

## Article

# A Correlated Multi-Pixel Inversion Approach for Aerosol Remote Sensing

Feng Xu <sup>1,\*</sup>, David J. Diner <sup>1</sup> , Oleg Dubovik <sup>2</sup>  and Yoav Schechner <sup>3</sup>

<sup>1</sup> Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA; David.J.Diner@jpl.nasa.gov

<sup>2</sup> Laboratoire d'Optique Atmosphérique, CNRS/Université Lille-1, 59655 Villeneuve d'Ascq, France; oleg.dubovik@univ-lille.fr

<sup>3</sup> Viterbi Faculty of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel; yoav@ee.technion.ac.il

\* Correspondence: Feng.Xu@jpl.caltech.edu

Received: 8 January 2019; Accepted: 19 March 2019; Published: 27 March 2019



**Abstract:** Aerosol retrieval algorithms used in conjunction with remote sensing are subject to ill-posedness. To mitigate non-uniqueness, extra constraints (in addition to observations) are valuable for stabilizing the inversion process. This paper focuses on the imposition of an empirical correlation constraint on the retrieved aerosol parameters. This constraint reflects the empirical dependency between different aerosol parameters, thereby reducing the number of degrees of freedom and enabling accelerated computation of the radiation fields associated with neighboring pixels. A cross-pixel constraint that capitalizes on the smooth spatial variations of aerosol properties was built into the original multi-pixel inversion approach. Here, the spatial smoothness condition is imposed on principal components (PCs) of the aerosol model, and on the corresponding PC weights, where the PCs are used to characterize departures from the mean. Mutual orthogonality and unit length of the PC vectors, as well as zero sum of the PC weights also impose stabilizing constraints on the retrieval. Capitalizing on the dependencies among aerosol parameters and the mutual orthogonality of PCs, a perturbation-based radiative transfer computation scheme is developed. It uses a few dominant PCs to capture the difference in the radiation fields across an imaged area. The approach is tested using 27 observations acquired by the Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) during multiple NASA field campaigns and validated using collocated AERONET observations. In particular, aerosol optical depth, single scattering albedo, aerosol size, and refractive index are compared with AERONET aerosol reference data. Retrieval uncertainty is formulated by accounting for both instrumental errors and the effects of multiple types of constraints.

**Keywords:** correlated aerosol inversion; radiative transfer; multiangle radiometry; polarimetry

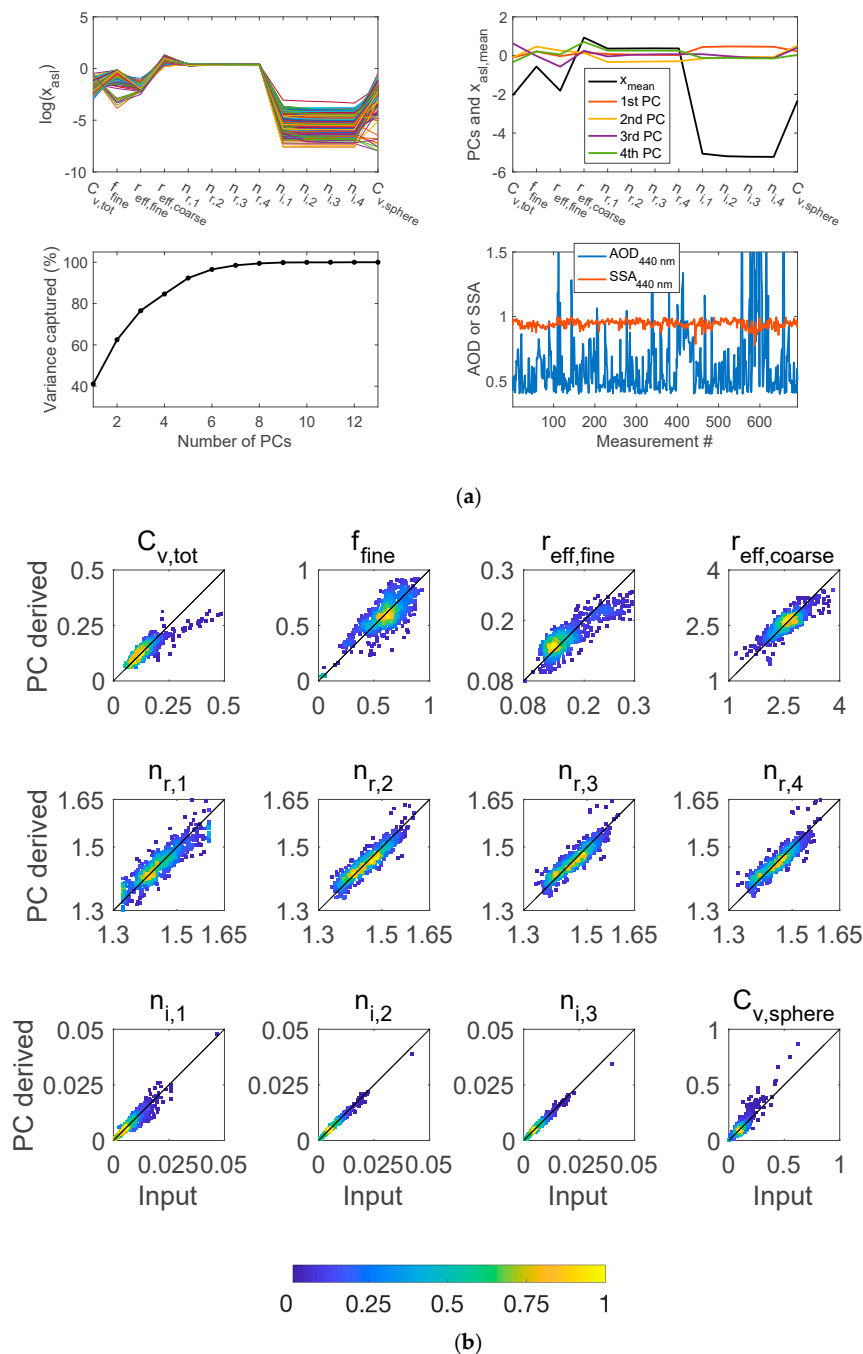
## 1. Introduction

Aerosol retrievals performed on remotely-sensed radiometric and polarimetric imagery are subject to non-uniqueness of the solutions and a large computational burden. The former is caused by insufficient information in the observations and the latter is associated with the high dimensionality of the parameter space. To stabilize the inversion process, optimization-based inversions are often informed by extra constraints that supplement the observations. Exploration and application of these constraints are primary tasks in designing a reliable inversion algorithm. A particularly effective physical constraint for remote sensing applications is the so-called smoothness condition (see [1–5]), which takes advantage of the general observation that some quantities (such as ozone vertical distribution and atmospheric temperature profile) vary smoothly in certain dimensions. For aerosols, their loading and properties

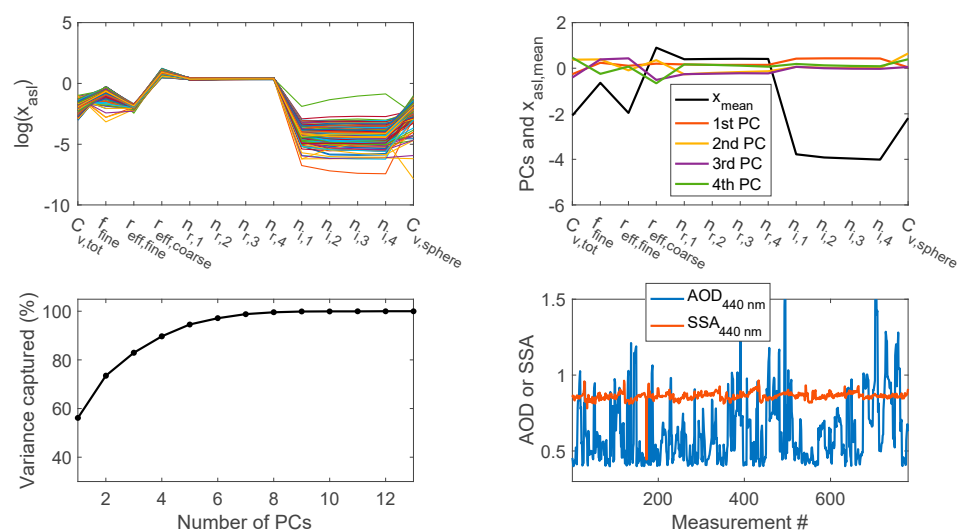
(e.g., size distribution or spectral real and imaginary refractive index) tend to vary smoothly in the spatial (horizontal and vertical) dimensions, while the underlying surface reflectance tends to vary smoothly in the temporal dimension. Imposition of this type of constraint leads to the “multi-pixel” inversion approach [6]. In this case, many pixels are stably inverted at one time and the dimensionality of the retrieval space is proportional to the number of inverted pixels.

Additional constraints can further reduce the large dimensionality of the parameter space, with the expected benefits of higher stability and accelerated retrievals. In this study, we introduce a “correlation constraint” on the aerosol inversions. This constraint takes advantage of empirical dependencies between different sets of aerosol parameters. The cross-correlation coefficient between a pair of aerosol fields quantifies the degree of correlation—as long as this coefficient is not zero, the aerosol variables are not independent of each other, i.e., they are correlated to some extent. However, development of a sophisticated physical model to parameterize these correlations is intractable due to the complexity of the real physical world. On the other hand, empirical methods, such as principal component analysis (PCA) or empirical orthogonal function (EOF) analysis, can be used to capture phenomenological correlations between aerosol fields. By implementing PCA over a training dataset, the original fields can be transformed to superpositions of a set of mutually orthogonal vectors, or principal components (PCs). Then, the original aerosol fields can be approximately reconstructed using a linear combination of a certain number of PCs, with accuracy dependent on the number of PCs employed.

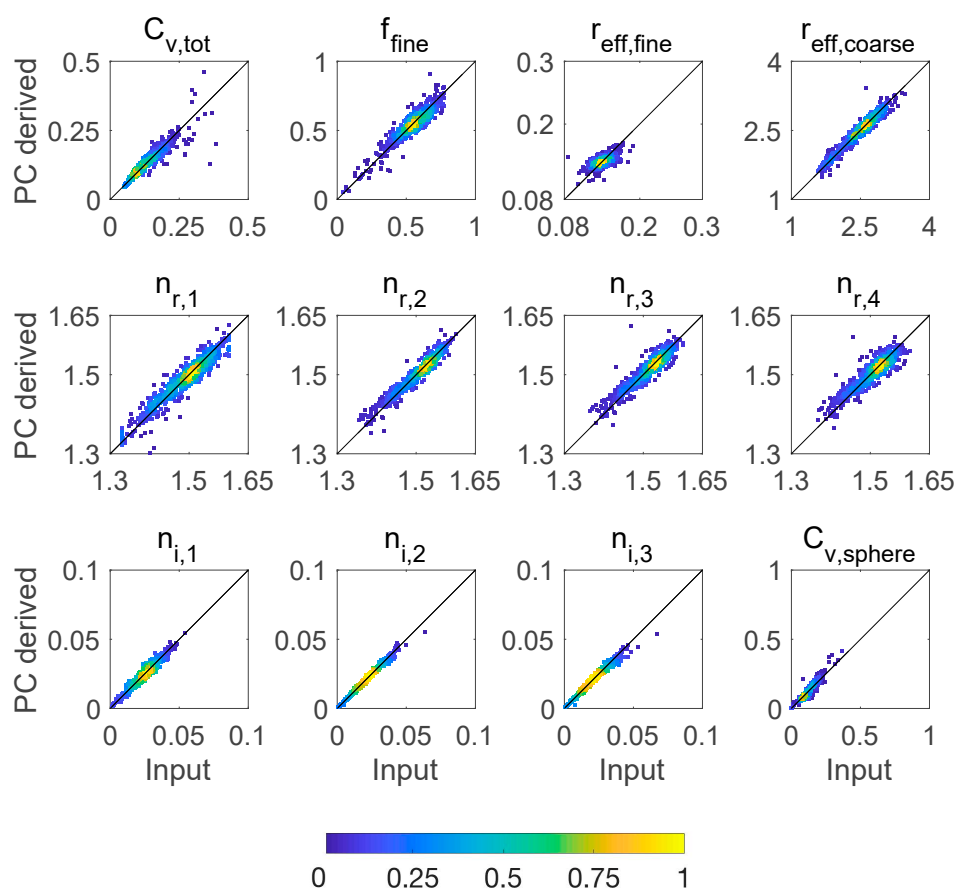
Figures 1 and 2 show a PC analysis of AERONET Level 2 aerosol inversions within a circular domain of diameter 2000 km around the Fresno, California, and Namibe, Angola, sites, respectively. The aerosol fields include total volume concentration, fraction of fine mode particles, effective radii of fine and coarse mode particles, spectral real and imaginary parts of refractive index, and volume concentration of the spherical particles. To generate Figure 1, a total of 691 retrievals from 60 AERONET stations within a circular domain of 2000 km around the AERONET site in Fresno, California, at longitude  $-119.787^\circ$  and latitude  $36.738^\circ$  for the years 1994–2018 were used in the PC analysis. For Figure 2, a total of 780 retrievals from 18 AERONET stations within a circular domain of 2000 km around the AERONET site in Namibe, Angola, at longitude  $12.178^\circ$  and latitude  $-15.159^\circ$  for the years 2000–2017 were used for the PC analysis. Although there are 13 parameters and the aerosol fields vary both temporally and spatially, a few PCs dominate the variations about the mean. For these two cases, 85–90% of the variance in the aerosol fields is captured by the first four PCs. Corresponding to the AERONET analysis in Figure 1 (Fresno domain), Table 1 lists the mutual correlation coefficients among all 13 parameters. Many different types of parameters show correlation coefficients exceeding  $\sim 0.3$ , and among related parameters, such as real and imaginary part of the refractive index, even stronger correlation is observed. While non-absorbing and weakly absorbing particles are dominant in the Fresno domain (Figure 1), absorbing (smoke) aerosols prevail in the Namibe domain (Figure 2). Nevertheless, mutual correlation among aerosol parameters is found in the latter domain as well, as indicated in Table 2. Moreover, comparison of Tables 1 and 2 shows dependence of correlation on specific sets of parameters and regions. For example, the correlation between total aerosol concentrations and spherical particle concentration in the 2000 km Fresno region is weaker than that in the Namibe region. The correlation between fine mode aerosol fraction and real part of refractive index, however, is stronger in Fresno region.



**Figure 1.** (a) Top left panel: AERONET inversion of aerosol fields and properties (in natural logarithm space) including total volume concentration ( $C_{v,\text{tot}}$ ), fraction of fine mode aerosols ( $f_{\text{fine}}$ ), effective radii of fine ( $r_{\text{eff, fine}}$ ) and coarse ( $r_{\text{eff, coarse}}$ ) mode aerosols, column effective real part of refractive indices ( $n_{r, 1-4}$ ) at 0.439, 0.675, 0.870, and 1.018  $\mu\text{m}$ , respectively, imaginary part of refractive indices ( $n_{i, 1-4}$ ), and spherical particle volume concentration ( $C_{v,\text{sphere}}$ ). Though not specified by use of legend, each color is associated with an independent set of retrieval parameters. Top right panel: spatial and temporal mean of all retrievals and the first four principal components (PCs). Bottom left panel: percentage variance of the aerosol fields captured by PCs, indicating that 41%, 62%, 77%, and 85% of variance is captured by the first one, two, three, and four PCs, respectively. Bottom right panel: Aerosol optical depth (AOD) and single scattering albedo (SSA) at 440 nm that are reported in the AERONET retrievals analyzed here. (b) Regression of the derived aerosol properties from four PC against input AERONET values. The scatter plot is colored by the density of the points. Each panel corresponds to a specific aerosol property indicated in the title.



(a)



(b)

**Figure 2.** Same as Figure 1 but analysis was performed for a circular domain of 2000 km around AERONET site in Namibe, Angola. The first four PCs capture 56%, 74%, 83%, and 90% variance of the aerosol fields.

**Table 1.** Correlation coefficient of AERONET retrieved aerosol properties analyzed in Figure 1 (for 2000 km circular around Fresno, California). The correlation is calculated for all parameters in logarithmic space.

	$C_{v,tot}$	$f_{fine}$	$r_{eff,fine}$	$r_{eff,coarse}$	$n_{r,1}$	$n_{r,2}$	$n_{r,3}$	$n_{r,4}$	$n_{i,1}$	$n_{i,2}$	$n_{i,3}$	$n_{i,4}$	$C_{v,sphere}$
$C_{v,tot}$	1.00	−0.34	−0.41	0.08	−0.12	−0.01	0.01	0.02	−0.01	−0.16	−0.23	−0.25	0.06
$f_{fine}$	−	1.00	0.18	0.30	−0.34	−0.36	−0.34	−0.35	0.22	0.32	0.31	0.30	0.67
$r_{eff,fine}$	−	−	1.00	−0.01	−0.17	−0.35	−0.41	−0.43	−0.16	−0.08	−0.01	−0.01	−0.06
$r_{eff,coarse}$	−	−	−	1.00	−0.01	−0.01	−0.01	−0.01	0.25	0.23	0.21	0.20	0.25
$n_{r,1}$	−	−	−	−	1.00	0.94	0.88	0.84	0.33	0.31	0.28	0.28	−0.29
$n_{r,2}$	−	−	−	−	−	1.00	0.98	0.96	0.28	0.22	0.18	0.17	−0.28
$n_{r,3}$	−	−	−	−	−	−	1.00	0.99	0.24	0.16	0.12	0.11	−0.26
$n_{r,4}$	−	−	−	−	−	−	−	1.00	0.21	0.12	0.08	0.08	−0.28
$n_{i,1}$	−	−	−	−	−	−	−	−	1.00	0.95	0.90	0.87	0.23
$n_{i,2}$	−	−	−	−	−	−	−	−	−	1.00	0.98	0.96	0.29
$n_{i,3}$	−	−	−	−	−	−	−	−	−	−	1.00	0.99	0.25
$n_{i,4}$	−	−	−	−	−	−	−	−	−	−	−	1.00	0.23
$C_{v,sphere}$	−	−	−	−	−	−	−	−	−	−	−	−	1.00

**Table 2.** Correlation coefficient of AERONET retrieved aerosol properties analyzed in Figure 2 (for 2000 km circular domain around Namibe, Angola).

	$C_{v,tot}$	$f_{fine}$	$r_{eff,fine}$	$r_{eff,coarse}$	$n_{r,1}$	$n_{r,2}$	$n_{r,3}$	$n_{r,4}$	$n_{i,1}$	$n_{i,2}$	$n_{i,3}$	$n_{i,4}$	$C_{v,sphere}$
$C_{v,tot}$	1.00	−0.32	−0.37	−0.05	−0.43	−0.41	−0.42	−0.40	−0.52	−0.54	−0.54	−0.54	0.39
$f_{fine}$	−	1.00	0.23	0.42	−0.12	−0.06	0.02	0.09	0.62	0.59	0.57	0.56	0.45
$r_{eff,fine}$	−	−	1.00	−0.07	0.20	0.18	0.17	0.14	0.26	0.28	0.27	0.26	−0.01
$r_{eff,coarse}$	−	−	−	1.00	0.14	0.19	0.25	0.33	0.36	0.41	0.41	0.42	0.26
$n_{r,1}$	−	−	−	−	1.00	0.97	0.92	0.86	0.51	0.58	0.58	0.58	−0.31
$n_{r,2}$	−	−	−	−	−	1.00	0.98	0.94	0.57	0.61	0.61	0.61	−0.28
$n_{r,3}$	−	−	−	−	−	−	1.00	0.98	0.61	0.63	0.63	0.62	−0.22
$n_{r,4}$	−	−	−	−	−	−	−	1.00	0.62	0.63	0.63	0.62	−0.17
$n_{i,1}$	−	−	−	−	−	−	−	−	1.00	0.96	0.94	0.92	0.19
$n_{i,2}$	−	−	−	−	−	−	−	−	−	1.00	0.99	0.99	0.12
$n_{i,3}$	−	−	−	−	−	−	−	−	−	−	1.00	1.00	0.09
$n_{i,4}$	−	−	−	−	−	−	−	−	−	−	−	1.00	0.08
$C_{v,sphere}$	−	−	−	−	−	−	−	−	−	−	−	−	1.00

This empirical correlation among aerosol parameters motivates reduction of the dimensionality of the retrieval parameter space and acceleration of the radiative transfer (RT) computations. Benefiting from the empirical aerosol correlations and orthogonality of PCs, increased computational efficiency is achieved by developing a PC-based fast multi-pixel RT computation scheme, as described in Section 4. A straightforward way to reduce dimensionality of the retrieval space is to first derive PCs from a training dataset, such that the retrieval problem reduces to determination of the pixel-dependent PC weights. To account for possible errors in the precomputed PCs, the correlation-based inversion should allow adjustment of the elements of the PC vectors. This is not straightforward, as the vector elements vary significantly in magnitude. A methodology for handling this issue is discussed in Section 3.

PCA has been used to significantly accelerate hyperspectral RT computations [7–9]. Following a determination of representative components from a training dataset of trace-gas and solar-induced chlorophyll fluorescence, the hyperspectral inversions of these quantities are sped up by retrieving PC weights [10–12]. In another retrieval application, Multi-angle Imaging SpectroRadiometer (MISR) operational aerosol retrieval over land uses spatial contrasts to derive PCs of the surface related contribution to the top-of-the-atmosphere (TOA) radiances [13,14]. The similarity in the angular shapes of surface bidirectional reflectance factors among MISR’s four spectral bands in the visible and near-infrared [15] is an example of a prior empirical constraint used to stabilize the aerosol retrievals. In contrast to MISR’s application of PCA to the surface boundary, this paper applies PCA to the aerosol parameters. An optimization approach is developed, which allows both PC vectors and PC weights to be retrieved.

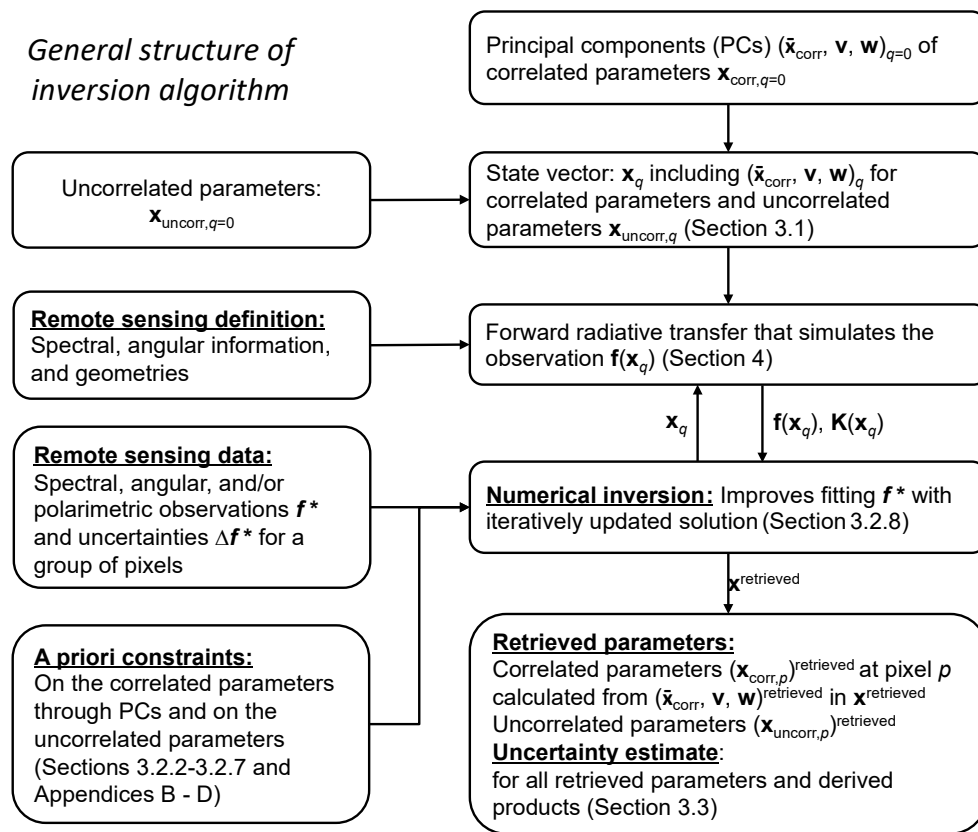
Our retrieval approach is referred to as the “correlated multi-pixel inversion”, where “multi-pixel” reflects that our approach simultaneously uses multiple pixels within image-based measurements (as in the original multi-pixel inversion [6]), while “correlated” captures the extra correlation constraint. In contrast to most optimization approaches that retrieve the individual aerosol fields (see reviews in [16,17]) from multi-spectral radiometric or polarimetric remote sensing observations, we retrieve PC weights and PC vectors of those fields that have spectral, spatial, or temporal correlation with each other, with the purpose of reducing parameter space and improving algorithm efficiency. The individual fields are then constructed from the PC weights and PC vectors. The smoothness constraints on those fields, as implemented in the original “multi-pixel algorithm” used with airborne data at JPL [18,19], are adapted to the current correlated multi-pixel inversion approach. Both “multi-pixel” and “correlated multi-pixel” retrieval approaches utilize a one-dimensional (1D) code based on the independent pixel approximation [20] to ensure forward modeling efficiency.

This paper is organized as follows. Following an algorithm overview in Section 2, the correlated multi-pixel retrieval algorithm and error analysis are formulated in Section 3. A faster forward vector (polarized) RT model for a coupled atmosphere-surface system is presented in Section 4. In Section 5, retrieval tests are performed using 27 datasets acquired by the Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) during multiple NASA field campaigns. A summary is presented in Section 6.

## 2. General Structure of the Algorithm

The expected advantages of correlated multi-pixel inversion in PC space are two-fold. First, it reduces the number of parameters to be retrieved and utilizes observed correlations as empirical constraints to improve the retrieval efficiency and accuracy. Specifically, the correlated parameter space reduces from  $N_{\text{corr}} \times N_{\text{pixel}}$  to  $(N_{\text{PC}} + 1) \times N_{\text{corr}} + N_{\text{PC}} \times N_{\text{pixel}}$ , where  $N_{\text{pixel}}$ ,  $N_{\text{corr}}$ , and  $N_{\text{PC}}$  are the number of image pixels in the scene, the number of correlated aerosol parameters per pixel, and the number of PC vectors, respectively. Reduction of parameter in conjunction with the imposition of proper constraints is expected to mitigate ill-posedness. Moreover, establishment of the PC vectors using ground-based measurements, climatology, or other data sources enables characterization of aerosol properties for which the remote-sensing observations may have insufficient sensitivity (such as aerosol chemical composition or vertical profile).

An overview of the correlated multi-pixel approach and its mapping to different sections of this paper is shown in Figure 3. Symbols and abbreviations used in this work are found in Appendix A (Table A1). A training dataset of correlated parameters is used to derive the spatial and temporal mean of correlated parameters over a targeted area, along with PCs of the correlated departures from the mean. Together with the uncorrelated retrieval parameters, a state vector is initialized with elements specified in Section 3.1. In addition to observational constraints provided by the remote-sensing observations used in the retrievals (Section 3.2.1), convergence and robustness retrieval are optimized by imposing additional constraints. For correlated fields, the additional constraints include: (a) the spectral variation of certain correlated fields (e.g., aerosol optical properties); (b) the spatial variation of certain correlated fields across neighboring image pixels; (c) mutual orthogonality and unit length of the PC vectors; and (d) zero sum of the PC weights. Formalisms for imposing these constraints in PC space are provided in Sections 3.2.2–3.2.7 and Appendix C. For any uncorrelated field (such as surface reflectance), constraints on the spectral, spatial, or temporal variations of the associated parameters are handled in the same way as is formulated in the original multi-pixel inversion approach [6] and repeated in Section 3.2.3 and Appendix B. Combining observational and *a priori* constraints, a system of equations is established at each iteration and then solved to increment the solution (Section 3.2.8) in an inversion process. Error analysis for all retrieved properties is discussed in Section 3.3.



**Figure 3.** General structure of the correlated multi-pixel inversion approach. The interpretation of symbols used in the figure can be found in Table A1 of Appendix A.

To preserve generality of the approach, we use the symbol  $\mathbf{x}_{\text{corr}}$  in the notations below to denote a column vector containing an arbitrary set of parameters that have correlation with each other. In the specific implementation of the correlated multi-pixel inversion for AirMSPI aerosol remote sensing (Section 5), we specify the correlation to be among aerosol parameters only, while the surface parameters are uncorrelated with each other or with aerosol properties (though in principle these assumptions could be relaxed [21–23]). Utilization of the aerosol correlations enables development of a fast PC-RT model for TOA radiance and polarization calculation over a group of pixels (Section 4).

As the retrieval output, the correlated aerosol fields for pixel  $p$  ( $\mathbf{x}_{\text{corr},p}$ ) are constructed from pixel-resolved PC weights in vector form ( $\mathbf{w}_p$ ), spatially and temporally effective PC matrix ( $\mathbf{v}$ ), and spatial and temporal mean vector  $\bar{\mathbf{x}}_{\text{corr}}$ , namely

$$\mathbf{x}_{\text{corr},p} = \bar{\mathbf{x}}_{\text{corr}} + \mathbf{v} \times \mathbf{w}_p \quad (1)$$

where  $\bar{\mathbf{x}}_{\text{corr}}$  is a single column vector and the matrix  $\mathbf{v}$  is made up of  $N_{\text{PC}}$  column vectors consisting of correlated aerosol fields, with  $N_{\text{PC}}$  denoting the number of retrieved principal components. By integrating all PCs into a matrix, we have

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{N_{\text{PC}}} \end{bmatrix} \quad (2)$$

where each PC is composed of  $N_{\text{corr}}$  elements, i.e.,

$$\mathbf{v}_k = [v_k(1); v_k(2); \dots; v_k(N_{\text{corr}})] \quad (3)$$



where “;” indicates that the elements  $v_k(1), v_k(2), \dots$  and  $v_k(N_{\text{corr}})$  are vertically arranged into a column vector. We can further define a PC weight matrix  $\mathbf{w}$  that consists of  $N_{\text{pixel}}$  columns, namely,

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_{N_{\text{pixel}}} \end{bmatrix} \quad (4)$$

where the vector containing PC weights for  $p$ -th pixel is

$$\mathbf{w}_p = [w_p(1); w_p(2); \cdots; w_p(N_{\text{PC}})] \quad (5)$$

The quantities  $(\bar{\mathbf{x}}_{\text{corr}}, \mathbf{v}, \mathbf{w})$  are contained in the solution vector  $\mathbf{x}$  from the last iteration of an inversion process.

