7. Frequency figure and bar charts of AERONET optical and microphysical parameters inverted for conditions of AOD at 440 nm, AOD_{440nm} greater than 0.4. The total number of data points for the analysis is $N = 2012$. (a) AOD frequency as a function of volumetric fine-mode fraction fmf_V; (b) the bar chart of Angstrom exponent (AE) as a function of the fmf_V; (c–d): the bar chart of real part (m_r) and imaginary part (m_i) of the complex refractive index at 440, 675, 875 and 1020 nm, respectively, as a function of fmf_V. (e–f): the bar chart of the effective radius (r_{eff}) and effective variance (v_{eff}) as function of the fmf_V, respectively. For (b–f), the standard deviations are shown as the error bar centered at the top of each bar for their corresponding mean values.

retrieval algorithm, the aerosol size distribution model is still dynamic because FMF_V is adjusted (e.g., indirectly retrieved) through direct retrieval of V_0^f and V_0^c.

Fig. 7(a) shows that fmf_V values are mainly distributed in the range of [0.3, 0.8], and the accumulative frequency in this range is more than 88%. For AE, the mean value of AE firstly increases significantly from 0.41 to 1.49 with the increasing fmf_V from 0.1 to 0.7. After that, AE value decreases to 1.36 with fmf_V reaching 0.9. At all wavelengths (440, 675, 875 and 1020 nm), the real part of refractive index, m_r, decreases as fmf_V increasing from 0.1 to 0.9; while the imaginary part, m_i, first increases with fmf_V and reaches nearly a constant after $fmf_V = 0.6$ (except at 440 nm where m_i shows slight decrease). In contrast, the values of effective radius (r_{eff}) and effective variance (v_{eff}) show nearly no changes as a function of fmf_V for both fine-mode and coarse-mode aerosols, which supports the assumption of a bi-lognormal size distribution for aerosols in the retrieval algorithm.

To determine the parameters of complex refractive index in the aerosol model, the mean values of m_r and m_i with $fmf_V > 0.8$ have been extracted as the results of the fine-mode aerosols at each wavelength, while the mean values of m_r and m_i with $fmf_V < 0.2$ are regarded as the results of the coarse-mode aerosols correspondingly. Based on the extracted m_r and m_i at 440, 675, 875 and 1020 nm, the coefficients $m_r(\lambda_0)$, b_r, $m_i(\lambda_0)$ and b_i can be fitted by Eq. (8) with the least squares fitting method for the reference wavelength $\lambda_0 = 550$ nm. Furthermore, to determine the parameters of particle size distribution in aerosol model, the mean values of r_{eff} and v_{eff} in fine-mode dominated (Fig. 7e) and coarse-mode dominated (Fig. 7f) are obtained respectively as the effective radius and effective variance for the fine-mode and coarse-mode aerosols. Table 5 lists the aerosol parameters derived from AERONET for the retrieval algorithm, in which the mean fine-mode and coarse-mode aerosol volume concentrations ($\overline{V_0}$) values are used for the first guess in inversion with the prior values $\overline{V_0^f} = 0.052$ and $\overline{V_0^c} = 0.061$,
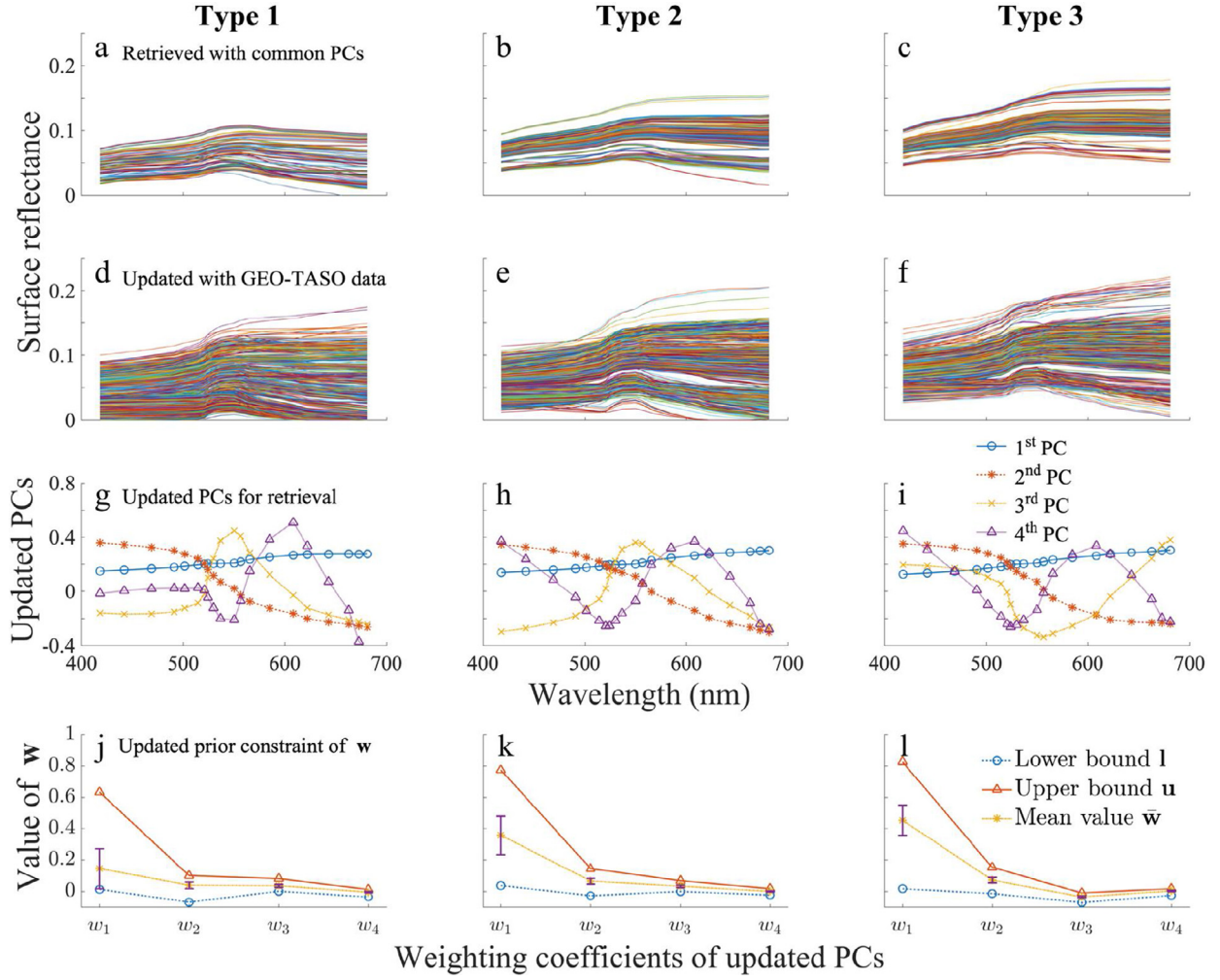
Fig. 8. The initial values of retrieved surface reflectance, the adjusted surface reflectance, and the updated PCs as a function of 20 common spectral bands for 3 different surface types respectively, as well as the corresponding prior constraints of weighting coefficients respectively. $(a-c)$: the initial surface reflectance reconstructed by common PCs; $(d-f)$: the adjusted surface reflectance updated with the GEO-TASO data; $(g-i)$: the updated PCs based on the results of adjusted surface reflectance; $(j-l)$: the prior constraint of weighting coefficients, including the lower bound, upper bound, mean and standard deviation.

Table 5
Aerosol model parameters extracted from AERONET ($\lambda_0 = 550$ nm).

Parameters	Fine-mode	Coarse-mode
$m_r(\lambda_0)$	1.412	1.506
b_r	-0.0065	-0.0261
$m_i(\lambda_0)$	0.0069	0.0037
b_i	0.1984	1.602
r_{eff} (μm)	0.160	2.185
v_{eff}	0.305	0.483
$\overline{V_0}$ (μm$^3\mu$m^{-2})	0.052	0.061

respectively, and the corresponding prior errors are both set to 80% [75,76].

4. Results and validations

4.1. OE retrieval results in the study area

Fig. 8 illustrates the retrieved surface reflectance at the initial step, the adjusted surface reflectance, as well as the updated PCs (for 20 selected bands) and corresponding prior constraint of weighting coefficients for 3 different surface types. We carried out the initial retrieval process with the common PCs (flag = 1)

based on surface types classified in Fig. 4(g), and consequently, the retrieved surface reflectance for these 3 different surface types were obtained and are shown in panel Fig. 8(a–c). After that, the adjusted surface reflectance vector $\boldsymbol{\rho}_{adju}^s$ with the real GEO-TASO data is calculated from Eq. (27) and are plotted in panel (d–f) of Fig. 8 for three surface types. Subsequently, the updated PCs can be extracted from the dataset of $\boldsymbol{\rho}_{adju}^s$, and are shown in panel (g–i) of Fig. 8. Correspondingly, the prior knowledge or constraint parameters of surface reflectance are also obtained, as shown in Fig. 8(j–l).

