Development of a Global Validation Package for Satellite Oceanic Aerosol Optical Thickness Retrieval Based on AERONET Observations and Its Application to NOAA/NESDIS Operational Aerosol Retrievals

TOM X.-P. ZHAO,*,& LARRY L. STOWE,* ALEXANDER SMIRNOV,+ DAVID CROSBY,* JOHN SAPPER,# AND CHARLES R. MCCLAIN@

*NOAA/NESDIS/ORA, Camp Springs, Maryland
+NASA Goddard Space Flight Center, and Science Systems and Applications, Inc., Greenbelt, Maryland
#NOAA/NESDIS/OSDPD, Suitland, Maryland
@NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 29 December 2000, in final form 16 May 2001)

ABSTRACT

In this paper, a global validation package for satellite aerosol optical thickness retrieval using the Aerosol Robotic Network (AERONET) observations as ground truth is described. To standardize the validation procedure, the optimum time–space match-up window, the ensemble statistical analysis method, the best selection of AERONET channels, and the numerical scheme used to interpolate/extrapolate these observations to satellite channels have been identified through sensitivity studies. The package is shown to be a unique tool for more objective validation and intercomparison of satellite aerosol retrievals, helping to satisfy an increasingly important requirement of the satellite aerosol remote sensing community. Results of applying the package to the second-generation operational aerosol observational data (AEROBS) from the *NOAA-14* Advanced Very High Resolution Radiometer (AVHRR) in 1998 and to the same year aerosol observation data [Clouds and the Earth's Radiant Energy System-Single Scanner Foodprint version 4 (CERES-SSF4)] from the Tropical Rainfall Measuring Mission (TRMM) Visible Infrared Scanner (VIRS) are presented as examples of global validation. The usefulness of the package for identifying improvements to the aerosol optical thickness τ retrieval algorithm is also demonstrated.

The principal causes of systematic errors in the current National Oceanic and Atmospheric Administration (NOAA)/National Environmental Satellite, Data, and Information Service (NESDIS) operational aerosol optical thickness retrieval algorithm have been identified and can be reduced significantly, if the correction and adjustment suggested from the global validation are adopted. Random error in the τ retrieval is identified to be a major source of error on deriving the effective Ångström wavelength exponent α and may be associated with regional differences in aerosol particles, which are not accounted for in the current second-generation operational algorithm. Adjustments to the nonaerosol and aerosol radiative transfer model parameters that reduce systematic errors in τ retrievals are suggested for consideration in the next-generation algorithm. Basic features that should be included in the next-generation algorithm to reduce random error in τ retrievals and the resulting error in the effective Ångström wavelength exponent have also been discussed.

Compared to the AERONET observation, the *NOAA-14* AVHRR (AEROBS) τ values for mean conditions are biased high by 0.05 and 0.08, with random errors of 0.08 and 0.05, at 0.63 and 0.83 μm, respectively. Correspondingly, the TRMM VIRS (CERES-SSF4) values for mean conditions are biased high by 0.06 and 0.02, with random errors of 0.06 and 0.04 at 0.63 and 1.61 μm, respectively. After corrections and adjustments to the retrieval algorithm, the biases in both channels of AVHRR and VIRS are reduced significantly to values close to zero, although random error is almost unchanged. The α exponent derived directly from the aerosol optical thicknesses (τs) has been shown to be poorly correlated both before and after adjustments, indicating that random error in the τ measurement (possibly related to aerosol model parameter variations or cloud–surface reflectance contamination) needs to be reduced.

& Currently a CIRA Visiting Scientist.

Corresponding author address: Dr. Tom X-P. Zhao, E/RA1, RM 711-C, WWBG, NOAA/NESDIS/ORA, 5200 Auth Road, Camp Springs, MD 20746-4304.
E-mail: xuepeng.zhao@noaa.gov

1. Introduction

It is generally recognized that the direct and indirect radiative effects of tropospheric aerosols on global climate are comparable to greenhouse gases but with opposite sign and with larger uncertainties (Hansen and Lacis 1990; Charlson et al. 1992; Lacis and Mishchenko 1995; IPCC 1996). Long-term global aerosol measurement based on satellite aerosol remote sensing may help us to practically reduce these uncertainties (King et al. 1999; Hansen et al. 2000) provided the performance of the retrieval algorithms and instruments is well evaluated and documented. Actually, many validations performed for spaceborne and airborne aerosol retrievals can be found in literature (e.g., Kaufman et al. 1990; Ignatov et al. 1995a,b; Stowe et al. 1997; Nakajima and Higurashi 1997; Tanre et al. 1997; Chu et al. 1999; Goloub et al. 1999; Durkee et al. 2000). One common characteristic of these validations is that they were performed only in a limited region and period covered either by an airborne or a ship cruise campaign, or both. Comparable validation results have been obtained from these validations, but with different validation concepts and procedures, which may generate many ambiguities for intercomparison of these retrieval algorithms.

Lacking global ground-based aerosol observations is the major obstacle for global validation of satellite aerosol retrievals. However, the Aerosol Robotic Network [AERONET (Holben et al. 1998)] initiated by the National Aeronautics and Space Administration's (NASA) Earth Observing System program, recently expanded through federation with many non-NASA institutions, now provides a unique opportunity for global validation of satellite aerosol retrievals for different sensors. Data from this network provide globally distributed and quality-controlled observations of aerosol spectral optical depths, aerosol size distributions, etc., in a manner suitable for integration with satellite data (see, e.g., Dubovik et al. 2000; Smirnov et al. 2000).

It is time to standardize procedures for a more objective global validation of satellite aerosol retrievals and their intercomparison. This is a very important step considering the fact that ''no one sensor system is capable of providing totally unambiguous information, and hence a careful intercomparison of derived products from different sensors, together with a comprehensive network of ground-based sunphotometer and sky radiometer systems, are required to advance our quantitative understanding of global aerosol characteristics'' (King et al. 1999).

The second-generation National Oceanic and Atmospheric Administration (NOAA)/National Environmental Satellite Data and Information Service (NESDIS) operational aerosol optical thickness τ retrieval algorithm is used to process data from *NOAA-14* Advanced Very High Resolution Radiometer (AVHRR) (Stowe et al. 1997) and the Tropical Rainfall Measuring Mission (TRMM) Visible Infrared Scanners (VIRS) (Ignatov and Stowe 2000) for global oceanic aerosol monitoring. The algorithm provides estimates of τ independently in each reflectance channel of AVHRR (0.63 and 0.83 μm) and VIRS (0.63 and 1.61 μm), assuming the molecular atmosphere, the aerosol microphysics, and surface reflectance are known and invariant. The effective Ångström wavelength exponent α can be derived from the two independent measurements of τ in two reflectance channels of AVHRR or VIRS. In operation, the relationship between aerosol optical thickness τ and dimensionless reflectance ρ (normalized to solar flux) is described by a four-dimensional lookup table (LUT), precalculated for different τ and view–sun–azimuth angles using Dave's radiative transfer model, assuming Eterman's vertical profile of aerosol concentration; midlatitude ozone; and water vapor profiles. Other parameters to be specified as input to the LUT are Lambertian oceanic reflectance with diffuse glint correction to the phase function, the volume scattering and absorption coefficients, and the aerosol phase function derived from Mie calculations with a prescribed aerosol microphysical model. The oceanic albedo is set to 0.002 and 0.0005, and 0 for 0.63-, 0.83-, and 1.61-μm channels, respectively. The aerosol model uses a monomodal lognormal distribution: $dN/d\ln r = N(2\pi)^{-1/2} \ln^{-1}(\sigma) \exp[-\ln(r/r_m)^2/2(\ln\sigma)^2]$, where mode radius $r_m = 0.1$ μm and variance of size distribution $\sigma = 2.03$; a complex index of refraction $n = 1.40–0.0i$ (or albedo of single scattering $\omega = 1$). More detailed descriptions on the retrieval algorithm can be found in Stowe et al. (1997).

This algorithm has not gone through a complete global validation. The documented performance of the algorithm was only based on three oceanic cruise validations in the tropical and North Atlantic Ocean for its operational one-channel (0.63 μm) version on *NOAA-9* and *NOAA-11* AVHRR (Ignatov et al. 1995a,b; Stowe et al. 1997). The global aerosol products of AVHRR are to be widely used for studies of radiative forcing on climate change due to its long-term availability, retrospectively back to 1981 in the AVHRR Pathfinder Atmosphere data (Stowe and Jacobowitz 1997). Thus, a more complete and objective global validation and error estimation of the retrieval algorithm applied to both channels is necessary.

Based on the above considerations, a global validation package based on AERONET observations has been developed. A prototype for using AERONET data objectively for global validation of satellite aerosol remote sensing on visible to midinfrared spectral channels is provided. In this paper, not only is our global validation package introduced and the results of applying it to *NOAA-14* AVHRR (AEROBS) and TRMM VIRS [Clouds and the Earth's Radiant Energy System-Single Scanner Footprint version 4 (CERES-SSF4)] aerosol data (derived with the second-generation NOAA/NESDIS operational aerosol retrieval algorithm) presented, but also the advantages of our global validation package (based on a comprehensive network of ground-based

TABLE 1. Selected 13 AERONET island stations and their location (latitude and longitude) for our global aerosol retrieval validation. The highlighted stations are those picked with sufficient level 2 data in 1998 for our validation.

No.	Station	Latitude, longitude
1	**Andros Island**	**24.68°, −77.78°**
2	Ascension Island	−7.97°, −14.40°
3	**Bahrain**	**26.32°, 50.50°**
4	Barbados	13.17°, −59.50°
5	**Bermuda**	**32.37°, −64.68°**
6	Cape Verde	16.72°, −22.93°
7	Dakar	14.38°, −16.95°
8	**Dry Tortuga**	**24.60°, −82.78°**
9	Guadeloup	16.32°, −61.50°
10	**Kaashidhoo**	**4.95°, 73.45°**
11	**Lanai**	**20.82°, 156.98°**
12	St. Nicolas	33.25°, −119.49°
13	**Surinam**	**5.78°, −55.20°**

TABLE 2. Selected time–space match-up windows used in sensitivity studies for determining optimal match-up window.