### 3. Inversion

#### 3.1. State Vector

Equations (1)–(5) demonstrate the basic operations for constructing correlated fields from PC weight and PC vectors. We define a state vector  $\mathbf{x}_{\text{state}}$  that includes both correlated and uncorrelated fields. The retrieval parameters are arranged in the following order,

$$\mathbf{x}_{\text{state}} = [\bar{\mathbf{x}}_{\text{corr}}; \mathbf{v}_{k=1}; \mathbf{v}_{k=2}; \cdots; \mathbf{v}_{k=N_{\text{PC}}}; \mathbf{w}_{p=1}; \mathbf{x}_{p=1, \text{uncorr}}; \mathbf{w}_{p=2}; \mathbf{x}_{p=2, \text{uncorr}}; \cdots; \mathbf{w}_{p=N_{\text{pixel}}}; \mathbf{x}_{p=N_{\text{pixel}}, \text{uncorr}}] \quad (6)$$

where the mean of aerosol fields takes the form of a column vector

$$\bar{\mathbf{x}}_{\text{corr}} = [\bar{x}_{\text{corr},1}; \bar{x}_{\text{corr},2}; \cdots; \bar{x}_{\text{corr},N_{\text{TP,corr}}}] \quad (7)$$

and  $N_{\text{TP,corr}}$  is the total number of types of parameters correlated with others, and each type may have a subset of values. For example, real and imaginary refractive indices are two types of parameters correlated with each other and each of them have a subset of values as a function of wavelength. For the  $j$ -th type of correlated parameter with  $L(j)$  elements,

$$\bar{\mathbf{x}}_{\text{corr},j} = [\bar{x}_{\text{corr},j}(1); \bar{x}_{\text{corr},j}(2); \cdots; \bar{x}_{\text{corr},j}(L(j))] \quad (8)$$

Similarly, an arbitrary ( $k^{\text{th}}$ ) PC vector composed of  $N_{\text{TP,corr}}$  types of correlated parameters can be arranged into the following column vector,

$$\mathbf{v}_k = [\mathbf{v}_{k,1}; \mathbf{v}_{k,2}; \cdots; \mathbf{v}_{k,N_{\text{TP,corr}}}] \quad (9)$$

where for the type of correlated parameter with  $L_{\text{corr}}(j)$  elements,

$$\mathbf{v}_{k,j} = [v_{k,j}(1); v_{k,j}(2); \cdots; v_{k,j}(L_{\text{corr}}(j))] \quad (10)$$

The PC weight vector for pixel  $p$  is

$$\mathbf{w}_p = [w_p(1); w_p(2); \cdots; w_p(N_{\text{PC}})] \quad (11)$$

For the state vector with  $N_{\text{TP,uncorr}}$  types of uncorrelated parameters, we have for an arbitrary pixel  $p$ ,

$$\mathbf{x}_{p, \text{uncorr}} = [\mathbf{x}_{p, \text{uncorr},1}; \mathbf{x}_{p, \text{uncorr},2}; \cdots; \mathbf{x}_{p, \text{uncorr},N_{\text{TP,uncorr}}}] \quad (12)$$

where the  $[n_{\text{TP,uncorr}}]$ -th type of uncorrelated parameter has  $L_{\text{uncorr}}(n_{\text{TP,uncorr}})$  elements, namely,

$$\mathbf{x}_{\text{uncorr},p,n_{\text{TP,uncorr}}} = [x_{\text{uncorr},p,n_{\text{TP,uncorr}}}(1); x_{\text{uncorr},p,n_{\text{TP,uncorr}}}(2); \cdots; x_{\text{uncorr},p,n_{\text{TP,uncorr}}}(L_{\text{uncorr}}(n_{\text{TP,uncorr}}))] \quad (13)$$



### 3.2. Constraints

Reduction of parameter space is critical for ensuring inversion efficiency and mitigating ill-posedness of large-scale optimization. Moreover, imposition of various types of constraints is important for stabilizing the retrievals and improving accuracy. The correlated multi-pixel inversion algorithm described here uses a total of ten types of constraints:

1. The first type of constraint, formulated in Section 3.2.1, consists of the observations provided directly by the remote sensing instrument(s).
2. The second type of constraint consists of *a priori* values for the retrieval parameters, and is described in Section 3.2.2.
3. For the uncorrelated fields (such as surface reflection properties), the pixel-resolved properties  $\mathbf{x}_{\text{uncorr}}$  can be subjected to both across-pixel and within-pixel constraints. Imposition of these types of constraints has been incorporated into the original multi-pixel inversion [6] and is repeated in Section 3.2.3 as well as in Appendix B as the third type of constraint.
4. When a set of parameters (e.g., aerosol properties) are correlated with each other, their mean can be subjected to smoothness constraints. This is referred to as the fourth type of constraint in Section 3.2.3.
5. Transformation of smooth variations of aerosol properties from regular aerosol parameters into the PC space forms the fifth type of constraint, discussed in Section 3.2.4.
6. In PC space, the smoothness constraints can be applied to the PC weights  $\mathbf{w}$  and vectors  $\mathbf{v}$  separately. Application to across-pixel weights  $\mathbf{w}$ , discussed in Section 3.2.3, forms the sixth type of constraint.
7. Similarly, application of the smoothness constraints to certain type of parameters within a PC vector  $\mathbf{v}_k$ , also discussed in Section 3.2.3, forms the seventh type of constraint. Although it appears that the sixth and seventh types of constraints applied to “ $\mathbf{w}$ ” and “ $\mathbf{v}$ ” separately are redundant with the fifth type of constraint that ensures a smooth variation of overall correlated field constructed from PCs; they are helpful when poor initial guesses of “ $\mathbf{w}$ ” and “ $\mathbf{v}$ ” are provided by a training dataset.
8. The eighth type of constraint, formulated in Section 3.2.5, imposes a zero sum of the PC weights.
9. The ninth type of constraint, formulated in Section 3.2.6, imposes mutual orthogonality of the PC vectors.
10. The tenth type of constraint, formulated in Section 3.2.7, imposes unit norm on all PC vectors.

Even with this set of constraints, numerical errors are unavoidable during the iterations. Hence, a post-correction is implemented after each iteration by reapplying PC analysis to the updated PCs. This way the intrinsic properties of PCs are strictly preserved.

With the ten types of constraints ( $M = 10$ ) introduced above, we now describe the statistical inversion of multi-source (or multi-constraint) data. It involves solving the following system of equations (see [24] for Equations (14)–(20) below):

$$\mathbf{f}_i^* = \mathbf{f}_i(\mathbf{x}) + \Delta \mathbf{f}_i^*, \quad 1 \leq i \leq M \quad (14)$$

where  $\mathbf{f}_i^*$  denotes the  $i^{\text{th}}$  type of constraint,  $\Delta \mathbf{f}_i^*$  is the error with this type of constraint, and  $\mathbf{x} = \mathbf{x}_{\text{state}}$ , as defined in Equation (6). Formally, the statistical independence of different sources of constraints means that the covariance matrix of joint constraint  $\mathbf{f}^* = [\mathbf{f}_1^*; \mathbf{f}_2^*; \dots; \mathbf{f}_M^*]$  has the following structure

$$\mathbf{C}_{\mathbf{f}^*} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \mathbf{0} & \mathbf{0} \\ \dots & \dots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & & \mathbf{C}_M \end{bmatrix} \quad (15)$$

where  $\mathbf{C}_i$  indicates the covariance matrix of  $i$ -th constraint ( $\mathbf{f}_i^*$ ). Following the expressions of  $\mathbf{f}_i^*$  and  $\mathbf{C}_{\mathbf{f}_i^*}$ , the probability distribution function (PDF) of joint data ( $1 \leq i \leq M$ ) can be derived by multiplying PDFs of data from all  $M$  sources, namely,

$$P(\mathbf{f}(\mathbf{x})|\mathbf{f}^*) = \prod_{i=1}^M P(\mathbf{f}_i(\mathbf{x})|\mathbf{f}_i^*) \sim \exp \left\{ -\frac{1}{2} \sum_{i=1}^M [\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_i^*]^T (\mathbf{C}_i)^{-1} [\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_i^*] \right\} \quad (16)$$

Further introducing the weight matrix ( $\mathbf{W}$ ) for multiple ( $M$ ) types of constraints, the objective cost function to be minimized has a quadratic form, namely,

$$\Psi_{\text{total}}(\mathbf{x}) = \sum_{i=1}^M \gamma_i \Psi_i(\mathbf{x}) \quad (17)$$

where

$$\Psi_i(\mathbf{x}) = \frac{1}{2} [\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_i^*]^T \mathbf{W}_i^{-1} [\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_i^*] \quad (18)$$

$$\mathbf{W}_i = \frac{1}{\varepsilon_i^2} \mathbf{C}_i \quad (19)$$

$$\gamma_i = \frac{\varepsilon_1^2}{\varepsilon_i^2} \quad (20)$$

In the above equations,  $\varepsilon_i^2$  is the first diagonal element of  $\mathbf{C}_i$  (i.e.,  $\varepsilon_i^2 = \{\mathbf{C}_i\}_{11}$ ) and the Lagrange factor  $\gamma_i$  weights the contribution of each type of constraint with respect to the first one ( $\gamma_1 = 1$ ).

Minimization of  $\Psi_{\text{total}}(\mathbf{x})$  in Equation (17) means its gradient with respect to the solution  $\mathbf{x}$  approaches zero, such that

$$\nabla \Psi_{\text{total}}(\mathbf{x}) = \sum_{i=1}^M \gamma_i \nabla \Psi_i(\mathbf{x}) = 0 \quad (21)$$

which can be ensured by enforcing the gradient of all components approach zero, namely

$$\nabla \Psi_i(\mathbf{x}) = \mathbf{K}_i^T \mathbf{W}_i^{-1} (\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_i^*) = 0 \quad (22)$$

where  $\mathbf{K}_i$  is the Jacobian matrix containing the derivatives of  $i$ -th type of constraint with respect to the retrieval parameters. To solve the above equation iteratively, we replace  $\mathbf{x}$  by  $\mathbf{x} - \Delta \mathbf{x}$  in Equation (22) and substitute

$$\mathbf{f}_i(\mathbf{x} - \Delta \mathbf{x}) = \mathbf{f}_i(\mathbf{x}) - \mathbf{K}_i \Delta \mathbf{x} \quad (23)$$

into it. This results in

$$\nabla \Psi_i(\mathbf{x} - \Delta \mathbf{x}) = \mathbf{K}_i^T \mathbf{W}_i^{-1} (\mathbf{f}_i(\mathbf{x}) - \mathbf{K}_i \Delta \mathbf{x} - \mathbf{f}_i^*) = 0 \quad (24)$$

or equivalently,

$$(\mathbf{K}_i^T \mathbf{W}_i^{-1} \mathbf{K}_i) \Delta \mathbf{x} = \mathbf{K}_i^T \mathbf{W}_i^{-1} (\mathbf{f}_i(\mathbf{x}) - \mathbf{f}_i^*) = \nabla \Psi_i(\mathbf{x}) \quad (25)$$

The Jacobian matrix  $\mathbf{K}_i$  in Equations (22)–(25) consists of the derivative of the  $l$ -th observational or *a priori* data with respect to the  $n$ -th unknown,

$$K_{i,(l,n)} = \left. \frac{\partial \mathbf{f}_{i,l}}{\partial \mathbf{x}_n} \right|_{\mathbf{x}} \quad (26)$$

More explicit evaluation of  $\mathbf{f}_i(\mathbf{x})$ ,  $\mathbf{f}_i^*$ , and  $\mathbf{K}$  and  $\mathbf{W}$  matrices for all ten types of constraints is discussed in the following subsections.

### 3.2.1. Observational Constraints ( $i = 1$ )

Assuming a total number of  $N_{\text{pixel}}$  pixels, where each pixel has  $Z$  observations, then after arranging all pixel data into a column vector we have

$$\mathbf{f}_{i=1}^*(\mathbf{x}) = [\mathbf{f}_{1,p_1}^*; \mathbf{f}_{1,p_2}^*; \cdots; \mathbf{f}_{1,N_{\text{pixel}}}^*]_{N \times 1} = [y_1; y_2; \cdots; y_{N_f}] \quad (27)$$

where  $N_f = N_{\text{pixel}} \times Z$ .

The parameters of the atmosphere-surface model are adjusted so that the model prediction of radiance and polarization  $\mathbf{f}_1(\mathbf{x})$  fit the observational constraints  $\mathbf{f}_1^*(\mathbf{x})$ . The calculation of  $\mathbf{f}_1(\mathbf{x})$  is introduced in Section 4. The Jacobian matrix  $\mathbf{K}_1$  consists of first order partial derivatives with respect to correlated and uncorrelated parameters in the vicinity of  $\mathbf{x}$ , namely,

$$\mathbf{K}_1 = \begin{bmatrix} \mathbf{K}_{1,\bar{\mathbf{x}}_{\text{corr}}} & \mathbf{K}_{1,\mathbf{v}} & \mathbf{K}_{1,\mathbf{w}}(p_1) & \mathbf{K}_{1,\text{uncorr}}(p_1) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{K}_{1,\bar{\mathbf{x}}_{\text{corr}}} & \mathbf{K}_{1,\mathbf{v}} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{1,\mathbf{w}}(p_2) & \mathbf{K}_{1,\text{uncorr}}(p_2) & \cdots & \mathbf{0} & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \ddots & \cdots & \cdots \\ \mathbf{K}_{1,\bar{\mathbf{x}}_{\text{corr}}} & \mathbf{K}_{1,\mathbf{v}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{K}_{1,\mathbf{w}}(N_{\text{pixel}}) & \mathbf{K}_{1,\text{uncorr}}(N_{\text{pixel}}) \end{bmatrix} \quad (28)$$

where  $\mathbf{K}_{1,\bar{\mathbf{x}}_{\text{corr}}}$  is the Jacobian matrix containing derivatives of observations with respect to spatial and temporal mean correlated parameters (total number is  $N_{\text{corr}}$ ). Variation of  $\bar{\mathbf{x}}_{\text{corr}}$  impacts observations in all pixels. Then  $\mathbf{K}_{1,\bar{\mathbf{x}}_{\text{corr}}}$  is evaluated by

$$\mathbf{K}_{1,\bar{\mathbf{x}}_{\text{corr}}} = \begin{bmatrix} \frac{\partial y_1}{\partial \bar{\mathbf{x}}_{\text{corr}}(1)} & \frac{\partial y_1}{\partial \bar{\mathbf{x}}_{\text{corr}}(2)} & \cdots & \frac{\partial y_1}{\partial \bar{\mathbf{x}}_{\text{corr}}(N_{\text{corr}})} \\ \frac{\partial y_2}{\partial \bar{\mathbf{x}}_{\text{corr}}(1)} & \frac{\partial y_2}{\partial \bar{\mathbf{x}}_{\text{corr}}(2)} & \cdots & \frac{\partial y_2}{\partial \bar{\mathbf{x}}_{\text{corr}}(N_{\text{corr}})} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\partial y_{N_f}}{\partial \bar{\mathbf{x}}_{\text{corr}}(1)} & \frac{\partial y_{N_f}}{\partial \bar{\mathbf{x}}_{\text{corr}}(2)} & \cdots & \frac{\partial y_{N_f}}{\partial \bar{\mathbf{x}}_{\text{corr}}(N_{\text{corr}})} \end{bmatrix} \quad (29)$$

The matrix  $\mathbf{K}_{1,\mathbf{v}}$  in Equation (28) is the Jacobian matrix containing derivatives of observations with respect to all PC elements. It is evaluated as follows:

$$\mathbf{K}_{1,\mathbf{v}} = \begin{bmatrix} \mathbf{K}_{1,\mathbf{v}}(k=1) & \mathbf{K}_{1,\mathbf{v}}(k=2) & \cdots & \mathbf{K}_{1,\mathbf{v}}(k=N_{\text{PC}}) \end{bmatrix} \quad (30)$$

with

$$\mathbf{K}_{1,\mathbf{v}}(k) = \begin{bmatrix} \frac{\partial y_1}{\partial v_k(1)} & \frac{\partial y_1}{\partial v_k(2)} & \cdots & \frac{\partial y_1}{\partial v_k(N_{\text{corr}})} \\ \frac{\partial y_2}{\partial v_k(1)} & \frac{\partial y_2}{\partial v_k(2)} & \cdots & \frac{\partial y_2}{\partial v_k(N_{\text{corr}})} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\partial y_{N_f}}{\partial v_k(1)} & \frac{\partial y_{N_f}}{\partial v_k(2)} & \cdots & \frac{\partial y_{N_f}}{\partial v_k(N_{\text{corr}})} \end{bmatrix} \quad (31)$$

The matrix  $\mathbf{K}_{1,\mathbf{w}}$  in Equation (28) is the Jacobian matrix containing derivatives of observations with respect to pixel-resolved PC weights and is evaluated as follows:

$$\mathbf{K}_{1,\mathbf{w}}(p) = \begin{bmatrix} \frac{\partial y_{p,1}}{\partial w_p(1)} & \frac{\partial y_{p,1}}{\partial w_p(2)} & \cdots & \frac{\partial y_{p,1}}{\partial w_p(N_{\text{PC}})} \\ \frac{\partial y_{p,2}}{\partial w_p(1)} & \frac{\partial y_{p,2}}{\partial w_p(2)} & \cdots & \frac{\partial y_{p,2}}{\partial w_p(N_{\text{PC}})} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\partial y_{p,Z}}{\partial w_p(1)} & \frac{\partial y_{p,Z}}{\partial w_p(2)} & \cdots & \frac{\partial y_{p,Z}}{\partial w_p(N_{\text{PC}})} \end{bmatrix} \quad (32)$$

Moreover, the derivatives with respect to pixel-resolved uncorrelated parameters are evaluated by,

$$\mathbf{K}_{1,\text{uncorr}}(p) = \begin{bmatrix} \frac{\partial y_{p,1}}{\partial x_p(1)} & \frac{\partial y_{p,1}}{\partial x_p(2)} & \cdots & \frac{\partial y_{p,1}}{\partial x_p(L(1)+L(2)+\dots+L(N_{\text{uncorr}}))} \\ \frac{\partial y_{p,2}}{\partial x_p(1)} & \frac{\partial y_{p,2}}{\partial x_p(2)} & \cdots & \frac{\partial y_{p,2}}{\partial x_p(L(1)+L(2)+\dots+L(N_{\text{uncorr}}))} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\partial y_{p,Z}}{\partial x_p(1)} & \frac{\partial y_{p,Z}}{\partial x_p(2)} & \cdots & \frac{\partial y_{p,Z}}{\partial x_p(L(1)+L(2)+\dots+L(N_{\text{uncorr}}))} \end{bmatrix} \quad (33)$$

The covariance matrix for the first constraint (i.e., the observations) assembles the sub-weighting matrix from all pixels, namely,

$$\mathbf{W}_1 = \begin{bmatrix} \mathbf{W}_1(1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_1(2) & \cdots & \mathbf{0} \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_1(N_{\text{pixel}}) \end{bmatrix} \quad (34)$$

### 3.2.2. A Priori Constraints ( $i = 2$ )

The *a priori* constraint is constructed in the same way as in a previous study [24]. Namely, Equation (14) becomes

$$\mathbf{f}_{i=2}^* = \mathbf{x}^{a \text{ priori}} = \mathbf{x} + \Delta \mathbf{x}^{a \text{ priori}} \quad (35)$$

Then in Equations (21)–(25),  $\mathbf{K}_{i=2} = \mathbf{I}$  (identity matrix) and  $\mathbf{W}_{i=2} = \frac{1}{\varepsilon_{a*}^2} \mathbf{C}_{a*}$ . More explicitly,  $\mathbf{W}_{i=2}$  can be constructed from estimated range of each parameter relative to the first one,

$$\mathbf{W}_{i=2} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \frac{(x_{2,\text{max}} - x_{2,\text{min}})^2}{(x_{1,\text{max}} - x_{1,\text{min}})^2} & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \frac{(x_{N,\text{max}} - x_{N,\text{min}})^2}{(x_{1,\text{max}} - x_{1,\text{min}})^2} \end{bmatrix} \quad (36)$$

### 3.2.3. Smoothness Constraints in Regular Parameter Space ( $i = 3, 4, 6$ , and $7$ )

The formulation of these types of constraints is similar, so we discuss them together. The third type of constraint reflects the smooth variation of an uncorrelated type of parameter (such as the variation of parameter of the surface bidirectional reflectance distribution function with wavelength). The fourth type of constraint reflects the smooth variation of certain type of parameter in the mean field (e.g., aerosol refractive index as a function of wavelength). The sixth type of constraint reflects the smooth variation of PC weights from pixel to pixel. The seventh equation reflects the smooth variation of certain type of parameters residing in a PC vector (such as the deviation of refractive index from the mean). The strengths of the sixth and seventh types of constraints depends on the rank of PC. Putting all four of these types of constraints together, we have

$$\begin{cases} \mathbf{f}_3^* = \mathbf{0}^* = \mathbf{S}_{3,m} \mathbf{x}_{\text{uncorr}} + \Delta_{g(x_{\text{uncorr}})}^* \\ \mathbf{f}_4^* = \mathbf{0}^* = \mathbf{S}_{4,m} \bar{\mathbf{x}} + \Delta_{g(\bar{\mathbf{x}}_{\text{corr}})}^* \\ \mathbf{f}_6^* = \mathbf{0}^* = \mathbf{S}_{6,m} \mathbf{w}_{\text{state}} + \Delta_{g(w)}^* \\ \mathbf{f}_7^* = \mathbf{0}^* = \mathbf{S}_{7,m} \mathbf{v}_{\text{state}} + \Delta_{g(v)}^* \end{cases} \quad (37)$$

where  $\mathbf{w}_{\text{state}}$  and  $\mathbf{v}_{\text{state}}$  are column vectors containing only PC weights and vectors, respectively, and are extracted from the overall state vector expressed in Equation (6), and  $\mathbf{S}_{i,m}$  is the differentiation matrix of  $m$ -th order for  $i$ -th type of constraint.

When a retrieval parameter varies smoothly as a function of some variable  $z$ , it is assumed to be locally approximated by a smooth function  $g(z)$ , such as a constant, a line, a parabola, etc. With a polynomial form, the  $m$ -th derivative approaches zero [6], such that

$$S_{i,m}(z)z = \frac{d^m g(z)}{dz^m} = 0 \Rightarrow \begin{cases} g_{m=1}(z) = \text{const} \\ g_{m=2}(z) = Az + B \\ g_{m=3}(z) = Az^2 + Bz + C \end{cases} \quad (38)$$

For a discretized grid of the variable  $z$ , the explicit form of  $S_{i,m}z$ , is,

$$S_{i,m}(z_j)z_j = \begin{cases} \frac{dg}{dz} \approx \frac{\Delta^1 g}{\Delta_1(z)} = \frac{g(z_{j+1}) - g(z_j)}{\Delta_1(z_j)}, \text{ for } m = 1 \\ \frac{d^m g}{dz^m} \approx \frac{\Delta^m g}{\Delta_m(z)} = \frac{\Delta^{m-1}g(z_{j+1})/\Delta_{m-1}(z_{j+1}) - \Delta^{m-1}g(z_j)/\Delta_{m-1}(z_j)}{[\Delta_{m-1}(z_j) + \Delta_{m-1}(z_{j+1})]/2}, \text{ for } m \geq 2 \end{cases} \quad (39)$$

Taking the orders of difference  $m = 1$  and  $2$  as examples, we have

$$\begin{cases} \Delta_{m=1}(z_j) = z_{j+1} - z_j \\ \Delta_{m=2}(z_j) = [\Delta_1(z_j) + \Delta_1(z_{j+1})]/2 \end{cases} \quad (40)$$

Application of the above equation to  $L$  discretized grids  $r_j$  (namely,  $1 \leq j \leq L$ ) leads to

$$\mathbf{f}_i(\mathbf{x}) = \mathbf{S}_{i,m}\mathbf{x} \quad (41)$$

so that by invoking Equation (26),

$$\mathbf{K}_i = \mathbf{S}_{i,m} \quad (42)$$

where the matrix  $\mathbf{S}_{i,m}$  is evaluated by,

$$\mathbf{S}_{i,m=1} = \begin{bmatrix} \frac{1}{\Delta_1(1)} & -\frac{1}{\Delta_1(1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{\Delta_1(2)} & -\frac{1}{\Delta_1(2)} & \cdots & 0 \\ 0 & 0 & \cdots & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{\Delta_1(L-1)} & -\frac{1}{\Delta_1(L-1)} \end{bmatrix} \quad (43)$$

and

$$\mathbf{S}_{i,m=2} = \begin{bmatrix} \frac{2}{\Delta_1(1)[\Delta_1(1)+\Delta_1(2)]} & \frac{-2}{\Delta_1(1)\Delta_1(2)} & \frac{2}{\Delta_1(2)[\Delta_1(1)+\Delta_1(2)]} & 0 & 0 & \cdots & 0 \\ 0 & \frac{-2}{\Delta_1(2)[\Delta_1(2)+\Delta_1(3)]} & \frac{2}{\Delta_1(2)\Delta_1(3)} & \frac{2}{\Delta_1(3)[\Delta_1(2)+\Delta_1(3)]} & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \ddots & \ddots & \cdots \\ 0 & 0 & \cdots & \frac{2}{\Delta_1(L-1)[\Delta_1(L-1)+\Delta_1(L)]} & \frac{-2}{\Delta_1(L-1)\Delta_1(L)} & \cdots & \frac{2}{\Delta_1(L)[\Delta_1(L-1)+\Delta_1(L)]} \end{bmatrix} \quad (44)$$

The same principle applied to higher orders of difference ( $m > 2$ ) ensures a smooth curve with  $\frac{d^m g(z)}{dz^m} = 0$ . Substitution of Equations (41), (42), and  $\mathbf{f}_i^* = \mathbf{0}^*$  into Equation (18) gives,

$$\Psi_i(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{S}_{i,m}^T \mathbf{W}_{i,m}^{-1} \mathbf{S}_{i,m} \mathbf{x} \quad (45)$$

where the weighting matrix  $\mathbf{W}$  has the following diagonal terms,

$$\{\mathbf{W}_{i,m}\}_{jj} = \frac{1}{\Delta_m(z_j)} \quad (46)$$

and  $\Delta_m(z_j)$  is specified in Equation (40) for  $m = 1$  and  $2$ , and can be generalized to an arbitrary higher order.