The results for the fitting with GEO-TASO hyperspectral measurements are compared in Fig. 9 between the two cases, one with the common and another with updated PCs, over the same pixel of the surface (clustering) type 1. To quantitatively evaluate the goodness of OE fitting, we compute the sum of absolute value of residuals, $\Sigma\Delta\rho$, and the values of the cost function. By comparing these OE fitting results from using common with using updated PCs, the sum of absolute value of residuals decrease form 0.064, 0.277, 0.061, 0.257, 0.057 and 0.274 (Fig. 9a–f) to 0.052, 0.049, 0.035, 0.068, 0.036 and 0.038 (Fig. 9g–l) for 1st to 6th flights, respectively. This comparison suggests that the fitting has been significantly improved with the updated PCs. Likewise, for the surface reflectance results retrieved respectively by using common (Fig. 9m) and updated PCs (Fig. 9n), their spectral shape and values have some mi-
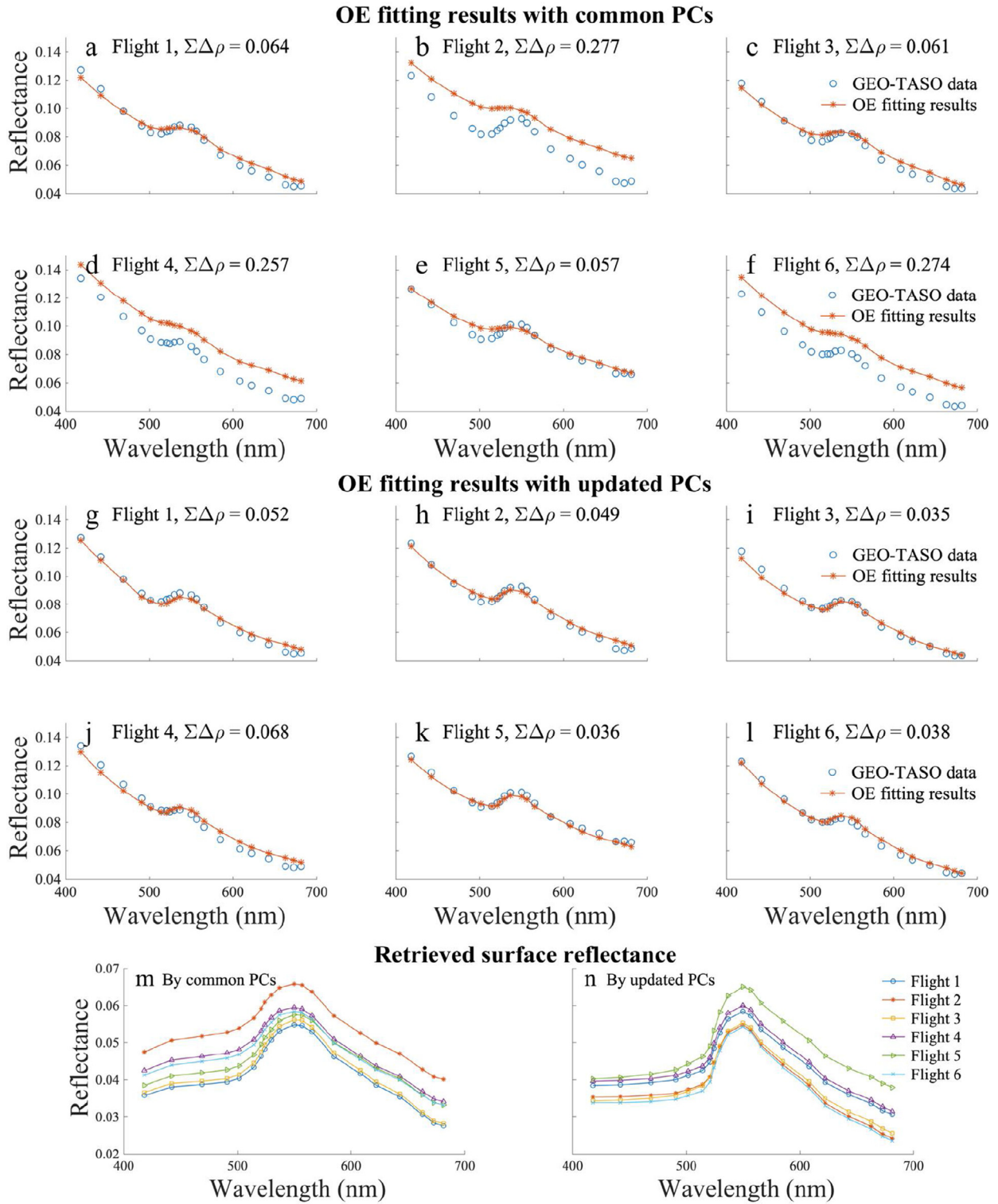
Fig. 9. Comparison of the GEO-TASO measured reflectance with the fitting results using either common or updated PCs for a pixel belonging to the surface type 1. The reflectance is described with 20 bands selected by this study. (a) − (f) represent the fitting results with common PCs for flight 1 to flight 6, respectively, while panels (g) − (l) represent the fitting results with updated PCs from flight 1 to flight 6, respectively. (m) − (n) show the summary of retrieved surface reflectance for each of 6 flights by common and updated PCs, respectively. The symbol $\Sigma\Delta\rho$ denotes the sum of absolute values of the residuals from the OE fitting.

nor changes and adjustments with different PCs, which further led to the changes of the OE fitting and retrieved aerosol results.

To further illustrate the importance of updated PCs for the OE fitting comprehensively for different surface type, we show in Fig. 10 the fitting results of GEO-TASO and retrieved surface re-

flectance for another pixel that belongs to the surface type cluster 2. For the fitting results, the sum of absolute value of residuals decreases significantly form 0.349, 0.123, 0.073, 0.051, 0.198 (Fig 10a–f) and 0.071 to 0.051, 0.024, 0.036, 0.035, 0.071 and 0.028 (Fig. 10g–1) for the sequence of 6 flights, respectively. Similar to the results
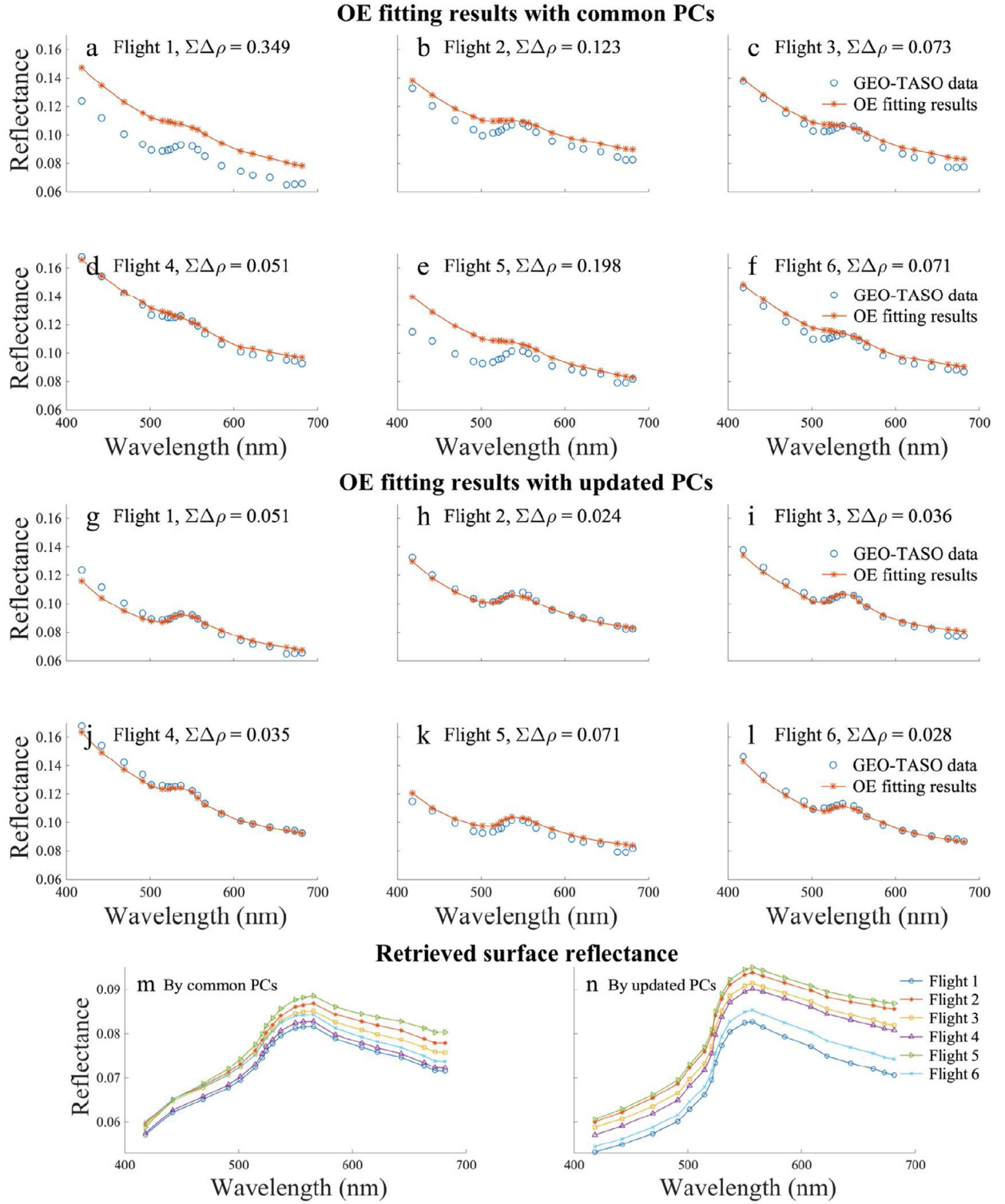
Fig. 10. Similar as Fig. 9 but for another a different pixel belong to the surface type 2.

in Fig. 9, the spectral shape and values of retrieved surface reflectance also have some changes for 2 different PCs (Fig. 10m–n).