Time window (h)	Space window (radius of the match-up circle around an AERONET station)				
	100 (km)	200 (km)	300 (km)	400 (km)	500 (km)
+/− 1	W11	W12	W13	W14	W15
+/− 2	W21	W22	W23	W24	W25
+/− 3	W31	W32	W33	W34	W35
+/− 4	W41	W42	W43	W44	W45

observations) to the improvement of satellite aerosol retrieval algorithms (as anticipated by, for example, King et al. 1999) is demonstrated. Development of the global validation package is described in section 2. Validation results on the second-generation NOAA/NESDIS operational aerosol retrieval algorithm, applied to channels 1 and 2, are given in section 3. The value of the package for the improvement of an aerosol retrieval algorithm is demonstrated in section 4. Some important issues are discussed in section 5. Summary and concluding remarks are given in section 6.

2. Development of the global validation package

a. Basic concept

Since aerosol optical thickness τ is derived from the measurement of spectral attenuation of the direct solar beam using the CIMEL sun–sky radiometers at AERONET stations (Holben et al. 1998). This measurement is equivalent to sun photometer measurement. Its accuracy is much higher than that derived from backward scattering radiance [''contaminated'' by varying surface (land, ocean, cloud) properties] measured from satellite (Tanre et al. 1996). For convenience, we subsequently refer to these solar extinction measurements as sun photometer (SP) data. Thus, AERONET measurements have been widely used as ground truth for the validation of both airborne and spaceborne aerosol observations (e.g., Chu et al. 1998; Higurashi et al. 2000).

We have selected 13 AERONET island stations (see Table 1) in our validation package, which cover the major regimes of global oceanic aerosol characteristics (cf. Husar et al. 1997). Quality-controlled level 2 AERONET aerosol optical thickness observations (e.g., Smirnov et al. 2000) are used as ground truth in the validation. Initially, the global validation package was applied to AVHRR and VIRS data from 1998 (the year VIRS data became available) to validate the NESDIS second-generation retrieval algorithm performance. After checking the level 2 AERONET SP data, it is found that there are no level 2 data for Dakar and Guadeloup in 1998. There are only 2 months of level 2 data in 1998 for Cape Verde (January and February), Ascension Island (November and December), and San Nicolas (November and December). Also, level 2 data in 1998 at Barbados are missing for some channels. Thus, only 7 of 13 AERONET stations (Andros Island, Bahrain, Bermuda, Dry Tortugas, Kaashidhoo, Surinam, and Lanai, which are highlighted in Table 1) have been kept in our validation for the 1998 satellite observations.

The aerosol optical thickness τ and its effective Ångström wavelength exponent α retrieved from these two satellite sensors (τ_{st} and α_{st}) are matched up with the corresponding ''truth'' values derived from the same-day surface AERONET SP observations (τ_{sp} and α_{sp}). They are statistically processed within an optimum space–time window from which scatter diagrams of τ_{st} versus τ_{sp} and α_{st} versus α_{sp} are produced. Linear regression analyses are performed, predicting the satellite-retrieved values of τ_{st} or α_{st} as a function of the SP values of τ_{sp} or α_{sp} in the form of $\tau_{st} = A + B\tau_{sp}$ (or $\alpha_{st} = A + B\alpha_{sp}$). Retrieval algorithm performance can be evaluated from resulting statistical parameters of the linear regression: A (intercept), B (slope), σ (standard error), and R^2 (square of correlation coefficient). For example, a nonzero intercept tells us the retrieval algorithm is biased at low τ values, which may be associated with sensor calibration error or improper assumptions about ocean surface reflection. A slope that is different from unity (proportional error) indicates that there may be some inconsistency between the aerosol microphysical model (such as refractive index) used in the retrieval algorithm and that in the real world. A very good diagnostic analysis of the physical rationale behind the errors represented by nonzero intercept and nonunity slope have been performed in Stowe et al. (1997) using the linearized single scattering approximation of the radiative transfer equation, which is also utilized in the discussion of section 5.

This validation has been performed not only for single AERONET stations (called regional validation) but also for the ensemble of all selected stations (hereafter called global validation). Since the number of match-up days found for a single AERONET station in 1998 is not sufficient (probably need at least 60 samples) for con-

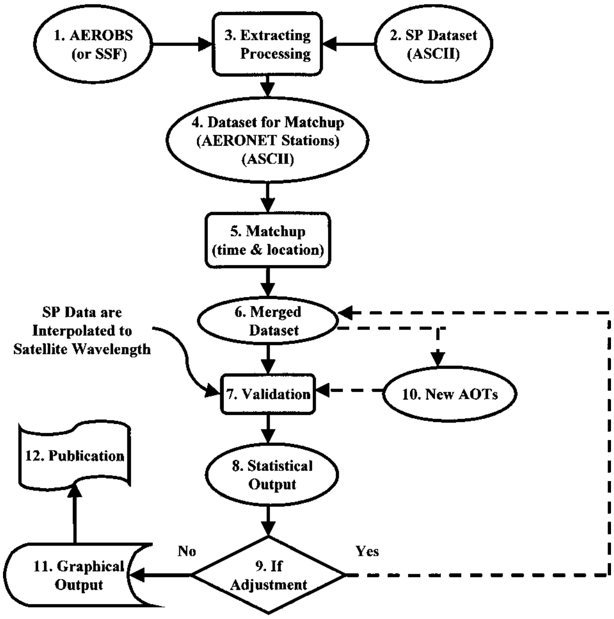

FIG. 1. Flowchart of global aerosol validation procedure.

clusive regional validation, the paper has been restricted to the global validation.

Our validation package is summarized with a flowchart in Fig. 1. Each step (1–12) is briefly described. The original satellite data (AEROBS or CERES-SSF4) and surface AERONET observations (steps 1, 2) are collected and reformatted (step 3) around the 13 baseline AERONET stations to produce a smaller archived dataset (step 4) for use in subsequent steps. The matchups are searched (step 5) according to an optimal time–space window to produce a merged, match-up (archived) dataset (step 6). This archived match-up dataset is used to do regression validation (step 7) and to generate statistical summaries (step 8). Based on interpretation of these results, conclusions can be published (steps 11, 12). If potential improvements to the retrieval algorithm have been identified (dash-line loop), a new match-up dataset is generated (step 10) for further evaluation. *This step (dash-line loop) could also be used to evaluate any other retrieval algorithm in a consistent and unambiguous way through the use of the same match-up dataset (step 6) produced here.*

Some important optimizing studies, which are critical to the reliability and consistency of the validation results, are discussed sequentially in section 2b.

b. Sensitivity studies and optimization

The noise in τ_{st} and τ_{sp} may result not only from their natural variability and measurement errors but also from errors associated with improper statistical treatment of the validation process, in particular, selection of match-up window size and sampling approach. If the validation of a retrieval algorithm is not performed according to an optimal procedure (based on solid quantitative analysis), the resulting performance may not be truly representative. An optimal validation procedure is also very important for the intercomparison of different satellite aerosol retrieval algorithms since subtle differences in their validation procedures may cause the performance of the algorithms to appear to be different. In other words, if validation procedures are not standardized, one can always adjust the procedures to obtain comparable performance with another algorithm. This is why one sees similar validation performance for different satellite aerosol retrieval algorithms in the literature. It is time to standardize validation procedures for a more objective validation of satellite aerosol retrievals and their intercomparison. This paper is an attempt to move toward this objective. Initially, the importance of aerosol validation optimization is elucidated since it is a necessary step for validation standardization and is not addressed sufficiently in the satellite aerosol remote sensing literature.

This optimization will be sensitive to the selection of match-up window size since it actually defines the space–time collocation for the two different observations (satellite and ground truth) being compared. In our validation operation, daily quality-controlled (level 2) SP data (multispectral optical thickness) at AERONET stations are used to compute τ and α for the solar reflectance spectral channels of the satellite instruments (0.63, 0.83, and 1.61 μm). Error may be introduced by using different schemes to interpolate (or extrapolate, as is the case at 1.61 μm) aerosol data from SP channels to satellite channels. Sensitivity studies have been performed mainly with respect to these uncertainties, which are discussed below in sequence. These sensitivity studies not only minimize the errors in the validation process but also optimize and standardize the global validation process.

1) TIME–SPACE MATCH-UP WINDOW

Selection of the match-up window is a critical part of the aerosol validation process, since it actually defines the collocation between satellite and SP observations used in the comparison. We have not found any quantitative analysis of this issue in the aerosol retrieval literature. The representativeness of a validation result without considering this issue carefully would be questionable.

Three AERONET stations with the most level 2 SP observations in 1998 (Dry Tortugas, Bermuda, and Kaashidhoo) have been chosen as the base stations for this investigation. AVHRR aerosol optical thickness retrievals have been evaluated against SP observations through linear regression analysis ($\tau_{st} = A + B\tau_{sp}$) (or scatter diagrams of τ_{st} versus τ_{sp}) for 20 different match-up windows summarized in Table 2. A second-order polynomial interpolation scheme [see section 2b(2) for

TABLE 3. (a) Regression parameters A, B, σ, R^2, and N (number of regression samples) for the optimal match-up window (1 h/100 km) and an extreme match-up window of 1 h/500 km for the 0.63-μm channel of AVHRR.

Match-up window	A	B	σ	R^2	N
1 h/500 km	0.1066	0.6712	7.3102E-2	0.4092	207
1 h/100 km	0.0623	0.9514	7.9069E-2	0.5181	120

(b) Regression parameters A, B, σ, R^2, and N (number of regression samples) for the optimal match-up window (1 h/100 km^{-1}) and an extreme match-up window of 1 h/500 km^{-1} for the 0.63-μm channel of VIRS.