Substitution of Equations (41), (42), and  $\mathbf{f}_i^* = \mathbf{0}^*$  into Equation (26) gives

$$(\mathbf{S}_{i,m}^T \mathbf{W}_{i,m}^{-1} \mathbf{S}_m) \Delta \mathbf{x} = \mathbf{S}_{i,m}^T \mathbf{W}_{i,m}^{-1} (\mathbf{S}_{i,m} \mathbf{x}) \quad (47)$$

Further defining

$$\Omega_i = \mathbf{S}_{i,m}^T \mathbf{W}_{i,m}^{-1} \mathbf{S}_{i,m} \quad (48)$$

we have

$$\Omega_i \Delta \mathbf{x} = \Omega_i \mathbf{x} \quad (49)$$

Note that for equidistant pixels on an imaged grid, the weighting matrix  $\mathbf{W}$  is a unity matrix (namely  $\mathbf{W} = \mathbf{I}$ ). To distinguish the smoothness constraints imposed on different types of parameters, we note here,

$$\Omega_i = \begin{cases} \Omega_{\text{uncorr}}, & i = 3 \\ \Omega_{\text{corr}}^{\bar{x}}, & i = 4 \\ \Omega_{\text{corr}}^w, & i = 6 \\ \Omega_{\text{corr}}^v, & i = 7 \end{cases} \quad (50)$$

### 3.2.4. Smoothness Constraints in PC Space ( $i = 5$ )

The fifth type of constraint reflects the smooth variation of correlated parameters in different dimensions. However, unlike the smoothness constraint directly imposed on relevant uncorrelated parameters, it has to be transformed to the PC vectors and weights. For simplicity, we put all PC weights into a single-column vector  $\mathbf{w}_{\text{state}} = [\mathbf{w}_{k=1}; \mathbf{w}_{k=2}; \dots; \mathbf{w}_{k=N_{\text{PC}}}]$  and all PC elements for  $j$ -th correlated parameter into another single-column vector  $\mathbf{v}_{\text{state}} = [\mathbf{v}_{k=1}; \mathbf{v}_{k=2}; \dots; \mathbf{v}_{k=N_{\text{PC}}}]$ . The smooth variation of the  $j$ -th correlated aerosol parameter is ensured by,

$$\mathbf{f}_{i=5}^* = \mathbf{0}^* = \mathbf{S}_{i,m} \mathbf{x}_{\text{wv}} + \Delta_{\text{g}([\mathbf{v}; \mathbf{w}])}^* \quad (51)$$

where  $\mathbf{x}_{\text{wv}} = \begin{bmatrix} \mathbf{v}_{\text{state}} \\ \mathbf{w}_{\text{state}} \end{bmatrix}$ . To evaluate the differentiation matrix  $\mathbf{S}_{i,m}$ , we recall that a pixel-resolved correlated parameter is constructed from PCs via Equation (1). An arbitrary correlated parameter  $x_{\text{corr},p}(j)$  associated with  $p$ -th pixel is derived as,

$$x_{\text{corr},p}(j) = \bar{x}_{\text{corr},j} + \sum_{k=1}^{N_{\text{PC}}} w_p(k) v_k(j) = \bar{x}_{\text{corr},j} + [w_p(1)v_1(j) + w_p(2)v_2(j) + \dots + w_p(N_{\text{PC}})v_{N_{\text{PC}}}(j)] \quad (52)$$

#### (a) across-pixel smoothness

The across-pixel variation of a correlated parameter is contributed by pixel-resolved PC weights, namely,

$$\Delta \mathbf{x}_{\text{corr},p} = \mathbf{v} \times (\Delta \mathbf{w}_p) \quad (53)$$

Taking the first order of difference ( $m = 1$ ) in the across-pixel variation of the field as an example, Equation (39) becomes,

$$S_{m=1}(z_j) z_j = \frac{dg}{dz} \approx \frac{\Delta_1 g}{\Delta_1 z} \approx \frac{\sum_{k=1}^{N_{\text{PC}}} \{v_k(j) \times [w_{j+1}(k) - w_j(k)]\}}{\Delta_1(z_j)} \quad (54)$$

where  $z$  measures the inter-pixel relative distance. As an example, we take the simplest case of a single PC ( $k = 1$ ), three pixels ( $p_1, p_2, p_3$ ) associated with PC weights  $[w_{p1}, w_{p2}, w_{p3}]$ , and two correlated parameters associated with PC elements  $v_{k=1}(1)$  and  $v_{k=1}(2)$ . The matrix  $\mathbf{S}_{i,m}$  is constructed

in the following way to ensure smooth variation of the correlated parameter across equidistant (e.g., unity spacing) pixels via  $\mathbf{f}_{i=5}(\mathbf{x}_{wv}) = \mathbf{S}_{i,m}\mathbf{x}_{wv}$ :

$$\mathbf{f}_{i=5}(\mathbf{x}_{wv}) = \mathbf{S}_{i,m}\mathbf{x}_{wv} = \left\{ \frac{1}{2} \begin{bmatrix} w_{p1}(k=1) - w_{p2}(k=1) & 0 & v_{k=1}(1) & -v_{k=1}(1) & 0 \\ 0 & w_{p1}(k=1) - w_{p2}(k=1) & v_{k=1}(2) & -v_{k=1}(2) & 0 \\ w_{p2}(k=1) - w_{p3}(k=1) & 0 & v_{k=1}(1) & -v_{k=1}(1) & 0 \\ 0 & w_{p2}(k=1) - w_{p3}(k=1) & 0 & v_{k=1}(2) & -v_{k=1}(2) \end{bmatrix} \right\} \times \begin{bmatrix} v_{k=1}(1) \\ v_{k=1}(2) \\ w_{p1}(k=1) \\ w_{p2}(k=1) \\ w_{p3}(k=1) \end{bmatrix} \quad (55)$$

(b) within-pixel smoothness

Within-pixel variation of a parameter (such as refractive index) describes its dependence on a non-spatial variable, e.g., wavelength. For an arbitrary pixel  $p$ , the within-pixel variation of a correlated parameter is contributed by PC vectors, namely,

$$\Delta \mathbf{x}_{\text{corr},p} = (\Delta \mathbf{v}) \times \mathbf{w}_p \quad (56)$$

Taking the first order of difference ( $m = 1$ ) as an example, the differentiation matrix  $\mathbf{S}_{i,m}$  is constructed in the following way to ensure smooth variation of the correlated parameter to  $x(j)$  as a function of  $z$  across which  $x(j)$  is smooth (for example,  $z$  is wavelength when  $x(j)$  is aerosol refractive index),

$$S_{i,m=1}(z_j)z_j = \frac{dg}{dz} \approx \frac{\Delta_m g}{\Delta_m z} \approx \frac{\sum_{k=1}^{N_{PC}} \{w_p(k) \times [v_k(j+1) - v_k(j)]\}}{\Delta_1 z_j} \quad (57)$$

Taking the simplest case of a single PC ( $k = 1$ ), three parameters associated with PC elements  $v_{k=1}(1)$ ,  $v_{k=1}(2)$  and  $v_{k=1}(3)$  as an example,

$$\mathbf{f}_{i=5}(\mathbf{x}_{wv}) = \mathbf{S}_{i,m=1}\mathbf{x}_{wv} = \left\{ \frac{1}{2} \begin{bmatrix} \frac{w_p(k=1)}{\Delta_1(z_1)} & -\frac{w_p(k=1)}{\Delta_1(z_1)} & 0 & \frac{v_{k=1}(2)-v_{k=1}(1)}{\Delta_1(z_1)} \\ 0 & \frac{w_p(k=1)}{\Delta_1(z_2)} & -\frac{w_p(k=1)}{\Delta_1(z_2)} & \frac{v_{k=1}(3)-v_{k=1}(2)}{\Delta_1(z_2)} \end{bmatrix} \right\} \times \begin{bmatrix} v_{k=1}(1) \\ v_{k=1}(2) \\ v_{k=1}(3) \\ w_p(k=1) \end{bmatrix} \quad (58)$$

Invoking  $\mathbf{f}_i^* = \mathbf{0}^*$ , the cost function in the form of Equation (45) is derived for both across-pixel and within-pixel smoothness constraints. Further, substitution of  $\mathbf{f}_{i=5}(\mathbf{x}_{wv}) = \mathbf{S}_{i,m=1}\mathbf{x}_{wv}$  into Equation (26) gives the expression for Jacobian matrix elements,

$$K_{i,(l,n)} = \frac{\partial(\mathbf{S}_{i,m}\mathbf{x}_{wv})_l}{\partial x_{wv,n}} \Big|_{\mathbf{x}_{wv}} \quad (59)$$

in which the matrix  $\mathbf{S}_{i,m}$  is a function of  $\mathbf{x}$  (see Equation (57)), therefore  $\mathbf{K}_i \neq \mathbf{S}_{i,m}$ . Intuitively, one can think of  $\mathbf{f}_{i=5}(\mathbf{x}_{wv}) = \mathbf{S}_{i,m=1}\mathbf{x}_{wv}$  as a model prediction to fit the “observation”  $\mathbf{f}_i^*$ , which is a zero vector. Therefore, the above Jacobian matrix is evaluated in the same way as is done for observational constraints by use of finite difference methodology. The finite difference method is used for evaluating Equation (59).

Invoking  $\mathbf{f}_i(\mathbf{x}_{wv}) = \mathbf{S}_{i,m}\mathbf{x}_{wv}$  and  $\mathbf{f}_i^* = \mathbf{0}^*$ , Equation (25) becomes

$$(\mathbf{K}_i^T \mathbf{W}_i^{-1} \mathbf{K}_i) \Delta \mathbf{x}_{wv} = \mathbf{K}_i^T \mathbf{W}_i^{-1} \mathbf{S}_{i,m} \mathbf{x}_{wv} \quad (60)$$

Further defining

$$\begin{cases} \Omega_{\text{corr},1}^{\text{wv}} = \mathbf{K}_i^T \mathbf{W}_i^{-1} \mathbf{K}_i \\ \Omega_{\text{corr},2}^{\text{wv}} = \mathbf{K}_i^T \mathbf{W}_i^{-1} \mathbf{S}_{i,m} \end{cases} \quad (61)$$

we have

$$\Omega_{\text{corr},1} \Delta \mathbf{x}_{wv} = \Omega_{\text{corr},2} \mathbf{x}_{wv} \quad (62)$$



Note that if  $\mathbf{S}_{i,m}$  is independent of  $\mathbf{x}$  (cases for 3, 4, 6, and 7), then  $\mathbf{K}$  can be analytically evaluated from Equation (59), namely,

$$\mathbf{K}_i = \mathbf{S}_{i,m} \quad (63)$$

so that

$$\Omega_{\text{corr},1}^{\text{wv}} = \Omega_{\text{corr},2}^{\text{wv}} = \mathbf{S}_{i,m}^T \mathbf{W}_i^{-1} \mathbf{S}_{i,m} \quad (64)$$

In this manner, the static smoothness matrices for cases for 3, 4, 6, and 7 (Section 3.2.3) are recovered.

Based on above demonstration, the evaluation of  $\mathbf{S}_{i,m}$ ,  $\mathbf{K}_i$ , and  $\Omega_i$  matrices is generalized from the case of regular (non-correlated) retrieval parameters to the case of correlated parameters represented by PCs in Appendix C. The  $\Omega$  matrix derived above is accounted for later to solve for an increment of the PC terms during optimization.

### 3.2.5. Zero-sum Constraint on PC Weights ( $i = 8$ )

As an intrinsic property of pixel-resolved PC weights, they are required to sum to zero for an arbitrary set of PC vectors. This constraint is ensured by multiplying the PC weights  $\mathbf{w}$  in the state vector by an  $\mathbf{O}$  matrix, such that,

$$\mathbf{f}_{i=8}^* = \mathbf{0}^* = \mathbf{f}_{i=8}(\mathbf{w}_{\text{state}}) + \Delta_{\mathbf{O}}^* \quad (65)$$

where  $\mathbf{w}_{\text{state}}$  is a column vector defined in Section 3.2.4 and contains the weights of all PCs, and  $\mathbf{f}_{i=8}(\mathbf{w}_{\text{state}})$  is expressed as

$$\mathbf{f}_8(\mathbf{w}_{\text{state}}) = \mathbf{O} \mathbf{w}_{\text{state}} \quad (66)$$

where

$$\mathbf{O} = \begin{bmatrix} \vec{1} & \vec{0} & \cdots & \vec{0} \\ \vec{0} & \vec{1} & \cdots & \vec{0} \\ \cdots & \cdots & \ddots & \vec{0} \\ \vec{0} & \vec{0} & \cdots & \vec{1} \end{bmatrix}_{N_{\text{PC}} \times (N_{\text{PC}} \times N_{\text{pixel}})}, \quad \vec{1} = [1, 1, \dots, 1]_{1 \times N_{\text{pixel}}} \quad (67)$$

Further substituting  $\mathbf{f}_8^* = \mathbf{0}^*$  and  $\mathbf{f}_8(\mathbf{w}_{\text{state}}) = \mathbf{O} \mathbf{w}_{\text{state}}$  into Equation (18), the cost function has the following quadratic form,

$$\Psi_i(\mathbf{w}_{\text{state}}) = \frac{1}{2} (\mathbf{w}_{\text{state}}^T \mathbf{O}^T \mathbf{O} \mathbf{w}_{\text{state}}) \quad (68)$$

Substitution of Equation (66) into Equation (26) gives the evaluation of Jacobian matrix elements,

$$K_{i,(l,n)} = \left. \frac{\partial (\mathbf{O} \mathbf{w}_{\text{state}})_l}{\partial w_{\text{state},n}} \right|_{\mathbf{w}_{\text{state}}} = O_{l,n} \quad (69)$$

Further defining a zero-sum matrix to be,

$$\Omega_{\mathbf{O}} = \Omega_i = \mathbf{O}^T \mathbf{O} \quad (70)$$

and invoking  $\mathbf{f}_8(\mathbf{w}_{\text{state}}) = \mathbf{O} \mathbf{w}_{\text{state}}$  and  $\mathbf{f}_8^* = \mathbf{0}^*$ , Equation (25) becomes,

$$\Omega_{\mathbf{O}} \Delta \mathbf{w}_{\text{state}} = \Omega_{\mathbf{O}} \mathbf{w}_{\text{state}} \quad (71)$$

### 3.2.6. Mutual Orthogonality Constraint among PC Vectors ( $i = 9$ )

An intrinsic property of PC vectors is that they are mutually orthogonal. This forms the orthogonality constraint imposed on all pairs of PC vectors via  $\mathbf{f}_i$  ( $i = 9$ ), namely

$$\mathbf{f}_{i=9}^* = \mathbf{0}^* = \mathbf{f}_{i=9}(\mathbf{v}_{\text{state}}) + \Delta_{\Gamma}^* \quad (72)$$

where  $\mathbf{v}_{\text{state}}$  is a column vector defined in Section 3.2.4 and contains all PC vectors, and  $\mathbf{f}_{i=9}(\mathbf{v}_{\text{state}})$  is expressed as

$$\mathbf{f}_9(\mathbf{v}_{\text{state}}) = \mathbf{\Gamma} \mathbf{v}_{\text{state}} \quad (73)$$

where,

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{v}_{k=2}^T & \mathbf{v}_{k=1}^T & \vec{0} & \vec{0} & \cdots & \vec{0} & \vec{0} \\ \mathbf{v}_{k=3}^T & \vec{0} & \mathbf{v}_{k=1}^T & \vec{0} & \cdots & \vec{0} & \vec{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{v}_{k=N_{\text{PC}}}^T & \vec{0} & \vec{0} & \vec{0} & \cdots & \vec{0} & \mathbf{v}_{k=1}^T \\ \vec{0} & \mathbf{v}_{k=3}^T & \mathbf{v}_{k=2}^T & \vec{0} & \cdots & \vec{0} & \vec{0} \\ \vec{0} & \mathbf{v}_{k=4}^T & \vec{0} & \mathbf{v}_{k=2}^T & \cdots & \vec{0} & \vec{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vec{0} & \mathbf{v}_{k=N_{\text{PC}}}^T & \vec{0} & \vec{0} & \cdots & \vec{0} & \mathbf{v}_{k=2}^T \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vec{0} & \vec{0} & \vec{0} & \vec{0} & \cdots & \mathbf{v}_{k=N_{\text{PC}}-1}^T & \mathbf{v}_{k=N_{\text{PC}}}^T \end{bmatrix} \quad (74)$$

Further substituting  $\mathbf{f}_9^* = \mathbf{0}^* = \left\{ \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T \right\}_{\frac{1}{2}[N_{\text{PC}} \times (N_{\text{PC}}-1)] \times 1}$  and  $\mathbf{f}_9(\mathbf{v}_{\text{state}}) = \mathbf{\Gamma} \mathbf{v}_{\text{state}}$  into Equation (18), the cost function is derived as,

$$\Psi_i(\mathbf{v}) = \frac{1}{2} \mathbf{v}_{\text{state}}^T \mathbf{\Gamma}^T \mathbf{\Gamma} \mathbf{v}_{\text{state}} \quad (75)$$

Substitution of Equation (73) into Equation (26) gives the evaluation of Jacobian matrix elements,

$$k_{i,(l,n)} = \left. \frac{\partial (\mathbf{\Gamma} \mathbf{v})_l}{\partial v_{\text{state},n}} \right|_{\mathbf{v}_{\text{state}}} = \mathbf{\Gamma}_{l,n} \quad (76)$$

Further defining a zero-sum matrix to be,

$$\mathbf{\Omega}_{\Gamma} = \mathbf{\Omega}_i = \mathbf{\Gamma}^T \mathbf{\Gamma} \quad (77)$$

and invoking  $\mathbf{f}_9(\mathbf{v}_{\text{state}}) = \mathbf{\Gamma} \mathbf{v}_{\text{state}}$  and  $\mathbf{f}_9^* = \mathbf{0}^*$ , Equation (25) becomes,

$$\mathbf{\Omega}_{\Gamma} \Delta \mathbf{v}_{\text{state}} = \mathbf{\Omega}_{\Gamma} \mathbf{v}_{\text{state}} \quad (78)$$

### 3.2.7. Unity-norm Constraint PC Vectors ( $i = 10$ )

Another intrinsic property of each PC vector is that its norm equals unity (i.e.,  $\|\mathbf{v}_k\|^2 = \mathbf{v}_k^T \mathbf{v}_k = 1$ ). This forms the unity constraint imposed on all PC vectors via  $\mathbf{f}_i$  ( $i = 10$ ), such that

$$\mathbf{f}_{i=10}^* = \mathbf{1}^* = \mathbf{f}_{i=10}(\mathbf{v}_{\text{state}}) + \Delta_{\text{U}}^* \quad (79)$$

where

$$\mathbf{f}_{10}(\mathbf{v}_{\text{state}}) = \mathbf{U} \mathbf{v}_{\text{state}} \quad (80)$$

and  $\mathbf{U}$  is expressed as,

$$\mathbf{U} = \begin{bmatrix} \mathbf{v}_{k=1}^T & \vec{0} & \cdots & \vec{0} \\ \vec{0} & \mathbf{v}_{k=2}^T & \cdots & \vec{0} \\ \cdots & \cdots & \ddots & \vec{0} \\ \vec{0} & \vec{0} & \cdots & \mathbf{v}_{k=N_{\text{PC}}}^T \end{bmatrix}_{N_{\text{PC}} \times (N_{\text{PC}} \times N_{\text{corr}})} \quad (81)$$

Then substituting Equation (80) and  $\mathbf{f}_{10}^* = \mathbf{1}^* = \left\{ \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T \right\}_{N_{PC} \times 1}$  into Equation (18), the cost function is derived as,

$$\Psi_i(\mathbf{v}_{\text{state}}) = \frac{1}{2}(\mathbf{v}_{\text{state}}^T \mathbf{U}^T \mathbf{U} \mathbf{v}_{\text{state}} - N_{PC}) \quad (82)$$

Moreover, substitution of Equation (80) into Equation (26) gives the evaluation of Jacobian matrix elements:

$$K_{i,(l,n)} = \left. \frac{\partial(\mathbf{U} \mathbf{v}_{\text{state}})_l}{\partial v_{\text{state},n}} \right|_{\mathbf{v}_{\text{state}}} \quad (83)$$

Further defining a unity constraining matrix to be,

$$\begin{cases} \Omega_{U,1} = \mathbf{K}_i^T \mathbf{K}_i \\ \Omega_{U,2} = \mathbf{K}_i^T \mathbf{U} \end{cases} \quad (84)$$

and invoking  $\mathbf{f}_{10}(\mathbf{v}_{\text{state}}) = \mathbf{U} \mathbf{v}_{\text{state}}$  and  $\mathbf{f}_{10}^* = \mathbf{1}^*$ , Equation (25) becomes equivalent to

$$\Omega_{U,1} \Delta \mathbf{v}_{\text{state}} = \Omega_{U,2} \mathbf{v}_{\text{state}} - \mathbf{K}_i^T \mathbf{f}_i^* \quad (85)$$

### 3.2.8. Construction of Overall Equation System

Accounting for the above ten types of above-specified constraints, the solution to minimizing Equation (17) is approached iteratively. The solution from iteration  $q$  is incremented as follows,

$$\mathbf{x}_{q+1} = \mathbf{x}_q - \Delta \mathbf{x}_q \quad (86)$$

where  $\Delta \mathbf{x}_q$  is obtained by accounting for both observational and *a priori* constraints derived in Sections 3.2.1–3.2.7,

$$\mathbf{A}_q \times \Delta \mathbf{x}_q = \nabla \Psi_{\text{total}}(\mathbf{x}_q) \quad (87)$$

As the explicit form, we have

$$\mathbf{A}_q = \mathbf{K}_{1,q}^T \mathbf{W}_1^{-1} \mathbf{K}_{1,q} + \gamma \Omega_{\text{total},1} + \gamma_a \mathbf{W}_a^{-1} \quad (88)$$

$$\nabla \Psi_{\text{total}}(\mathbf{x}_q) = \mathbf{K}_{1,q}^T \mathbf{W}_1^{-1} [\mathbf{f}_1(\mathbf{x}_q) - \mathbf{f}_1^*] + \gamma \Omega_{\text{total},2} \mathbf{x}_q + \gamma_a \mathbf{W}_a^{-1} (\mathbf{x}_q - \mathbf{x}^{\text{a priori}}) + \begin{bmatrix} \mathbf{0}^* \\ -\gamma \mathbf{U} \mathbf{K}_{10,q}^T \mathbf{f}_{10}^* \\ \mathbf{0}^* \end{bmatrix} \quad (89)$$

where “ $\gamma \Omega_{\text{total}}$ ” incorporates  $i = 3$ –10 types of constraints, namely,

$$\begin{cases} \gamma \Omega_{\text{total},1} = \gamma_{\text{uncorr}} \Omega_{\text{uncorr}}^{\text{Ra}} + \gamma_{\text{corr}}^{\bar{x}} \Omega_{\text{corr}}^{\bar{x},\text{Ra}} + \gamma_{\text{corr}}^{\text{wv}} \Omega_{\text{corr}}^{\text{wv},\text{Ra}} + \gamma_{\text{corr}}^{\text{w}} \Omega_{\text{corr}}^{\text{w},\text{Ra}} + \gamma_{\text{corr}}^{\text{v}} \Omega_{\text{corr}}^{\text{v},\text{Ra}} + \gamma_{\text{O}} \Omega_{\text{O}}^{\text{Ra}} + \gamma_{\text{T}} \Omega_{\text{T}}^{\text{Ra}} + \gamma_{\text{U}} \Omega_{\text{U},1}^{\text{Ra}} \\ \gamma \Omega_{\text{total},2} = \gamma_{\text{uncorr}} \Omega_{\text{uncorr}}^{\text{Ra}} + \gamma_{\text{corr}}^{\bar{x}} \Omega_{\text{corr}}^{\bar{x},\text{Ra}} + \gamma_{\text{corr}}^{\text{wv}} \Omega_{\text{corr}}^{\text{wv},\text{Ra}} + \gamma_{\text{corr}}^{\text{w}} \Omega_{\text{corr}}^{\text{w},\text{Ra}} + \gamma_{\text{corr}}^{\text{v}} \Omega_{\text{corr}}^{\text{v},\text{Ra}} + \gamma_{\text{O}} \Omega_{\text{O}}^{\text{Ra}} + \gamma_{\text{T}} \Omega_{\text{T}}^{\text{Ra}} + \gamma_{\text{U}} \Omega_{\text{U},2}^{\text{Ra}} \end{cases} \quad (90)$$

where the eight terms on the right-hand-side of the above equation are equal to  $\Omega_i$  with  $i = 3$ –10, respectively, and the multipliers  $\gamma_{\dots}$  control the strength of these constraints. The smoothness matrices  $\gamma_{\text{uncorr}} \Omega_{\text{uncorr}}^{\text{Ra}}$ ,  $\gamma_{\text{corr}}^{\bar{x}} \Omega_{\text{corr}}^{\bar{x},\text{Ra}}$ ,  $\gamma_{\text{O}} \Omega_{\text{O}}^{\text{Ra}}$ ,  $\gamma_{\text{T}} \Omega_{\text{T}}^{\text{Ra}}$ ,  $\gamma_{\text{U}} \Omega_{\text{U}}^{\text{Ra}}$ ,  $\gamma_{\text{corr}}^{\text{wv}} \Omega_{\text{corr}}^{\text{wv},\text{Ra}}$ , and  $\gamma_{\text{corr}}^{\text{w/v}} \Omega_{\text{corr}}^{\text{w/v},\text{Ra}}$  act on uncorrelated fields, multi-pixel mean of correlated fields, PC weights, PC vectors, PC vectors, combined PC weights and vectors, and separate PC weights and vectors, respectively. They are essentially equal to  $\gamma_{\text{uncorr}} \Omega_{\text{uncorr}}$ ,  $\gamma_{\text{corr}}^{\bar{x}} \Omega_{\text{corr}}^{\bar{x}}$ ,  $\gamma_{\text{O}} \Omega_{\text{O}}$ ,  $\gamma_{\text{T}} \Omega_{\text{T}}$ ,  $\gamma_{\text{U}} \Omega_{\text{U}}$ ,  $\gamma_{\text{corr}}^{\text{wv}} \Omega_{\text{corr}}^{\text{wv}}$ , and  $\gamma_{\text{corr}}^{\text{w/v}} \Omega_{\text{corr}}^{\text{w/v}}$ , respectively, as evaluated in Sections 3.2.3–3.2.7. However, the elements of these matrices are rearranged into the same dimension to add to each other in assembling  $\gamma \Omega_{\text{total}}$  in Equation (90). In performing the rearrangement, specific locations of relevant parameters in the state vector (see Equation (6)) have to be accounted for, and zero values have to be filled in to accommodate the parameters not subjected to a specific constraint. For the uncorrelated parameters and fields, the explicit forms of  $\gamma_{\text{uncorr}} \Omega_{\text{uncorr}}$  as well as  $\gamma_{\text{corr}}^{\bar{x}} \Omega_{\text{corr}}^{\bar{x}}$  are given in Appendix B. For the correlated parameters and fields, the smoothness constraints imposed

on PC vector and weight retrieval incorporates two types: the coupled type ( $\gamma_{\text{corr}}^{\text{wv}} \Omega_{\text{corr}}^{\text{wv}}$ ), which acts on combined PC weights and vectors, and the decoupled type  $\gamma_{\text{corr}}^{\text{w/v}} \Omega_{\text{corr}}^{\text{w/v}}$ , which act on PC weights and vectors separately. Based on previous demonstration of principles for deriving  $\Omega_i$  for  $i = 5$ , a comprehensive evaluation of  $\gamma_{\text{corr}}^{\text{wv}} \Omega_{\text{corr}}^{\text{wv}}$  is provided in Appendix C. Comprehensive evaluations of  $\gamma_{\text{corr}}^{\text{w}} \Omega_{\text{corr}}^{\text{w}}$  and  $\gamma_{\text{corr}}^{\text{v}} \Omega_{\text{corr}}^{\text{v}}$  are provided in Appendix D. The imposition of  $\gamma_{\text{corr}}^{\text{w}} \Omega_{\text{corr}}^{\text{w}}$  and  $\gamma_{\text{corr}}^{\text{v}} \Omega_{\text{corr}}^{\text{v}}$  are effective when pixel-resolved PC weights themselves or PC elements associated with a type of parameter present certain smooth behavior (usually appearing in low order PCs). Our retrieval tests indicate that a combined use of  $\gamma_{\text{corr}}^{\text{wv}} \Omega_{\text{corr}}^{\text{wv}}$  and  $\gamma_{\text{corr}}^{\text{w/v}} \Omega_{\text{corr}}^{\text{w/v}}$  stabilize the inversion of PCs.

Ideally, a retrieval is deemed successful when the minimization of the cost function is achieved, such that

$$\Psi_{\text{total}} \leq (N_f + N_c + N_{a^*} - N_a) \varepsilon_f^2 \quad (91)$$

where  $N_f$ ,  $N_c$ ,  $N_a$ , and  $N_{a^*}$  are the total number of observations, total number of constraints imposed on retrieval (including smoothness constraints in different dimensions, zero-sum constraint over PC weights and orthogonality and unity constraints over PC vectors), total number of retrieval parameters, and total number of *a priori* estimates of parameters, respectively; and  $\varepsilon_f^2$  is the expected variance due to measurement errors. In practice, forward RT modeling error and other unmodeled effects can impede realization of the required cost function minimization. Therefore, the retrieval is also terminated when the relative difference of fitting residues with solutions from two successive iterations drops below a user-specified threshold value,  $\varepsilon_c^2$ , namely,

$$\frac{|\Psi_{\text{total}}(\mathbf{x}_{q+1}) - \Psi_{\text{total}}(\mathbf{x}_q)|}{\Psi_{\text{total}}(\mathbf{x}_q)} \leq \varepsilon_c^2. \quad (92)$$

### 3.2.9. Determination of Lagrange Multipliers

The Lagrange multipliers reflect the strength of smoothness constraints for a given parameter to be retrieved. Each multiplier is defined as,

$$\gamma_i = \varepsilon_{i=1}^2 / \varepsilon_i^2 \quad (93)$$

where  $\varepsilon_i^2$  are the first diagonal elements of the covariance matrices corresponding to  $i$ -th type of constraints. To estimate  $\varepsilon_i^2$  for a given parameter to be retrieved ( $f = x$ ), which is a function of  $z$ , the most unsmooth known solution  $f^{\text{us}}(z)$  over the target domain is used, as suggested by Dubovik and King [25], namely,

$$\varepsilon^2 = \int_{z_{\min}}^{z_{\max}} \left( \frac{d^m[f^{\text{us}}(z)]}{d^m z} \right)^2 dz \quad (94)$$

where  $z_{\min}$  and  $z_{\max}$  specify the lower and upper bound of  $z$ . For  $i = 3, 4, 5$ ,  $\gamma_i$  is evaluated using least smooth known solution of correlated parameter  $x$  over the target domain. For  $i = 6$  and  $7$ ,  $\gamma_i$  is evaluated using the least smooth known solution of PC weights and PC vectors, respectively, which are derived from a training dataset over the target domain.

In practical implementation of our algorithm, the Lagrange multipliers are modified in the following way:

$$\gamma_i^{\text{Final}} = \frac{N_f \varepsilon_f^2}{N_i \varepsilon_i^2} \gamma_i \quad (95)$$

There are two differences between  $\gamma_i^{\text{Final}}$  and  $\gamma_i$ :

1. The multipliers " $N_f/N_i$ " are introduced to account for possible redundancy of the measured and *a priori* data. Considering that  $\varepsilon_i^2$  is a variance of the error in a single measured or estimated *a priori* value, if we have  $N$  values of a similar kind, the total variance increases in proportion to  $N$ . Introducing this coefficient ensures that when there are several kinds of data, the data with fewer values are given comparable weight, as the data type for which there is a greater number of available values.