The bar chart in Fig. 11 shows the normalized cost function $J(\mathbf{x})$ as a function of the iteration sequence for 2 different pixel cases in 6 flights, in which the cost function values of the 6th iteration in panels (a–d) correspond to the fitting cases shown in Fig. 9 (a–f), (g–l), Fig. 10 (a–f) and (g–l), respectively. In general, regardless of the inversion with common or updated PCs, the optimized iteration achieves the convergence after the 6th iteration in most cases. That is to say, the cost function no longer changes discernably and approaches a constant value after the 6th iteration in most cases. Therefore, we only consider the iteration sequence from the initial setting to the 6th iteration in this study.
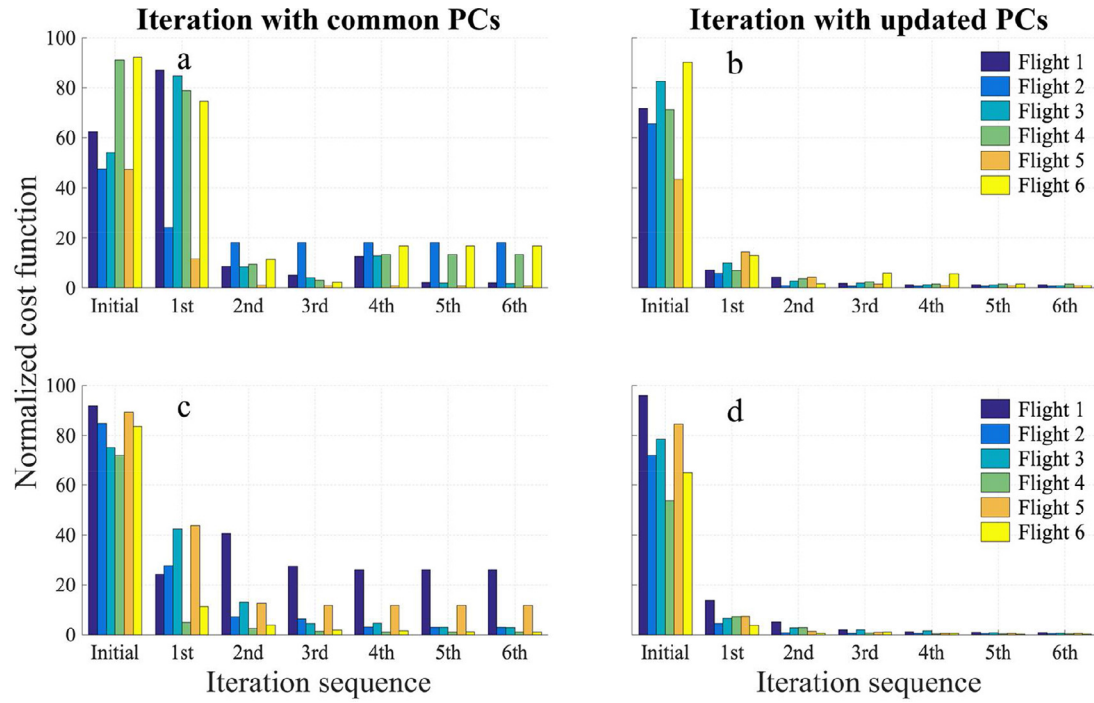
Fig. 11. Bar charts of the normalized cost function values as a function of the iteration sequence for 2 different pixels in 6 flights. Results from using common and updated PCs in the optimization are shown in the left and right column, respectively. The cost function results of the 6th iteration in (a−d) correspond to the fitting cases in Fig. 9(a−f), Fig. 9(g−l), Fig. 10(a−f) and Fig. 10(g−l), respectively.

While Fig. 9 to Fig. 11 are examples of the fitting results, they reveal the following interesting details that we find are common for the retrievals of different pixels in this study.

(1) With the common PCs, the GEO-TASO hyperspectral measurement cannot be well fitted in part of the spectral wavelength bands, especially in the spectral reflectance valley around 500 nm and the spectral peak around 550 nm.

(2) With the updated PCs, the GEO-TASO data can be well fitted in most cases, and the spectral details can be covered and reconstructed finely. Correspondingly, the values of cost function are much smaller than those results with the common PCs when the iteration converges.

(3) The spectral shapes and values of retrieved surface reflectance have some difference between the reconstructed results by common and updated PCs, which further influences the fitting effect of GEO-TASO data and the convergent cost functions. Obviously, some spectral details cannot be covered by common PCs, and thus the fitting results of GEO-TASO with updated PCs are much better than those results with common PCs.

(4) The regularization parameter γ in the Eq. (1) play an important role in the optimized inversion, which is used to adjust the cost function's contribution of measurement and a priori. If the value of γ is not suitable ($\gamma = 7.5$ in this study), the value of cost function $J(\mathbf{x})$ cannot be effectively reduced along the iteration sequence, which further results in outliers in the validations of AOD and AOD$_f$ in Section 4.2. In other words, OE optimization has the risk of hitting a local rather than global minimum. Therefore, the iteration should be terminated when the $J(\mathbf{x})$ reach the minimum along the iterated sequence by additional intervention in order to get the better spectral fittings. Furthermore, in each interaction, the hyperspectral surface reflectance values are updated, so that the retrieval here is not solely to minimize the cost function, but also to ensure the constraints of physics, i.e.,

the convergence of surface reflectance and aerosol retrievals at the end of the retrieval.

Based on the results of retrieved state vector, the spectral AOD values are derived as the part of the retrieval out as well. Fig. 12 shows the comparison of AOD at 550 nm retrieved with common and updated PCs in the study area for flight 1 to flight 6. From the map results, the spatial distribution of AOD retrieved with common PCs appears to be smeared by the surface types; the AOD results over the vegetated surface are much lower than those over bright surface type (Fig. 12a−f). By the contrast, the spatial distribution of retrieved AOD with updated PCs is much smoother and is less affected by surface distribution (Fig. 12g−l), which is thought to be more reasonable.