Match-up window	A	B	σ	R^2	N
1 h/500 km	0.1786	0.4507	7.3237E-2	0.2254	68
1 h/100 km	0.0910	0.7917	6.2746E-2	0.5978	25

details] with four SP channels (0.87, 0.67, 0.50, and 0.44 μm) is used to estimate τ_{sp} at satellite wavelengths.

The match-up window of Table 2 consists of an outer circle with variable radial distance from the site, excluding an inner circle (with a fixed radius of 25 km) to reduce the effects of coastline or shallow water influences. It is also defined by the time difference (in hours) between the satellite and SP observations. Linear regression coefficients (A, B, σ, and R^2) have been derived for all 20 match-up windows listed in Table 2 at the three selected stations. Plots of the resulting regression coefficients at Dry Tortugas are given as an example in Fig. 2. Results from the other two sites are similar. The window with intercept A closest to zero, slope B closest to unity, highest correlation R^2 or lowest standard error σ is considered optimal. If there are any inconsistencies between the coefficients, the highest value of R^2 determines the optimum window. This is because only when the correlation of two different observations is sufficiently large does the relationship between them have meaning. Based on these criteria, it is obvious from Fig. 2 that a 1 h/100 km match-up window is optimal for Dry Tortugas.

A similar study with 1-h time and variable spatial match-up window has been performed for aerosol retrievals from CERES-SSF version 4 data at Bermuda and Kaashidhoo (which have most of the SSF–SP match-ups in 1998). At Bermuda, the optimal space window is 200 km, if judgment is based on the values of coefficients A and B, but it is 100 km when judgment is based on the correlation coefficient R^2. Thus, as previously mentioned, we chose 100 km as the optimal match-up window for Bermuda. No ambiguity exists for Kaashidhoo and 100 km is optimal.

Regression parameters A, B, σ, and R^2 for the two extreme spatial match-up windows (1 h/100 km and 1 h/500 km) are listed in Tables 3a and 3b for the global ensemble of AEROBS and SSF4 data from the 0.63-μm channel, respectively. Again, a second-order interpolation scheme for the four SP channels (0.87, 0.67, 0.50, and 0.44 μm) has been employed. It is obvious that the regression coefficients are closer to ideal with the 100-km match-up window. This shows how critical the selection of an optimal match-up window can be for aerosol validation.

2) INTERPOLATION/EXTRAPOLATION OF SP OBSERVATION

The sensitivity study on interpolation/extrapolation of SP wavelength-dependent observations to satellite channels (0.63, 0.83, and 1.61 μm) has been performed on two aspects of the problem using the optimal match-up window (1 h/100 km) regression statistics. The first study is on channel selection and the second is on the interpolation scheme. Although sensitivity studies have been performed for both regional and global validations, only results from the global validation, with its larger statistical sample, are presented. Since AERONET SP observations are not available from the same set of channels for all stations, we have examined the validation procedure sensitivity to three selected sets of wavelength channels (listed in Table 4). Channel set I is considered the default set, since it is available at almost all selected AERONET stations. Two interpolation schemes, using first-order and second-order polynomial fits (in natural logarithm of wavelength) to each set of observations, have been selected for testing.

Validation results from the linear regression formula ($\tau_{st} = A + B\tau_{sp}$) for the two SP channel sets (I and II) and the two interpolation schemes are summarized in Tables 5a and 5b for AEROBS. Since extrapolation is involved for the validation of SSF4 1.61-μm channel data, the 1.02-μm channel is added to SP channel set II to form channel set III. Sensitivity results with SSF4 data are summarized in Tables 6a and 6b.

In Tables 5a and 5b, the regression coefficients for each AVHRR channel change only slightly by choosing different SP channel sets and interpolation schemes. We have concluded that these small differences are not statistically significant with the following statistical testing approach, using SP channel set I along with first- and second-order interpolation for the 0.63-μm channel as an example.

First, half of the match-up points (e.g., 60 out of 120 for AEROBS) are randomly picked to do one regression analysis and the other half are used to do a second regression analysis. Then, three difference statistics (DSP_A for intercept A, DSP_B for slope B, and DSP_σ for standard error) are computed from the following three equations:

$$DSP_A = \frac{A_1 - A_2}{\sqrt{s_{A_1}^2 + s_{A_2}^2}}, \quad (1a)$$

$$DSP_B = \frac{B_1 - B_2}{\sqrt{s_{B_1}^2 + s_{B_2}^2}}, \quad \text{and} \quad (1b)$$

$$DSP_\sigma = \frac{\sigma_1^2}{\sigma_2^2}, \quad (1c)$$

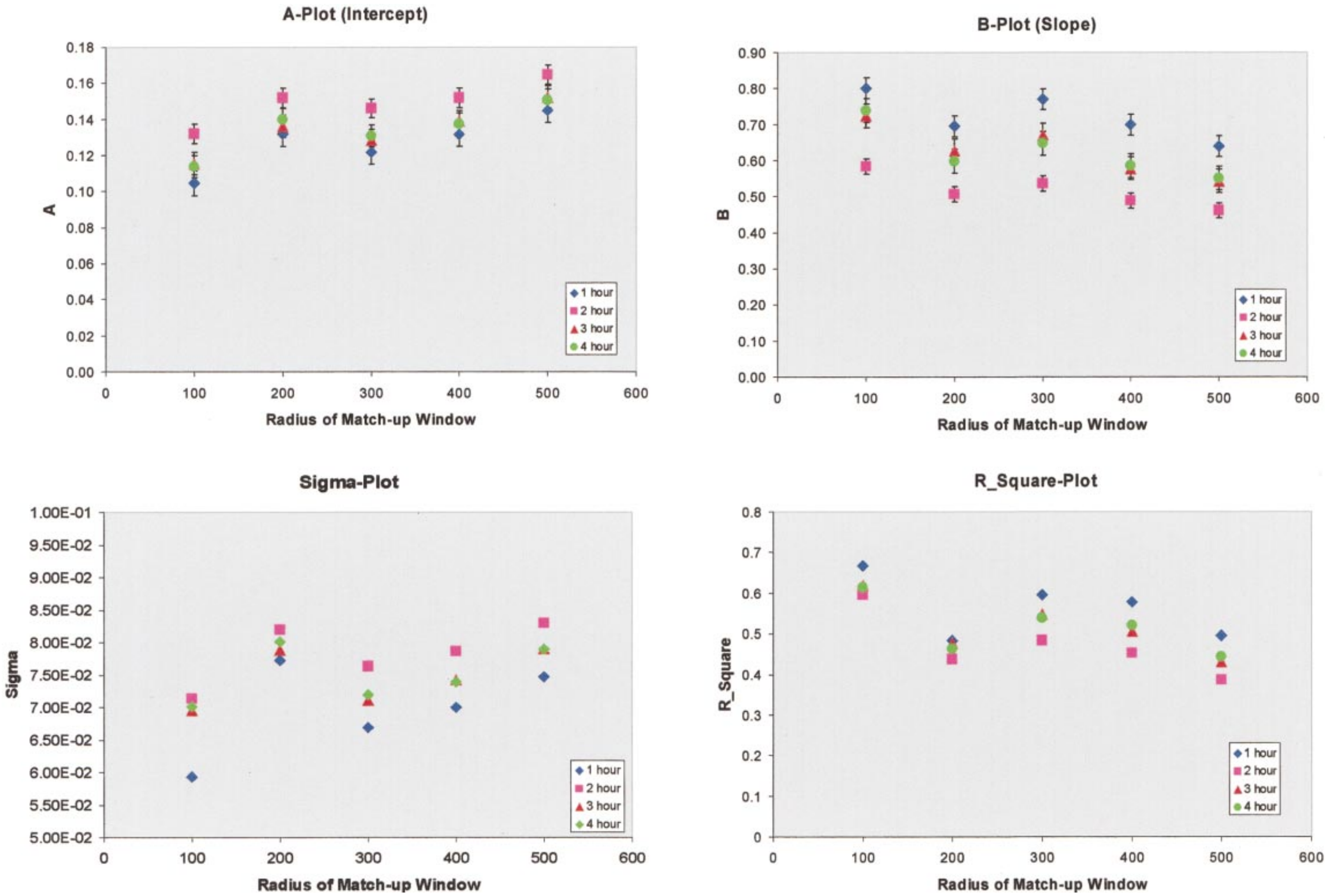

FIG. 2. Plots of regression coefficients (intercept, slope, standard error, and square of correlation coefficient) for 20 different match-up window sizes at the Dry Tortugas AERONET site for AVHRR (AEROBS) aerosol optical thickness from the 0.63-μm channel.

TABLE 4. Three sets of SP wavelength channels for sensitivity study on channel interpolation.

Channel set	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5	Channel 6	Channel 7
I	—	870 nm	670 nm	500 nm	440 nm	—	—
II	—	870 nm	670 nm	500 nm	440 nm	380 nm	—
III	1020 nm	870 nm	670 nm	500 nm	440 nm	380 nm	—

where A_1, B_1, and σ_1 are intercept, slope, and standard error of the regression line for the first half of match-up points, which uses first-order interpolation of set I SP data. Note that A_2, B_2, and σ_2 are from the regression line for the second half of the match-up points, which uses second-order interpolation. Further s_{A_i} and s_{B_i} are standard errors for A_i and B_i for order $i = 1, 2$. If the errors are approximately normal and the two true intercepts and the two true slopes for the regressions of the first- and the second-half match-up data are the same, then DSP_A and DSP_B will have approximately a normal distribution with mean value of zero and standard deviation one. If the errors of the two regressions have the same standard deviation, then DSP_σ will have approximately an *F* distribution (Ostle and Mensing 1988; Beyer 1991). The sample size (60) is large enough to ensure that the approximations are reasonable. If $-1.96 < \mathrm{DSP}_A$ (or DSP_B) $< +1.96$, it suggests there is no significant difference (at the 95% confidence level) between the intercept (or slope) using first- and second-order polynomial interpolation regressions. Similarly, if $0.60 < \mathrm{DSP}_\sigma < 1.67$, it means there is no significant difference (at the 95% confidence level) in the standard error of regression between the first- and second-order polynomial interpolation schemes. For this example, $\mathrm{DSP}_A = -0.218$, $\mathrm{DSP}_B = 0.655$, and $\mathrm{DSP}_\sigma = 0.923$, supporting the conclusion.