2. The multiplier  $\tilde{\varepsilon}_f^2/\varepsilon_f^2$  is introduced with  $\tilde{\varepsilon}_f^2$  estimated as the dynamic fitting residual during iterations:

$$\tilde{\varepsilon}_f^2(\mathbf{x}_q) \approx \frac{\Psi_{\text{total}}(\mathbf{x}_q)}{(N_f + N_c + N_{a^*} - N_a)\varepsilon_f^2} \quad (96)$$

With the multiplier  $\tilde{\varepsilon}_f^2/\varepsilon_f^2$ , the fitting residual is used as an estimate of measurement error variance. As a result, during the first few iterations, the contribution of the *a priori* term is strongest, and its influence decreases as the retrieval progresses. This is done to ensure mostly monotonic convergence, as in the Levenberg-Marquardt procedure [26,27]. However, the Levenberg-Marquardt approach does not specify a particular scheme for introducing these terms, rather it relies on the implementer's intuition. Our algorithm requires the fitting errors in the initial iterations to be dominated by model linearization errors as opposed to random measurement errors, because at each iterative step the full forward model is replaced by its linear approximation, the "errors of linearization" decrease as convergence toward the final solution progresses, and they practically disappear, so that  $\tilde{\varepsilon}_f^2$  becomes equal to  $\varepsilon_f^2$ . As a result of this adjustment of the Lagrange multiplier, the non-linear iteration becomes significantly more monotonic. The Lagrange multiplier that controls the strength of *a priori* ( $i = 2$ ), zero sum ( $i = 8$ ), orthogonality ( $i = 9$ ), and unity ( $i = 10$ ) constraints are dynamically updated using Equations (95) and (96). When the solution approaches the truth, the constraining effect of smoothness constraints via the Lagrange factor decreases during the optimization process.

### 3.3. Retrieval Error Estiamte

Instrumental errors, forward modeling errors, and physical constraints imposed on the retrievals need to be taken into account when evaluating retrieval errors. Instrumental errors include random and systematic effects, such as absolute bias, band-to-band error, and camera-to-camera error. If all of these error sources are well-characterized, their error propagation to the retrievals can be derived in a comprehensive way [19]. If not, then a practical method is to use the square root of the diagonal terms of the following covariance matrix,

$$\Delta x_j = \sqrt{(\mathbf{C}_{\Delta\mathbf{x}})_{jj}} \quad (97)$$

with

$$\mathbf{C}_{\Delta\mathbf{x}} = (\Delta\mathbf{x}_{\text{syst}}) \times (\Delta\mathbf{x}_{\text{syst}})^T + \mathbf{C}_{\Delta\mathbf{x},\text{rand}} \quad (98)$$

where the covariance matrix  $\mathbf{C}_{\Delta\mathbf{x},\text{rand}}$  accounts for the contributions by random errors in the measurement ( $\varepsilon_{\text{rand}}$ ),

$$\mathbf{C}_{\Delta\mathbf{x},\text{rand}} = [\mathbf{A}(\mathbf{x}^{\text{true}})]^{-1} \varepsilon_{\text{rand}}^2 \quad (99)$$

and systematic error  $\Delta\mathbf{x}_{\text{syst}}$  is estimated by

$$\Delta\mathbf{x}_{\text{syst}} = [\mathbf{A}(\mathbf{x}^{\text{true}})]^{-1} \nabla \Psi_{\text{total}}(\mathbf{x}^{\text{true}}) \quad (100)$$

where  $\mathbf{A}$  and  $\nabla \Psi_{\text{total}}$  are calculated by substituting  $\mathbf{x}^{\text{true}}$  into Equations (88) and (89), respectively.

As the retrieved solution is the closest estimate of the "true" solution  $\mathbf{x}^{\text{true}}$ , we adopt  $\mathbf{x}^{\text{true}} = \mathbf{x}^{\text{retrieved}}$  for uncertainty estimation. Uncertainties with uncorrelated parameters are available from Equation (97). Uncertainties with pixel-resolved correlated parameters rely on the propagation of retrieval uncertainties of PCs and PC weights, so that,

$$(\Delta\mathbf{x}_{p,\text{corr}}^{\text{retrieved}})^2 = (\Delta\mathbf{x}_{\text{corr}}^{\text{retrieved}})^2 + \sum_{k=1}^{N_{\text{PC}}} \left\{ [w_p^{\text{retrieved}}(k)]^2 (\Delta\mathbf{v}_k^{\text{retrieved}})^2 + [\Delta w_p^{\text{retrieved}}(k)]^2 (\mathbf{v}_k^{\text{retrieved}})^2 \right\} \quad (101)$$

To estimate errors for functions of the retrieved parameters (namely  $f(\mathbf{x}_p^{\text{retrieved}})$ , e.g., aerosol single-scattering albedo, (SSA)), the following chain rule is applied,

$$\Delta f(\mathbf{x}_p^{\text{retrieved}}) = \sqrt{\mathbf{K}_2^T(\mathbf{x}_p^{\text{retrieved}}) \mathbf{C}_{\Delta \mathbf{x}_p^{\text{retrieved}}} \mathbf{K}_2(\mathbf{x}_p^{\text{retrieved}})} \quad (102)$$

where  $\mathbf{K}_2$  denotes the Jacobian array containing derivatives of  $f(\mathbf{x}_p^{\text{retrieved}})$  with respect to  $\mathbf{x}_p^{\text{retrieved}}$ .

When  $\mathbf{x}^{a \text{ priori}}$  is not available, we assume  $\mathbf{x}^{a \text{ priori}} = \mathbf{x}_q$  so that the *a priori* term on the right-hand-side of Equation (89) for  $\nabla \Psi_{\text{total}}$  is neglected. Indeed, instrument errors are already contained in the observation vector. Therefore, by implementing Equations (97) and (98), it is possible that errors are double-counted in the case that they bias the solution in the same direction as the modeling errors, resulting in a conservative error estimate.

Most error estimates involving Jacobians assume that the calculation is representative of the whole solution space and that the retrieval error is linear with measurement error. These two assumptions can be problematic in situations where model or observation errors are large. In these cases, closure tests using synthetic data with combined random and systematic errors to obtain improved error estimates are recommended.

### 3.4. Retrieval Options

*A priori* or first guess of PCs can be obtained from running PCA on a climatology database, field measurements, or model output. In the case of PCs for which confidence in their contents is low, simultaneous retrieval of PC weights and PC vectors has to be performed. For high confidence PCs, a two-step retrieval can be used to speed up the retrieval. That is, we run PC weight retrieval using the predetermined PCs first and then relax PCs during the retrieval to refine the solution. This relaxation is one way to capture unexpected variance that exceeds that represented by the predetermined PCs. In the extreme situation, where PCs are not representative of the actual aerosol properties, the scene model can be reproduced by using  $N$  PCs, where  $N$  equals to the number of aerosol parameters. In the absence of confidence in an *a priori* set of PCs, retrieval efficiency is still desired. In this case, one can run the correlated multi-pixel inversion with 2–3 PC vectors to capture the major variation of the correlated fields and then (a) relax more PCs into the retrieval to capture unexpected variance, or (b) return to the original multi-pixel retrieval to refine the solution. These different options have been built into our algorithm. In addition, to stabilize the retrieval of correlated fields varying significantly in magnitude (even in logarithmic space), we built into the algorithm an option of retrieving scaled PC vectors and weights. In this case, the representation of correlated fields by PCs for  $p$ -th pixel (namely Equation (1)) is modified as,

$$\mathbf{x}_{\text{corr},p} = \bar{\mathbf{x}}_{\text{corr}} + \mathbf{c} \circ (\mathbf{v} \times \mathbf{w}_p) \quad (103)$$

where  $\circ$  in the above equation denotes an operation of element-wise multiplication, and the constant column vector  $\mathbf{c}$  has  $N_{\text{corr}}$  elements and is introduced here to balance or weight the contribution of all types of retrieval parameters in the PC analysis. It is calculated from (a) the standard deviation of a correlated field  $x$  varying in spatial or temporal scales ( $\sigma_{\text{spatio-temporal}, x}$ ), and (b) its uncertainty estimate ( $\sigma_{e, x}$ ), namely,

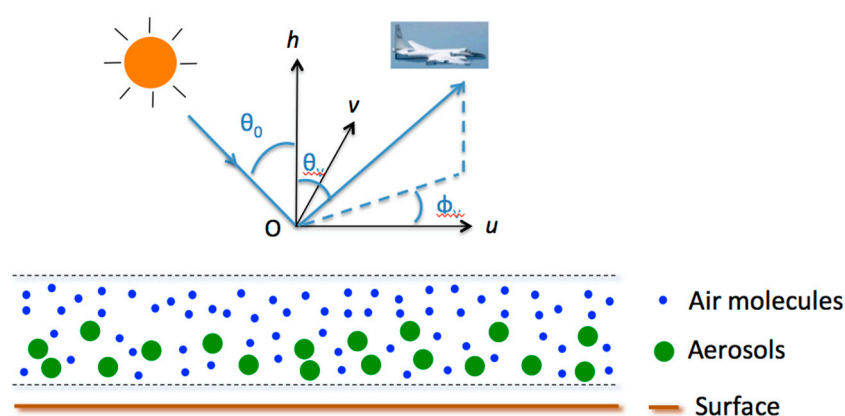
$$C_x = \sqrt{\sigma_{\text{spatio-temporal}, x}^2 + \sigma_{e, x}^2} \quad (104)$$

Such a scaled PC analysis is used to analyze the AERONET retrievals in Figures 1 and 2. To accommodate the constant column vector, the smoothness matrices for the correlated fields and Lagrange multipliers need a slight modification, which is straightforward and is not discussed here.

#### 4. Radiative Transfer in a Coupled Atmosphere-Surface System

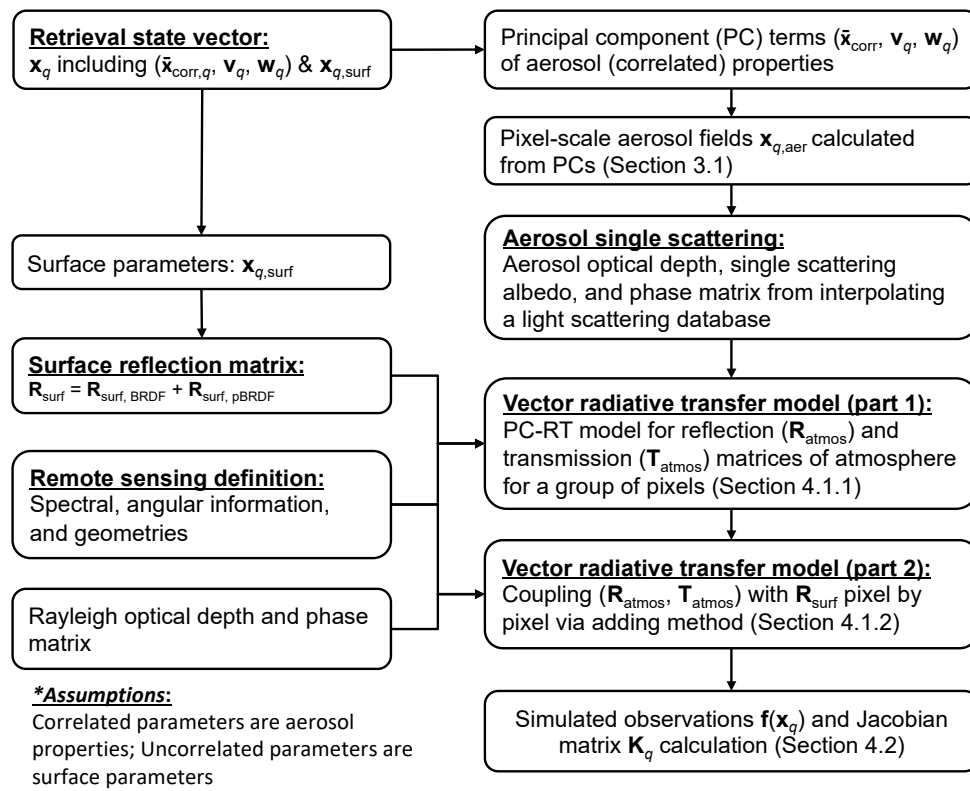
A fast radiative transfer model for a coupled atmosphere-surface (CAS) system plays a critical role in ensuring inversion efficiency. Where correlation among aerosol parameters exists, this section formulates the development of such a model for simulating TOA radiance for a group of pixels simultaneously. Figure 4 shows a CAS system consisting of a land surface and an aerosol/air-molecule mixed layer. The Markov chain model [28] is adopted for modeling RT in the atmospheric part of the CAS system. Coupling of atmospheric radiation and surface reflection is performed by use of the adding method [29]. As outlined in Figure 5, the principal-component based radiative transfer (PC-RT) model includes two steps: (1) calculating aerosol light scattering properties (aerosol optical depth, single scattering albedo, and phase matrix), and the reflection and transmission matrices of the atmosphere for a group of pixels by utilizing correlations in aerosol properties (Section 4.1.1); (2) coupling reflection/transmission matrices of the atmosphere with surface reflection pixel-by-pixel to model the TOA observations (Section 4.1.2). For retrieval use, evaluation of the Jacobian matrix that contains the derivatives of TOA observations with respect to all aerosol and surface parameters is described in Section 4.2.

To further ensure modeling efficiency, the correlated aerosol microphysical properties derived from their PCs are input to a light scattering database [30] for determination of aerosol scattering properties using interpolation. The modeled scattering properties include aerosol optical depth (AOD), single scattering albedo (SSA), and phase matrix. To model surface reflection, the Rahman-Pinty-Verstraete (RPV) model [31] is used to calculate the unpolarized surface bidirectional reflectance distribution function (BRDF). The parameters of the unpolarized part of the surface model include spectral weights and a few angular shape parameters. A modified microfacet model [32] is used to calculate the polarized surface BRDF (pBRDF). The pBRDF model parameters include spectral weight, shadowing width, and slope variance of the microfacets. Combination of the two components gives the full surface reflection model used in AirMSPI aerosol retrievals over land [19]. With the AOD and single scattering properties, a fast RT model that utilizes the correlation in aerosol fields and orthogonality of PC vectors is used to model the radiometric and polarimetric observations.



**Figure 4.** Depiction of the coupled atmosphere-surface (CAS) system model. The sun illuminates the top-of-atmosphere with solar zenith angle  $\theta_0$  and azimuthal plane  $\phi_0$ . The sensor views the atmosphere at view zenith angle  $\theta_v$  and azimuthal angle  $\phi_v$ . A Gaussian vertical distribution profile for aerosols. The Markov chain model is used for computing polarized RT in the atmosphere. It is then coupled with surface reflection using an adding method.



*Forward model\**

**Figure 5.** The scheme of PC-based forward radiative transfer modeling of remote sensing observations from an airborne or spaceborne sensor. The interpretation of symbols used in the figure can be found in Table A1 of Appendix A.

#### 4.1. Fast Multi-Pixel Polarized RT Modeling in the Atmosphere

At a given iteration step, aerosol optical depth (AOD) and light scattering properties, including SSA and phase matrices, are established for all layers. Then, along with the parameters of the surface reflection, RT modeling is performed to evaluate the TOA radiation field. The basic multi-pixel algorithm, which does not take advantage of correlations between aerosol parameters, runs the RT pixel by pixel. Making use of correlation in aerosol fields and orthogonality of PC vectors, the next section describes a fast way to compute radiation in the atmosphere, which is then coupled with surface reflection to derive the TOA radiation field.

##### 4.1.1. Fast Multiple-Pixel Radiative Transfer Modeling Utilizing Correlation

The spatial and spectral variation of aerosol fields, such as volume concentration as a function of aerosol size, aerosol refractive index, vertical distribution profile, and volume concentration of spherical particles, are captured by a few ( $N_{\text{PC}}$ ) dominant PC vectors. Taking advantage of orthogonality of the PCs, the quantities ( $\mathbf{Y}$ ), including total AOD ( $\tau_{\text{aer,tot}}$ ), absorption AOD ( $\tau_{\text{aer,abs}}$ ), and reflection and transmission matrices ( $\mathbf{R}$  and  $\mathbf{T}$ ) for the above-surface atmosphere, can be expanded into Taylor series for multi-variable vector-valued functions  $\mathbf{v}_k$ . Adopting the second order of approximation (justified later in this section) and finite difference method to calculate derivatives, we have,

$$\mathbf{Y}(\mathbf{x}_p) \approx \mathbf{Y}(\bar{\mathbf{x}}) + \sum_{k=1}^{N_{\text{PC}}} \left[ \frac{\mathbf{Y}(\bar{\mathbf{x}} + \delta_s \times \mathbf{v}_k) - \mathbf{Y}(\bar{\mathbf{x}} - \delta_s \times \mathbf{v}_k)}{2\delta_s} w_p(k) + \frac{\mathbf{Y}(\bar{\mathbf{x}} + \delta_s \times \mathbf{v}_k) - 2\mathbf{Y}(\bar{\mathbf{x}}) + \mathbf{Y}(\bar{\mathbf{x}} - \delta_s \times \mathbf{v}_k)}{2\delta_s^2} w_p^2(k) \right] \quad (105)$$

where  $\bar{\mathbf{x}}$  is state vector containing multi-pixel mean aerosol properties and  $\delta_s$  is the scale factor that perturbs a PC vector. The scale factor is empirically determined by accounting for: (a) the accuracy of

the two Jacobian related terms on the right-hand-side of the above equation; and (b) the sufficiency of expanding the radiation fields to the second order and using finite number of PCs to capture the variation of radiation across the imaging area. For an AirMSPI imaging area 10 km by 10 km, an empirical value of 0.1 is adopted to use the above equation.

The efficiency gain of implementing the above PC based RT modeling is obvious: only  $(2 \times N_{PC} + 1)$  RT runs are necessary for evaluating the radiance fields for all pixels, which is much smaller than required for pixel-by-pixel RT runs. If the variation of aerosol fields across an imaging area is extremely high, we use an alternative method built into the algorithm, which applies PCA to the difference in the RT fields evaluated by RT runs with low and high numbers of streams (“streams” represent the quadrature points for integrating the contribution of light from different directions to an event of scattering in a new direction), namely,

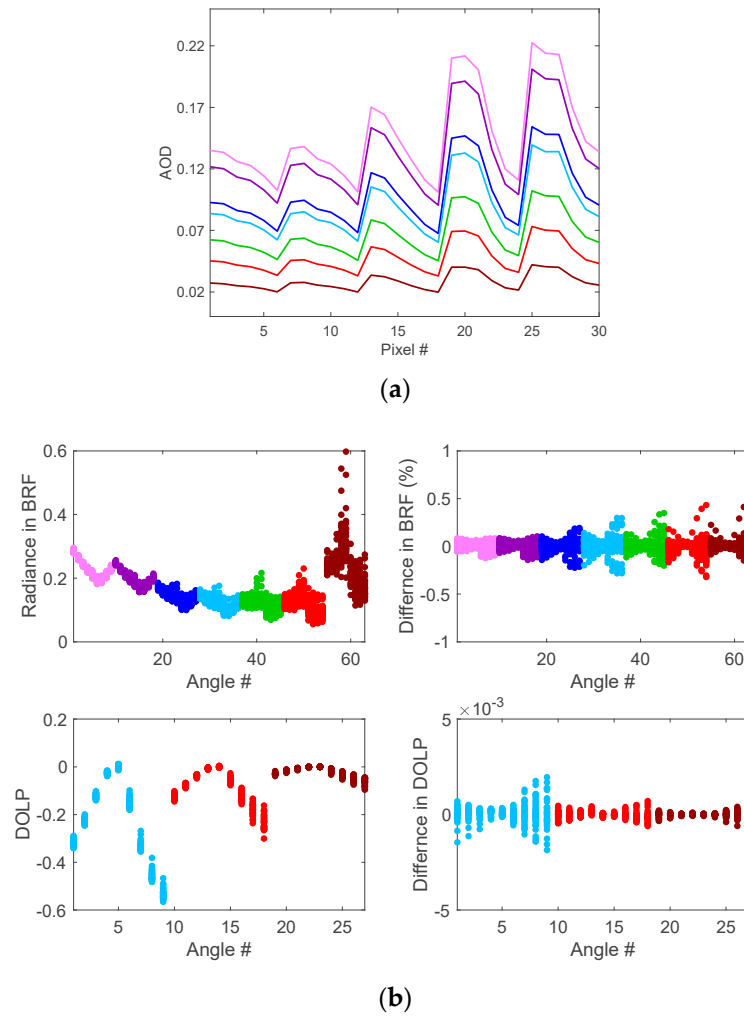
$$\mathbf{Y}(\mathbf{x}_p) - \mathbf{Y}^{LS}(\mathbf{x}_p) \approx \left[ \mathbf{Y}^{HS}(\bar{\mathbf{x}}) - \mathbf{Y}^{LS}(\bar{\mathbf{x}}) \right] + \sum_{k=1}^{N_{PC}} \left[ \frac{\Delta \mathbf{Y}(\bar{\mathbf{x}} + \delta_s \times \mathbf{v}_k) - \Delta \mathbf{Y}(\bar{\mathbf{x}} - \delta_s \times \mathbf{v}_k)}{2\delta_s} w_p(k) + \frac{\Delta \mathbf{Y}(\bar{\mathbf{x}} + \delta_s \times \mathbf{v}_k) - 2\Delta \mathbf{Y}(\bar{\mathbf{x}}) + \Delta \mathbf{Y}(\bar{\mathbf{x}} - \delta_s \times \mathbf{v}_k)}{2\delta_s^2} w_p^2(k) \right] \quad (106)$$

where “HS” and “LS” denote high and low stream-based RT runs, respectively, and  $\Delta \mathbf{Y} = \mathbf{Y}^{HS} - \mathbf{Y}^{LS}$ . In this case, RT runs with low order streams are performed for all pixels and then corrected by Equation (106). This methodology was proposed for a hyperspectral RT simulation [8] and adopted here for the multi-pixel RT simulation. Compared to directly applying PCA to RT via Equation (105), implementation of Equation (106) increases computational cost but captures more variance in the image-scale radiation fields and achieves higher modeling accuracy.

Note that the above evaluation assumes the same view and azimuthal angles for all pixels across the whole image, which is not strictly correct, as view angle varies pixel-by-pixel. To account for such an effect, the  $\mathbf{R}$  and  $\mathbf{T}$  matrices are calculated for a few view and azimuthal angular grids for each viewed image. Interpolation is then used to obtain the reflection and transmission matrices for an arbitrary pixel before coupling them with the surface reflection (as we assume no correlation between aerosol and surface reflection properties in this paper). Coupling of atmospheric radiation and surface reflection to get the TOA radiance for fitting observation is formulated in the next section.

Using three PCs ( $N_{PC} = 3$ ), Figure 6 shows a comparison of PC based RT computation of radiance (expressed as bidirectional reflectance factor, BRF) and degree of linear polarization (DOLP) for 30 super-pixels of an AirMSPI image against the RT computation obtained without using PCs. Here, BRF is defined as  $\pi I_{\text{meas}} d_{\text{ES}}^2 / \mu_0 E_0$ , where  $I_{\text{meas}}$  is the measured radiance,  $d_{\text{ES}}$  is the Earth-Sun distance,  $\mu_0$  is the cosine of solar zenith angle, and  $s$  is the exo-atmospheric solar irradiance. A super-pixel has the resolution of 1 km and is generated by aggregating 100 by 100 original AirMSPI pixels with 10-m resolution. Running RT model over super-pixels can mitigate the errors from the independent pixel approximation [20]. For a variation of aerosol loading up to 100%, the error in computed TOA, BRF, and DOLP from PC-based RT computation is found to be less than ~0.5% and ~0.0025, respectively. These values are smaller than typical instrument errors (e.g., ~4% for radiance and 0.005 for DOLP, which are the instrument requirements for AirMSPI [33]).

Note that the PC-RT model formulated here assumes that aerosol properties are correlated. It is possible, however, that some aerosol fields are not highly correlated. In this case, one can use more PCs to capture the variance. In the worst case scenario of no correlation, the RT calculation defaults to the pixel-by-pixel approach.



**Figure 6.** (a) Pixel-resolved spectral aerosol optical depth (AOD) used in simulating the radiance and degree of linear polarization (DOLP) in (b). The AODs for seven Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) spectral bands are plotted in different colors: pink (355 nm), purple (380 nm), dark blue (445 nm), light blue (470 nm), green (555 nm), red (660 nm), and brown (865 nm). b) The principal component based RT (PC-RT) simulation of radiance using three PCs (upper left panel) and DOLP (lower left panel) at AirMSPI's nine viewing angles within  $(-66^\circ, +66^\circ)$  range around nadir, and the relative error of PC based RT simulation of radiance (top right) and absolute error of DOLP (bottom right) as compared to direct RT computation pixel-by-pixel. Errors are estimated by  $100 \times (Y_{\text{PC-RT}} - Y_{\text{Direct-RT}})/Y_{\text{Direct-RT}}$  for  $Y = \text{radiance}$  and by  $(Y_{\text{PC-RT}} - Y_{\text{Direct-RT}})$  for  $Y = \text{DOLP}$ . The viewing geometries are adopted from one of the scenes acquired by AirMSPI during the ACEPOL field campaign. The unpolarized and polarized surface reflectance is calculated from the retrieved BRDF and pBRDF parameters, respectively.

#### 4.1.2. Coupling Atmospheric Radiation with Surface Reflection

Using the deterministic Markov chain RT model for a plane-parallel atmosphere ([28], or other RT models), we calculate two sets of reflection and transmission matrices for the atmosphere,  $(\mathbf{R}_{\text{atmos}}, \mathbf{T}_{\text{atmos}})$  and  $(\mathbf{R}_{\text{atmos}}^*, \mathbf{T}_{\text{atmos}}^*)$ , which result from illumination at the top and bottom of the atmosphere, respectively. We denote the surface reflection matrix by  $\mathbf{R}_{\text{surf}}$ , which consists of a BRDF component  $(\mathbf{R}_{\text{surf}}, \text{BRDF})$  and a pBRDF component  $(\mathbf{R}_{\text{surf}}, \text{pBRDF})$ . In accordance with the adding

method [29], two matrices  $\mathbf{Q}$  and  $\mathbf{S}$  are defined to account for the interaction between the surface and atmosphere, respectively,

$$\mathbf{S} = \sum_{n=1}^{N_{\max}} \mathbf{Q}_n \quad (107)$$

$$\mathbf{Q}_n = \mathbf{Q}_1 \mathbf{Q}_{n-1} \quad (108)$$

$$\mathbf{Q}_1 = \mathbf{R}_{\text{atmos}}^* \mathbf{R}_{\text{surf}} \quad (109)$$

Then the matrices for downwelling and upwelling diffuse light at the atmosphere-land interface are given by

$$\mathbf{D} = \mathbf{T}_{\text{atmos}} + \mathbf{S} \exp\left(-\frac{\tau_{\text{atmos}}}{\mu_0}\right) + \mathbf{S} \mathbf{T}_{\text{atmos}} \quad (110)$$

where  $\tau_{\text{atmos}}$  is the optical thickness of the whole atmosphere, as contributed by all atmospheric constituents, and the diffuse upwelling light from the surface is calculated as

$$\mathbf{U} = \mathbf{R}_{\text{surf}} \exp\left(-\frac{\tau_{\text{atmos}}}{\mu_0}\right) + \mathbf{R}_{\text{surf}} \mathbf{D} \quad (111)$$

The reflection matrix of the full CAS is,

$$\mathbf{R}_{\text{CAS}} = \mathbf{R}_{\text{atmos}} \exp\left(-\frac{\tau_{\text{atmos}}}{\mu}\right) \mathbf{U} + \mathbf{T}_{\text{atmos}}^* \mathbf{U} \quad (112)$$

For simplicity in describing the conceptual scheme, the superscript “ $m$ ” that denotes Fourier series order is not shown in the above expressions. In actuality, the TOA radiation fields for the CAS system are reconstructed from all orders of Fourier terms, namely,

$$\text{BRF}_{\text{TOA}} = \pi \sum_{m=0}^{\infty} (2 - \delta_{0m}) \mathbf{R}_{\text{CAS},11}^{(m)} \cos m(\phi_v - \phi_0) \quad (113)$$

$$\text{qBRF}_{\text{TOA}} = \pi \sum_{m=0}^{\infty} (2 - \delta_{0m}) \mathbf{R}_{\text{CAS},21}^{(m)} \cos m(\phi_v - \phi_0) \quad (114)$$

$$\text{uBRF}_{\text{TOA}} = \pi \sum_{m=0}^{\infty} (2 - \delta_{0m}) \mathbf{R}_{\text{CAS},31}^{(m)} \cos m(\phi_v - \phi_0) \quad (115)$$

$$\text{vBRF}_{\text{TOA}} = \pi \sum_{m=0}^{\infty} (2 - \delta_{0m}) \mathbf{R}_{\text{CAS},41}^{(m)} \cos m(\phi_v - \phi_0) \quad (116)$$

where  $(\phi_v - \phi_0)$  is the relative azimuth angle between the view and illumination directions. Equations (113)–(116) describe TOA BRFs for total, linearly polarized, and circularly polarized radiation. When DOLP is used to fit the observations, it is calculated as  $\text{DOLP} \approx \sqrt{\text{qBRF}_{\text{TOA}}^2 + \text{uBRF}_{\text{TOA}}^2} / \text{BRF}_{\text{TOA}}$ , where we neglect the minor contribution of circular polarization related term ( $\text{vBRF}_{\text{TOA}}$ ) when it is excluded from polarimetric measurements.