4.2. Validation at 7 aeronet sites

The retrieved spectral AOD results are compared with the counterparts of AEROENT at 7 sites (that are listed in Table 3) during May 17 to 18 in 2016 (https://espo.nasa.gov/korus-aq). In the comparison, for each site, the retrieved spectral AODs are derived by averaging the GEO-TASO retrievals for each pixel in 1km × 1km area centered over that site. For AERONET, the mean of AOD values within the time window of 15 min before and after the GEO-TASO observation time are used for the comparison. This method yields a total of 26 pairs between GEO-TASO and AERONET AODs that are spatially and temporally matched. For AERONET, its AOD values at 550 nm are interpolated by the Ångström exponent equation [90]. For each AOD pair, the corresponding values of fine-mode AOD (AOD$_f$), coarse-mode AOD (AOD$_c$) and fine-mode fraction (FMF) are also obtained from GEO-TASO retrieval and from the AEROENT's spectral deconvolution algorithm (SDA) products. It is worth noting that SDA reports AOD at 500 nm [91], and therefore this is also the wavelength we extract GEO-TASO retrievals for the comparison with SDA products. In addition, the Ångström exponent (AE) is calculated by the AODs at 440 nm and 675 nm from AERONET to validate the counterparts retrieved from the GEO-TASO.
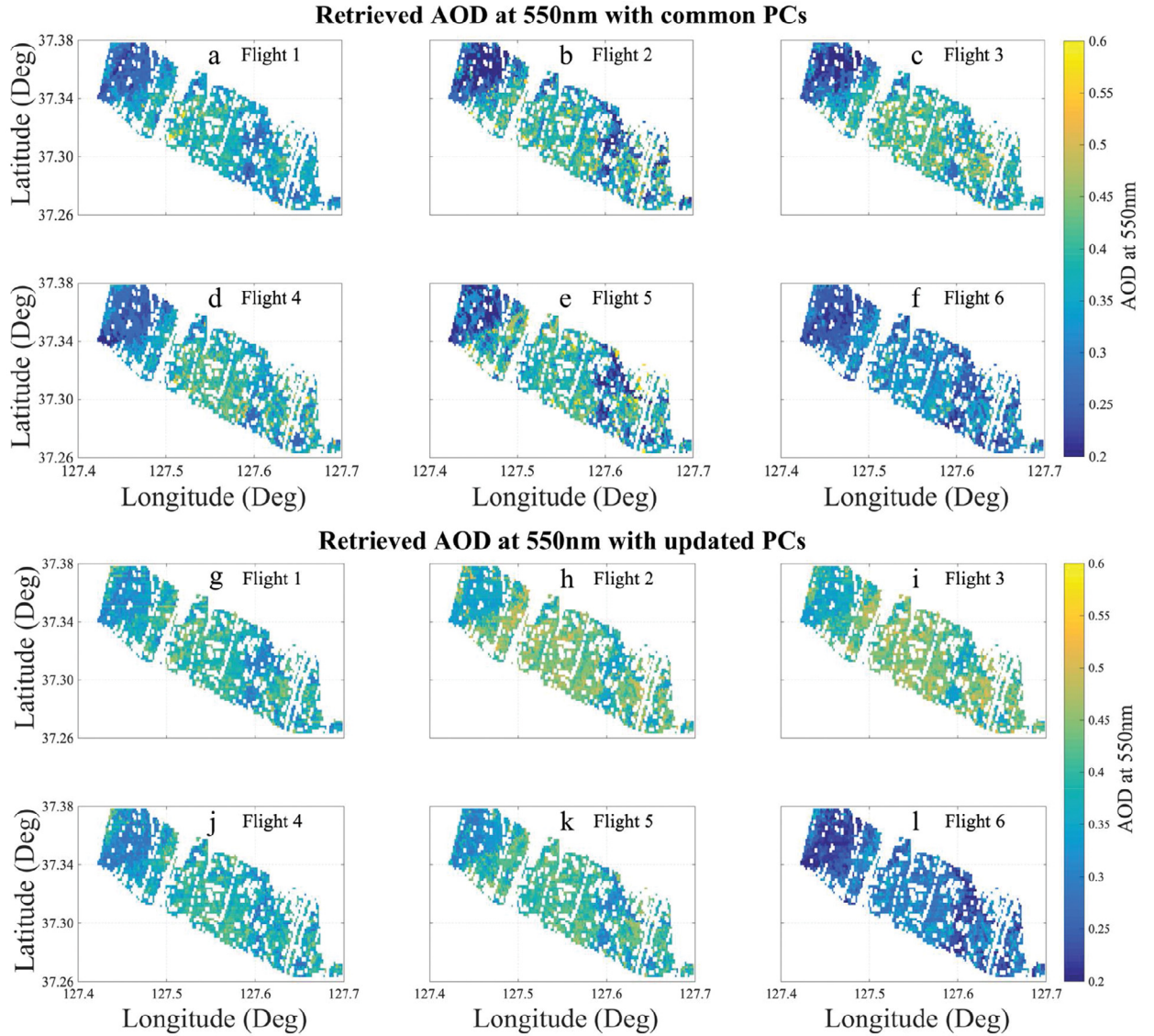
Fig. 12. Spatial distribution of AODs retrieved at 550 nm from GEO-TASO measurements taken in 6 different flights. (a−f): the results retrieved with common PCs for the 6 flights; (g−l): the results retrieved with updated PCs.

Fig. 13 shows the scatterplot of retrieved AODs by common and updated PCs versus the AERONET AODs in 440, 500, 550, and 675 nm, respectively. As a reference of expectation, we also show in Fig. 13 the uncertainty envelope of AOD retrieved from MODIS Dark-Target Algorithm [92,93]. By definition, ~75% of the AOD retrievals would fall into this expected uncertainty envelope. Regardless of using common or updated PCs in the retrieval, Fig. 13 shows that nearly 92% AODs values from GEO-TASO fall within this uncertainty envelope, suggesting that our retrievals have a slightly reduced uncertainty (as compared to MODIS AOD global retrieval statistics). Comparison of GEO-TASO AODs (y) retrieved from using common PCs with AERONET AODs (x) reveals the linear least squared regression lines of $y = 0.72x + 0.11$, $y = 0.72x + 0.10$, $y = 0.75x + 0.08$ and $y = 0.75x + 0.06$ and the corresponding Pearson correlation coefficients (R^2) in range of $0.53-0.54$ for 440, 550, 550, 675 nm, respectively. In contrast, the counterparts for evaluating GEO-TASO AODs retrieved from using updated PCs are $y = 1.05x + 0.007$, $y = 1.04x + 0.015$, $y = 1.08x + 0.008$ and $y = 1.05x + 0.013$ with R^2 in range of $0.75-0.77$. This contrast suggests that the approach of using updated PCs much improved

GEO-TASO retrievals of AODs, yielding a 20% increase in R^2 values (as well as better fitting equations with AERONET AODs).

While the Ordinary Least Square (OLS) fitting, as used in Fig. 13, is commonly used in the literature, it might not be best method for this study to fit the AODs retrievals with the AERONET AODs because the homoscedasticity (a requirement for using OLS [9]) in the retrieval uncertainty is difficult to quantify with limited number of data samples here. Nevertheless, the statistics from OLS in Fig. 13 can be viewed as the first-order assessment of retrieved AODs. To further validate the retrieval, we compute mean absolute error (MAE) and root mean squared error (RMSE). For the validations of AODs at 440, 550, 550, and 675 nm, the MAE values decrease from 0.051, 0.041, 0.034 and 0.025 for the retrievals with common PCs (left panes in Fig. 13) to 0.033, 0.032, 0.031 and 0.024 for the retrievals with updated PCs (right panels in Fig. 13) respectively; meanwhile, the RMSE values decrease from 0.067, 0.056, 0.047 and 0.034 to 0.049, 0.045, 0.042 and 0.032 (Fig. 13). From the results of these statistic metrics, we can find that retrieval accuracies of spectral AODs is significantly improved by the updated PCs rather than by the common PCs.
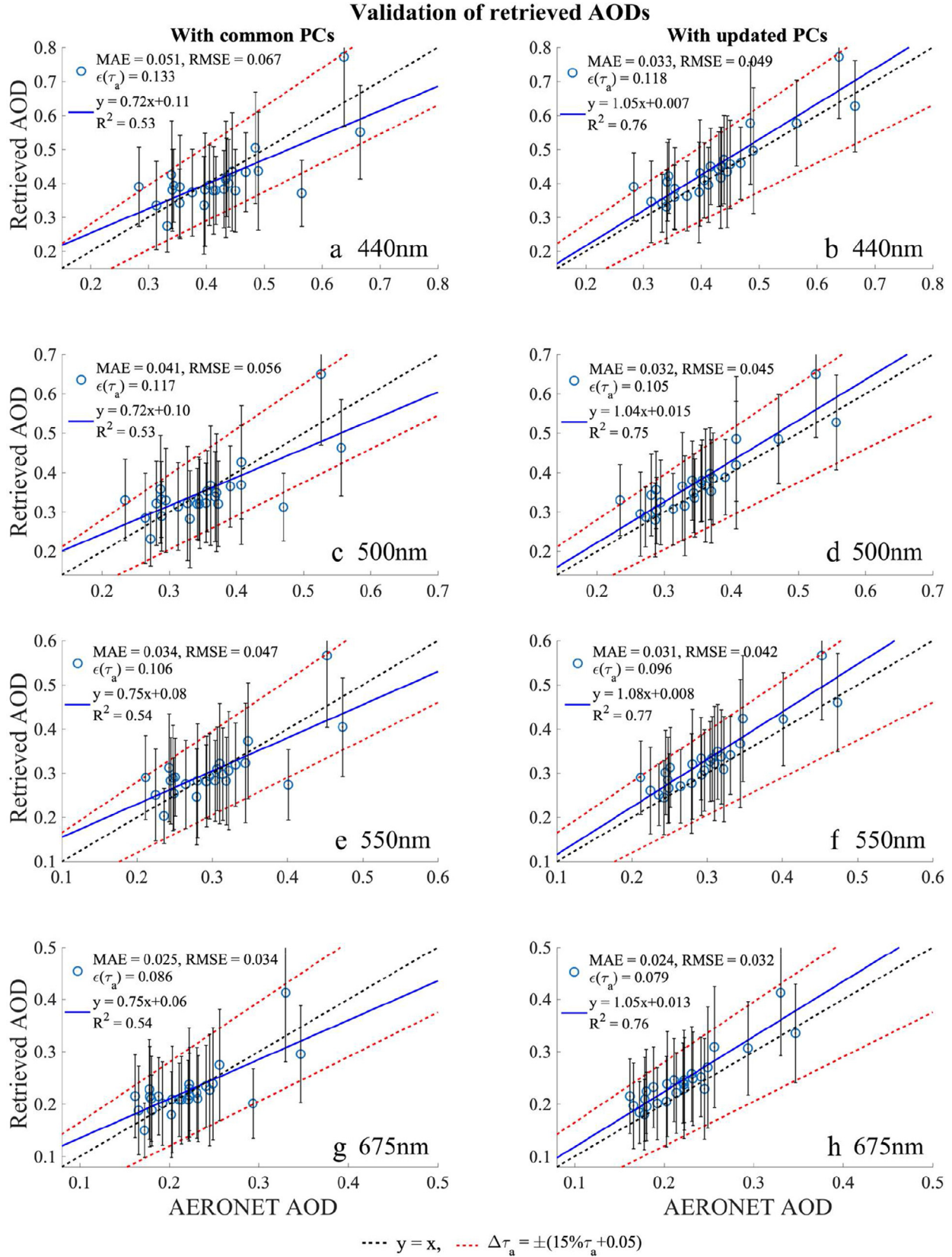
Fig. 13. Inter-comparison of GEO-TASO AODs and the AERONET AODs at 7 AERONET sites. The GEO-TASO AODs are retrieved with common PCs and updated PCs, and are shown the left and right column, respectively. The comparison is conducted separately at 440, 500, 550, and 675 nm, and is shown in the 1st, 2nd, 3rd, and 4th row, respectively. In each panel, the retrieval errors estimated at the pixel level by OE are shown as error bars in the vertical, and their mean value $\epsilon(\tau_a)$, is shown on the top left. For details, see the text in Section 4.2.
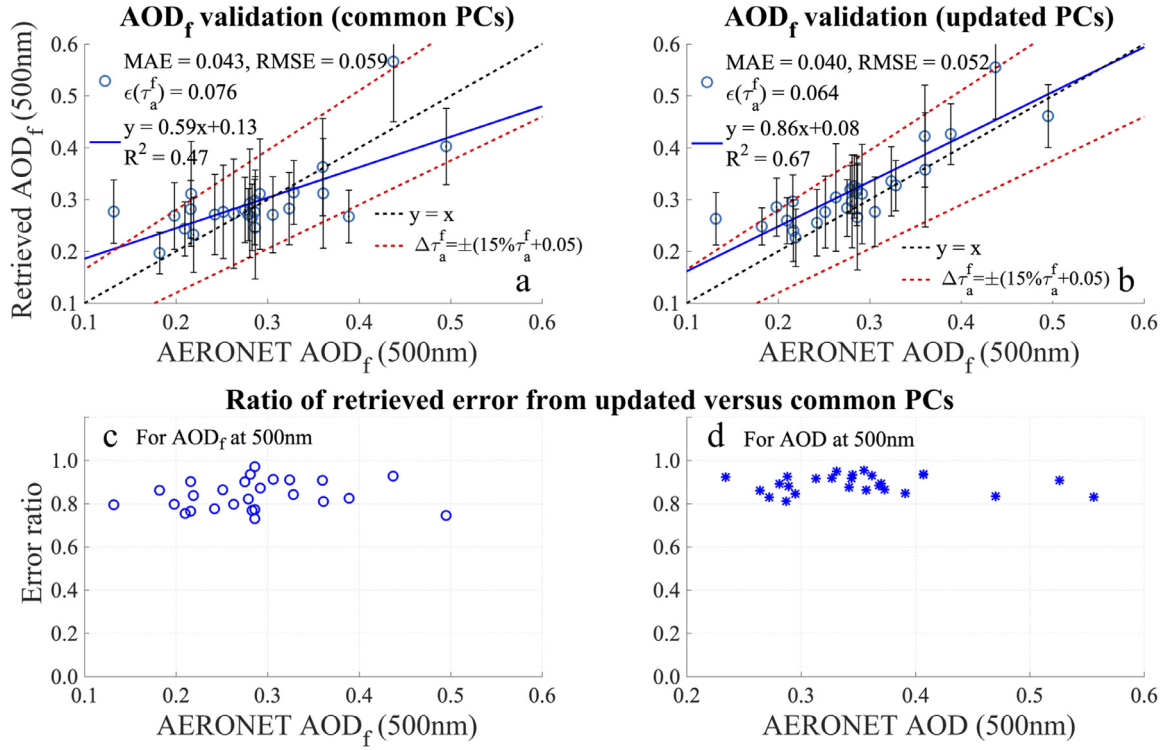
Fig. 14. (a): similar as Fig. 13a but for the inter-comparison of GEO-TASO fine-mode AOD (AOD$_f$) and the AERONET counterparts derived from AERONET SDA product at 500 nm at 7 AERONET sites. (b): similar as (a) but for the GEO-TASO AOD$_f$ retrieved with updated PCs. In both (a) and (b), $\epsilon(\tau_a^f)$ represents the mean absolute error of retrieved AOD$_f$. (c): the ratio of AOD$_f$ error (y-axis) between retrievals using updated PCs and common PCs at the pixel level as a function of corresponding AERONET AOD$_f$ at 7 AERONET sites (x-axis). (d): similar as (c) but for AOD.