The regression statistics not being sensitive to the selection of SP channel sets or interpolation schemes may be due to the fact that the spectral range covered by all SP channel sets always contains the AVHRR channels (0.63 and 0.83 μm). Thus, only interpolation (no extrapolation) is involved. Since second-order interpolation has traditionally been used for the validation of satellite aerosol retrievals in our past research (see Ignatov et al. 1995a,b; Stowe et al. 1997) as well as there being no statistically significant differences between the first- and second-order interpolations, it was decided to continue the use of second-order interpolation with set I SP channels for AEROBS validation.

Results of the above sensitivity study applied to TRMM VIRS (CERES-SSF4) data are presented in Tables 6a and 6b. At 0.63 μm, the change of regression statistics for different SP channel sets and interpolation schemes is minor. Thus, for channel 1 of VIRS, the regression results are not sensitive to the selection of SP channel set or interpolation scheme for the same reason as for AEROBS data. However, for the 1.61-μm channel, the regression statistics are more sensitive, particularly to the selection of interpolation scheme. This is probably because channel 2 of VIRS is beyond the spectral range of the SP channels. Extrapolation is required to obtain SP observations at 1.61 μm. Therefore, the selection of the optimum SP channel set and interpolation scheme is based on the validation performance of channel 2 (1.61 μm) rather than channel 1 (0.63 μm) for SSF data. From visual inspection of regression statistics in Tables 6a and 6b, first-order interpolation with SP channel set I (see highlighted row in Table 6a) appears to be the optimum choice for SSF validation. This choice is not anticipated since one would expect that the best agreement should be achieved for the channel

TABLE 5. (a) Regression statistics for validation of AEROBS data using two SP channel sets interpolated with a first-order polynomial in natural logarithm of wavelength.

Satellite wavelength	SP channel set	*A* value	ΔA value	*B* value	ΔB value	σ value	R^2 value
0.63 (μm)	I	0.0591	0.0147	1.0113	0.0870	7.7959E-2	0.5315
	II	0.0583	0.0148	1.0166	0.0876	7.8002E-2	0.5310
0.83 (μm)	I	0.0862	0.0090	0.9831	0.0721	5.1038E-2	0.6099
	II	0.0855	0.0089	0.9902	0.0711	5.0383E-2	0.6199

(b) Same as above but for second-order polynomial interpolation.

Satellite wavelength	SP channel set	*A* value	ΔA value	*B* value	ΔB value	σ value	R^2 value
0.63 (μm)	I	0.0623	0.0148	0.9514	0.0841	7.9069E-2	0.5181
	II	0.0627	0.0146	0.9682	0.0842	7.8388E-2	0.5263
0.83 (μm)	I	0.0862	0.0090	0.9913	0.0726	5.1016E-2	0.6102
	II	0.0862	0.0090	0.9907	0.0726	5.1036E-2	0.6099

TABLE 6. (a) Same as Table 5a, but for TRMM/VIRS (CERES-SSF4) data.*

Satellite wavelength	SP channel set	A value	ΔA value	B value	ΔB value	σ value	R^2 value
0.63 (μm)	I	0.0858	0.0253	0.8359	0.1361	6.0892E-2	0.6213
	II	0.0890	0.0251	0.8134	0.1337	6.1265E-2	0.6166
	III	0.0890	0.0251	0.8112	0.1336	6.1320E-2	0.6159
1.61 (μm)	**I**	**0.0255**	**0.0131**	**0.9002**	**0.1679**	**3.6199E-2**	**0.5556**
	II	0.0314	0.0133	0.7797	0.1608	3.8186E-2	0.5055
	III	0.0312	0.0132	0.7964	0.1634	3.8084E-2	0.5081

(b) Same as Table 5b, but for TRMM/VIRS (CERES-SSF4) data.*

Satellite wavelength	SP channel set	A value	ΔA value	B value	ΔB value	σ value	R^2 value
0.63 (μm)	I	0.0910	0.0257	0.7917	0.1354	6.2746E-2	0.5978
	II	0.0841	0.0258	0.8425	0.1388	6.1339E-2	0.6157
	III	0.0862	0.0255	0.8335	0.1370	6.1254E-2	0.6167
1.61 (μm)	I	0.0452	0.0115	0.5520	0.1188	3.8994E-2	0.4844
	II	0.0378	0.0124	0.6477	0.1360	3.8536E-2	0.4964
	III	0.0362	0.0129	0.6482	0.1393	3.8970E-2	0.4850

* Match-up data at Dry Tortugas has been eliminated in this regression computation since there is no SP observation on 1.61-μm channel at the station.

set requiring the least amount of spectral extrapolation. We suspect this probably is due to insufficient match-up samples found for the SSF4 data (only 25). More match-up points are expected with the next version of the data (SSF-ED1 uses higher-resolution land mask) and will be used to reexamine this conclusion. There are some water vapor absorption features at 1.02 μm that have not been accounted for in the AERONET aerosol optical thickness. This may also affect the result. It can be examined soon since AERONET data will include wavelength at 1.60 and 2.2 μm in the near future (B. Holben 2000, personal communication).

In summary, for AEROBS validation, second-order interpolation of the SP channel set I is selected, while, for SSF validation, the same SP channel set I but with first-order interpolation is optimal. One additional advantage of selecting SP channel set I is that it gives maximal global SP data coverage since these SP channels are available at all selected AERONET stations.

3) SAMPLING APPROACH

Sensitivity of the validation procedure to satellite data sampling approaches has also been studied. AEROBS data for channel 1 (0.63 μm) is used because it provides many more match-up points for validation than SSF4 data. Four sampling approaches have been investigated, using AERONET SP channel set I with second-order interpolation as truth. The first is the ''ensemble'' approach, which uses all AEROBS aerosol optical thickness values (or 500 closest in distance, if total is more than 500) in the optimum match-up window (1 h/100 km) to determine the aerosol optical thickness mean and variance for that AERONET site and day. The second is the ''best'' approach, where the AEROBS with τ_{st} closest to τ_{sp} is selected from the optimum match-up window. The third is the ''closest'' approach, where the closest (in distance) AEROBS to the SP location is selected from the optimum match-up window. The last is the ''ten closest'' approach, which is the same as the closest but using the 10 closest (in distance) AEROBS to compute mean and variance statistics (used in Ignatov et al. 1995a,b; Stowe et al. 1997). Three AERONET stations, Dry Tortugas, Bermuda, and Kaashidhoo, are used in this study.

Scatterplots and associated linear regression lines of τ_{st} versus τ_{sp} for the three selected stations and four τ_{st} sampling approaches are displayed in Fig. 3. Generally, the regression lines for the best and closest approaches are at the two extremes of meeting the optimum regression criteria. Regression lines associated with the ensemble approach are usually between these two extremes. The performance of the ten closest approach is unstable, which is worst over Dry Tortugas and Kaashidhoo but close to the ensemble approach over Bermuda. This suggests that 10 satellite observations are not sufficient for a reliable statistical analysis of AEROBS data. Since AEROBS and AERONET data are not exactly collocated or invariant in space and time, it is preferable to compare τ_{st} and τ_{sp} averaged over some space–time window. Furthermore, the final output products of AVHRR (and TRMM VIRS) aerosol optical thickness are in gridded format (i.e., averaged over a space–time window). Thus, it is concluded that the ensemble approach is most appropriate for validation of these aerosol products. Although the above study is performed only for AEROBS data, it is likely that this conclusion is generally applicable to validation of other satellite sensor aerosol retrievals, such as TRMM VIRS.