#### 4.2. Jacobian Evaluation

In an iterative way, the optimization algorithm adjusts the state vector that parameterizes aerosol and surface properties to approach the solution. At each iterative step, RT calculations are performed to obtain both the modeled radiation fields and their Jacobians, which describe how the radiation fields vary in response to the small perturbation of state vector components and determine the direction and step size to move for the iterative solution to converge. For an observation vector composed of a series

of  $N_f$  measurements ( $\mathbf{f}_1^* = [y_1; y_2; \dots; y_{N_f}]$ ) and a state vector of  $N_a$  components ( $\mathbf{x} = [x_1; x_2; \dots; x_{N_a}]$ ), the Jacobian matrix has the following structure,

$$\mathbf{K} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_{N_a}} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_{N_a}} \\ \dots & \dots & \ddots & \dots \\ \frac{\partial y_{N_f}}{\partial x_1} & \frac{\partial y_{N_f}}{\partial x_2} & \dots & \frac{\partial y_{N_f}}{\partial x_{N_a}} \end{bmatrix}_{N_f \times N_a} \quad (117)$$

Each matrix element is evaluated by use of the finite difference method. Namely, the derivative of modeled  $i$ -th data with respect to  $n$ -th state vector component is evaluated by,

$$\frac{\partial y_i}{\partial x_n} = \frac{y_i(x_n + \Delta x_n) - y_i(x_n - \Delta x_n)}{2\Delta x_n} \quad (118)$$

The above equation applies to the calculation of the derivatives of  $\tau_{\text{aer,tot}}$ ,  $\tau_{\text{aer,abs}}$ ,  $\mathbf{R}$ , and  $\mathbf{T}$  with respect to the mean aerosol properties and pixel-resolved PC weights. Then, without applying the finite difference method again, the derivative of quantity  $\mathbf{Y} = \{\tau_{\text{aer,tot}}, \tau_{\text{aer,abs}}, \mathbf{R}, \mathbf{T}\}$  at pixel  $p$  with respect to an element in  $k$ -th PC vector can be derived in a fast manner, namely, from the derivatives with respect to the mean and to the PC weights associated with  $p$ -th pixel,

$$\frac{\partial \mathbf{Y}_p}{\partial v_k(i)} \approx w_p(k) \frac{\partial \mathbf{Y}_p}{\partial \bar{x}(i)} \quad (119)$$

Such a strategy further improves the Jacobian evaluation efficiency when multiple PCs are retrieved.

## 5. Inversion of Aerosol and Surface Properties

Section 3 provided a general algorithm formulation to retrieve correlated and uncorrelated parameters. This section describes practical specifics in using AirMSPI data to retrieve aerosol and surface properties. In doing so, we assume correlations in aerosol properties, no correlation in surface properties, and no correlation between aerosol and surface properties.

The first guesses of PCs, PC weights, and mean aerosol fields are derived from a training dataset, which contains all historical AirMSPI retrievals for selected scenes with the criteria of well-calibrated measurements, clear-sky conditions, and collocated ground observations. The aerosol fields include spectrally dependent real and imaginary parts of the aerosol refractive index, volume concentrations of multiple aerosol size components, nonspherical particle fraction, and Gaussian profile-based parameterization of aerosol layer height and standard deviation [19]. Constrained by the information content in polarimetric observations, we retrieve aerosol properties by assuming a single layer, and an “effective” set of aerosol properties for the single layer that fits the polarimetric observations is derived.

With these assumptions, the correlated aerosol parameters include volume concentration as a function of five size components ( $C_{v1-N_{SC}}$  with  $N_{SC} = 5$  as adopted in [19]), real and imaginary parts of refractive index ( $n_r$  and  $n_i$ ), Gaussian distribution-based vertical profile parameterized by central height  $h_a$  and standard deviation  $s_a$ , and the volume fraction of spherical particles ( $f_{v, \text{sphere}}$ ). Then, the parameter spaces described at the beginning of Section 3 are specified as follows,

$$\bar{\mathbf{x}}_{\text{corr}} = \log \left[ \overbrace{\bar{C}_{v1} \dots \bar{C}_{vN_{SC}}}^{\bar{\mathbf{x}}_{\text{corr},1}^T}, \overbrace{\bar{n}_r, \lambda_1 \dots \bar{n}_r, \lambda_7}^{\bar{\mathbf{x}}_{\text{corr},2}^T}, \overbrace{\bar{n}_i, \lambda_1 \dots \bar{n}_i, \lambda_7}^{\bar{\mathbf{x}}_{\text{corr},3}^T}, \overbrace{\bar{h}_a}^{\bar{\mathbf{x}}_{\text{corr},4}^T}, \overbrace{\bar{s}_a}^{\bar{\mathbf{x}}_{\text{corr},5}^T}, \overbrace{\bar{f}_{v, \text{sphere}}}^{\bar{\mathbf{x}}_{\text{corr},6}^T} \right]^T \quad (120)$$

$$\mathbf{v} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_{N_{PC}}] \quad (121)$$

with

$$\mathbf{v}_k = \left[ \underbrace{\overline{v}_{k,C_{v1}} \cdots \overline{v}_{k,C_{v5}}}_{\mathbf{v}_{k,1}^T}, \underbrace{\overline{v}_{k,n_r, \lambda 1} \cdots \overline{v}_{k,n_r, \lambda 7}}_{\mathbf{v}_{k,2}^T}, \underbrace{\overline{v}_{k,n_i, \lambda 1} \cdots \overline{v}_{k,n_i, \lambda 7}}_{\mathbf{v}_{k,3}^T}, \underbrace{\overline{v}_{k,h_a}}_{\mathbf{v}_{k,4}}, \underbrace{\overline{v}_{k,s_a}}_{\mathbf{v}_{k,5}}, \underbrace{\overline{v}_{k,f_v, \text{sphere}}}_{\mathbf{v}_{k,6}} \right]^T \quad (122)$$

$$\mathbf{w} = [\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_{N_{\text{pixel}}}], \text{ with } \mathbf{w}_p = [w_p(1); w_p(2); \cdots; w_p(N_{\text{PC}})] \quad (123)$$

$$\mathbf{x}_{p,\text{uncorr}} = [\mathbf{x}_{p,\text{uncorr},1}; \mathbf{x}_{p,\text{uncorr},2}; \cdots; \mathbf{x}_{p,\text{uncorr},N_{\text{TP,uncorr}}}] \quad (124)$$

with  $N_{\text{TP,uncorr}} = 6$  types of uncorrelated parameters, including the BRDF spectral weight  $a_\lambda$  ( $j = 1$ ), anisotropy parameter  $k_\lambda$  ( $j = 2$ ), anisotropy parameter  $g_\lambda$  ( $j = 3$ ), pBRDF weight  $\epsilon_\lambda$  ( $j = 4$ ), shadowing width  $k_\gamma$  ( $j = 5$ ), and slope variance  $\sigma_s$  ( $j = 6$ ), namely,

$$\mathbf{x}_{p,\text{uncorr},j=1} = [a_p(\lambda_1); a_p(\lambda_2); \cdots; a_p(\lambda_7)] \quad (125)$$

$$\mathbf{x}_{p,\text{uncorr},j=2} = [k_p(\lambda_1); k_p(\lambda_2); \cdots; k_p(\lambda_7)] \quad (126)$$

$$\mathbf{x}_{p,\text{uncorr},j=3} = [g_p(\lambda_1); g_p(\lambda_2); \cdots; g_p(\lambda_7)] \quad (127)$$

$$\mathbf{x}_{p,\text{uncorr},j=4} = [\epsilon_p(\lambda_1); \epsilon_p(\lambda_2); \cdots; \epsilon_p(\lambda_7)] \quad (128)$$

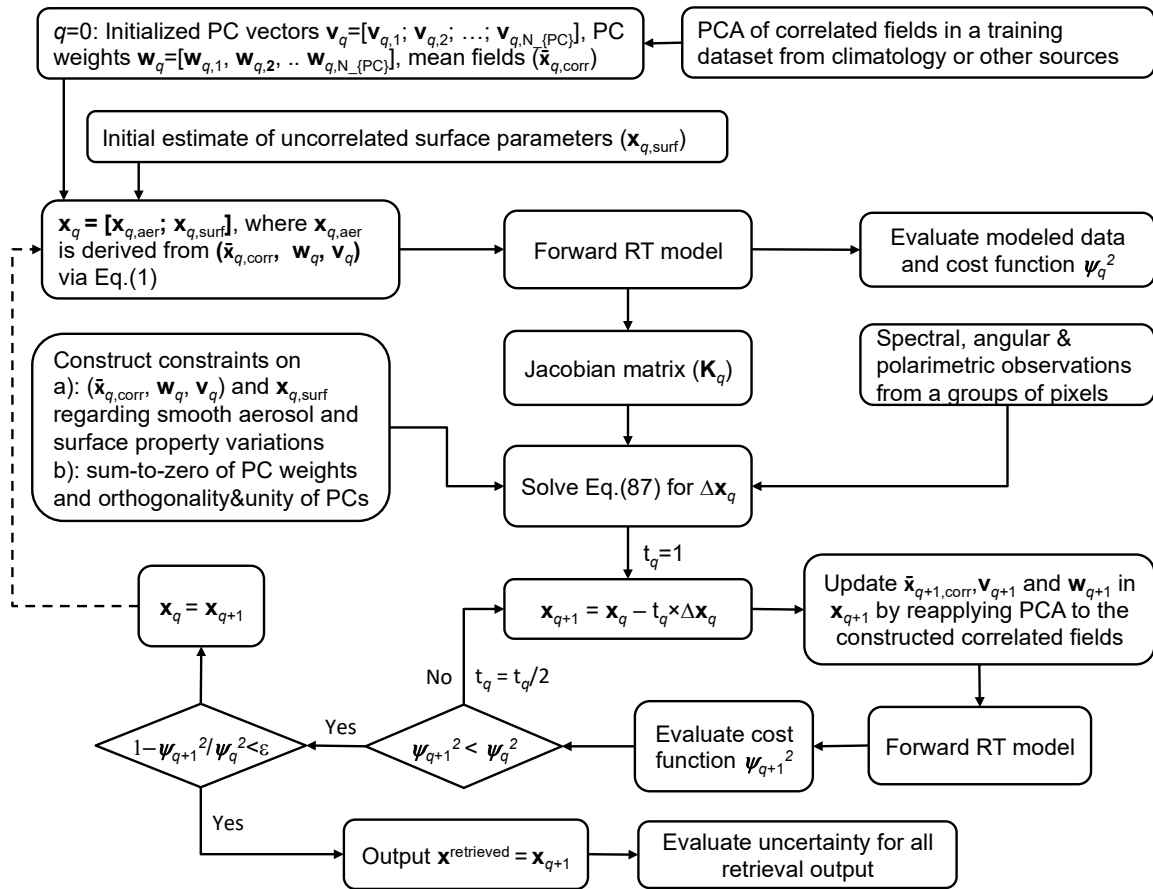
$$\mathbf{x}_{p,\text{uncorr},j=5} = [k_{\gamma,p}] \quad (129)$$

$$\mathbf{x}_{p,\text{uncorr},j=6} = [\sigma_{s,p}^2] \quad (130)$$

In Equation (120), the natural logarithm is used to ensure non-negativity of the solution after dynamic updates during the iterative optimization process. Though not noted here, the angular shape parameter “ $g$ ” with the RPV model needs to be offset by a constant before being transformed into logarithmic space.

An overview of correlated multi-pixel inversion of aerosol properties and surface reflection algorithm flow is shown in Figure 7. As part of the state vector, the PCs of the correlated aerosol microphysical properties are derived from a training dataset from climatology or other sources. As another part of the state vector, uncorrelated parameters (such as surface reflection properties) are initialized to be static [19]. The inversion is stabilized by applying *a priori* constraints on smooth variations of certain aerosol and surface properties in spatial and spectral dimensions. Iterations repeat until convergence is achieved. For the retrieval tests in this study, it takes five to seven iterations for the solution to converge.

As demonstrated in earlier POLDER and AirMSPI retrievals [6,19], imposition of temporal constraints on the variation of surface reflectance can improve aerosol and surface retrievals. In this case, multiple scenes acquired from revisits of a target have to be used. This functionality is not turned on in the following retrieval tests to simplify the demonstration.



**Figure 7.** Algorithm flowchart for correlated multi-pixel retrieval of aerosol and land surface reflection properties. The interpretation of symbols used in the figure can be found in Table A1 of Appendix A.

### 5.1. AirMSPI Datasets

The retrieval algorithm is designed to retrieve column aerosol and surface reflection properties from observations by AirMSPI, which flies on NASA's ER-2 aircraft at an altitude of 20 km and operates in eight spectral channels: 355, 380, 445, 470\*, 555, 660\*, 865\*, and 935 nm, with the asterisk denoting polarimetric bands, in which the Stokes parameters Q and U are measured in addition to radiance I. Images of each targeted area were obtained at 9 viewing angles: 0° (nadir), ±29°, ±48°, ±59°, and ±66° in AirMSPI's step-and-stare mode. At nadir, the imaged area covers a 10 km × 11 km region and the data are mapped to a 10-m spatial grid. Without using the water-vapor influenced band (935 nm), a total of 117 signals per pixel are used, which include 63 radiances (transformed to logarithmic space in retrieval) at nine angles and seven spectral bands, and 27 signals of  $q = Q/I$  and another 27 signals of  $u = U/I$  in the three polarimetric bands. Retrievals for all pixels of a surface area viewed from all 9 angles are performed simultaneously. The measurement errors are adopted as 4% for radiance (to account for angle-to-angle and pixel-to-pixel uncertainties) and 0.005 for DOLP.

A wide range of atmospheric conditions and terrestrial environments have been covered by AirMSPI during over a hundred flights in several airborne campaigns. In this paper, we use AirMSPI data from the Polarimeter Definition Experiment (PODEX) (January to February 2013), Studies of Emissions, Atmospheric Composition, Clouds, and Climate Coupling by Regional Surveys (SEAC<sup>4</sup>RS) (August to September 2013), CalWater-2 (January to March 2015), and Imaging Polarimetric Assessment and Characterization of Tropospheric Particulate Matter (ImPACT-PM) (July 2016) campaigns. From these, 27 AirMSPI step-and-stare data collection sequences were identified to be cloud-free and collocated with AERONET sun photometers for retrieval validation. Locations of these AERONET sites and AirMSPI/AERONET measurement times can be found in a previous study [19]. To control



the strength of multiple types of constraints, the initial values of Lagrange multipliers for the two PCs used in our retrieval are provided in Tables 3–5. Note that the difference between Tables 3 and 4 and Table 5 is that Tables 3 and 4 are for constructing the smoothness constraints over the correlated fields constructed from the combined set of PCs and PC weights. Table 5 is for constructing the smoothness constraints over PC weights and PC vectors separately. Tables 3–5 also list the first guesses of the relevant state vector components and the order of difference for imposing the smoothness constraints on these components.

**Table 3.** Initial guess of image-effective (multi-pixel mean) aerosol parameters and uncorrelated pixel-resolved surface parameters, and the order of difference and Lagrange multipliers for imposing within-pixel smoothness constraints.

	Range	First Guess	Order of Finite Difference for Spectral Smoothness Constraints ( $m_{\text{uncorr},\phi}$ )	Lagrange Regularization Factor ( $\gamma_{\text{uncorr},\phi}$ )
<b>Aerosol parameters (scene-averaged)</b>				
Volume concentration of size components ( $C_{v,1-5}$ , $\mu\text{m}^3/\mu\text{m}^2$ )	$[1.0 \times 10^{-6}, 5]$	0.002	-	-
Central height of aerosol distribution profile ( $h_a$ , km)	[0.05,10]	1	-	-
Standard deviation of aerosol distribution profile ( $s_a$ )	[0.5,2.5]	0.75	-	-
Real part of refractive index ( $n_r(\lambda)$ )	[1.33,1.60]	1.50	1	0.1
Imaginary part of refractive index ( $n_i(\lambda)$ )	$[5.0 \times 10^{-7}, 5.0 \times 10^{-1}]$	0.005	2	0.01
Spherical particle volume fraction ( $f_{v,\text{sphere}}$ )	[0.5,1.0]	0.95	-	-
<b>Surface parameters (pixel-resolved)</b>				
BRDF spectral weight ( $a_\lambda$ )	[0,0.7]	0.015–0.1	3	0.1
Anisotropy parameter ( $k_\lambda$ )	[0,1]	0.6	1	0.5
Anisotropy parameter ( $g_\lambda$ )	[-1,1]	0.1	1	0.5
pBRDF weight ( $\epsilon_\lambda$ )	[0,10]	0.01	-	-
Shadowing width ( $k_\gamma$ )	[0,1]	0.75	1	0.1
Slope variance ( $\sigma_s$ )	[0.05,0.5]	0.075	-	-

**Table 4.** Initial guess and the order of difference and Lagrange multipliers for imposing within-pixel and across-pixel smoothness constraints on the correlated aerosol fields through the first two PCs.

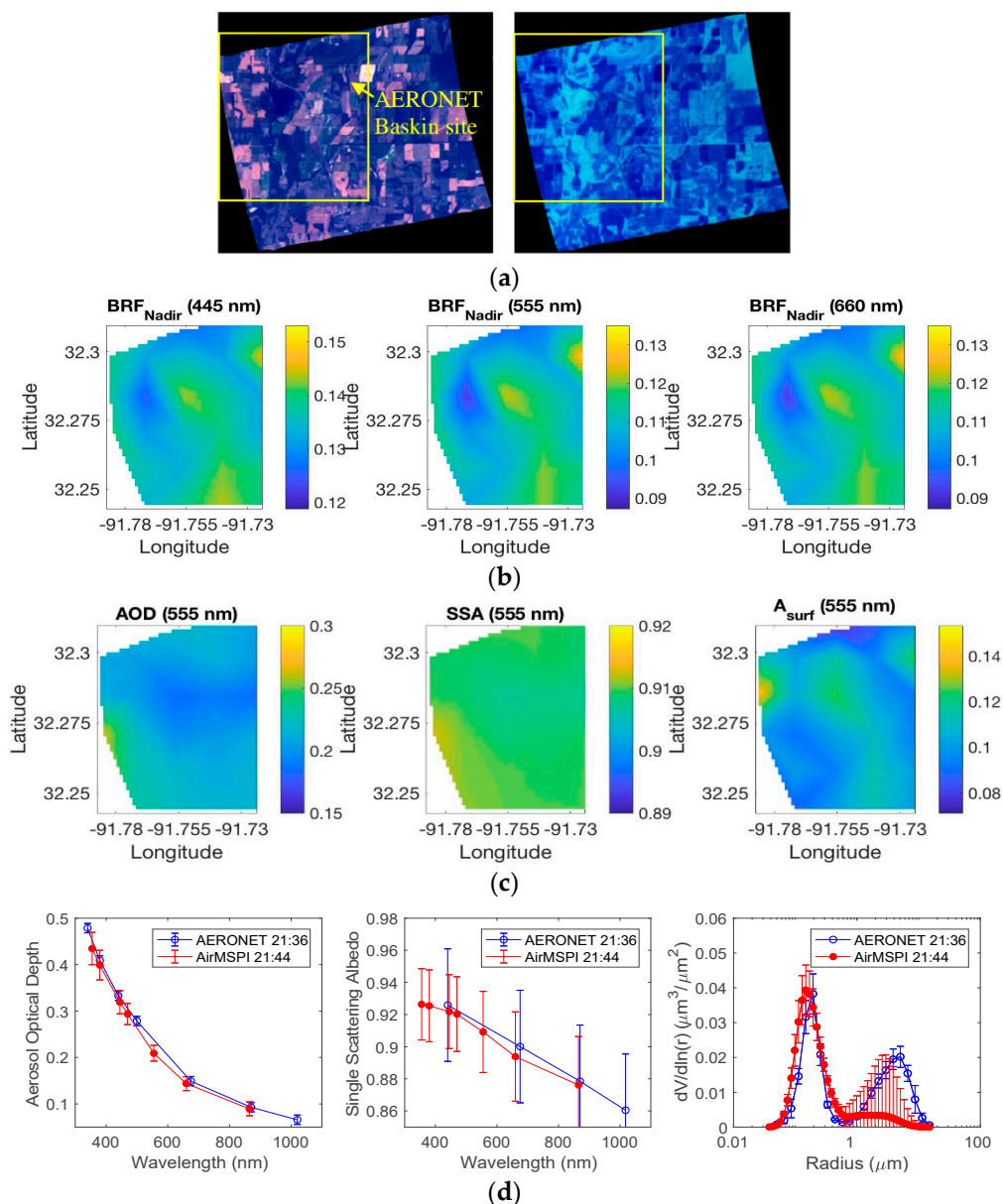
	Initial Guess (PC 1)	Range (PC 1)	Initial Guess (PC 2)	Range (PC 2)	Order of Finite Difference for Spectral Smoothness Constraints ( $m_{\text{corr},\phi}^{\text{wv}}$ )	Lagrange Multiplier ( $\gamma_{\text{corr},\phi}^{\text{wv}}$ )	Order of Finite Difference for Spatial Smoothness Constraints ( $m_{\text{corr},\Delta}^{\text{wv}}$ )	Lagrange Multiplier ( $\gamma_{\text{corr},\Delta}^{\text{wv}}$ )
$\log(C_{v,1-5})$	0.1–0.6	[-0.75,+0.75]	$-3 \times 10^{-1}$ – $7 \times 10^{-1}$	[-1,+1]	-	-	1	1
$\log(h_a)$	$\approx -1 \times 10^{-1}$	[-0.9,+0.9]	$\approx 5 \times 10^{-1}$	[-1,+1]	-	-	1	0.01
$\log(s_a)$	$\approx -2 \times 10^{-2}$	[-0.4,+0.4]	$\approx 7 \times 10^{-2}$	[-0.4,+0.4]	-	-	1	0.01
$\log(n_r(\lambda))$	$\approx 3 \times 10^{-3}$	[-0.1,+0.1]	$\approx 5 \times 10^{-3}$	[-0.1,+0.1]	1	0.1	1	10
$\log(n_i(\lambda))$	$\approx -2 \times 10^{-2}$	[-0.1,+0.1]	$\approx 2 \times 10^{-2}$	[-0.1,+0.1]	2	0.01	1	1
$\log(f_{v,\text{sphere}})$	$\approx 6 \times 10^{-3}$	[-0.05,+0.05]	$\approx 4 \times 10^{-3}$	[-0.05,+0.05]	-	-	1	0.1

**Table 5.** The order of difference and Lagrange multipliers for imposing within-PC constraints on the first two PC vectors and for imposing across-pixel constraint on the PC weights.

	Initial Guess	Range (in log-space)	Order of Finite Difference for Spectral Smoothness Constraints on First (Second) PC Vectors ( $m_{\text{corr},\phi}^{\text{v}}$ )	Lagrange Multiplier on First (Second) PC Vectors ( $\gamma_{\text{corr},\phi}^{\text{v}}$ )	Order of Finite Difference for Spatial Smoothness Constraints on First (Second) PC weight ( $m_{\text{corr},\Delta}^{\text{w}}$ )	Lagrange Multiplier on PC Weights ( $\gamma_{\text{corr},\Delta}^{\text{w}}$ )
$\log(C_{v,1-5})$	-0.3–0.7	[-1,+1]	-	-	-	-
$\log(h_a)$	$\approx 5 \times 10^{-1}$	[-1,+1]	-	-	-	-
$\log(s_a)$	$\approx 7 \times 10^{-2}$	[-0.4,+0.4]	-	-	-	-
$\log(n_r(\lambda))$	$\approx 5 \times 10^{-3}$	[-0.1,+0.1]	1(1)	0.01(0.001)	-	-
$\log(n_i(\lambda))$	$\approx 2.5 \times 10^{-2}$	[-0.01,+0.01]	2(2)	0.001(0.00001)	-	-
$\log(f_{v,\text{sphere}})$	$\approx 4 \times 10^{-3}$	[-0.05,+0.05]	-	-	1(1)	0.05(0.005)
$w_p$	0	[-10,10]	-	-	1(1)	0.05(0.005)

## 5.2. Retrieval Validation against AERONET Products

The retrieved AirMSPI AOD, SSA, size distribution, and refractive index are validated against AERONET Level 1.5 aerosol products. We choose Level 1.5 AERONET product (cloud-screened and quality controlled), as it reports more fields to validate our retrievals. Though the Level 2.0 AERONET product (quality-assured) has higher confidence, the SSA, refractive index, and aerosol size distribution fields were not all generated in 27 test cases. However, for the fields reported by both versions, negligible differences were observed. As a first check, a retrieval is performed using AirMSPI observations acquired on September 9, 2013, over the AERONET Baskin, Louisiana, site which is located at longitude =  $-91.738^\circ$  and latitude =  $32.282^\circ$ . The left image in Figure 8a shows nadir radiance using the spectral band combination of 445, 555, and 660 nm, while the right image displays DOLP at 470, 660, and 865 nm. The retrievals are performed over the area viewed in common at all 9 AirMSPI view angles, outlined by the yellow box. Figure 8b shows TOA BRDF at 445, 555, and 660 nm in the left, middle, and right panels, respectively, for the retrieval area with spatial resolution  $\sim 1.0$  km. Figure 8c shows maps of retrieved AOD, SSA, and surface albedo ( $A_{\text{surf}}$ ) at 555 nm in the left, middle, and right panels, respectively. The retrieved AOD, SSA, and volume weighted aerosol size distribution for the atmosphere above the super-pixel closest to the Baskin site is compared to the AERONET reference data in the left, middle, and right panels of Figure 8d. Good agreement (quantified below) is obtained for all of these quantities, except for the coarse particle size distribution, likely due to the lack of bands longward of 1000 nm in AirMSPI. Generally, the difference between AirMSPI retrievals of AOD, SSA, and size distribution and AERONET reference data is within their retrieval uncertainties. The AirMSPI uncertainties plotted in Figure 8d are estimated as the root mean square of the retrieval uncertainties of these aerosol quantities and the standard deviation of their variations over the whole image. The AERONET uncertainties consist of two parts: temporal variation within the  $\pm 1$ -h window centered on the AirMSPI nadir overpass time, and aerosol measurement and retrieval error [34]. A temporally closest AERONET reference data was identified compare to AirMSPI retrieval at the spatially closest pixel. To account for airmass change during the measurements, the  $\pm 1$ -h window centered on AirMSPI nadir overpass time is used to calculate the AERONET uncertainty from temporal variation.



**Figure 8.** (a) High resolution AirMSPI nadir imagery of Baskin, Louisiana, acquired on September 9, 2013. The left image is of radiance at 445, 555, and 660 nm. The location of the Baskin AERONET site is marked. The right image displays DOLP in the three polarimetric bands (470, 660, and 865 nm). The yellow box indicates the area viewed at all 9 AirMSPI view angles, and where data were used for retrieval algorithm testing. (b) Lower-resolution imagery (~1.0 km) of the retrieval area after pixel aggregation. The left, middle, and right panels are images of BRF at 445, 555, and 660 nm, respectively. (c) Retrieved AOD, SSA, and surface albedo (A) maps at 555 nm in the left, middle, and right panels, respectively. (d) The AirMSPI retrieved AOD, SSA, and volume weighted aerosol size distribution at the pixel closest to the Baskin AERONET site, compared to the AERONET-derived values.

Figure 9 shows a comparison of retrieved pixel-resolved AODs, SSAs, and aerosol size distributions ( $dV/d\ln(r)$ , in  $\mu\text{m}^3/\mu\text{m}^2$ ) from the correlated multi-pixel inversion with those retrieved using original multi-pixel algorithm adapted for AirMSPI [19]. The AOD results for seven AirMSPI spectral bands are plotted in different colors: pink (355 nm), purple (380 nm), dark blue (445 nm), light blue (470 nm), green (555 nm), red (660 nm), and brown (865 nm). Linear regression is performed to obtain slope  $a$ , intercept  $b$ , as well as the coefficient of determination  $R^2$ . The mean absolute difference

(MAD) in AERONET and AirMSPI results is also calculated to measure the overall deviation. Strong correlation and low bias ( $R^2 \geq 0.88$ ,  $a \sim 0.90$ ,  $b \leq 0.05$ , and  $MAD < 0.02$ ) are observed. It can also be observed that a variation of aerosol loading by  $\sim 30\%$  around the mean (namely in the range  $0.28 \leq AOD_{445nm} \leq 0.40$  with mean value 0.32) across the retrieval area is captured by the correlated multi-pixel inversion. Implementation of our approach using several datasets with even higher ( $\sim 90\%$ ) variation of aerosol loading over several smoke scenes acquired by AirMSPI during the recent Aerosol Characterization from Polarimeter and Lidar (ACEPOL) campaign showed the algorithm to be capable of capturing this variation. The regression in Figure 9b shows correlations and low bias of SSAs ( $R^2 > 0.40$ ,  $a > 0.60$ ,  $b < 0.030$ , and  $MAD \leq 0.004$ ) from the two inversions as well. Figure 9c shows basic consistency in the retrieved aerosol size distributions: both algorithms find the peaks of fine and coarse mode aerosol size to be around  $\sim 0.15 \mu m$  and  $\sim 2 \mu m$ , respectively. Due to the lower sensitivity of AirMSPI's longest wavelength 865 nm to coarse mode aerosols, some differences in coarse mode aerosol size can be observed. This indicates the impact of insufficient observational information about certain aerosol properties. Comparisons of pixel-scale AOD, SSA, and size distribution at other retrieval cases show a similar quality of agreement.