One advantage of using OE is its ability to compute posterior error at the pixel level. The uncertainty estimates of retrieved AODs are equal to the absolute posterior errors (ϵ_{τ_a}) calculated by the linear combination of $\epsilon_{\tau_a^f}$ and $\epsilon_{\tau_a^c}$ with the FMF as listed in Eq. (13), after considering 80% prior (relative) errors for the aerosol parameters, V_0^f and V_0^c, in the OE inversion. These posterior errors at the pixel level are shown as the vertical bars for each data points in Fig. 13. We found that the mean absolute errors (and the mean relative errors) have a slight decrease from 0.133 (33%), 0.117 (34%), 0.106 (35%) and 0.086 (39%) in the retrievals of using common PCs to 0.118 (31%), 0.105 (32%), 0.096 (34%) and 0.079 (37%) in the retrievals of using updated PCs, for the wavelength of 440, 500, 550, 675 nm, respectively (Fig. 13).

To further evaluate the retrieval, Fig. 14 compares AOD$_f$ by common and updated PCs respectively against AERONET SDA products at 500 nm, and shows the ratio of retrieved error from updated PCs verses common PCs respectively for AOD$_f$ and AOD. The scatterplots for the validation of AOD$_c$, FMF and AE are not shown, because these parameters have a much smaller variation in our case studies that are dominated by fine-mode aerosols. In the OE inversion, the mean absolute errors (mean relative errors) for AOD$_f$, and AOD$_c$ have slightly decreased from 0.076 (26%) and 0.041 (79%) to 0.064 (22%) and 0.038 (72%) at 500 nm for the inversion with common and updated PCs, respectively. In general, because values of AOD$_c$ are all smaller than 0.1 at 500 nm in this study, the relative errors of AOD$_c$ at the pixel level are much larger than those results of AOD$_f$. That is to say, in the visible spectrum, the coarse-mode AODs are more difficult to be retrieved than fine-mode AODs from measurements in the atmosphere dominated by fine-mode aerosols in the VIS. Furthermore, both FMF and AE are the indirectly-retrieved parameters that are inferred from either the GEO-TASO retrievals or AERONET measurement of AODs. Hence, using them to assess the retrievals of spectral AODs some-

times can lead to the confusion and difficulty in understanding or interpretation of the results. This is understood because retrievals of AOD may have large uncertainty at each wavelength, but still can yield a perfect FMF or AE retrieval to match AERONET's counterpart. In essence, the variations of FMF and AE values may not be directly driven by the variation of AOD (that is the main objective of the AOD retrieval); for example, the same value of FMF or AE can correspond to a wide range of paired AOD values at a different combination of two wavelengths. To avoid the aforementioned non-ideality in assessing FMF and AE retrieval at the pixel, we only discuss the their RMSE and MAE values against the AEROENT counterpart in below, while the linear regression and scattering plots for FMF and AE are not pursued further.

For the AOD$_f$ retrieved by common and updated PCs, the values of MAE and RMSE decrease from 0.043 and 0.059 to 0.04 and 0.052, and the linear least squared regression lines are from $y = 0.59x + 0.13$ to $y = 0.86x + 0.08$, with the correlation coefficient R^2 increasing from 0.47 to 0.67, respectively (Fig. 14a−b). Besides, the ratios of retrieved errors from updated verses common PCs for AOD$_f$ and AOD at 500 nm are all smaller than 1 and in the value range of 0.7−0.98, which yield an average of ~10% reduction of error by the updated PCs. For the validations of retrieved AOD$_c$ (figure not shown), the values of MAE and RMSE decrease from 0.024 and 0.02 to 0.021 and 0.017 respectively after the updated PCs are used. For FMF, the corresponding values of MAE and RMSE are both 0.055 and 0.072 regardless of using common or updated PCs. Similarly, for AE, the MAE and RMSE values show no change with the updated PCs, and the corresponding values are 0.12 and 0.13 respectively. Overall, the retrieval accuracies of AOD$_f$ and AOD$_c$ are improved by the updated PCs. In contrast, the retrieval accuracies of FMF and AE showed no improvement with updated PCs. Again, this is understood because the variations of FMF and AE are not necessarily associated with the change of AOD while the AOD

is the aerosol parameter that the GEO-TASO instrument is most sensitivity to [35].

For the reason that there are only few GEO-TASO measurements that were carried out with the consideration to mimic observations from TEMPO, the cases we had in KORUS-AQ is among the first of its kind for us to test our algorithm from a view point of TEMPO. Therefore, more case studies are needed to further evaluate the application of the algorithm developed here to TEMPO that carries an enhanced version of GEO-TASO instrument.

5. Summary and conclusion

As the third part of a series of studies to retrieve aerosol properties form the hyperspectral radiance measurements by new instrument GEO-TASO and future geostationary satellite TEMPO, we developed and conducted the PC-based OE algorithm to retrieve spectral AODs from the GEO-TASO data collected in the KORUS-AQ field campaign. We can summarize our work as follows.