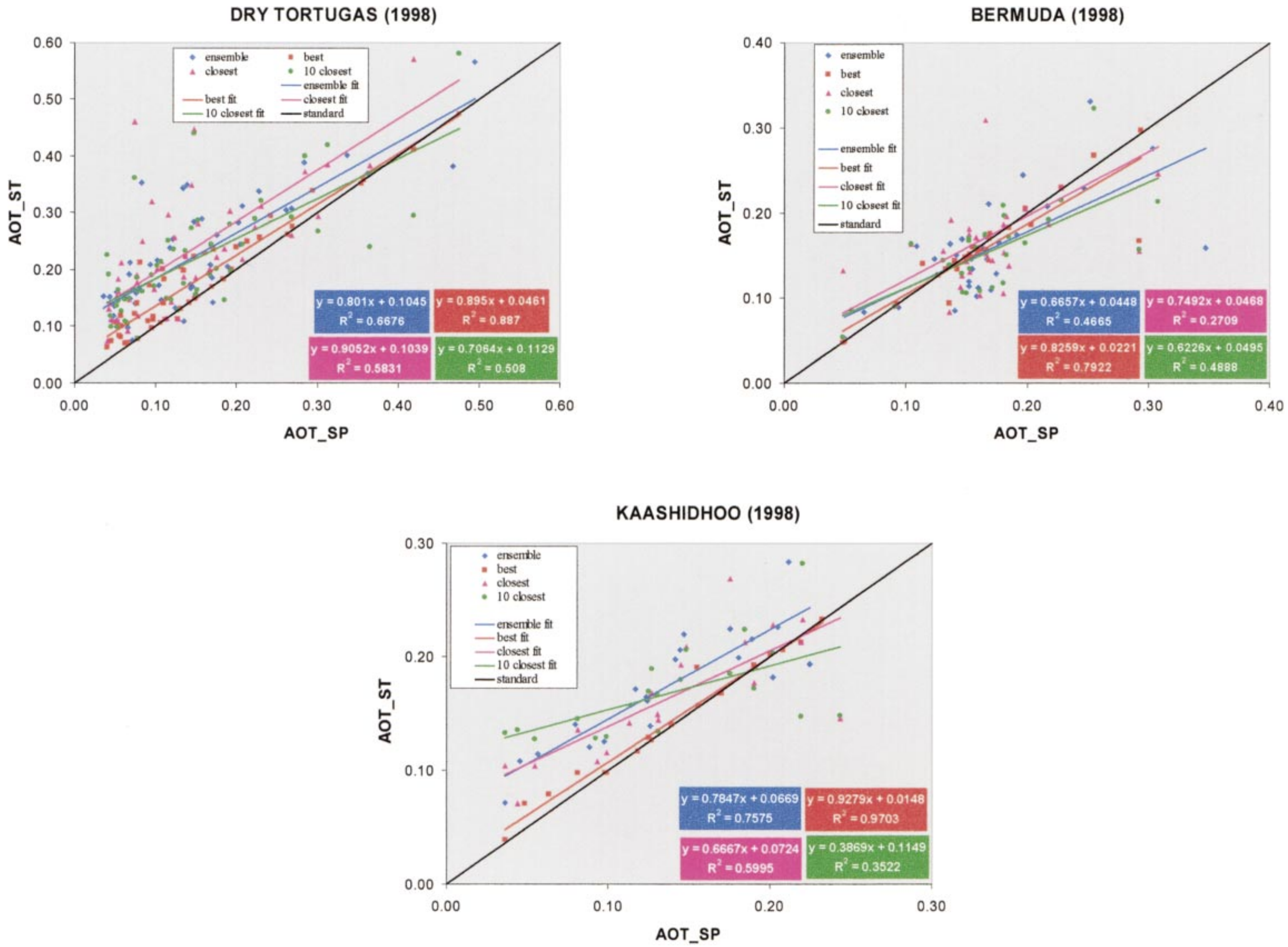

FIG. 3. Scatterplots and regression lines for τ at 0.63 μm for four averaging approaches applied to *NOAA-14* AVHRR (AEROBS) 1998 operational data with a match-up window of 1 h/100 km at Dry Tortugas, Bermuda, and Kaashidhoo AERONET sites. Regression formula $y = Ax + B$ and correlation parameter R^2 have also been presented for the four approaches, where x represents τ_{sp} and y represents τ_{st}. Colors are used to identify the match-up points, regression lines, and regression formulas from the four approaches.

3. Validation results

Using the standard procedures optimized through the above sensitivity studies, global (ensemble) validation on the second-generation NOAA/NESDIS operational aerosol retrieval algorithm was performed for 1998 *NOAA-14* AVHRR (AEROBS) data and TRMM VIRS (CERES-SSF4) data with AERONET observations.

The global linear regression equations predicting τ_{st} for channels 1 (0.63 μm) and 2 (0.83 μm) of *NOAA-14* AVHRR (AEROBS) from τ_{sp} are

$$\tau^1_{st} = 0.062(\pm 0.015) + 0.95(\pm 0.08)\tau^1_{sp}, \quad (2a)$$

$$\tau^2_{st} = 0.086(\pm 0.009) + 0.99(\pm 0.007)\tau^2_{sp}, \quad \text{and} \quad (2b)$$

$$\alpha_{st} = -0.259(\pm 0.107) + 0.26(\pm 0.07)\alpha_{sp}, \quad (2c)$$

with standard errors of $\sigma_1 = 0.079$, $\sigma_2 = 0.051$, and $\sigma_\alpha = 0.432$, and correlation coefficients of $R_1 = 0.72$, $R_2 = 0.78$, and $R_\alpha = 0.34$. The values in parentheses are plus and minus one standard error.

The same equations for channels 1 (0.63 μm) and 2 (1.61 μm) of TRMM VIRS (CERES-SSF4) observations are

$$\tau^1_{st} = 0.086(\pm 0.025) + 0.84(\pm 0.14)\tau^1_{sp}, \quad (3a)$$

$$\tau^2_{st} = 0.026(\pm 0.013) + 0.90(\pm 0.17)\tau^2_{sp}, \quad \text{and} \quad (3b)$$

$$\alpha_{st} = 1.167(\pm 0.821) - 0.25(\pm 0.74)\alpha_{sp}, \quad (3c)$$

with standard errors of $\sigma_1 = 0.061$, $\sigma_2 = 0.036$, and $\sigma_\alpha = 1.278$, and correlation coefficients of $R_1 = 0.79$, $R_2 = 0.75$, and $R_\alpha = 0.07$. The *NOAA-14* AVHRR (AEROBS) validation is much more reliable than that of TRMM VIRS (CERES-SSF4) because there are 120 match-ups for AVHRR but only 25 for VIRS. This may be the principal reason for the difference in regression coefficients for the two satellite observations at 0.63 μm. Also, the CERES-SSF dataset is being reprocessed with improved VIRS cloud and land masks, so this validation is clearly very preliminary and is only shown for comparison to AVHRR. It will not be discussed further.

The *NOAA-14* AVHRR validation at 0.63 μm is somewhat worse than three previous validations for *NOAA-9* and -11 AVHRR using ship-borne sun photometers (see Ignatov et al. 1995a,b; Stowe et al. 1997). This is mainly due to differences in the validation procedures, which is a good illustration of the importance of standardizing these procedures. More subjective decisions were involved in these earlier validations. For example, AEROBS with large spatial variability ($\sigma >$ 0.05) within a larger spatial (300 km) and time ($\pm$2 h) match-up window, and τ_{st} values more than 2σ larger than the mean were thrown out. These subjective adjustments (especially the later one) can misrepresent the true statistical validation of the τ_{st} values produced by the algorithm. Moreover, days when the ship's sun photometer records exhibited large temporal variability were also rejected. In the *NOAA-14* AVHRR validation, only the last filter (SP records with large temporal variability) is applied to exclude rare defects in the AERONET data.

Figure 4 shows scatterplots of the 120 AEROBS–AERONET match-up days for τ_1, τ_2, and α (means and std devs) with the corresponding linear regression equations. One can see from the figure how the τ_{st} filtering used in the earlier validation would improve the results. The cause of these outliers (far-off regression line and large $\sigma_{\tau_{st}}$) are thought to be due to cloud contamination but this has not yet been established. It is obvious from Fig. 4 [and Eqs. (2c) and (3c)] that there is larger uncertainty in the derived values of α from the retrieved τs. Ignatov et al. (1998) and Ignatov and Stowe (2002a,b) have shown that random errors in τ are amplified when deriving α, particularly as τ approaches zero. This is a result of the defining logarithm ratio relationship ($\alpha = -\ln[\tau_1/\tau_2]/\ln[\lambda_1/\lambda_2]$). It is of interest to mention that both forward numerical retrieval sensitivity studies (Mischenko et al. 1999) and theoretical analyses (Ignatov and Stowe 2000) confirm that α is less subject to error in the assumed aerosol retrieval model than are the τs from which it is derived. Thus, they conclude that α, which is least sensitive to uncertainties in the atmosphere–ocean model, should be retrieved along with τ as a second aerosol parameter. However, our validation indicates that one has to be careful in deriving α to fully benefit from this low sensitivity to errors in the retrieval model. This reduced error sensitivity in α results from assuming either directly (in the theoretical study) or indirectly (in the forward sensitivity test) that the retrieved τs are mainly subject to multiplicative errors, which cancel when taking their ratio. This is clearly not substantiated by the actual observations used in our validation (bottom panel in Fig. 4). Actually, there are sufficient random errors in the retrieved τs (may be from error in the aerosol retrieval model or other errors in the measurement) that are nonmultiplicative and therefore do not cancel when taking their ratio (see also Higurashi et al. 2000). It appears that an algorithm, in which α is derived directly from τs in two separated channels, cannot give quantitatively dependable values (qualitatively, perhaps). The qualitative information is still useful for separating broad categories of aerosol types (dust, haze, smoke, etc.) as has been demonstrated by others (Mishchenko et al. 1999; King et al. 1999; Higurashi et al. 2000). This issue is revisited in the following section along with results of investigations to identify possible sources of systematic error in τ as implied by the regression lines in Fig. 4. Thus, the validation procedure is shown to be a useful tool for adjusting aerosol retrieval algorithm parameters to reduce these errors (through the dash-line loop in Fig. 1).