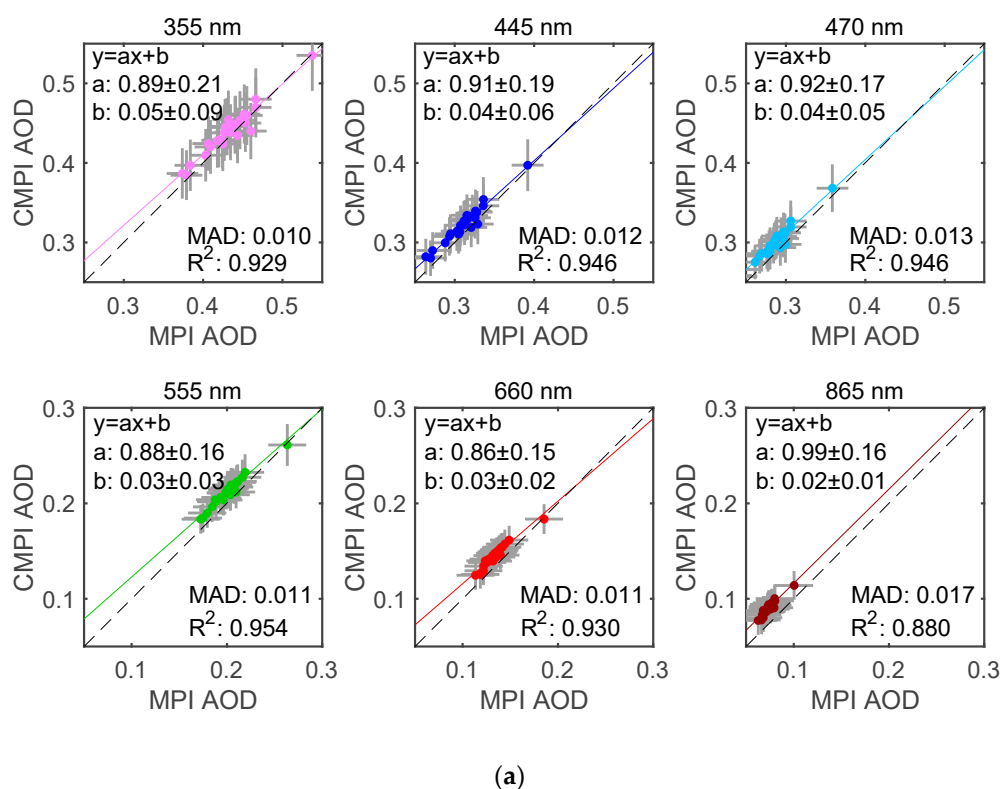
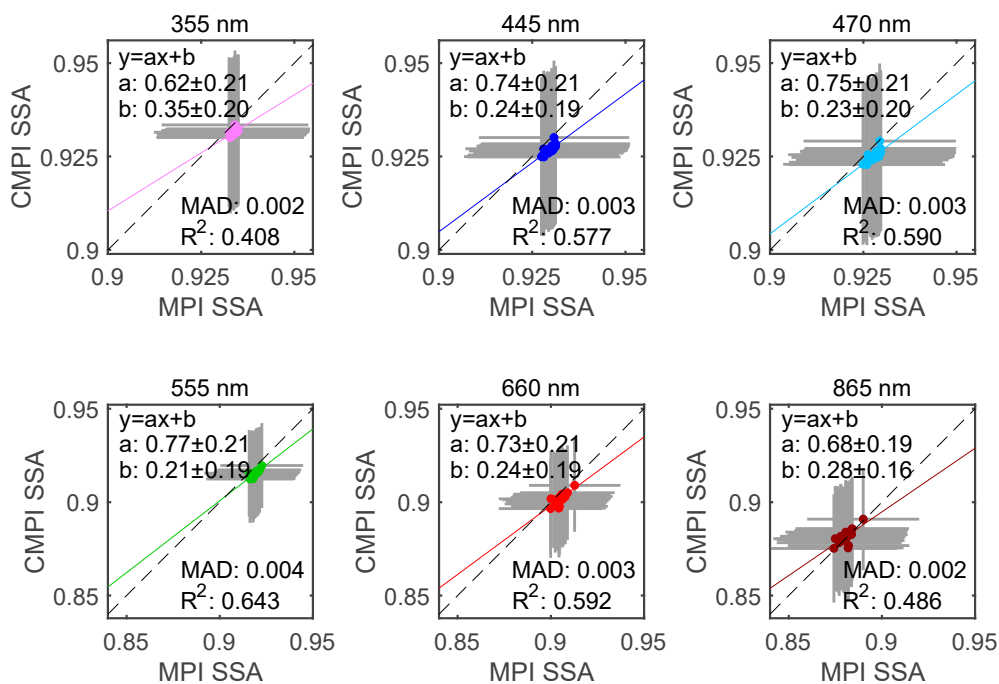
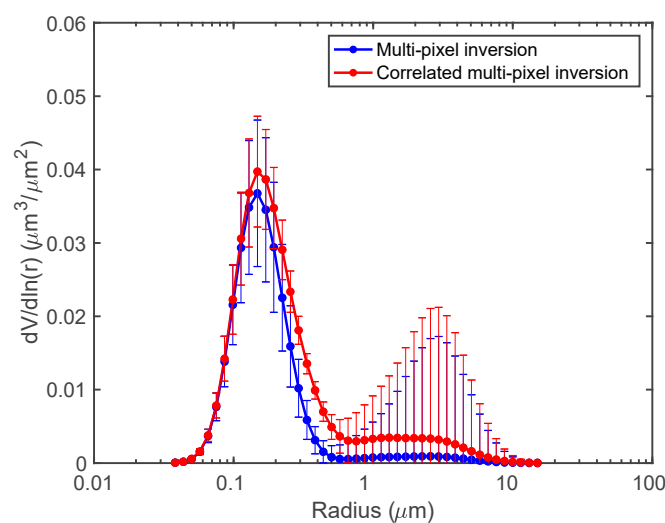


Figure 9. Cont.



(b)

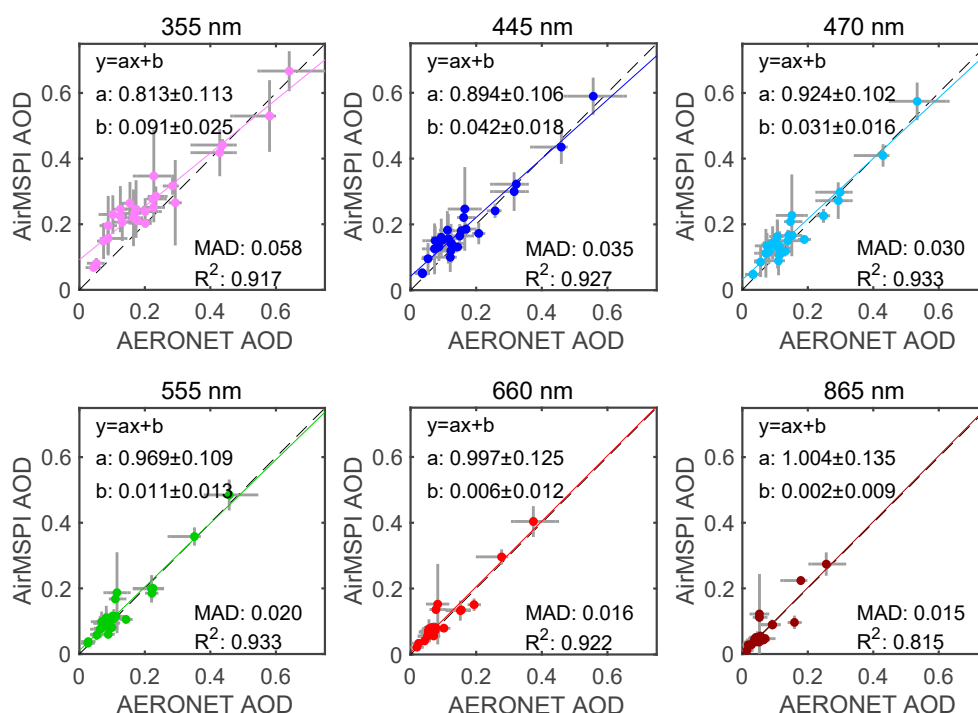


(c)

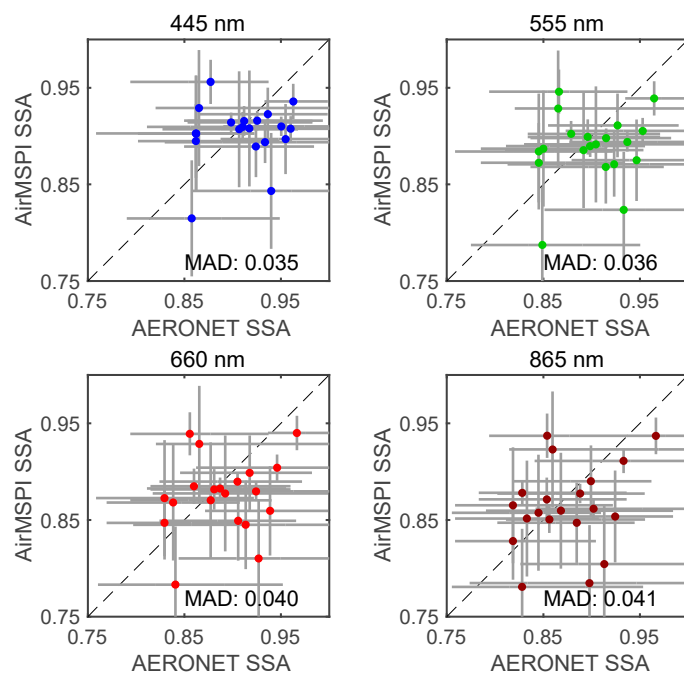
**Figure 9.** (a) Regression of pixel-scale AOD retrieved from correlated multi-pixel inversion (CMPI) against the previous retrievals using multi-pixel inversion (MPI, [19]). The AirMSPI dataset used for retrieval is the same as in Figure 5 over the AERONET Baskin site. The results for seven AirMSPI spectral bands are plotted in different colors: pink (355 nm), purple (380 nm), dark blue (445 nm), light blue (470 nm), green (555 nm), red (660 nm), and brown (865 nm). (b) The same as Figure 9a but for SSA. (c) Comparison of image-mean volume-weighted aerosol size distribution retrieved from correlated multi-pixel inversion and multi-pixel inversion.

Figures 10 and 11 compare AirMSPI and AERONET retrievals of AOD and SSA respectively. Since the temporal variation of aerosol loading and properties is not constant, non-symmetric AERONET error bars can be observed in Figures 10 and 11. Under the circumstance of no AERONET

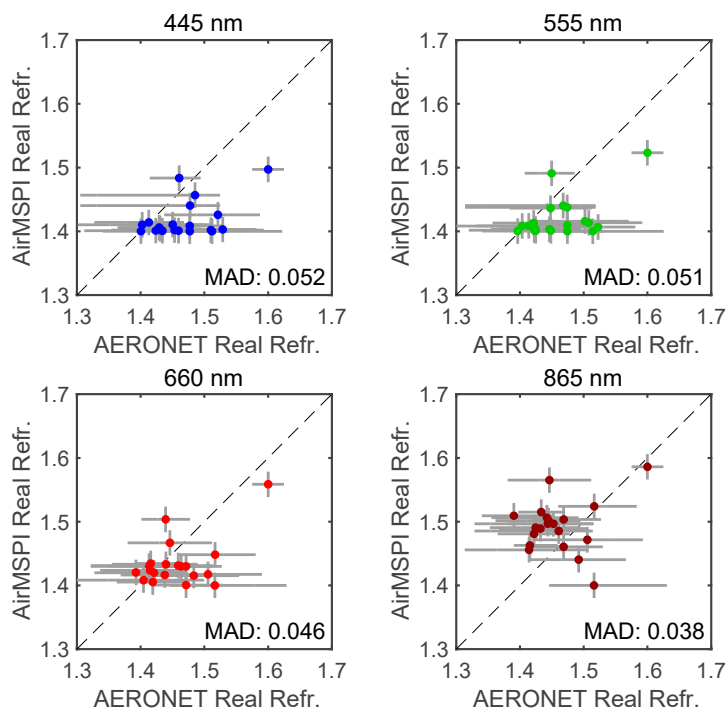
reference data before or after AirMSPI measurement, only AERONET measurement/retrieval error is reported. Figures 12 and 13 compare real and imaginary parts of aerosol refractive index, respectively. The AERONET spectral aerosol products were linearly interpolated in wavelength to match the AirMSPI band centers. Comparisons of fine and coarse mode effective radii are shown in Figure 14. To facilitate the comparison of aerosol size, an effective radius was calculated for fine and coarse mode aerosols from AirMSPI retrievals using Equation (35) in a previous study [19]. For AOD, linear regression is performed. Values of regression related parameters ( $a$ ,  $b$ ,  $R^2$ ) and MAD are indicated in all panels. The AOD regression shows a spectral means of coefficient of determination 0.91, slope 0.93, and intercept 0.03, reflecting high retrieval quality. While SSA and refractive index in Figures 11–13 show relatively larger differences between the AirMSPI and AERONET retrievals, the differences are generally within their respective uncertainties, which in turn depend on AirMSPI and AERONET observation errors and the sensitivities of the respective retrieval algorithms. Figure 14 shows a maximum difference of 30% between AirMSPI and AERONET retrieved fine mode aerosol size, whereas larger differences (up to 80%) are observed in coarse mode aerosol size. As noted above, shortwave infrared spectral bands, which AirMSPI lacks, are necessary to constrain the coarse mode aerosol size.



**Figure 10.** Regression of AirMSPI retrieved aerosol optical depth (AOD) against AERONET measured values. The upper three panels are for 355, 445, and 470 nm, and the lower three panels are for 555, 660, and 865 nm. Linear interpolation is used to obtain AERONET AOD values at the AirMSPI wavelengths. The AERONET uncertainties are from the  $\pm 1$ -h window around the time of AirMSPI overflight plus measurement uncertainties (0.01), while the AirMSPI uncertainties are the root mean square of the retrieval uncertainties and standard deviation of pixel-resolved AODs over the whole image. Linear regression analysis yields values of slope  $a$ , intercept  $b$ , coefficient of determination  $R^2$ , and mean absolute difference (MAD). Values of each are indicated in all panels.



**Figure 11.** Comparison of AirMSPI retrieved single scattering albedo (SSA) against AERONET retrievals. The upper left and right panels are for 445 and 555 nm and the lower left and right panels are for 660 and 865 nm. Linear interpolation is used to obtain AERONET SSA values at the AirMSPI wavelengths. The AirMSPI errors are computed from statistics obtained over the whole image plus the errors evaluated using the method in Section 3.3. Values of MAD are shown.



**Figure 12.** Same as Figure 11 but for the real part of aerosol refractive index.



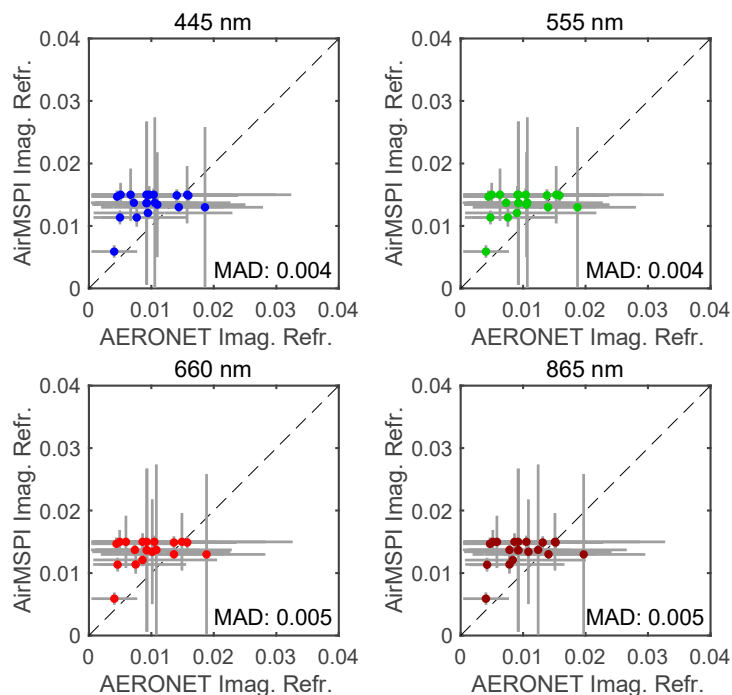


Figure 13. Same as Figure 11 but for the imaginary part of aerosol refractive index.

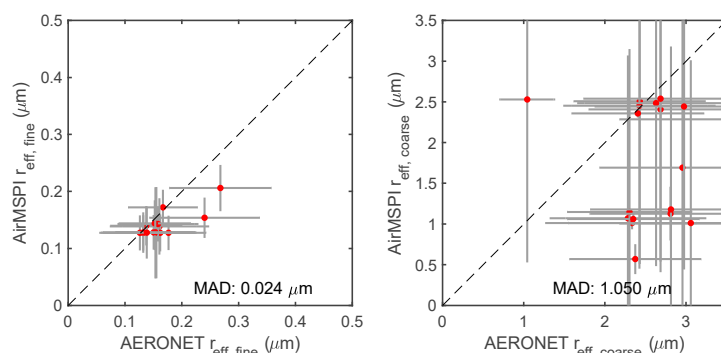


Figure 14. Same as Figure 11 but for the effective radii of fine and coarse mode aerosols.

Table 6 summarizes MAD in several key aerosol properties from correlated multi-pixel inversion approach (this paper) and from the original multi-pixel inversion approach adapted to AirMSPI [19]. These properties include AOD, SSA, real and imaginary parts of refractive index ( $n_r$ ,  $n_i$ ), and effective radii of fine and coarse particles ( $r_{\text{eff, fine}}$ ,  $r_{\text{eff, coarse}}$ ). The bias is evaluated by taking the mean of the absolute difference between AirMSPI and AERONET retrievals at collocated pixels. The correlated multi-pixel inversion has slightly higher MAD than that of original multi-pixel retrieval, namely by  $\sim 0.01$ ,  $0.015$ ,  $0.001$ ,  $0.002$ , and  $0.05$  for AOD and SSA in visible,  $n_r$ ,  $n_i$ ,  $r_{\text{eff, fine}}$ , and  $r_{\text{eff, coarse}}$ , respectively. The deviation of both correlated and original multi-pixel retrievals from AERONET reference data are mostly within the retrieval uncertainties from our algorithms (see Sections 3.3 and 4.2 of the previous study [19]) and those estimated for AERONET, as observed in Figures 10–14 of this work, as well as Figures 4–8 in the previous study [19].

Note that to enhance the retrieval efficiency, two PCs are adopted to perform the retrieval and achieve the retrieval quality in Table 6. Including more PCs in our retrieval would allow capturing more spatial variations of aerosol properties across the imaged area and improve accuracy, but at a cost of decreased retrieval efficiency. In correlated multi-pixel inversion, the retrieval efficiency is gained from two aspects: (a) reduction of parameter space by retrieving PCs of correlated fields; (2) use of PC-based RT model. For a case study with 22 correlated aerosol parameters and 30 uncorrelated (assumed)

surface parameters per super-pixel, the correlated multi-pixel saves 50% CPU time by using two PCs. The speed gain can be greater if we further capitalize the correlation in surface parameters, as well as the correlation between aerosol and surface parameters. It is anticipated that correlated multi-pixel inversion will save 90% CPU time if an image of  $100 \times 100$  pixels is retrieved simultaneously and all 52 parameters are correlated with each other.

**Table 6.** Mean absolute bias (MAD) in key aerosol properties from correlated multi-pixel inversion approach using two PCs (this paper) and from original multi-pixel inversion approach adapted to AirMSPI [19]. The bias is calculated by taking the mean of the absolute difference between AirMSPI and AERONET retrievals at collocated pixels. A total of 27 AirMSPI datasets are used. Interpolation is used to calculate some AERONET parameters at central wavelengths of the AirMSPI spectral bands. By performing regression against AERONET reference AOD, the coefficients of determination ( $R^2$ ) are given in columns 3 and 5. The ranges of non-AOD parameters are too small and the sample size is limited. Therefore a reliable regression analysis is not established and  $R^2$  is reported only for AOD in the table.

Parameters	Correlated Multi-pixel Inversion		Original Multi-pixel Inversion	
	MAD	$R^2$	MAD	$R^2$
AOD <sub>355 nm</sub>	0.058	0.917	0.040	0.925
AOD <sub>445 nm</sub>	0.035	0.927	0.024	0.955
AOD <sub>470 nm</sub>	0.030	0.933	0.021	0.960
AOD <sub>555 nm</sub>	0.020	0.933	0.016	0.960
AOD <sub>660 nm</sub>	0.016	0.922	0.013	0.959
AOD <sub>865 nm</sub>	0.015	0.815	0.014	0.851
SSA <sub>445 nm</sub>	0.035	-	0.030	-
SSA <sub>555 nm</sub>	0.036	-	0.030	-
SSA <sub>660 nm</sub>	0.040	-	0.032	-
SSA <sub>865 nm</sub>	0.041	-	0.035	-
$n_{r,445 nm}$	0.052	-	0.039	-
$n_{r,555 nm}$	0.051	-	0.039	-
$n_{r,660 nm}$	0.046	-	0.036	-
$n_{r,865 nm}$	0.038	-	0.037	-
$n_{i,445 nm}$	0.004	-	0.004	-
$n_{i,555 nm}$	0.004	-	0.004	-
$n_{i,660 nm}$	0.005	-	0.004	-
$n_{i,865 nm}$	0.005	-	0.005	-
$r_{eff,fine} (\mu m)$	0.024	-	0.022	-
$r_{eff,coarse} (\mu m)$	1.050	-	0.993	-

## 6. Summary and Outlook

Without utilizing correlations among aerosol parameters, optimization-based retrievals are confronted with a high-dimensional parameter space. However, certain types of aerosols or certain combinations of aerosol fields generally prevail within a targeted area, and consequently some aerosol parameters are correlated with each other (in other words, have high linear dependency, as captured by PC analysis). Due to the lack of accurate physical models, however, it is hard to quantify the physical processes and accurately quantify the correlations between all parameters. To mitigate the influence of model assumptions, *a priori* information about aerosol correlation informed by ground or other types of measurements is helpful. This motivates our development of PC-based aerosol inversion approach to improve the inversion stability and efficiency by reducing the number of retrieval parameters. The algorithm makes use of multiple types of constraints, including across-pixel smoothness constraints transformed to be imposed on the PCs to retrieve zero sum of the PC weights, and orthogonality and unit norm conditions on PC weights and PC vectors, respectively. While applying smoothness constraints to the PCs instead of the individual aerosol parameters, the regularization (smoothness) remains faithful to both aerosol correlations and smooth spatial and spectral variation within the

scene. To accelerate the multi-pixel retrieval, a PC-based RT model is developed, which capitalizes on the aerosol correlations and mutual orthogonality of the PC vectors. The retrieval methodology was tested by comparing aerosol retrievals from 27 AirMSPI datasets acquired between 2012 and 2016 with collocated aerosol reference data reported by AERONET. Mean absolute differences between AirMSPI and AERONET retrievals are found to be  $\sim 0.029$  and  $0.038$  for AOD and SSA, respectively.

The correlated multi-pixel inversion established in this paper is informed by prior estimates of aerosol properties within the retrieval in order to generate initial guesses for the PC vectors and to calculate the co-variance of *a priori*. Potential sources of such information can include AERONET climatologies, chemical transport model results, or satellite-based aerosol inversion output obtained, for example, from the MISR operational aerosol retrieval [13,14], and the near-real time POLDER aerosol retrieval [35]. Though not implemented in this study, it would be interesting to further impose on retrieval the correlations in surface properties and the correlations between aerosol and surface properties. To derive *a priori* of the correlations in surface properties, PC analysis can be performed over a surface reflectance dataset, such as the one based on the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm [36,37], the ASTER spectral library [38], and the U.S. Geological Survey (USGS) digital spectral library [39]. To impose the correlation between aerosol and surface, however, one might encounter a complex surface condition over a targeted area (e.g., with co-existence of multiple types of surfaces such as soil, water, snow, forest, etc.). Under such circumstances, surface pixel classification can be performed to identify surface types first, either from existing surface climatology or from various indices for vegetation, snow, soil, water, etc. based on their spectral difference. Then one can assign surface-type dependent sets of PCs to pixels, each set containing a small number of PCs. Finally, these surface-type dependent sets of PCs are retrieved simultaneously with the imposition of the constraints regarding smooth variations of relevant aerosol/surface properties in spatial, spectral and/or temporal directions. To account for dependent sets of PCs in a state vector, some modifications are necessary in formulating the smoothness constraints. Using such a strategy is expected to gain more retrieval efficiency than if one directly relaxes more PCs in retrievals to capture strong spatial variations of surface properties.

While the correlated multi-pixel approach developed here allows a retrieval of both PC vectors and PC weights, another way to capture regionally limited variability in aerosol type is to use a traditional lookup table (LUT) based aerosol retrieval. This approach is implemented in some operational aerosol retrievals employed by, e.g., Multi-angle Imaging SpectroRadiometer (MISR) [13,14] and Moderate resolution Imaging Spectroradiometer (MODIS) [40], and has extremely high retrieval efficiency. With reliable estimates of aerosol properties from other sources as noted above, a “smart” LUT can be generated, which then serves as the basis for a reliable set of PC vectors. If there is high confidence in the representativeness of these PCs, the retrieval could be confined to the PC weights only, which will be faster than the combined inversion of PC vectors and PC weights. Such an approach would compensate for the weakness in traditional LUT approach, namely that aerosol mixtures are confined to a discretized aerosol parameter space. Further testing of these ideas are planned using multi-angle satellite observations from MISR [41].

**Author Contributions:** F.X. formulated the correlation-based multi-pixel inversion approach and radiative transfer model for fast multi-pixel polarized radiance modeling, and tested the approach using AirMSPI data. D.J.D. and O.D. reviewed the whole formalism and provided editorial changes. Y.S. provided discussions on imposing correlation constraint to improve aerosol retrieval and editorial suggestions.

**Funding:** Feng Xu and David J. Diner’s work was performed at the Jet Propulsion Laboratory (JPL), California Institute of Technology under contract with the National Aeronautics and Space Administration. Oleg Dubovik is supported by the CaPPA (Chemical and Physical Properties of the Atmosphere) project funded by the French National Research Agency through PIA (Programme d’Investissement d’Avenir) program under contract (ANR-11-LABX-0005-01), the Hauts-de-France Regional Council, and the European Funds for Regional Economic Development. Yoav Schechner is a Landau Fellow supported by the Taub Foundation. His research, supported by the US-Israel Binational Science Foundation (BSF grant 2016325), is partly conducted in the Ollendorff Minerva Center. Minerva is funded through the BMBF.

**Acknowledgments:** We are grateful to Carol J. Bruegge, Michael J. Garay, Gerard van Harten, Veljko M. Jovanovic, Olga V. Kalashnikova, Brian E. Rheingans, Felix C. Seidel, Irina N. Tkatcheva, and Mika Tosca of JPL for their support of AirMSPI data acquisition and calibration in multiple NASA field campaigns. We also thank Xu Liu at NASA Langley Research Center and Vijay Natraj at JPL for helpful discussions on principal component analysis to accelerate radiative transfer modeling and atmospheric remote sensing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A Symbols and Abbreviations

**Table A1.** List of symbols and abbreviations.

Symbol/Abbreviation	Description
$a$	Slope of a regression line
$\bar{a}$	Mean slope of a set of regression lines
$a_\lambda$	Spectral weight of surface BRDF
$A$	Coefficient for a basic smooth function shape, e.g., " $g(z) = Az^2 + Bz + C$ " for a parabola
$\mathbf{A}$	Fischer matrix
AERONET	Aerosol Robotic Network
AirMSPI	Airborne Multiangle SpectroPolarimetric Imager
AOD	Aerosol optical depth
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
$A_{\text{surf}}$	Surface albedo
$b$	Intercept of a regression line
$\bar{b}$	Mean intercept of a set of regression lines
$B$	Coefficient for a basic smooth function shape, e.g., " $g(z) = Az^2 + Bz + C$ " for a parabola
BRDF	Bidirectional reflectance distribution function
BRF	Bidirectional reflectance factor (BRF) associated with Stokes vector component I
$\mathbf{c}$	Constant column vector that weight correlated fields in PC analysis
$C$	Coefficient for a basic smooth function shape, e.g., " $g(z) = Az^2 + Bz + C$ " for a parabola
$\mathbf{C}_i$	Covariance matrix of $i$ -th type of constraint
$\mathbf{C}_{\Delta x, \text{rand}}$	Covariance matrix of random errors in the measurements
$C_{v, \text{sphere}}$	Column volume concentration of spherical aerosols
$C_{v, \text{tot}}$	Total column volume concentration of aerosols
$\mathbf{d}$	Differentiation array
$d_{\text{ES}}$	Earth-Sun distance
DOLP	Degree of linear polarization
$dV/d\ln(r)$	Volume weighted aerosol size distribution
$E_0$	Exo-atmospheric solar irradiance
$\mathbf{f}_i^*$	Column vector of $i$ -th type of constraint
$\mathbf{f}_i$	Column vector that contains model prediction to fit $i$ -th type of constraint
$\Delta \mathbf{f}_i^*$	Column vector that contains the errors of $i$ -th type of constraint
$f_{\text{fine}}$	Fine mode aerosol fraction
$f_{v, \text{sphere}}$	Volume fraction of spherical aerosols
$g(z)$	Smooth function with variable $z$
$g_\lambda$	Spectral anisotropy parameter of surface BRDF
GRASP	Generalized Retrieval of Aerosol and Surface ProperEes
$h$	Cartesian coordinate in the direction $h$
$h_a$	Central height of aerosol vertical profile (constrained by Gaussian profile)
$I$	First Stokes vector component
$\mathbf{I}$	Identity matrix
$I_{\text{meas}}$	Measured radiance
$k_\lambda$	Spectral anisotropy parameter of surface BRDF
$k_\gamma$	Shadowing width of polarized BRDF
$\mathbf{K}$	Jacobian matrix
$\mathbf{K}_i$	The Jacobian matrix associated with $i$ -th type of constraint
$L(j)$	Length of $j$ -th type of correlated parameters (fields)
$m$	Order of difference in constructing smoothness matrix
$M$	Number of types of constraints imposed on retrieval
MAD	Mean absolute difference
MAIAC	MultiAngle Implementation of Atmospheric Correction
MISR	Multi-angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
$n_{r, j}$	Real part of refractive index at $j$ -th wavelength
$n_{i, j}$	Imaginary part refractive index at $j$ -th wavelength
$N_a$	Total number of retrieval parameters
$N_{a^*}$	Total number of <i>a priori</i> estimate of parameters
$N_c$	Total number of constraints imposed on retrieval

Table A1. Cont.