(1) With the combination of the band selection method presented in our previous work [35] and the real data from GEO-TASO, those best bands containing rich information content and less noise for the retrieval of aerosols are selected to reduce the dimensionality of hyperspectral measurements and to improve the computational efficiency of the retrieval algorithm. Using this approach together with the consideration of the necessity to cover spectral details of surface reflectance, we select 20 bands that are used throughout this study for the inversion of aerosol properties from GEO-TASO.

(2) By applying the Rayleigh correction from GEO-TASO data, the datasets of approximated (or the first guess of) surface reflectance are obtained, and the surface types can be classified by the k-mean clustering method in the study region. Based on this initial classification, prior constraints are also obtained for surface reflectance parameters, such as the lower and upper boundary as well as the mean and standard deviation for the weighting coefficients, which are very important for the initial inversion.

(3) For the aerosol models, the fine- and coarse- mode parameters of aerosol size distribution as well as the complex refractive index are obtained from the statistical analysis of the long-term AERONET level-2 inversion data of aerosol properties. While the climatology of aerosol parameters, as obtained by AERONET in the study region, is used here, it is not the essential component in the inversion. Even without the local AERONET data for aerosol model, our inversion can still work with the priori aerosol model from the literature, such as the global aerosol properties used in MODIS [94] and other aerosol models based on cluster analysis [95,96].

(4) For the initial inversion, the common PCs of surface spectra are extracted from typical surface spectral dataset in USGS and ASTER spectral libraries. However, the datasets in these libraries are limited and not able to cover all of the surface materials over the world. Correspondingly, the results of OE-based fitting with GEO-TASO data show that some spectral details at the pixel level are not able to be well reconstructed by the common PCs even when the cost function may be well reduced to the global minimum; this caveat renders unsatisfactory retrievals of spectral AODs and surface reflectances.

(5) Based on the retrieved surface reflectance by common PCs in initial inversion and the GEO-TASO data, the PCs can be updated for each surface type of clustering, and the corresponding constraints for the weighting coefficients can be

further obtained. With the updated PCs for the inversion, the OE fitting with GEO-TASO can lead to the improved retrievals of both spectral AODs and surface reflectance. When the optimized iteration gets convergence after 6 iterations in most cases, both the sum of absolute value of residuals and the value of cost function are decreased significantly in comparison with the counterparts in the retrievals by common PCs.

(6) The spectral AODs retrieved from GEO-TASO data are validated over 7 AERONET sites for the inversion by common and updated PCs in sequence, and quantitatively evaluated against the AERONET AODs in 440, 550, 550, 675 nm, respectively. The mean absolute errors (MSE) decrease from $0.025-0.051$ to $0.023-0.033$ with updated PCs, while the root mean squared errors (RMSE) decrease from $0.034-0.067$ to $0.032-0.049$. The comparison reveals that the updates PCs (and so surface reflectance) yield significant enhancement in AOD retrievals, improving the mean linear fitting equations from $y = 0.735x + 0.088$ to $y = 1.055x + 0.01$, and the Pearson correlation coefficient (R^2) from 0.54 to 0.76, respectively.

(7) The fine-mode AOD (AOD_f), coarse-mode (AOD_c), and fine-mode fraction (FMF) retrieved respectivley by common and updated PCs methods are validated against the SDA products of AERONET at 500 nm; also compared are the Ångström exponent (AE) calculated by the AODs at 440 nm and 675 nm. The validation results show that the retrieval accuracies of AOD_f and AOD_c are improved by the updated PCs, and the retrieval accuracies of FMF and AE exhibit no change with updated PCs. The values of MAE and RMSE are about 0.055 and 0.072 for the validation of FMF, and 0.12 and 0.13 for the validation of AE.

The PC-based OE approach for AOD retrieval, as shown by using GEO-TASO data collected during KORUS-AQ in this study, has important implications for TEMPO. TEMPO will provide hourly observation eight times per day for the North America. This frequent sampling can facilitate the principal component analysis at the pixel level, and hence, the surface type clustering analysis developed here as part of the aerosol algorithm for GEO-TASO may not be needed. Since the surface reflectance and its corresponding PCs vary in a much slower pace than the aerosol loading at the same location, dozens of cloud-free measurements by TEMPO over the same pixel within a week can be used as a combination to extract the prior constraints of surface reflectance as well as to further obtain the updated PCs for conducting the OE inversion. In this manner, although further studies are needed, the parameters of aerosol and surface reflectance may be retrieved simultaneously at each daytime hour from TEMPO for each of its cloud-free pixels over the North America.

Declaration of Competing Interest

The authors declare that they have no known competing financial interestsor personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Weizhen Hou: Methodology, Formal analysis, Writing - original draft. **Jun Wang:** Conceptualization, Funding acquisition, Supervision, Methodology, Formal analysis, Writing - original draft. **Xiaoguang Xu:** Software, Validation, Visualization. **Jeffrey S. Reid:** Writing - review & editing. **Scott J. Janz:** Data curation. **James W. Leitch:** Data curation.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jqsrt.2020.107161.

Reference

[1] Kokhanovsky AA. The modern aerosol retrieval algorithms based on the simultaneous measurements of the intensity and polarization of reflected solar light: a review. Frontiers in Environmental Science 2015;3.

[2] Zubko V, Kaufman YJ, Burg RI, Martins JV. Principal component analysis of remote sensing of aerosols over oceans. IEEE T Geosci Remote 2007;45:730–45.

[3] Kaufman YJ, Tanre D, Boucher O. A satellite view of aerosols in the climate system. Nature 2002;419:215–23.

[4] Mishchenko MI, Cairns B, Hansen JE, Travis LD, Burg R, Kaufman YJ, et al. Monitoring of aerosol forcing of climate from space: analysis of measurement requirements. J Quant Spectrosc Ra 2004;88:149–61.

[5] IPCC. Climate change 2014: synthesis report. Geneva, Switzerland: IPCC; 2014.

[6] Kokhanovsky AA, Breon FM, Cacciari A, Carboni E, Diner D, Di Nicolantonio W, et al. Aerosol remote sensing over land: a comparison of satellite retrievals using different algorithms and instruments. Atmos Res 2007;85:372–94.

[7] Mishchenko MI, Geogdzhayev IV, Cairns B, Carlson BE, Chowdhary J, Lacis AA, et al. Past, present, and future of global aerosol climatologies derived from satellite observations: a perspective. J Quant Spectrosc Ra 2007;106:325–47.

[8] Mishchenko MI, Liu L, Geogdzhayev IV, Travis LD, Cairns B, Lacis AA. Toward unified satellite climatology of aerosol properties. J Quant Spectrosc Ra 2010;111:540–52.

[9] Sayer AM, Govaerts Y, Kolmonen P, Lipponen A, Luffarelli M, Mielonen T, et al. A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing. Atmos Meas Tech 2020;13:373–404.

[10] Fioletov VE, McLinden CA, Krotkov N, Yang K, Loyola DG, Valks P, et al. Application of OMI, SCIAMACHY, and GOME-2 satellite SO2 retrievals for detection of large emission sources. J Geophys Res: Atmos 2013;118(11):418 399-11.

[11] Kim D, Lee H, Hong H, Choi W, Lee Y, Park J. Estimation of surface NO$_2$ volume mixing ratio in four metropolitan cities in korea using multiple regression models with OMI and AIRS Data. Remote Sens (Basel) 2017;9:627.

[12] Wang J, Aegerter C, Xu X, Szykman JJ. Potential application of VIIRS Day/Night Band for monitoring nighttime surface PM2.5 air quality from space. Atmos Environ 2016;124:55–63.

[13] Wang J, Xu X, Spurr R, Wang Y, Drury E. Improved algorithm for MODIS satellite retrievals of aerosol optical thickness over land in dusty atmosphere: implications for air quality monitoring in China. Remote Sens Environ 2010;114:2575–83.

[14] Dubovik O, Li Z, Mishchenko MI, Tanré D, Karol Y, Bojkov B, et al. Polarimetric remote sensing of atmospheric aerosols: instruments, methodologies, results, and perspectives. J Quant Spectrosc Ra 2019;224:474–511.

[15] Kokhanovsky AA, Davis AB, Cairns B, Dubovik O, Hasekamp OP, Sano I, et al. Space-based remote sensing of atmospheric aerosols: the multi-angle spectro-polarimetric frontier. Earth-sci Rev 2015;145:85–116.

[16] Li Z, Eck T, Zhang Y, Zhang Y, Li D, Li L, et al. Observations of residual submicron fine aerosol particles related to cloud and fog processing during a major pollution event in Beijing. Atmos Environ 2014;86:187–92.

[17] Wang J, Christopher SA, Brechtel F, Kim J, Schmid B, Redemann J, et al. Geostationary satellite retrievals of aerosol optical thickness during ACE-Asia. J Geophys Res: Atmos 2003;108.

[18] Wang J, Christopher SA, Reid JS, Maring H, Savoie D, Holben BN, et al. GOES 8 retrieval of dust aerosol optical thickness over the Atlantic Ocean during PRIDE. J Geophys Res: Atmos 2003;108.