4. Application of the validation procedure to algorithm improvement

a. Specular reflectance

To eliminate the impact of specular reflectance from the rough oceanic surface, aerosol retrievals in the sec-

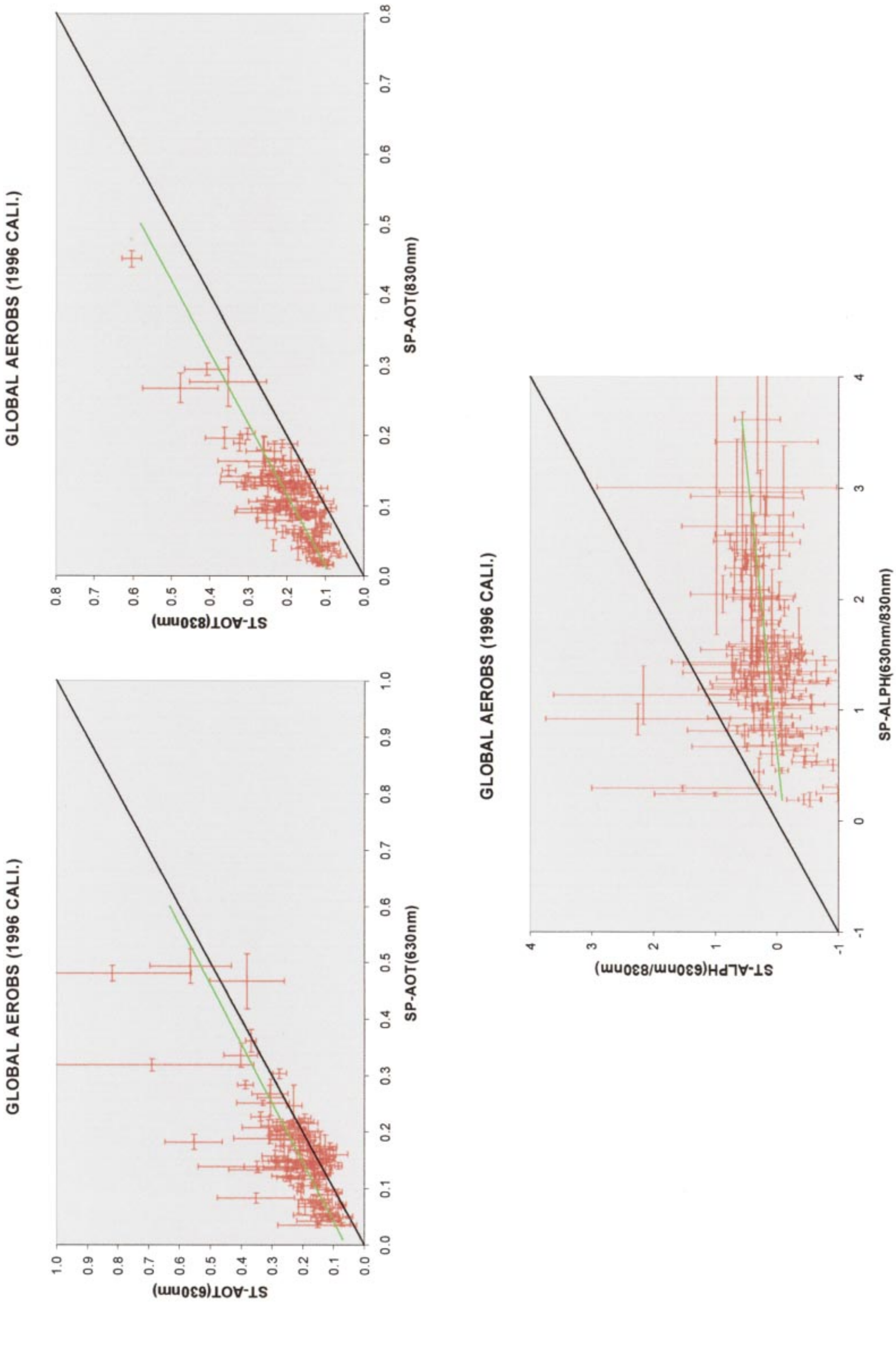

FIG. 4. Scatterplots of τ_1, τ_2, and α and linear regression lines from the optimum global (all seven AERONET sites from Table 1) validation of 1998 AVHRR (AEROBS) data which used the operational (1996) calibration scheme. Horizontal and vertical error bars are $\pm$ one std dev long.

TABLE 7. (a) Regression statistics for determining the sensitivity of AEROBS retrieval errors to specular reflectance.

Satellite channels	γ angle (°)	A value	B value	σ value	R^2 value
0.63 μm	>40	0.0623	0.9514	7.9069E-2	0.5181
	>50	0.0607	0.9661	8.0571E-2	0.5377
	>60	0.0612	0.9680	8.0389E-2	0.5459
0.83 μm	>40	0.0862	0.9913	5.1016E-2	0.6102
	>50	0.0860	1.0023	5.0504E-2	0.6407
	>60	0.0844	1.0279	5.1070E-2	0.6528

(b) Same as above, but for SSF4 data.

Satellite channels	γ angle (°)	A value	B value	σ value	R^2 value
0.63 μm	>40	0.0858	0.8359	6.0892E-2	0.6213
	>50	0.0985	0.6690	5.2072E-2	0.6229
	>60	0.0982	0.6621	5.7342E-2	0.5976
1.61 μm	>40	0.0255	0.9002	3.6199E-2	0.5556
	>50	0.0363	0.5753	2.7938E-2	0.5017
	>60	0.0433	0.4841	3.4773E-2	0.3195

ond-generation NOAA/NESDIS operational algorithm are limited to gamma (γ) angles (angle between viewing angle and specular ray from the flat ocean) larger than 40°. However, consistency checks of AEROBS and VIRS (CERES-SSF4) data suggest a possible impact of specular reflectance at $\gamma > 40°$ (Ignatov and Stowe 2002). To investigate this speculation, a sensitivity study has been performed by including all observations with $\gamma > 50°$ and $\gamma > 60°$ in two subsets of the original match-up database for both AEROBS–SP and SSF4–SP data. Accordingly, these datasets have been subjected to the validation procedure and results are summarized in Tables 7a and 7b, respectively, for AEROBS and SSF4.

The improvement of regression results for both channels of AEROBS data is minor with increases in the γ angle limit from 40° to 60°. However, for SSF data, the change of regression coefficients is more irregular. For both 0.63-μm and 1.61-μm channels, the τ_{st} retrieval does appear to be sensitive to increases in the γ angle limit. However, all regression parameters become worse with increasing γ angle limit, opposite to what is expected if specular reflection were affecting the retrievals. This is again probably due to the much lower number of regression points (25 in SSF4 compared to 120 in AEROBS for $\gamma > 40°$) involved in the SSF4 analysis. Thus, based solely on the AEROBS analysis, it is concluded that increasing the γ angle limit beyond the nominal 40° does not significantly remove errors in aerosol optical thickness, and therefore, specular reflection of radiation beyond 40° is probably not a serious problem.

b. Calibration, Rayleigh scattering, and diffuse surface reflectance

Lowering the systematic high bias at low aerosol optical thickness in the above validation results (positive regression line intercept in Fig. 4) has been sought by checking the values of nonaerosol related elements (including calibration, Rayleigh scattering, and diffuse surface reflectance) used in the current operational retrieval algorithm. This investigation has been performed on both AEROBS and SSF4 data, but because of the small sample of SSF4 match-up points, the analysis of SSF4 data is used only as a check on the conclusions drawn from the analysis of AEROBS data and is not shown here. In Table 8, the original AEROBS operational values of calibration, Rayleigh scattering, and diffuse surface reflectance are listed as well as values after correction or adjustment based on sensitivity studies using our validation procedure. The imaginary part of the aerosol refractive index is also listed, corresponding to the discussion in section 4c).

The regression statistics before and after the adjustments (or corrections) listed in Table 8 are summarized in Table 9. First, the most recent calibration slope drift correction coefficients (C. R. N. Rao 1998, personal communication) are used to correct AEROBS reflectances. These are then used to derive new τ_1 and τ_2 values (dash-line loop in Fig. 1), to which the global validation procedure is again applied. The high bias at low aerosol optical thickness (or intercept A) is lowered in both channels 1 and 2 as a result. However, the proportional error (slope B) and correlation (R^2) are moved further away from their ideal values, while there is a small improvement in the standard error.

The Rayleigh scattering optical thickness used in our operational algorithm has been found to be in error, based on 6S radiative transfer model calculations for the exact response functions of channels 1 and 2 for *NOAA-14* AVHRR (A. Ignatov 2000, personal communication). It is too large in both channels (cf. Table 8). Using these correct values in the retrieval LUTs actually increases the bias at low aerosol optical thickness. The change of the other regression parameters (except for slope in channel 2) is minor. Since more aerosols (larger τ) are now required to match the observed reflectance, the only remaining free model parameter,

TABLE 8. Corrections and adjustments to the AVHRR calibration and operational algorithm to reduce positive bias at low τ (intercept > 0) and negative bias at large τ (slope < 1).

Channel	Operation (1996)	Correction (1998)
	Calibration slope (CS) (d = day after launch)	
0.63 μm	S1 = 0.109 + 2.32 × 10^{-5} d	S1 = 0.1107 + 1.35 × 10^{-t} d
0.83 μm	S2 = 0.129 + 3.73 × 10^{-5} d	S2 = 0.1343 + 1.33 × 10^{-t} d
	Rayleigh optical thickness (ROT)	
Channel	Operation	Correction
0.63 μm	0.0607	0.0554
0.83 μm	0.0205	0.0180
	Diffuse surface reflectance (DSR)	
Channel	Operation	Adjustment
0.63 μm	0.002	0.01
0.83 mm	0.0005	0.0006
	Imaginary part of the aerosol refractive index (IPARI)	
0.63 μm	0.000	0.005
0.83 μm	0.000	0.008

which can effectively lower this bias, is the diffuse surface reflectance ρ_{dsr}.

Sensitivity studies have been performed with the validation procedure to identify an optimal diffuse surface reflectance that, together with the correct calibration and Rayleigh optical thickness, reduces the bias at low τ to near zero. For each retrieval channel of AVHRR, several values of surface diffuse reflectance have been used to construct new LUTs for deriving τ_1 and τ_2 from the reflectances, to which the global validation is again applied (dash-line loop in Fig. 1). The values of ρ_{dsr} that yield an intercept closest to zero in each channel are shown in Table 8 (DSR heading) and the corresponding regression statistics are listed in Table 9. The new values for ρ_{dsr} for both 0.63 μm and 0.83 μm are larger than before, so less aerosol is required to match the observed reflectances. While greatly improving the bias at low τ, it is apparent from Table 9 that the other regression statistics are only slightly degraded. The new diffuse surface reflectances are somewhat larger than expected for open oceans. This may suggest that there are coastal (e.g., shallow water) effects within 25–100 km of each AERONET site or that the simple Lambertian surface model assumption, with its adjustment for diffuse glint used in the current operational algorithm (Stowe et al. 1997), has to be raised to nonphysically high values in order for the retrievals to be unbiased at low aerosol optical thickness. The 6S radiation transfer model, which treats the surface reflectance as a wavy Fresnel surface with wind-driven slopes, will provide our future LUTs (see Ignatov and Stowe 2002a,b). It thus may allow a more physically consistent value to again be used for this diffuse reflectance component.

c. *Imaginary part (n_i) of the aerosol refractive index*

The above improvements from nonaerosol parameters, which reduce the bias at low optical thickness, actually worsen the bias at large optical thickness (slope of regression line further away from unity). The only way to remove this bias and not affect the intercept is

TABLE 9. Regression statistics for AEROBS data before and after sequentially adjusting calibration and aerosol retrieval algorithm parameters (see Table 8) to reduce retrieval errors.