Symbol/Abbreviation	Description
$N_{\text{corr}}$	Number of correlated parameters (fields)
$N_f$	Total number of observations (all pixels are accounted)
$N_i$	Total number of $i$ -th type of constraint
$N_{\text{PC}}$	Total number of principle components
$N_{\text{pixel}}$	Total number of pixels
$N_{\text{SC}}$	Number of aerosol size components
$N_{\text{TP,uncorr}}$	Total number of types of uncorrelated parameters
$\mathbf{O}$	Constraining matrix that reflects zero-sum of PC weights
pBRDF	Polarized BRDF
$P$	Probability distribution function (PDF)
PC	Principal component
PCA	Principal component analysis
POLDER	Polarization and Directionality of Earth's Reflectance
$q$	Ratio of Stokes components Q and I
$q$	Iterative step during optimization
$q\text{BRF}_{\text{TOA}}$	Top-of-atmosphere BRF associated with Stokes vector component Q
$Q$	Second Stokes vector component
$r$	Radius of aerosol
$r_{\text{eff,coarse}}$	Effective radius of coarse mode aerosols
$r_{\text{eff,fine}}$	Effective radius of fine mode aerosols
$\mathbf{R}$	Reflection matrix
$R^2$	Coefficient of determination
$\mathbf{R}_{\text{atmos}}$	Reflection matrix for atmosphere associated with light illumination from top of the atmosphere
$\mathbf{R}_{\text{atmos}}^*$	Reflection matrix for atmosphere associated with light illumination from bottom of the atmosphere
$\mathbf{R}_{\text{CAS}}$	Reflection matrix for the coupled atmosphere-surface system (CAS)
RPV	Rahman-Pinty-Verstraete (surface BRDF model)
$\mathbf{R}_{\text{surf}}$	Surface reflection matrix
$\mathbf{R}_{\text{surf, BRDF}}$	Depolarizing part of surface reflection matrix
$\mathbf{R}_{\text{surf, pBRDF}}$	Polarizing part of surface reflection matrix
RT	Radiative transfer
$s_a$	Standard deviation of aerosol vertical profile (constrained by Gaussian profile)
$\mathbf{S}_{i,m}$	Differentiation matrix of $m$ -th order for $i$ -th type of constraint
SSA	Single scattering albedo
$t$	Temporal coordinate
$\mathbf{T}$	Transmission matrix
$\mathbf{T}_{\text{atmos}}$	Transmission matrix for atmosphere associated with light illumination from top of the atmosphere
$\mathbf{T}_{\text{atmos}}^*$	Transmission matrix for atmosphere associated with light illumination from bottom of the atmosphere
TOA	Top-of-atmosphere
$u$	Cartesian coordinate in the direction $u$
$u$	Ratio of Stokes components U and I
$u\text{BRF}_{\text{TOA}}$	Top-of-atmosphere BRF associated with Stokes vector component U
$U$	Third Stokes vector component
$\mathbf{U}$	Constraining matrix that reflects the unit length of a PC
USGS	U.S. Geological Survey
$v$	Cartesian coordinate in the direction $v$
$\mathbf{v}$	PC matrix containing $N_{\text{PC}}$ columns PC vectors
$v\text{BRF}_{\text{TOA}}$	Top-of-atmosphere BRF associated with Stokes vector component V
$\mathbf{v}_k$	The $k$ -th PC vector
$\mathbf{v}_{\text{state}}$	Column vector containing all PC vectors
$\mathbf{w}$	PC weight matrix containing $N_{\text{pixel}}$ column vectors containing PC weights
$\mathbf{w}_i$	Weight matrix for $i$ -th type of constraint
$\mathbf{w}_p$	Column vector containing PC weights for $p$ -th pixel
$\mathbf{w}_{\text{state}}$	Column vector containing all PC weights
$\mathbf{x}$	Column state vector including all retrieval parameters
$\mathbf{x}^a \text{ priori}$	<i>a priori</i> of state vector
$\bar{\mathbf{x}}_{\text{corr}}$	Column vector containing spatial and temporal mean of correlated parameters (fields)
$\mathbf{x}_{\text{corr},p}$	Column vector containing correlated parameters (fields) for $p$ -th pixel
$\mathbf{x}_{\eta,\text{aer}}$	The vector consisting of correlated aerosol properties – calculated from the solution at $q$ -th iteration
$\mathbf{x}_{\eta,\text{surf}}$	The vector consisting of uncorrelated surface reflection properties – containing in the solution at $q$ -th iteration
$\mathbf{x}^{\text{retrieved}}$	Retrieved column state vector
$\Delta\mathbf{x}_{\text{syst}}$	Systematic error in retrieval
$\mathbf{x}_{\text{uncorr},p}$	Column vector containing uncorrelated parameters (fields) for $p$ -th pixel
$\mathbf{x}^{\text{true}}$	Column state vector associated with true solution
$\mathbf{x}_{\text{wv}}$	Column vector including PC weights and vectors
$(\Delta x)_j$	The retrieval error in $j$ -th parameter

Table A1. Cont.

Symbol/Abbreviation	Description
$y_i$	$i$ -th observational signal
$\mathbf{Y}^{\text{HS}}$	Output of RT calculation with high stream approximation
$\mathbf{Y}^{\text{LS}}$	Output of RT calculation with low stream approximation
$z$	Variable of a smooth function
$z_{\min}$	Lower bound of $z$
$z_{\max}$	Upper bound of $z$
$Z$	Number of observations per pixel
$\delta$	Kronecker delta
$\delta_s$	Scale factor that perturbs a PC vector
$\epsilon_{\text{rand}}$	Random error in measurements
$\epsilon_\lambda$	Spectral weight of pBRDF
$\epsilon_1^2$	First diagonal element of $\mathbf{C}_i$
$\epsilon_c^2$	User-specified threshold value to diagnose the convergence of optimization
$\epsilon_f^2$	Expected variance due to measurement errors
$\theta_0$	Solar zenith angle
$\theta_v$	View zenith angle
$\lambda$	Wavelength
$\mu_0$	Cosine of solar zenith angle
$\gamma_i$	Lagrange factor for $i$ -th type of constraint
$\phi_0$	Solar azimuthal angle
$\Psi_i$	Objective cost function for $i$ -th type of constraint
$\Psi_{\text{total}}$	Overall objective cost function
$\nabla \Psi_i$	Gradient of the objective cost function for $i$ -th type of constraint
$\nabla \Psi_{\text{total}}$	Gradient of the overall objective cost function
$\Omega_i$	The smoothness matrix associated with $i$ -th type of constraints
$\Omega^{\text{Ra}}$	The rearranged smoothness matrix from $\Omega_{\dots}$
$\Omega_{\text{uncorr}}$	The smoothness matrix imposed on uncorrelated parameters (fields)
$\bar{\Omega}_{\text{corr}}$	The smoothness matrix imposed on spatial and temporal mean of correlated parameters (fields)
$\Omega_{\text{corr}}^{\text{w}}$	Smoothness matrix imposed on pixel resolved PC weights
$\Omega_{\text{corr}}^{\text{v}}$	Smoothness matrix imposed on a PC vector
$\Omega_{\text{corr}}^{\text{wv}}$	Smoothness matrix imposed on correlated parameters (fields)
$\sigma_{\text{spatio-temporal}, x}$	Standard deviation of a correlated field $x$
$\sigma_{e, x}$	Uncertainty estimate of a correlated field $x$
$\sigma_s$	Slope variance of polarized BRDF
$\tau_{\text{aer,tot}}$	Total aerosol optical depth
$\tau_{\text{aer,abs}}$	Total absorption aerosol optical depth
$\tau_{\text{atmos}}$	Atmospheric optical depth
$\Gamma$	Constraining matrix that reflects the mutual orthogonality in PCs

## Appendix B Smoothness Matrix to Constrain Uncorrelated Parameter Retrieval

To explain construction of the smoothness matrix for PC vectors and weights, we start by describing the smoothness matrix used in the original multi-pixel algorithm formulated by Dubovik et al. [6]. This appendix forms the basis for extension to PC-based smoothness matrix, described in Appendices C and D. Two major classes of constraints are imposed on the PC retrieval: across-pixel (spatial) constraints and within-pixel (e.g., spectral) constraints. The following matrix incorporates both across-pixel and within-pixel constraints for a set of uncorrelated aerosol parameters:

$$\gamma_{\text{uncorr}} \mathbf{\Omega}_{\text{uncorr}} = \gamma_{\text{uncorr}, \Delta} \mathbf{\Omega}_{\text{uncorr}, \Delta} + \begin{pmatrix} \gamma_{\text{uncorr}, \diamond} \mathbf{\Omega}_{\text{uncorr}, 1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \gamma_{\text{uncorr}, \diamond} \mathbf{\Omega}_{\text{uncorr}, 2} & \cdots & \mathbf{0} \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \gamma_{\text{uncorr}, \diamond} \mathbf{\Omega}_{\text{uncorr}, N_{\text{pixel}}} \end{pmatrix} \quad (\text{A1})$$

where  $\gamma_{\text{uncorr}, \Delta} \mathbf{\Omega}_{\text{uncorr}, \Delta}$  is a block matrix that includes across-pixel smoothness constraints,  $\gamma_{\text{uncorr}, \diamond} \mathbf{\Omega}_{\text{uncorr}, \diamond, p}$  is a block matrix that includes within-pixel smoothness constraints over the parameters associated with  $p$ -th pixel, and  $\mathbf{0}$  is the zero block matrix. Pixel resolved matrices  $\gamma_{\text{uncorr}, \diamond} \mathbf{\Omega}_{\text{uncorr}, \diamond, p}$  do not interfere with each other, so they are arranged along the diagonal axis of the large matrix on the right-hand-side of the above equation. To facilitate the use of  $\gamma_{\text{uncorr}} \mathbf{\Omega}_{\text{uncorr}}$  calculated via Equation (A1), the uncorrelated parameters are grouped together in the order of  $\mathbf{x}_{\text{uncorr}} = [\mathbf{x}(t_1); \mathbf{x}(t_2); \cdots; \mathbf{x}(t_{N_t})]$ , where  $\mathbf{x}(t_j) = [\mathbf{x}(v_1; t_j); \mathbf{x}(v_2; t_j); \cdots; \mathbf{x}(v_{N_v}; t_j)]$ ,  $\mathbf{x}(v_i; t_j) =$

$[\mathbf{x}(u_1; v_i; t_j); \mathbf{x}(u_2; v_i; t_j); \dots; \mathbf{x}(u_{N_u}; v_i; t_j)]$ , and  $\mathbf{x}(u_k; v_i; t_j)$  is a vector that contains uncorrelated parameters for the pixel  $(u_k, v_i)$  observed at temporal point  $t_j$ . Evaluations of  $\gamma_{\text{uncorr}, \Delta} \mathbf{\Omega}_{\text{uncorr}, \Delta}$  and  $\gamma_{\text{uncorr}, \diamond} \mathbf{\Omega}_{\text{uncorr}, \diamond, p}$  are discussed in Appendices B.1 and B.2, respectively.

### Appendix B.1 Across-Pixel Smoothness Matrix

The across-pixel constraints that ensure smooth variation of a parameter in space and time is expressed by,

$$\gamma_{\text{uncorr}, \Delta} \mathbf{\Omega}_{\text{uncorr}, \Delta} = \gamma_{\text{uncorr}, u} \mathbf{\Omega}_{\text{uncorr}, u} + \gamma_{\text{uncorr}, v} \mathbf{\Omega}_{\text{uncorr}, v} + \gamma_{\text{uncorr}, t} \mathbf{\Omega}_{\text{uncorr}, t} \quad (\text{A2})$$

To simplify the explanation, consider an image of  $N_u$  pixels along the horizontal spatial dimension  $u$ , two pixels along the horizontal spatial dimension  $v$  ( $N_v = 2$ ) and two successive measurements in time ( $N_t = 2$ ) of some parameter  $x$ . The construction of a smoothness matrix along the vertical direction is neglected for simplicity. In this case, the single-column state vector is arranged as,

$$\mathbf{x}_{\text{uncorr}} = \begin{bmatrix} x(u_1; v_1; t_1) \\ x(u_2; v_1; t_1) \\ x(u_3; v_1; t_1) \\ \vdots \\ x(u_{N_u}; v_1; t_1) \\ x(u_1; v_2; t_1) \\ x(u_2; v_2; t_1) \\ x(u_3; v_2; t_1) \\ \vdots \\ x(u_{N_u}; v_2; t_1) \\ \text{to continue} \end{bmatrix} \begin{array}{l} \text{continued} \\ x(u_1; v_1; t_2) \\ x(u_2; v_1; t_2) \\ x(u_3; v_1; t_2) \\ \vdots \\ x(u_{N_u}; v_1; t_2) \\ x(u_1; v_2; t_2) \\ x(u_2; v_2; t_2) \\ x(u_3; v_2; t_2) \\ \vdots \\ x(u_{N_u}; v_2; t_2) \end{array} \quad (\text{A3})$$

The corresponding smoothness matrix constraining the horizontal and temporal variation of  $x$  is given by,

$$\gamma_{\text{uncorr}, u/v/t} \mathbf{\Omega}_{\text{uncorr}, u/v/t} = \left[ \text{diag}(\sqrt{\gamma_{\text{uncorr}, u/v/t}} \mathbf{s}_{\text{uncorr}}^{(m_{\text{uncorr}, u/v/t})}) \right]^T \left[ \text{diag}(\sqrt{\gamma_{\text{uncorr}, u/v/t}} \mathbf{s}_{\text{uncorr}}^{(m_{\text{uncorr}, u/v/t})}) \right] \quad (\text{A4})$$

where  $\gamma_{\text{uncorr}, u/v/t}$  controls the strength of smoothness constraint, which varies for different types of parameters. Specifically, the component that ensures the smooth variation of  $x$  in the direction  $u$  is expressed as,

$$\sqrt{\gamma_{\text{uncorr}, u}} \mathbf{s}_{\text{uncorr}, u}^{(m_{\text{uncorr}, u})} = \sqrt{\gamma_{\text{uncorr}, u}} \begin{bmatrix} \mathbf{s}_{\text{uncorr}, u}^{(m_{\text{uncorr}, u})} & \mathbf{0} \\ \mathbf{0} & \mathbf{s}_{\text{uncorr}, u}^{(m_{\text{uncorr}, u})} \end{bmatrix} \quad (\text{A5})$$

where

$$\mathbf{s}_{\text{uncorr}, u}^{(m_{\text{uncorr}, u})} = \begin{bmatrix} d_1^{(m_{\text{uncorr}, u})}(1) & d_1^{(m_{\text{uncorr}, u})}(2) & \dots & d_1^{(m_{\text{uncorr}, u})}(m_{\text{uncorr}, u} + 1) & 0 & \dots & 0 \\ 0 & d_2^{(m_{\text{uncorr}, u})}(1) & d_2^{(m_{\text{uncorr}, u})}(2) & \dots & d_2^{(m_{\text{uncorr}, u})}(m_{\text{uncorr}, u} + 1) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_{N_u - m_{\text{uncorr}, u}}^{(m_{\text{uncorr}, u})}(1) & d_{N_u - m_{\text{uncorr}, u}}^{(m_{\text{uncorr}, u})}(2) & \dots & d_{N_u - m_{\text{uncorr}, u}}^{(m_{\text{uncorr}, u})}(m_{\text{uncorr}, u} + 1) \end{bmatrix} \quad (\text{A6})$$

As an example, the differentiation array for the first order of difference in above equation is,

$$\mathbf{d}_j^{(m_{\text{uncorr}, u}=1)} = \left[ \frac{1}{\Delta_1(j)}, -\frac{1}{\Delta_1(j)} \right] \quad (\text{A7})$$

where  $\Delta_1(j)$  accounts for the distance between neighboring pixels in the direction  $u$  and is evaluated by Equation (40), and for the second order of difference,

$$\mathbf{d}_j^{(m_{\text{uncorr},u}=2)} = \begin{bmatrix} \frac{2}{\Delta_1(j)[\Delta_1(j)+\Delta_1(j+1)]}, & -\frac{2}{\Delta_1(j)\Delta_1(j+1)}, & \frac{2}{\Delta_1(j+1)[\Delta_1(j)+\Delta_1(j+1)]} \end{bmatrix} \quad (\text{A8})$$

Moreover, the smoothness matrix in the direction  $v$  is,

$$\sqrt{\gamma_{\text{uncorr},v}} \mathbf{S}_{\text{uncorr},v}^{(m_{\text{uncorr},v})} = \sqrt{\gamma_{\text{uncorr},v}} \begin{bmatrix} \mathbf{S}_{\text{uncorr},v}^{(m_{\text{uncorr},v})} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\text{uncorr},v}^{(m_{\text{uncorr},v})} \end{bmatrix} \quad (\text{A9})$$

where

$$\mathbf{S}_{\text{uncorr},v}^{(m_{\text{uncorr},v}=1)} = \begin{bmatrix} d_1^{(m_{\text{uncorr},v})(1)} & 0 & \dots & 0 & d_1^{(m_{\text{uncorr},v})(2)} & 0 & \dots & 0 \\ 0 & d_1^{(m_{\text{uncorr},v})(1)} & \dots & 0 & 0 & d_1^{(m_{\text{uncorr},v})(2)} & \dots & 0 \\ \dots & \dots & \ddots & \dots & \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & d_1^{(m_{\text{uncorr},v})(1)} & 0 & 0 & \dots & d_1^{(m_{\text{uncorr},v})(2)} \end{bmatrix} \quad (\text{A10})$$

where we assume  $m_{\text{uncorr},v} = 1$  and the calculation of the differentiation array  $\mathbf{d}$  accounts for the distance between neighboring pixels in the direction  $v$  and is evaluated by Equation (40).

The smoothness matrices in the direction  $t$  is,

$$\mathbf{S}_{\text{uncorr},t}^{(m_{\text{uncorr},t}=1)} = \begin{bmatrix} d_1^{(m_{\text{uncorr},t})(1)} & 0 & \dots & 0 & 0 & 0 & \dots & 0 & d_1^{(m_{\text{uncorr},t})(2)} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & d_1^{(m_{\text{uncorr},t})(1)} & \dots & 0 & 0 & 0 & \dots & 0 & 0 & d_1^{(m_{\text{uncorr},t})(2)} & 0 & \dots & 0 & 0 & \dots & 0 \\ d_1^{(m_{\text{uncorr},t})(1)} & 0 & \dots & 0 & 0 & 0 & \dots & 0 & d_1^{(m_{\text{uncorr},t})(2)} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & d_1^{(m_{\text{uncorr},t})(1)} & \dots & 0 & 0 & 0 & \dots & 0 & 0 & d_1^{(m_{\text{uncorr},t})(2)} & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \quad (\text{A11})$$

where we assume  $m_{\text{uncorr},t} = 1$  and the calculation of the differentiation array  $\mathbf{d}$  accounts for the temporal gap between successive measurements and is evaluated by Equation (40).

Note that the above formalism applies to all uncorrelated but smoothly varying parameters. The incorporation of these smoothness matrices for different parameters into an overall matrix  $\gamma_{\text{uncorr},\Delta} \mathbf{\Omega}_{\text{uncorr},\Delta}$  is designed to account for the locations of these parameters in a retrieval state vector.

## Appendix B.2 Within-Pixel Smoothness Matrix

Certain types of parameters, such as the spectral weight of the microfacet model-based pBRDF function and spectral shape parameters in the RPV model-based BRDF function as a function of wavelength, are subjected to inherent smoothness. The smooth variation of such a type of within-pixel parameters is ensured by

$$\gamma_{\text{uncorr},\diamond} \mathbf{\Omega}_{\text{uncorr},\diamond} = \begin{bmatrix} (\gamma_{\text{uncorr},\diamond} \mathbf{\Omega}_{\text{uncorr},\diamond})_{p1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & (\gamma_{\text{uncorr},\diamond} \mathbf{\Omega}_{\text{uncorr},\diamond})_{p2} & \dots & \mathbf{0} \\ \dots & \dots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & (\gamma_{\text{uncorr},\diamond} \mathbf{\Omega}_{\text{uncorr},\diamond})_{N_{\text{pixel}}} \end{bmatrix} \quad (\text{A12})$$

where for any pixel  $p$ ,

$$(\gamma_{\text{uncorr},\diamond} \mathbf{\Omega}_{\text{uncorr},\diamond})_p = \left[ \sqrt{\gamma_{\text{uncorr},\diamond}} \mathbf{S}_{\text{uncorr},\diamond}^{(m_{\text{uncorr},\diamond})} \right]_p^T \left[ \mathbf{W}^{(m_{\text{uncorr},\diamond})} \right]^{-1} \left[ \sqrt{\gamma_{\text{uncorr},\diamond}} \mathbf{S}_{\text{uncorr},\diamond}^{(m_{\text{uncorr},\diamond})} \right]_p \quad (\text{A13})$$

where

$$\left[ \sqrt{\gamma_{\text{uncorr},\diamond}} \mathbf{S}_{\text{uncorr},\diamond}^{(m_{\text{uncorr},\diamond})} \right]_p = \begin{bmatrix} \mathbf{C}_p^{(m_{\text{uncorr},\diamond}(1))}(1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_p^{(m_{\text{uncorr},\diamond}(2))}(2) & \dots & \mathbf{0} \\ \dots & \dots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{C}_p^{(m_{\text{uncorr},\diamond}(N_{\text{TP,uncorr}}))}(N_{\text{TP,uncorr}}) \end{bmatrix} \quad (\text{A14})$$



with the following submatrix for  $j$ -th type of uncorrelated parameters,

$$C_p^{(m_{\text{uncorr},\phi}(j))}(j) = \sqrt{\gamma_{\text{uncorr},\phi}(j)} \begin{bmatrix} d_1^{(m_{\text{uncorr},\phi}(j))}(1) & d_2^{(m_{\text{uncorr},\phi}(j))}(2) & \dots & d_1^{(m_{\text{uncorr},\phi}(j))}(m_{\text{uncorr},\phi}(j)+1) & 0 & \dots & 0 \\ 0 & d_2^{(m_{\text{uncorr},\phi}(j))}(1) & d_2^{(m_{\text{uncorr},\phi}(j))}(2) & \dots & d_2^{(m_{\text{uncorr},\phi}(j))}(m_{\text{uncorr},\phi}(j)+1) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_{\text{TP,uncorr}(j)-m_{\text{uncorr},\phi}(j)}^{(m_{\text{uncorr},\phi}(j))}(1) & d_{\text{TP,uncorr}(j)-m_{\text{uncorr},\phi}(j)}^{(m_{\text{uncorr},\phi}(j))}(2) & \dots & d_{\text{TP,uncorr}(j)-m_{\text{uncorr},\phi}(j)}^{(m_{\text{uncorr},\phi}(j))}(m_{\text{uncorr},\phi}(j)+1) \end{bmatrix} \quad (\text{A15})$$

where we assume  $j$ -th parameter belongs to  $[\text{TP}_{\text{uncorr}}(j)]$ -th type of uncorrelated parameters, the calculation of the differentiation array  $\mathbf{d}$  accounts for the distance between neighboring wavelengths and proceeds in the way as expressed in Equations (A7) and (A8).

The weight matrix in Equation (A13) is expressed as,

$$\mathbf{W}^{(m_{\text{uncorr},\phi})} = \begin{bmatrix} \mathbf{W}^{(m_{\text{uncorr},\phi}(1))}(1) & 0 & 0 & \dots & 0 \\ 0 & \mathbf{W}^{(m_{\text{uncorr},\phi}(2))}(2) & 0 & \dots & 0 \\ 0 & 0 & \mathbf{W}^{(m_{\text{uncorr},\phi}(3))}(3) & \dots & 0 \\ \dots & \dots & \dots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \mathbf{W}^{(m_{\text{uncorr},\phi}(N_{\text{TP,uncorr}}))}(N_{\text{TP,uncorr}}) \end{bmatrix} \quad (\text{A16})$$

where

$$\mathbf{W}^{(m_{\text{uncorr},\phi}(i))}(i) = \begin{bmatrix} \frac{1}{\Delta_{m_{\text{uncorr},\phi}(i)}(1)} & 0 & \dots & 0 \\ 0 & \frac{1}{\Delta_{m_{\text{uncorr},\phi}(i)}(2)} & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \frac{1}{\Delta_{m_{\text{uncorr},\phi}(i)}(L(i)-m_{\text{uncorr},\phi}(i))} \end{bmatrix} \quad (\text{A17})$$

where  $\Delta_m(j)$  is evaluated by Equation (40).

To generate  $\gamma_{\text{corr}}^{\bar{x}} \mathbf{\Omega}_{\text{corr}}^{\bar{x}}$  (see Equation (50) for  $\mathbf{\Omega}_{\text{corr}}^{\bar{x}}$ ), the evaluation of smoothness matrix of the correlated fields in the column vector  $\bar{x}$  containing multi-pixel mean fields proceeds in a similar way as formulated for the within-pixel constraint.

### Appendix C Smoothness Matrix for Correlated Parameters

As shown in Appendix B, the original multi-pixel inversion algorithm imposes smoothness constraints directly on the aerosol fields. In the correlated multi-pixel inversion, the correlated parameters are not directly retrieved; the retrieval involves the PC vectors, PC weights, and multi-pixel mean fields. Therefore, the smoothness constraints must be imposed on PCs and PC weights. Like the implementation for uncorrelated parameter retrievals, the overall smoothness matrix includes both across-pixel ( $\gamma_{\text{corr},\Delta}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\Delta}^{\text{wv}}$ ) and within-pixel constraints ( $\gamma_{\text{corr},\phi}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\phi}^{\text{wv}}$ ), namely,

$$\begin{cases} \gamma_{\text{corr}}^{\text{wv}} \mathbf{\Omega}_{\text{corr},1}^{\text{wv}} = \gamma_{\text{corr},\Delta}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\Delta,1}^{\text{wv}} + \gamma_{\text{corr},\phi}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\phi,1}^{\text{wv}} \\ \gamma_{\text{corr}}^{\text{wv}} \mathbf{\Omega}_{\text{corr},2}^{\text{wv}} = \gamma_{\text{corr},\Delta}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\Delta,2}^{\text{wv}} + \gamma_{\text{corr},\phi}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\phi,2}^{\text{wv}} \end{cases} \quad (\text{A18})$$

The explicit forms of  $\gamma_{\text{corr},\Delta}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\Delta}^{\text{wv}}$  and  $\gamma_{\text{corr},\phi}^{\text{wv}} \mathbf{\Omega}_{\text{corr},\phi}^{\text{wv}}$  will be given in the Appendices C.1 and C.2.

#### Appendix C.1 Across-Pixel Smoothness Constraints

To simplify the discussion, the formalism given in this and the next section assumes that the smoothness constraint applied to a parameter  $x$  occurs in the dimension  $u$ . It can be easily extended to enable construction of smoothness matrices in other spatial dimensions.

To constrain across-pixel variation of a certain parameter, the  $\gamma_{\text{corr},\Delta}^{\text{wv}} \Omega_{\text{corr},\Delta}^{\text{wv}}$  matrix in Equation (A18) is given by

$$\begin{cases} \gamma_{\text{corr},\Delta}^{\text{wv}} \Omega_{\text{corr},\Delta}^{\text{wv}} = \sum_{i=1}^{N_{\text{corr}}} \left[ \sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}} \frac{d(\mathbf{S}_{\text{wv},i}^{(m_{\text{corr},\Delta}^{\text{wv}})} \mathbf{x}_i)}{d\mathbf{x}_i} \right]^T \left[ \sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}} \frac{d(\mathbf{S}_{\text{wv},i}^{(m_{\text{corr},\Delta}^{\text{wv}})} \mathbf{x}_i)}{d\mathbf{x}_i} \right] \\ \gamma_{\text{corr},\Delta}^{\text{wv}} \Omega_{\text{corr},\Delta}^{\text{wv}} = \sum_{i=1}^{N_{\text{corr}}} \left[ \sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}} \frac{d(\mathbf{S}_{\text{wv},i}^{(m_{\text{corr},\Delta}^{\text{wv}})} \mathbf{x}_i)}{d\mathbf{x}_i} \right]^T \left[ \sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}} \mathbf{S}_{\text{wv},i}^{(m_{\text{corr},\Delta}^{\text{wv}})} \right] \end{cases} \quad (\text{A19})$$

where the column vector  $\mathbf{x}_i$  has the same length as that of the state vector and contains all-pixel PC weights and the  $i$ -th elements of all PC vectors (all rest elements of  $\mathbf{x}_i$  are zero), and  $\sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}} \mathbf{S}_{\text{wv},i}^{(m_{\text{corr},\Delta}^{\text{wv}})}$  includes two matrix components  $\mathbf{B}$  and  $\mathbf{C}$  that account for the contributions by PCs and PC weights, respectively, in the following form,

$$\sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}} \mathbf{S}_{\text{wv},i}^{(m_{\text{corr},\Delta}^{\text{wv}})} = \begin{bmatrix} 0 & \mathbf{B}_{w,p1}^{(m_{\text{corr},\Delta}^{\text{wv}})}(i) & \mathbf{C}_{p1}^{(m_{\text{corr},\Delta}^{\text{wv}})}(1,i) & \mathbf{0}_{\text{uncorr},p1} & \mathbf{C}_{p2}^{(m_{\text{corr},\Delta}^{\text{wv}})}(2,i) & \mathbf{0}_{\text{uncorr},p2} & \dots & \dots & \mathbf{C}_{p1}^{(m_{\text{corr},\Delta}^{\text{wv}})}(m_{\text{corr},\Delta}^{\text{wv}}+1,i) & \mathbf{0}_{\text{uncorr},p1+m_{\text{corr},\Delta}^{\text{wv}}} & \mathbf{0} & \dots & \mathbf{0} \\ 0 & \mathbf{B}_{w,p2}^{(m_{\text{corr},\Delta}^{\text{wv}})}(i) & \mathbf{0} & \mathbf{0} & \mathbf{C}_{p2}^{(m_{\text{corr},\Delta}^{\text{wv}})}(1,i) & \mathbf{0}_{\text{uncorr},p2} & \dots & \dots & \dots & \mathbf{C}_{p2}^{(m_{\text{corr},\Delta}^{\text{wv}})}(m_{\text{corr},\Delta}^{\text{wv}}+1,i) & \mathbf{0}_{\text{uncorr},p2+m_{\text{corr},\Delta}^{\text{wv}}} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \mathbf{B}_{w,N_{\text{psud}}-m_{\text{corr},\Delta}^{\text{wv}}}(i) & \mathbf{0} & \dots & \dots & \dots & \mathbf{0} & \mathbf{C}_{N_{\text{psud}}-m_{\text{corr},\Delta}^{\text{wv}}}^{(m_{\text{corr},\Delta}^{\text{wv}})}(1,i) & \mathbf{0}_{\text{uncorr},N_{\text{psud}}-m_{\text{corr},\Delta}^{\text{wv}}} & \mathbf{C}_{N_{\text{psud}}-m_{\text{corr},\Delta}^{\text{wv}}}^{(m_{\text{corr},\Delta}^{\text{wv}})}(2,i) & \mathbf{0}_{\text{uncorr},N_{\text{psud}}-m_{\text{corr},\Delta}^{\text{wv}}+1} & \dots & \mathbf{C}_{N_{\text{psud}}-m_{\text{corr},\Delta}^{\text{wv}}}^{(m_{\text{corr},\Delta}^{\text{wv}})}(m_{\text{corr},\Delta}^{\text{wv}}+1,i) & \mathbf{0}_{\text{uncorr},N_{\text{psud}}} \end{bmatrix} \quad (\text{A20})$$

where the first column of zero matrices accommodate the multi-pixel mean of correlated parameters, and the explicit form of  $\mathbf{B}$  matrix is expressed as,

$$\mathbf{B}_{w,p}^{(m_{\text{corr},\Delta}^{\text{wv}})}(i) = \begin{bmatrix} \mathbf{B}_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i, w_{\dots}(1)) & \mathbf{B}_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i, w_{\dots}(2)) & \dots & \mathbf{B}_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i, w_{\dots}(N_{\text{PC}})) \end{bmatrix} \quad (\text{A21})$$

with

$$\mathbf{B}_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i, w_{\dots}(k)) = \sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}(i)} \begin{bmatrix} \sum_{m=0}^{m_{\text{corr},\Delta}^{\text{wv}}(i)} d_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i)(m+1)w_{p+m}(k) & 0 & \dots & 0 \\ 0 & \sum_{m=0}^{m_{\text{corr},\Delta}^{\text{wv}}(i)} d_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i)(m+1)w_{p+m}(k) & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \sum_{m=0}^{m_{\text{corr},\Delta}^{\text{wv}}(i)} d_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i)(m+1)w_{p+m}(k) \end{bmatrix} \quad (\text{A22})$$

where the  $i$ -th correlated parameter.