[19] Wang J, Xu X, Ding S, Zeng J, Spurr R, Liu X, et al. A numerical testbed for remote sensing of aerosols, and its demonstration for evaluating retrieval synergy from a geostationary satellite constellation of GEO-CAPE and GOES-R. J Quant Spectrosc Ra 2014;146:510–28.

[20] Zhang Y, Li Z, Zhang Y, Hou W, Xu H, Chen C, et al. High temporal resolution aerosol retrieval using Geostationary Ocean Color Imager: application and initial validation. J Appl Remote Sens 2014;8:083612.

[21] Zhang H, Hoff RM, Kondragunta S, Laszlo I, Lyapustin A. Aerosol optical depth (AOD) retrieval using simultaneous GOES-East and GOES-West reflected radiances over the western United States. Atmos Meas Tech 2013;6:471–86.

[22] Bousserez N, Henze DK, Rooney B, Perkins A, Wecht KJ, Turner AJ, et al. Constraints on methane emissions in North America from future geostationary remote-sensing measurements. Atmos Chem Phys 2016;16:6175–90.

[23] Choi M, Kim J, Lee J, Kim M, Park Y-J, Jeong U, et al. GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during the DRAGON-NE Asia 2012 campaign. Atmos Meas Tech 2016;9:1377–98.

[24] Kim M, Kim J, Torres O, Ahn C, Kim W, Jeong U, et al. Optimal estimation-based algorithm to retrieve aerosol optical properties for GEMS measurements over Asia. Remote Sens (Basel) 2018;10:162.

[25] Bak J, Kim JH, Liu X, Chance K, Kim J. Evaluation of ozone profile and tropospheric ozone retrievals from GEMS and OMI spectra. Atmos Meas Tech 2013;6:239–49.

[26] Zoogman P, Liu X, Suleiman RM, Pennington WF, Flittner DE, Al-Saadi JA, et al. Tropospheric emissions: monitoring of pollution (TEMPO). J Quant Spectrosc Ra 2017;186:17–39.

[27] Chance K., Liu X., Suleiman R.M., Flittner D.E., Al-Saadi J., Janz S.J. Tropospheric emissions: monitoring of pollution (TEMPO). Proc SPIE 8866, Earth Observing Systems XVIII2013. p. 88660D.

[28] Noël S, Bramstedt K, Bovensmann H, Gerilowski K, Burrows JP, Standfuss C, et al. Quantification and mitigation of the impact of scene inhomogeneity on Sentinel-4 UVN UV-VIS retrievals. Atmos Meas Tech 2012;5:1319–31.

[29] Nowlan CR, Liu X, Leitch JW, Chance K, González Abad G, Liu C, et al. Nitrogen dioxide observations from the Geostationary Trace gas and Aerosol Sensor Optimization (GeoTASO) airborne instrument: retrieval algorithm and measurements during DISCOVER-AQ Texas 2013. Atmos Meas Tech 2016;9:2647–68.

[30] Leitch JW, Delker T, Good W, Ruppert L, Murcray F, Chance K, et al. The GeoTASO airborne spectrometer project. Proc. SPIE 9218. Earth Observ Syst XIX 2014:92181H1–19.

[31] Crumeyrolle S, Chen G, Ziemba L, Beyersdorf A, Thornhill L, Winstead E, et al. Factors that influence surface PM2.5 values inferred from satellite observations: perspective gained for the US Baltimore–Washington metropolitan area during DISCOVER-AQ. Atmos Chem Phys 2014;14:2139–53.

[32] Kim H, Zhang Q, Heo J. Influence of intense secondary aerosol formation and long-range transport on aerosol chemistry and properties in the Seoul Metropolitan Area during spring time: results from KORUS-AQ. Atmos Chem Phys 2018;18:7149–68.

[33] Lennartson EM, Wang J, Gu J, Castro Garcia L, Ge C, Gao M, et al. Diurnal variation of aerosol optical depth and PM2.5 in South Korea: a synthesis from AERONET, satellite (GOCI), KORUS-AQ observation, and the WRF-Chem model. Atmos Chem Phys 2018;18:15125–44.

[34] Hou W, Wang J, Xu X, Reid JS, Han D. An algorithm for hyperspectral remote sensing of aerosols: 1. Development of theoretical framework. J Quant Spectrosc Ra 2016;178:400–15.

[35] Hou W, Wang J, Xu X, Reid JS. An algorithm for hyperspectral remote sensing of aerosols: 2. Information content analysis for aerosol parameters and principal components of surface spectra. J Quant Spectrosc Ra 2017;192:14–29.

[36] Lenoble J, Remer LA, Tanré D. Aerosol remote sensing. Berlin Heidelberg: Springer; 2013.

[37] Li W, Stamnes K, Spurr R, Stamnes J. Simultaneous retrieval of aerosol and ocean properties by optimal estimation: seaWiFS case studies for the Santa Barbara Channel. Int J Remote Sens 2008;29:5689–98.

[38] Govaerts YM, Wagner S, Lattanzio A, Watts P. Joint retrieval of surface reflectance and aerosol optical depth from MSG/SEVIRI observations with an optimal estimation approach: 1. Theory. J Geophys Res 2010;115.

[39] Thomas GE, Poulsen CA, Curier RL, de Leeuw G, Marsh SH, Carboni E, et al. Comparison of AATSR and SEVIRI aerosol retrievals over the Northern Adriatic. Q J Roy Meteor Soc 2007;133:85–95.

[40] Wagner SC, Govaerts YM, Lattanzio A. Joint retrieval of surface reflectance and aerosol optical depth from MSG/SEVIRI observations with an optimal estimation approach: 2. Implementation and evaluation. J Geophys Res 2010;115.

[41] Govaerts Y, Luffarelli M. Joint retrieval of surface reflectance and aerosol properties with continuous variation of the state variables in the solution space – Part 1: theoretical concept. Atmos Meas Tech 2018;11:6589–603.

[42] Jeong U, Kim J, Ahn C, Torres O, Liu X, Bhartia PK, et al. An optimal-estimation-based aerosol retrieval algorithm using OMI near-UV observations. Atmos Chem Phys 2016;16:177–93.

[43] Wurl D, Grainger RG, McDonald AJ, Deshler T. Optimal estimation retrieval of aerosol microphysical properties from SAGE~II satellite observations in the volcanically unperturbed lower stratosphere. Atmos Chem Phys 2010;10:4295–317.

[44] Clarisse L, Coheur PF, Prata F, Hadji-Lazaro J, Hurtmans D, Clerbaux C. A unified approach to infrared aerosol remote sensing and type specification. Atmos Chem Phys 2013;13:2195–221.

[45] Thomas GE, Carboni E, Sayer AM, Poulsen CA, Siddans R, Grainger RG. Oxford-RAL Aerosol and Cloud (ORAC): aerosol retrievals from satellite radiometers. In: Kokhanovsky AA, de Leeuw G, editors. Satellite aerosol remote sensing over land. Chichester, UK: Springer; 2009. p. 193–225.

[46] Sayer AM, Thomas GE, Grainger RG, Carboni E, Poulsen C, Siddans R. Use of MODIS-derived surface reflectance data in the ORAC-AATSR aerosol retrieval algorithm: impact of differences between sensor spectral response functions. Remote Sens Environ 2012;116:177–88.

[47] Diner DJ, Hodos RA, Davis AB, Garay MJ, Martonchik JV, Sanghavi SV, et al. An optimization approach for aerosol retrievals using simulated MISR radiances. Atmos Res 2012;116:1–14.

[48] Dubovik O, Herman M, Holdak A, Lapyonok T, Tanré D, Deuzé JL, et al. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. Atmos Meas Tech 2011;4:975–1018.

[49] Li L, Dubovik O, Derimian Y, Schuster GL, Lapyonok T, Litvinov P, et al. Retrieval of aerosol components directly from satellite and ground-based measurements. Atmos Chem Phys 2019;19:13409–43.

[50] Hasekamp OP, Litvinov P, Butz A. Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements. J Geophys Res 2011;116:D14204.

[51] Fu G, Hasekamp O. Retrieval of aerosol microphysical and optical properties over land using a multimode approach. Atmos Meas Tech 2018;11:6627–50.

[52] Chen X, Wang J, Liu Y, Xu X, Cai Z, Yang D, et al. Angular dependence of aerosol information content in CAPI/TanSat observation over land: effect of polarization and synergy with A-train satellites. Remote Sens Environ 2017;196:163–77.

[53] Chen X, Yang D, Cai Z, Liu Y, Spurr R. Aerosol retrieval sensitivity and error analysis for the cloud and aerosol polarimetric imager on board TanSat: the effect of multi-angle measurement. Remote Sens (Basel) 2017;9:183.

[54] Li Z, Hou W, Hong J, Zheng F, Luo D, Wang J, et al. Directional Polarimetric Camera (DPC): monitoring aerosol spectral optical properties over land from satellite observation. J Quant Spectrosc Ra 2018;218:21–37.