AVHRR channels	Adjusted parameter	A value	ΔA value	B value	ΔB value	σ value	R^2 value
0.63 μm	Operation	0.0623	0.0148	0.9512	0.0841	0.07907	0.5180
	1998 CS	0.0332	0.0145	0.8404	0.0821	0.07716	0.4683
	ROT	0.0654	0.0144	0.8465	0.0819	0.07707	0.4725
	DSR	0.0030	0.0147	0.8308	0.0838	0.07875	0.4525
	IPARI	−0.0019	0.0174	0.9903	0.0987	0.09284	0.4579
0.83 μm	Operation	0.0861	0.0090	0.9913	0.0726	0.05101	0.6103
	1998 CS	0.0538	0.0079	0.8578	0.0641	0.04506	0.6004
	ROT	0.0679	0.0076	0.8219	0.0616	0.04326	0.5995
	DSR	0.0173	0.0083	0.8199	0.0666	0.04679	0.5601
	IPARI	0.0178	0.0099	0.9996	0.0803	0.05639	0.5658

by adjusting parameters of the aerosol microphysical model used in the retrieval algorithm.

The imaginary part of the aerosol refractive index is the obvious first choice for adjustment since the aerosol is assumed to be nonabsorbing in the current operational retrieval algorithm. If it is absorbing, as most tropospheric aerosol with large optical thickness is (e.g., Nakajima et al. 1996; Kaufman et al. 1997; Tanre et al. 1997; Nakajima and Higurashi 1997; Mishchenko et al. 1999; Schmid et al. 1999; Dubovik and King 2000), this would cause the subunity slope observed in the validation results (see, Ignatov 1995a; Stowe et al. 1997). Several values of n_i have been selected for each channel of AVHRR and new LUTs were generated. Again, the resulting aerosol optical thicknesses (τ_1 and τ_2) are then subjected to the global validation procedure (dash-line loop in Fig. 1). The n_i that yields a slope closest to unity in each channel is 0.005 for channel 1 (0.63 μm) and 0.008 for channel 2 (0.83 μm; cf. Table 8). A similar wavelength dependence of n_i for these two AVHRR channels has also been observed by Higurashi et al. (2000) in the validation of their two-channel AVHRR aerosol retrieval algorithm.

It is seen in Table 9, that the random error of the retrieved τ (represented by the standard error of linear regression, σ) is not reduced by any of the above corrections and adjustments. These random errors in τ_1 and τ_2 are amplified when used to derive α as already discussed. This random error is likely to be at least partially due to regional variability in aerosol type. The second-generation algorithm uses a uniform aerosol model (a lognormal distribution with a 0.1-μm mode radius and 2.03-μm variance and refractive index $m = 1.4 - i0.0$) globally. The current match-up sample size for a single site of AERONET in 1998 is too small to verify this hypothesis with the validation procedure. AVHRR (AEROBS) data are being collected for 1999 and 2000 so that the regional validation can be done.

A new retrieval algorithm that accounts for regional difference in aerosol particles may be required to reduce this random error and possibly some of the systematic errors as well. As mentioned before, this kind of algorithm needs to be designed carefully to minimize the effects of τ errors on a derived size parameter, like α. There are at least two approaches. The first is to derive α iteratively from a set of LUTs corresponding to different α values. These LUTs should cover the range of possible atmospheric values of α with sufficient resolution. The α derived by using τs from the second-generation algorithm can be used as a first guess. The first guess α is then used to find a closely corresponding LUT for deriving new τs and α. This process will continue until the derived α converges. The second approach is to make α another dimension in an LUT and to retrieve τs and α simultaneously. The first approach should yield more accurate τs by including regional size information of aerosols through α on the LUTs, which in turn may reduce errors passed to α from the τ retrieval. The second approach should be expected to yield more accurate retrievals of α since it avoids using retrieved τs directly to derive it. These two approaches are basically the structured and unstructured approaches defined by Ignatov and Stowe (2000) for the third-generation retrieval algorithm presently under development at NOAA/NESDIS. The new algorithm allows the aerosol parameter most affecting the Ångström wavelength exponent to vary in the LUTs. This makes the retrieval algorithm sensitive to regional differences in aerosol in some senses and should reduce random error accordingly. Further discussion of this topic is outside the scope of this paper and will be presented later by its developers. The validation procedure presented here will play a critical role in assessing the performance of the new algorithm.

As a summary of this section, improvements of the validation results at low aerosol optical thickness (by adjusting nonaerosol model parameters) and at high aerosol optical thickness (by correcting the imaginary part of the aerosol refractive index) are demonstrated if one compares the scatterplots of Fig. 5 with those of Fig. 4 for AEROBS data.

5. Discussion

We feel the need for discussions on some important issues related to our validation before we draw final conclusions. First is the representativeness of the AERONET-measured aerosol to the global tropospheric aerosols. Actually, very careful study on this issue has been performed recently by the AERONET scientists (e.g., Kaufman et al. 2001, manuscript submitted to *Geophys. Res. Lett.*; Smirnov et al. 2002). They have compared AERONET measurement (spanning over 2–5 yr) of τ and size parameters (such as Ångström wavelength α and effective radius $R_{\rm eff}$) over selected AERONET marine sites in both the Pacific (Nauru, Lanai, Tahiti) and the Atlantic (Bermuda, Ascension Island) with 150 historical ship measurements from the last 3 decades. Their results convincingly indicate that the AERONET measured maritime aerosol is a good representation of the tropospheric background marine aerosols, especially for the sites in the Pacific. So, we believe AERONET aerosol measurements can be used to represent the climatologies of tropospheric maritime aerosols as long as the AERONET sites are carefully selected.

The next issue is how to reasonably extend the adjustment or correction we identified through sensitivity studies described in section 4 to global implementation, especially for the diffuse surface reflectance and the aerosol model used in the retrieval algorithm. Regional validation based on sufficient match-up samples is a feasible next step. Three years of match-up data (1998, 1999, and 2000) are being collected now for reliable regional validation. First, the diffuse surface reflectance determined from the sensitivity studies for the AERO-

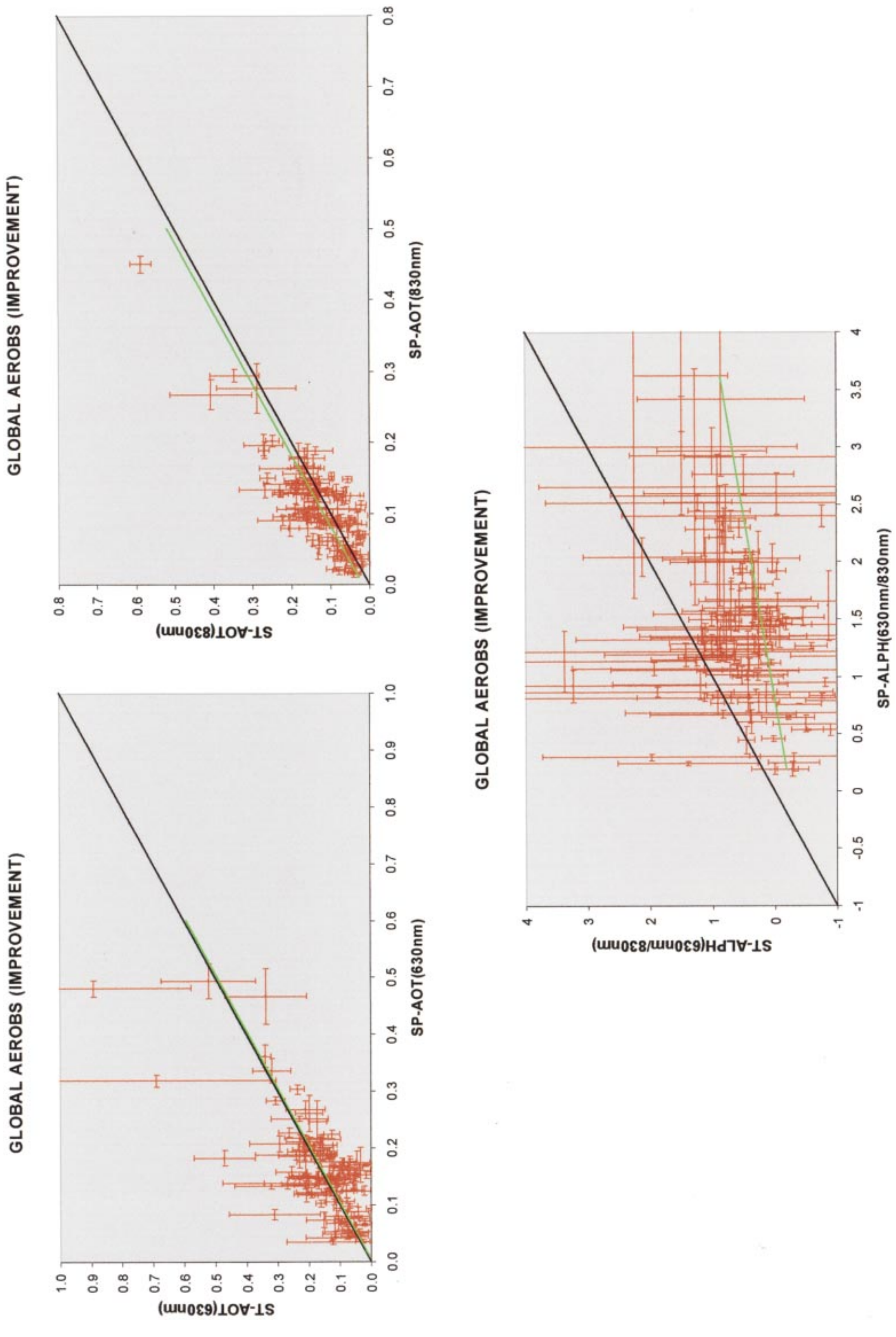

FIG. 5. Same as Fig. 4, but with an improved AVHRR calibration scheme (1998) a retrieval algorithm incorporating a correct Rayleigh optical thickness, and adjusted values of surface diffuse reflectance and imaginary part of the aerosol refractive index.