The matrix  $\mathbf{C}$  is a null matrix except its  $i$ -th row has fill-in values, namely

$$\mathbf{C}_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(m', i) = \sqrt{\gamma_{\text{corr},\Delta}^{\text{wv}}(i)} d_p^{(m_{\text{corr},\Delta}^{\text{wv}})}(i)(m') \begin{bmatrix} 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ v_1(i) & v_2(i) & \dots & v_{N_{\text{PC}}}(i) \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix} \quad (\text{A23})$$

Moreover, Equation (59) is implemented to evaluate  $d(\mathbf{S}_{\text{wv},i}^{(m_{\text{corr},\Delta}^{\text{wv}})} \mathbf{x}_i) / d\mathbf{x}_i$  in Equation (A19).

## Appendix C.2 Within-Pixel Smoothness Matrix

The overall within-pixel smoothness matrix  $\gamma_{\text{corr},\diamond}^{\text{wv}} \Omega_{\text{corr},\diamond}^{\text{wv}}$  in Equation (A18) is expressed as,

$$\begin{cases} \gamma_{\text{corr},\diamond}^{\text{wv}} \Omega_{\diamond,1} = \sum_{i=1}^{N_{\text{TP},\text{corr}}} \left[ \sqrt{\gamma_{\text{corr},\diamond}^{\text{wv}}} \frac{d(\mathbf{S}_{\text{wv}}^{(m_{\text{corr},\diamond}^{\text{wv}})} \mathbf{x}_i)}{d\mathbf{x}_i} \right]^T \left[ \mathbf{W}^{(m_{\text{corr},\diamond}^{\text{wv}})} \right]^{-1} \left[ \sqrt{\gamma_{\text{corr},\diamond}^{\text{wv}}} \frac{d(\mathbf{S}_{\text{wv}}^{(m_{\text{corr},\diamond}^{\text{wv}})} \mathbf{x}_i)}{d\mathbf{x}_i} \right] \\ \gamma_{\text{corr},\diamond}^{\text{wv}} \Omega_{\diamond,2} = \sum_{i=1}^{N_{\text{TP},\text{corr}}} \left[ \sqrt{\gamma_{\text{corr},\diamond}^{\text{wv}}} \frac{d(\mathbf{S}_{\text{wv}}^{(m_{\text{corr},\diamond}^{\text{wv}})} \mathbf{x}_i)}{d\mathbf{x}_i} \right]^T \left[ \mathbf{W}^{(m_{\text{corr},\diamond}^{\text{wv}})} \right]^{-1} \left[ \sqrt{\gamma_{\text{corr},\diamond}^{\text{wv}}} \mathbf{S}_{\text{wv}}^{(m_{\text{corr},\diamond}^{\text{wv}})} \right] \end{cases} \quad (\text{A24})$$

where the column vector  $\mathbf{x}_i$  has the same length as that of the state vector and contains all-pixel PC weights and the  $i$ -th type of correlated fields included in all PC vectors (all rest elements of  $\mathbf{x}_i$  are zero), and the weight matrix  $\mathbf{W}$  that has a similar structure as Equations (A16) and (A17) (but for the  $i$ -th type of correlated parameters and is expanded to account for the number of pixels). To account for the

smooth variation of a type of correlated parameters, the contributions of PC vectors and weights are coupled in the following way,

$$\sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{WV}}^{(m_{\text{corr},\phi}^{\text{WV}})} = \begin{bmatrix} \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}})} & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}}^{(m_{\text{corr},\phi}^{\text{WV}})} & \mathbf{0} & \cdots & \mathbf{0} \\ \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}})} & \mathbf{0} & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}}^{(m_{\text{corr},\phi}^{\text{WV}})} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}})} & \mathbf{0} & \cdots & \cdots & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}}^{(m_{\text{corr},\phi}^{\text{WV}})} \end{bmatrix} \quad (\text{A25})$$

where

$$\sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}})} = \begin{bmatrix} \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},k=1)} & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},k=2)} & \cdots & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},k=N_{\text{PC}})} \end{bmatrix} \quad (\text{A26})$$

with

$$\sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},k)} = \begin{bmatrix} \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},(1),k)}(1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},(2),k)}(2) & \cdots & \mathbf{0} \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},(N_{\text{TP},\text{corr}}),k)}(N_{\text{TP},\text{corr}}) \end{bmatrix} \quad (\text{A27})$$

$$\mathbf{S}_{\text{V}}^{(m_{\text{corr},\phi}^{\text{WV}},k)}(j) = w_p(k) \times \begin{bmatrix} d_1^{(m_{\text{corr},\phi}^{\text{WV}})}(1) & d_1^{(m_{\text{corr},\phi}^{\text{WV}})}(2) & \cdots & d_1^{(m_{\text{corr},\phi}^{\text{WV}})}(m_{\text{corr},\phi}^{\text{WV}}+1) & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & d_2^{(m_{\text{corr},\phi}^{\text{WV}})}(1) & d_2^{(m_{\text{corr},\phi}^{\text{WV}})}(2) & \cdots & d_2^{(m_{\text{corr},\phi}^{\text{WV}})}(m_{\text{corr},\phi}^{\text{WV}}+1) & \cdots & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \cdots & \cdots & d_{L(j)-m_{\text{corr},\phi}^{\text{WV}}}^{(m_{\text{corr},\phi}^{\text{WV}})}(1) & d_{L(j)-m_{\text{corr},\phi}^{\text{WV}}}^{(m_{\text{corr},\phi}^{\text{WV}})}(2) & \cdots & d_{L(j)-m_{\text{corr},\phi}^{\text{WV}}}^{(m_{\text{corr},\phi}^{\text{WV}})}(m_{\text{corr},\phi}^{\text{WV}}+1) \end{bmatrix} \quad (\text{A28})$$

In Equation (A25),  $\sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}}^{(m_{\text{corr},\phi}^{\text{WV}})}$  is expressed as,

$$\sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}}^{(m_{\text{corr},\phi}^{\text{WV}})} = \begin{bmatrix} \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_1}^{(m_{\text{corr},\phi}^{\text{WV}})} & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_2}^{(m_{\text{corr},\phi}^{\text{WV}})} & \cdots & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_{N_{\text{pixel}}}}^{(m_{\text{corr},\phi}^{\text{WV}})} \end{bmatrix} \quad (\text{A29})$$

with

$$\sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}})} = \begin{bmatrix} \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(1),k=1)}(1) & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(1),k=2)}(1) & \cdots & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(1),k=N_{\text{PC}})}(1) & \mathbf{0}_{\text{uncorr},p} \\ \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(2),k=1)}(2) & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(2),k=2)}(2) & \cdots & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(2),k=N_{\text{PC}})}(2) & \mathbf{0}_{\text{uncorr},p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(N_{\text{TP},\text{corr}}),k=1)}(N_{\text{TP},\text{corr}}) & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(N_{\text{TP},\text{corr}}),k=2)}(N_{\text{TP},\text{corr}}) & \cdots & \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}_p}^{(m_{\text{corr},\phi}^{\text{WV}},(N_{\text{TP},\text{corr}}),k=N_{\text{PC}})}(N_{\text{TP},\text{corr}}) & \mathbf{0}_{\text{uncorr},p} \end{bmatrix} \quad (\text{A30})$$

with the column vector for  $j$ -th type of correlated parameter described by,

$$\sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \mathbf{S}_{\text{W}}^{(m_{\text{corr},\phi}^{\text{WV}},(j),k)}(j) = \sqrt{\gamma_{\text{corr},\phi}^{\text{WV}}} \times \begin{bmatrix} \sum_{i=1}^{m_{\text{corr},\phi}^{\text{WV}}(j)+1} d_1^{(m_{\text{corr},\phi}^{\text{WV}})}(i) \times v_k \left( \left[ \sum_{n=1}^{j-1} L(n) \right] + i \right) \\ \sum_{i=1}^{m_{\text{corr},\phi}^{\text{WV}}(j)+1} d_2^{(m_{\text{corr},\phi}^{\text{WV}})}(i) \times v_k \left( \left[ \sum_{n=1}^{j-1} L(n) \right] + i \right) \\ \vdots \\ \sum_{i=1}^{m_{\text{corr},\phi}^{\text{WV}}(j)+1} d_{L(j)-m_{\text{corr},\phi}^{\text{WV}}}^{(m_{\text{corr},\phi}^{\text{WV}})}(i) \times v_k \left( \left[ \sum_{n=1}^{j-1} L(n) \right] + i \right) \end{bmatrix} \quad (\text{A31})$$

Moreover, Equation (59) is implemented to evaluate  $d(\mathbf{S}_{\text{WV}}^{(m_{\text{corr},\phi}^{\text{WV}})} \mathbf{x}_i) / d\mathbf{x}_i$  in Equation (A24).

## Appendix D Decoupled Smoothness Constraints

Smooth variations are often directly observed in PC weights and elements in a PC vector associated with a certain type of parameter. Direct imposition of smoothness helps stabilize the first few iterations when *a priori* information about the PC vectors and weights is insufficient. In an integrated form, the across-pixel and within-PC smoothness matrix is expressed as

$$\gamma_{\text{corr}}^{\text{W/V}} \mathbf{\Omega}_{\text{corr}}^{\text{W/V}} = \gamma_{\text{corr}}^{\text{W}} \mathbf{\Omega}_{\text{corr}}^{\text{W}} + \gamma_{\text{corr}}^{\text{V}} \mathbf{\Omega}_{\text{corr}}^{\text{V}} \quad (\text{A32})$$

where  $\gamma_{\text{corr}}^w \mathbf{\Omega}_{\text{corr}}^w$  and  $\gamma_{\text{corr}}^v \mathbf{\Omega}_{\text{corr}}^v$  represent the matrix form of smoothness constraints imposed on PC weights and PC vectors, respectively. The explicit forms of  $\gamma_{\text{corr}}^w \mathbf{\Omega}_{\text{corr}}^w$  and  $\gamma_{\text{corr}}^v \mathbf{\Omega}_{\text{corr}}^v$  are given in the Appendices D.1 and D.2, respectively.

#### Appendix D.1 Across-Pixel Smoothness Constraints on PC Weights

The smoothness constraints imposed directly on PC weights is expressed as

$$\gamma_{\text{corr}}^w \mathbf{\Omega}_{\text{corr}}^w = \sum_{k=1}^{N_{\text{PC}}} \left[ \sqrt{\gamma_{\text{corr},k}^w} \mathbf{S}_w^{(m_{\text{corr},k}^w)}(k) \right]^T \left[ \sqrt{\gamma_{\text{corr},k}^w} \mathbf{S}_w^{(m_{\text{corr},k}^w)}(k) \right] \quad (\text{A33})$$

where

$$\sqrt{\gamma_{\text{corr},k}^w} \mathbf{S}_w^{(m_{\text{corr},k}^w)}(k) = \begin{bmatrix} 0 & 0 & \mathbf{C}_{p1}(1,k) & \mathbf{0}_{\text{uncorr},p1} & \mathbf{C}_{p1}(2,k) & \mathbf{0}_{\text{uncorr},p2} & \dots & \dots & \mathbf{C}_{p1}(m_{\text{corr},k}^w+1,k) & \mathbf{0}_{\text{uncorr},p1+m_{\text{corr},k}^w} & \mathbf{0} & \dots & \mathbf{0} \\ 0 & 0 & 0 & 0 & \mathbf{C}_{p2}(1,k) & \mathbf{0}_{\text{uncorr},p2} & \mathbf{C}_{p2}(2,k) & \mathbf{0}_{\text{uncorr},p3} & \dots & \mathbf{C}_{p2}(m_{\text{corr},k}^w+1,k) & \mathbf{0}_{\text{uncorr},p2+m_{\text{corr},k}^w} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & \mathbf{0} & \mathbf{C}_{N_{\text{pixel}}-m_{\text{corr},k}^w}(1,k) & \mathbf{0}_{\text{uncorr},N_{\text{pixel}}-m_{\text{corr},k}^w} & \mathbf{C}_{N_{\text{pixel}}-m_{\text{corr},k}^w}(2,k) & \mathbf{0}_{\text{uncorr},N_{\text{pixel}}-m_{\text{corr},k}^w+1} & \dots & \mathbf{C}_{N_{\text{pixel}}-m_{\text{corr},k}^w}(m_{\text{corr},k}^w+1,k) & \mathbf{0}_{\text{uncorr},N_{\text{pixel}}} \end{bmatrix} \quad (\text{A34})$$

where the first two columns of zero matrices are to accommodate the multi-pixel mean of correlated parameters and PC vectors, and  $\mathbf{C}(i,k)$  is calculated by

$$\mathbf{C}_p(m',k) = d_{k,p}^{(m_{\text{corr},k}^w)}(m') \sqrt{\gamma_{\text{corr},k}^w} \begin{bmatrix} 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \delta(k,1) & \delta(k,2) & \dots & \delta(k,N_{\text{PC}}) \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix}_{N_{\text{Pixel}} \times N_{\text{PC}}} \quad (\text{A35})$$

where the delta function  $\delta(k,k') = 1$  for  $k = k'$  and 0 for  $k \neq k'$ , respectively, and the differentiation arrays are expressed similarly as Equations (A7) and (A8), except that an extra subscript “ $k$ ” is introduced to the differentiation arrays  $\mathbf{d}$  to allow different strength and orders of difference of smoothness constraints to impose on a specific ( $k$ -th) principal component.

#### Appendix D.2 Within-PC Smoothness Constraints

The matrix form of within-PC smoothness constraints imposed directly upon a PC is evaluated by,

$$\gamma_{\text{corr}}^v \mathbf{\Omega}_{\text{corr}}^v = \sum_{k=1}^{N_{\text{PC}}} \left[ \sqrt{\gamma_{\text{corr},k}^v} \mathbf{S}_v^{(m_{\text{corr},k}^v)}(k) \right]^T \left[ \mathbf{W}^{(m_{\text{corr},k}^v)} \right]^{-1} \left[ \sqrt{\gamma_{\text{corr},k}^v} \mathbf{S}_v^{(m_{\text{corr},k}^v)}(k) \right] \quad (\text{A36})$$

where

$$\left[ \sqrt{\gamma_{\text{corr},k}^v} \mathbf{S}_v^{(m_{\text{corr},k}^v)}(k) \right]_v = \begin{bmatrix} \mathbf{0} & \mathbf{T}^{(m_{\text{corr},k}^v)}(1) & \mathbf{T}^{(m_{\text{corr},k}^v)}(2) & \dots & \mathbf{T}^{(m_{\text{corr},k}^v)}(N_{\text{PC}}) & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (\text{A37})$$

where the first zero matrix accommodates the multi-pixel mean of correlated parameters, the last two zero matrices accommodate the PC weights and uncorrelated parameters, and  $\mathbf{T}^{(m_{\text{corr},k}^v)}(k)$  is evaluated by,

$$\mathbf{T}^{(m_{\text{corr},k}^v)}(k) = \begin{bmatrix} \mathbf{C}^{(m_{\text{corr},k}^v(1),k)}(1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{(m_{\text{corr},k}^v(2),k)}(2) & \dots & \mathbf{0} \\ \dots & \dots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{C}^{(m_{\text{corr},k}^v(N_{\text{TP},\text{corr}},k)}(N_{\text{TP},\text{corr}}) \end{bmatrix} \quad (\text{A38})$$

where

$$\mathbf{C}^{(m_{\text{corr},k}^v(j),k)}(j) = \sqrt{\gamma_{\text{corr},k}^v(j)} \begin{bmatrix} d_{k,1}^{(m_{\text{corr},k}^v(j))}(1) & d_{k,1}^{(m_{\text{corr},k}^v(j))}(2) & \dots & d_{k,1}^{(m_{\text{corr},k}^v(j))}(m_{\text{corr},k}^v(j)+1) & 0 & \dots & 0 \\ 0 & d_{k,2}^{(m_{\text{corr},k}^v(j))}(1) & d_{k,2}^{(m_{\text{corr},k}^v(j))}(2) & \dots & d_{k,2}^{(m_{\text{corr},k}^v(j))}(m_{\text{corr},k}^v(j)+1) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & d_{k,L(j)-m_{\text{corr},k}^v(j)}^{(m_{\text{corr},k}^v(j))}(1) & d_{k,L(j)-m_{\text{corr},k}^v(j)}^{(m_{\text{corr},k}^v(j))}(2) & \dots & d_{k,L(j)-m_{\text{corr},k}^v(j)}^{(m_{\text{corr},k}^v(j))}(m_{\text{corr},k}^v(j)+1) \end{bmatrix} \quad (\text{A39})$$

where  $\mathbf{d}$  depends on the type of correlated parameters and is expressed in Equations (A7) and (A8). It can also vary as a function of a specific principal component. The weights matrix  $\mathbf{W}^{(m_{\text{corr},k}^v)}$  in Equation (A36) is evaluated in the same way as for within-pixel constraints for uncorrelated parameters (see Equation (A16)).

## References

1. Phillips, D.L. A technique for numerical solution of certain integral equation of first kind. *J. ACM* **1962**, *9*, 84–97. [\[CrossRef\]](#)
2. Twomey, S. On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. *J. ACM* **1963**, *10*, 97–101. [\[CrossRef\]](#)
3. Twomey, S. *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*; Elsevier: New York, NY, USA, 1977.
4. Tikhonov, A.N. On the solution of incorrectly stated problems and a method of regularization. *Dokl. Akad. Nauk SSSR* **1963**, *151*, 501–504.
5. Tikhonov, A.N.; Arsenin, V.Y. *Solution of Ill-Posed Problems*; Wiley: New York, NY, USA, 1977.
6. Dubovik, O.; Herman, M.; Holdak, A.; Lapyonok, T.; Tanré, D.; Deuzé, J.L.; Ducos, F.; Sinyuk, A.; Lopatin, A. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. *Atmos. Meas. Tech.* **2011**, *4*, 975–1018. [\[CrossRef\]](#)
7. Liu, X.; Smith, W.L.; Zhou, D.K.; Larar, A. Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept. *Appl. Opt.* **2006**, *45*, 201–209. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Natraj, V.; Spurr, R.J.D. A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering-absorbing media. *J. Quant. Spectrosc. Radiat. Transf.* **2007**, *107*, 263–293. [\[CrossRef\]](#)
9. Spurr, R.J.D.; Natraj, V.; Lerot, C.; van Roozendaal, M.; Loyola, D. Linearization of the principal component analysis method for radiative transfer acceleration: Application to retrieval algorithms and sensitivity studies. *J. Quant. Spectrosc. Radiat. Transf.* **2013**, *125*, 1–17. [\[CrossRef\]](#)
10. Liu, X.; Zhou, D.K.; Larar, A.M.; Smith, W.L.; Mango, S.A. Case-study of a principal-component-based radiative transferforward model and retrieval algorithm using EAQUATE data. *Q. J. R. Meteorol. Soc.* **2007**, *133*, 243–256. [\[CrossRef\]](#)
11. Liu, X.; Zhou, D.K.; Larar, A.M.; Smith, W.L.; Schluessel, P.; Newman, S.M.; Taylor, J.P.; Wu, W. Retrieval of atmospheric profiles and cloud properties from IASI spectra using super-channels. *Atmos. Chem. Phys.* **2009**, *9*, 9121–9142. [\[CrossRef\]](#)
12. Liu, X.; Liu, L.; Zhang, S.; Zhou, X. New spectral fitting method for full-spectrum solar-induced chlorophyll fluorescence retrieval based on principal components analysis. *Remote Sens.* **2015**, *7*, 10626–10645. [\[CrossRef\]](#)
13. Martonchik, J.V.; Diner, D.J.; Kahn, R.A.; Verstraete, M.M.; Pinty, B.; Gordon, H.R.; Ackerman, T.P. Techniques for the retrieval of aerosol properties over land and ocean using multiangle data. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1212–1227. [\[CrossRef\]](#)
14. Martonchik, J.V.; Kahn, R.A.; Diner, D.J. Retrieval of aerosol properties over land using MISR observations. In *Satellite Aerosol Remote Sensing over Land*; Kokhanovsky, A., de Leeuw, G., Eds.; Springer: Berlin, Germany, 2009.
15. Diner, D.J.; Martonchik, J.V.; Kahn, R.A.; Pinty, B.; Gobron, N.; Nelson, D.L.; Holben, B.N. Using angular and spectral shape similarity constraints to improve MISR aerosol and surface retrievals overland. *Remote Sens. Environ.* **2005**, *94*, 155–171. [\[CrossRef\]](#)
16. Dubovik, O.; Li, Z.; Mishchenko, M.I.; Tanré, D.; Karol, Y.; Bojkov, B.; Cairns, B.; Diner, D.J.; Espinosa, W.R.; Goloub, P.; et al. Polarimetric remote sensing of atmospheric aerosols: Instruments, methodologies, results, and perspectives. *J. Quant. Spectrosc. Radiat. Transf.* **2019**, *224*, 474–511. [\[CrossRef\]](#)
17. Kokhanovsky, A.A. The modern aerosol retrieval algorithms based on the simultaneous measurements of the intensity and polarization of reflected solar light: A review. *Front. Environ. Sci.* **2015**, *3*, 4. [\[CrossRef\]](#)

18. Xu, F.; Dubovik, O.; Zhai, P.-W.; Diner, D.J.; Kalashnikova, O.V.; Seidel, F.C.; Litvinov, P.; Bovchaliuk, A.; Garay, M.J.; van Harten, G.; et al. Joint retrieval of aerosol and water-leaving radiance from multispectral, multiangular and polarimetric measurements over ocean. *Atmos. Meas. Tech.* **2016**, *9*, 2877–2907. [[CrossRef](#)]
19. Xu, F.; van Harten, G.; Diner, D.J.; Kalashnikova, O.V.; Seidel, F.C.; Bruegge, C.J.; Dubovik, O. Coupled retrieval of aerosol properties and land surface reflection using the Airborne Multiangle Spectropolarimetric Imager. *J. Geophys. Res. Atmos.* **2017**, *122*, 7004–7026. [[CrossRef](#)]
20. Cahalan, R.F.; Ridgway, W.; Wiscombe, W.J.; Bell, T.L.; Snider, J.B. The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.* **1994**, *51*, 2434–2455. [[CrossRef](#)]
21. Hou, W.; Wang, J.; Xu, X.; Reid, J.S.; Han, D. An algorithm for hyperspectral remote sensing of aerosols: 1. Development of theoretical framework. *J. Quant. Spectrosc. Radiat. Transf.* **2016**, *178*, 400–415. [[CrossRef](#)]
22. Hou, W.; Wang, J.; Xu, X.; Reid, J.S. An algorithm for hyperspectral remote sensing of aerosols: 2. Information content analysis for aerosol parameters and principal components of surface spectra. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *192*, 14–29. [[CrossRef](#)]
23. Dubovik, O.; University of Lille; Xu, F.; Jet Propulsion Laboratory. Personal communication, 2018.
24. Dubovik, O. Optimization of numerical inversion in photopolarimetric remote sensing. In *Photopolarimetry in Remote Sensing*; Videen, G., Yatskiv, Y., Mishchenko, M., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2004; pp. 65–106.
25. Dubovik, O.; King, M.D. A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. *J. Geophys. Res.* **2000**, *105*, 673–696. [[CrossRef](#)]
26. Levenberg, K. A method for the solution of certain non-linear problems in Least Squares. *Quart. Appl. Math.* **1944**, *2*, 164–168. [[CrossRef](#)]
27. Marquardt, D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **1963**, *11*, 431–441. [[CrossRef](#)]
28. Xu, F.; Davis, A.B.; West, R.A.; Esposito, L.W. Markov chain formalism for polarized light transfer in plane-parallel atmospheres, with numerical comparison to the Monte Carlo method. *Opt. Express* **2010**, *19*, 946–967. [[CrossRef](#)] [[PubMed](#)]
29. Lacis, A.A.; Hansen, J.E. A parameterization for the absorption of solar radiation in the Earth's atmosphere. *J. Atmos. Sci.* **1974**, *31*, 118–133. [[CrossRef](#)]
30. Dubovik, O.; Sinyuk, A.; Lapyonok, T.; Holben, B.N.; Mishchenko, M.; Yang, P.; Eck, T.F.; Volten, H.; Muñoz, O.; Veihelmann, B.; et al. Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust. *J. Geophys. Res.* **2006**, *111*, D11208. [[CrossRef](#)]
31. Rahman, H.; Pinty, B.; Verstraete, M.M. Coupled surface-atmosphere reflectance (CSAR) model 2. Semiempirical surface model usable with NOAA advanced very high resolution radiometer data. *J. Geophys. Res.* **1993**, *98*, 20791–20801. [[CrossRef](#)]
32. Litvinov, P.; Hasekamp, O.; Cairns, B. Models for surface reflection of radiance and polarized radiance: Comparison with airborne multi-angle photopolarimetric measurements and implications for modeling top-of-atmosphere measurements. *Remote Sens. Environ.* **2011**, *115*, 781–792. [[CrossRef](#)]
33. Van Harten, G.; Diner, D.J.; Daugherty, B.J.S.; Rheingans, B.E.; Bull, M.A.; Seidel, F.C.; Chipman, R.A.; Cairns, B.; Wasilewski, A.P.; Knobelspiesse, K.D. Calibration and validation of airborne multiangle spectropolarimetric imager (AirMSPI) polarization measurements. *Appl. Opt.* **2018**, *57*, 4499–4513. [[CrossRef](#)]
34. Dubovik, O.; Smirnov, A.; Holben, B.N.; King, M.D.; Kaufman, Y.J.; Eck, T.F.; Slutsker, I. Accuracy assessment of aerosol optical properties retrieval from AERONET sun and sky radiance measurements. *J. Geophys. Res.* **2000**, *105*, 9791–9806. [[CrossRef](#)]
35. Dubovik, O.; Lapyonok, T.; Litvinov, P.; Herman, M.; Fuertes, D.; Ducos, F.; Lopatin, A.; Chaikovsky, A.; Torres, B.; Derimian, Y.; et al. GRASP: A versatile algorithm for characterizing the atmosphere. *SPIE Newsroom* **2014**, *25*. [[CrossRef](#)]
36. Lyapustin, A.; Martonchik, J.; Wang, Y.; Laszlo, I.; Korkin, S. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *J. Geophys. Res. Atmos.* **2011**, *116*, D03210. [[CrossRef](#)]
37. Lyapustin, A.; Wang, Y.; Laszlo, I.; Kahn, R.; Korkin, S.; Remer, L.; Levy, R.; Reid, J.S. Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *J. Geophys. Res. Atmos.* **2011**, *116*, D03211. [[CrossRef](#)]

38. Baldridge, A.M.; Hook, S.J.; Grove, C.I.; Rivera, G. The ASTER spectral library version 2.0. *Remote Sens. Environ.* **2009**, *113*, 711–715. [[CrossRef](#)]
39. Clark, R.N.; Swayze, G.A.; Wise, R.; Livo, E.; Hoefen, T.; Kokaly, R.; Sutley, S.J. USGS digital spectral library splib06a. *USA Geol. Surv. Digit. Data Ser.* **2007**, *231*, 2007.
40. Levy, R.C.; Remer, L.A.; Mattoo, S.; Vermote, E.F.; Kaufman, Y.J. Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance. *J. Geophys. Res.* **2007**, *112*, D13211. [[CrossRef](#)]
41. Diner, D.J.; Beckert, J.C.; Reilly, T.H.; Bruegge, C.J.; Conel, J.E.; Kahn, R.; Martonchik, J.V.; Ackerman, T.P.; Davies, R.; Gerstl, S.A.W.; et al. Multiangle imaging spectro-radiometer (MISR) description and experiment overview. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1072–1087. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).