[55] Hou W, Li Z, Wang J, Xu X, Goloub P, Qie L. Improving remote sensing of aerosol microphysical properties by near-infrared polarimetric measurements over vegetated land: information content analysis. J Geophys Res: Atmos 2018;123:2215–43.

[56] Davis AB, Merlin G, Cornet C, Labonnote LC, Riédi J, Ferlay N, et al. Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: an optimal estimation approach. J Quant Spectrosc Ra 2018;216:6–16.

[57] Waquet F, Cairns B, Knobelspiesse K, Chowdhary J, Travis LD, Schmid B, et al. Polarimetric remote sensing of aerosols over land. J Geophys Res 2009;114.

[58] Di Noia A, Hasekamp OP, Wu L, van Diedenhoven B, Cairns B, Yorks JE. Combined neural network/Phillips–Tikhonov approach to aerosol retrievals over land from the NASA Research Scanning Polarimeter. Atmos Meas Tech 2017;10:4235–52.

[59] Wu L, Hasekamp O, van Diedenhoven B, Cairns B. Aerosol retrieval from multiangle, multispectral photopolarimetric measurements: importance of spectral range and angular resolution. Atmos Meas Tech 2015;8:2625–38.

[60] Stamnes S, Hostetler C, Ferrare R, Burton S, Liu X, Hair J, et al. Simultaneous polarimeter retrievals of microphysical aerosol and ocean color parameters from the "MAPP" algorithm with comparison to high-spectral-resolution lidar aerosol and ocean products. Appl Opt 2018;57:2394–413.

[61] Gao M., Zhai P.-.W., Franz B., Hu Y., Knobelspiesse K., Werdell P.J., et al. Inversion of multi-angular polarimetric measurements over open and coastal ocean waters: a joint retrieval algorithm for aerosol and water leaving radiance properties. Atmos Measur Tech Discuss 2019;1–31.

[62] Xu F, Diner D, Dubovik O, Schechner Y. A Correlated Multi-Pixel Inversion Approach for Aerosol Remote Sensing. Remote Sens (Basel) 2019;11:746.

[63] Xu F, Dubovik O, Zhai P-W, Diner DJ, Kalashnikova OV, Seidel FC, et al. Joint retrieval of aerosol and water-leaving radiance from multispectral, multiangular and polarimetric measurements over ocean. Atmos Meas Tech 2016;9:2877–907.

[64] Xu F, van Harten G, Diner DJ, Kalashnikova OV, Seidel FC, Bruegge CJ, et al. Coupled retrieval of aerosol properties and land surface reflection using the Airborne Multiangle SpectroPolarimetric Imager. J Geophys Res: Atmos 2017;122:7004–26.

[65] Fan C, Fu G, Noia AD, Smit M, Rietjens JHH, Ferrare RA, et al. Use of a neural network-based ocean body radiative transfer model for aerosol retrievals from multi-angle polarimetric measurements. Remote Sens (Basel) 2019;11:1–19.

[66] Fu G, Hasekamp O, Rietjens J, Smit M, Di Noia A, Cairns B, et al. Aerosol retrievals from different polarimeters during the ACEPOL campaign using a common retrieval algorithm. Atmos Meas Tech 2020;13:553–73.

[67] Kuhlmann G, Hueni A, Damm A, Brunner D. An Algorithm for In-Flight Spectral Calibration of Imaging Spectrometers. Remote Sens (Basel) 2016;8:1017.

[68] Thompson DR, Natraj V, Green RO, Helmlinger MC, Gao B-C, Eastwood ML. Optimal estimation for imaging spectrometer atmospheric correction. Remote Sens Environ 2018;216:355–73.

[69] Thompson DR, Babu KN, Braverman AJ, Eastwood ML, Green RO, Hobbs JM, et al. Optimal estimation of spectral surface reflectance in challenging atmospheres. Remote Sens Environ 2019;232:111258.

[70] Rodgers CD. Inverse methods for atmospheric sounding: theory and practice. Singapore: World Scientific; 2000.

[71] Thorpe AK, Frankenberg C, Aubrey AD, Roberts DA, Nottrott AA, Rahn TA, et al. Mapping methane concentrations from a controlled release experiment using the next generation airborne visible/infrared imaging spectrometer (AVIRIS-NG). Remote Sens Environ 2016;179:104–15.

[72] Thorpe AK, Frankenberg C, Thompson DR, Duren RM, Aubrey AD, Bue BD, et al. Airborne DOAS retrievals of methane, carbon dioxide, and water vapor concentrations at high spatial resolution: application to AVIRIS-NG. Atmos Meas Tech 2017;10:3833–50.

[73] Xu X, Wang J. UNL-VRTM, a testbed for aerosol remote sensing: model developments and applications. Springer series in light scattering, volume 4: light scattering and radiative transfer. Kokhanovsky A, editor. Springer; 2019.

[74] Ding S, Wang J, Xu X. Polarimetric remote sensing in oxygen A and B bands: sensitivity study and information content analysis for vertical profile of aerosols. Atmos Meas Tech 2016;9:2077–92.

[75] Xu X, Wang J. Retrieval of aerosol microphysical properties from AERONET photopolarimetric measurements: 1. Information content analysis. J Geophys Res: Atmos 2015;120:7059–78.

[76] Xu X, Wang J, Zeng J, Spurr R, Liu X, Dubovik O, et al. Retrieval of aerosol microphysical properties from AERONET photopolarimetric measurements: 2. A new research algorithm and case demonstration. J Geophys Res: Atmos 2015;120:7079–98.

[77] Yu J, Li M, Wang Y, He G. A decomposition method for large-scale box constrained optimization. Appl Math Comput 2014;231:9–15.

[78] Byrd R, Lu P, Nocedal J, Zhu C. A limited memory algorithm for bound constrained optimization. Siam J Sci Comput 1995;16:1190–208.

[79] Zhu C, Byrd RH, Lu P, Nocedal J. L-BFGS-B - Fortran Subroutines for large-scale bound constrained optimization. Evanston: Northwestern University; 1994.

[80] Holben BN, Eck TF, Slutsker I, Tanré D, Buis JP, Setzer A, et al. AERONET—A federated instrument network and data archive for aerosol characterization. Remote Sens Environ 1998;66:1–16.

[81] Li S, Qiu J, Yang X, Liu H, Wan D, Zhu Y. A novel approach to hyperspectral band selection based on spectral shape similarity analysis and fast branch and bound search. Eng Appl Artif Intel 2014;27:241–50.

[82] Xu X, Wang J, Zeng J, Hou W, Meyer KG, Platnick SE, et al. A pilot study of shortwave spectral fingerprints of smoke aerosols above liquid clouds. J Quant Spectrosc Ra 2018;221:38–50.

[83] Kokaly RF, Clark RN, Swayze GA, Livo KE, Hoefen TM, Pearson NC, et al.. USGS spectral library version 7. Reston, VA: Data Series; 2017. p. 68.

[84] Baldridge AM, Hook SJ, Grove CI, Rivera G. The ASTER spectral library version 2.0. Remote Sens Environ 2009;113:711–15.

[85] Spurr R. VLIDORT: a linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. J Quant Spectrosc Ra 2006;102:316–42.

[86] Park S, Lee J-J, Yun C-B, Inman DJ. Electro-mechanical impedance-based wireless structural health monitoring using PCA-data compression and k-means clustering algorithms. J Intell Mater Syst Struct 2007;19:509–20.

[87] Akhtar N, Mian A. RCMF: robust Constrained Matrix Factorization for Hyperspectral Unmixing. IEEE T Geosci Remote 2017;55:3354–66.

[88] Bioucas-Dias JM, Plaza A, Dobigeon N, Parente M, Du Q, Gader P, et al. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. IEEE J Sel Top Appl Earth Observat Remote Sens 2012;5:354–79.

[89] Fishman J, Iraci LT, Al-Saadi J, Chance K, Chavez F, Chin M, et al. The United States' next generation of atmospheric composition and coastal ecosystem measurements: NASA's geostationary coastal and air pollution events (GEO-CAPE) mission. B Am Meteorol Soc 2012;93:1547–66.

[90] Schuster GL, Dubovik O, Holben BN. Angstrom exponent and bimodal aerosol size distributions. J Geophys Res 2006;111.

[91] O'Neill NT, Eck TF, Smirnov A, Holben BN, Thulasiraman S. Spectral discrimination of coarse and fine mode optical depth. J Geophys Res 2003;108.

[92] Levy RC, Remer LA, Kleidman RG, Mattoo S, Ichoku C, Kahn R, et al. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. Atmos Chem Phys 2010;10:10399–420.

[93] Levy RC, Remer LA, Mattoo S, Vermote EF, Kaufman YJ. Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. J Geophys Res 2007;112.

[94] Levy RC, Remer LA, Dubovik O. Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land. J Geophys Res 2007;112.

[95] Omar AH, Won J-G, Winker DM, Yoon S-C, Dubovik O, McCormick MP. Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements. J Geophys Res 2005;110.

[96] Li Z, Zhang Y, Xu H, Li K, Dubovik O, Goloub P. The fundamental aerosol models over china region: a cluster analysis of the ground-based remote sensing measurements of total columnar Atmosphere. Geophys Res Lett 2019;46:4924–32.