NET sites that are surrounded by deep water and fewer waves will be analyzed to determine an appropriate value of the adjustment that can be applied globally in the algorithm. For example, Lanai is a good candidate site due to the fact that it is surrounded by deep water, is sheltered from wind by the surrounding islands, receives very little rain and is not known for large waves (Price 1983; Smirnov et al. 2002). Second, baseline AERONET sites can be further divided into different categories according to the prevailing origins of the aerosols over them (such as maritime, mineral dust, biomass burning, and urban–industrial, or a combination of them) and the aerosol properties derived from the skylight analyses. Sensitivity studies used in the global validation for aerosol model adjustment can be similarly applied to each category of sites. This will allow us to evaluate the aerosol regional representativeness of a retrieval algorithm.

One would naturally expect a variable regional performance for different categories of sites if the retrieval algorithm assumes a global uniform aerosol model (such as our second-generation algorithm). Correspondingly, more stable regional performance is anticipated for a retrieval algorithm that is sensitive to some kind of aerosol type (such as our third-generation algorithm). The purpose of the regional validation is to identify and document this variable performance of a retrieval algorithm in a quantitative way. The validation procedures proposed in this paper form a basis for achieving this purpose.

One may have noticed that we have basically made two kinds of manipulation to maximize the agreement between the satellite retrieval and the ground truth. The first is the optimization procedures based purely on the statistic regression analyses. The second is the fine adjustments to the intercept and slope of the linear regression formula based on solid physical rationale, which is illustrated by using the following linearized single-scattering approximation of the radiative transfer equation (see also Stowe et al. 1997):

$$\tau_{\mathrm{ST}} = 4\mu_s\mu_v\frac{\rho - \rho^R - \rho^S T}{\omega P^A}, \qquad (4)$$

where ρ is an apparent reflectance of the ocean–atmosphere system; ρ^R is the Rayleigh scattering contribution; ρ^S is the diffuse surface reflectance; T is the total atmospheric transmittance; P^A and ω are the aerosol phase function and single-scattering albedo, respectively; and μ_s and μ_v are the cosine of solar and view zenith angles, respectively. The errors in aerosol optical thickness τ_{ST} may result from ρ^S, P^A, and ω since all other terms are reasonably well known. Note that ρ and ρ^S participate in the equation as additive terms, and (ωP^A) as a multiplicative one. This suggests that the nonzero intercept ($A \neq 0$) in the regression formula $\tau_{\mathrm{ST}} = A + B\tau_{\mathrm{SP}}$ comes from calibration error or from the improper treatment of ocean surface reflection, and the nonunity slope ($B = d\tau_{\mathrm{ST}}/d\tau_{\mathrm{SP}} \neq 1$) comes from incorrect assumptions in the aerosol model. It is this physical rationale that leads us to make adjustments performed with sensitivity studies in the previous section 4. In summary, our validation approach is based on a combination of purely statistical regression with fine adjustments to the intercept and slope of the regression formula based on physical rationale.

One can see that after sufficient correlation has been established from the statistical analysis, further delicate adjustment on the slope and intercept helps us identify possible sources of error and reduce them in the retrieval algorithm with only minor degradation on the correlation as shown in Table 9. We think this combination of pure statistical analysis with the intercept and slope adjustments (supported by physical rationale) is the unique aspect of the validation procedure used in the paper. Of course, we have to admit that more investigations are needed to justify the quantity of the adjustment suggested from the sensitivity studies. This is very difficult to do using the Dave radiative transfer model used in our second-generation operational algorithm due to its limitations. This is better done with our third-generation algorithm based on the more complete 6S radiative transfer model, which will give more precise answers to these kinds of questions. The new features in the 6S code, such as a wavy Fresnel surface with wind-driven slopes and more widely used aerosol model (e.g., bimodel lognormal distribution), will make the further physical quantitative studies possible. It can be seen that the validation procedures adopted in this paper and the datasets collected form a basis for testing the new algorithm.

6. Summary and concluding remarks

For more objective global validation and intercomparison of aerosol retrieval algorithms, a validation package utilizing AERONET observations as ''ground truth'' has been developed. Based on detailed sensitivity studies, a 1 h 100 km^{-1} time–space match-up window has been identified to assure sufficient collocation and comparable space–time variance statistics for aerosols measured from two different platforms (satellite and ground). Also, the ensemble method of averaging on this time–space scale was found to be best for identifying biases in satellite aerosol retrieval products (such as optical thickness and Ångström wavelength exponent) especially important for climate research. This space–time scale is also close to the scales used when gridding these pixel level retrieval products for statistical analyses. For validation of aerosol retrievals from *NOAA-14* AVHRR with AERONET measurements, the final result is not sensitive to the choice of interpolation scheme and SP channel sets. Thus, for historical continuation, a second-order polynomial fit in logarithm of wavelength to the four SP channels is used to determine the AERONET optical thickness for the two AVHRR channels. For validation of TRMM/VIRS (CE-

TABLE 10. Systematic and random errors for 1998 AEROBS τ data before and after making calibration, Rayleigh optical thickness, surface diffuse reflectance, and imaginary part of the aerosol refractive index corrections and adjustments listed in Table 8.

Correction status	Channel (μm)	Systematic errors: Minimum (τ = 0.00)	Systematic errors: Mean (τ = 0.15 at λ_1) (τ = 0.11 at λ_2)	Systematic errors: Maximum (τ = 1.00)	Random error ($\pm$)
Before correction	λ_1 = 0.63	+0.06	+0.05	+0.01	0.08
	λ_2 = 0.83	+0.09	+0.08	+0.08	0.05
After correction	λ_1 = 0.63	−0.01	−0.003	−0.01	0.08
	λ_2 = 0.83	+0.02	+0.02	+0.02	0.05

RES-SSF4) retrievals, the same four SP channels, but with first-order polynomial interpolation, are optimal, even though extrapolation is required for the 1.61-μm channel.

This global validation package has been applied to the second-generation NOAA/NESDIS operational aerosol retrieval algorithm by using 1998 observations from *NOAA-14* AVHRR (AEROBS) and TRMM/VIRS (CERES-SSF4). It has been shown to be useful for improving the satellite aerosol retrieval by making adjustments to aerosol retrieval algorithm parameters. Resulting linear regression slopes and intercepts are used to estimate systematic errors at mean τ values and two extremes. The standard error of regression provides an estimate of random error. The systematic and random errors from global validation for 1998 AVHRR (AEROBS) τ data for channel 1 (0.63 μm) and 2 (0.83 μm) are summarized in Table 10. A similar summary for 1998 VIRS (CERES-SSF4) τ data for channel 1 (0.63 μm) and 2 (1.61 μm) is given in Table 11. Values, both before and after correcting and adjusting nonaerosol model parameters and the imaginary part of the aerosol refractive index, are listed for comparison.

The *NOAA-14* AVHRR (AEROBS) τ values for mean conditions are biased high by 0.05 and 0.08, with random errors of 0.08 and 0.05, at 0.63 μm and 0.83 μm, respectively. Correspondingly, the TRMM/VIRS (CERES-SSF4) τ values for mean conditions are biased high by 0.06 and 0.02, with random errors of 0.06 and 0.04, at 0.63 μm and 1.61 μm, respectively. After corrections and adjustments, the biases in both channels of AVHRR and VIRS have been reduced significantly and are close to zero, although random error is almost unchanged. The effective Ångström wavelength exponent α, derived directly from the τs, has been shown to be poorly correlated both before and after adjustments, indicating that random error in the τ measurement (possibly related to aerosol model parameter variations or cloud–surface reflectance contamination) needs to be reduced. The results of TRMM/VIRS validation need to be viewed with caution due to the small match-up sample size resulting from use of a preliminary version 4 of the CERES-SSF dataset. However, it is interesting to see that the algorithm adjustments and statistical results of SSF4 aerosol validation are not in conflict with those from AVHRR–AEROBS, even though the sample size of SSF4 match-ups is much smaller. The γ angle limit of 40° used in the NOAA/NESDIS second-generation retrieval algorithm to exclude the contaminating effects of ocean specular reflection appears to be adequate.

Currently, validation is restricted to the global scale to obtain sufficient match-up sample size. In the future, multiple year data will be collected for regional validation analysis, which will help to identify algorithm biases due to regionally different aerosol types. These global validation studies suggest that the random errors in the current NOAA/NESDIS operational aerosol retrieval algorithm could very well be due to regional differences in aerosol particles. These differences are to be included in the third-generation algorithm, currently under development at NESDIS. The new algorithm, similar to those already developed by Mishchenko et al. (1999), Higurashi and Nakajima (1999), and Tanre et al. (1997), will derive aerosol optical thickness and a size parameter (such as Ångström wavelength exponent

TABLE 11. Same as Table 10, but for 1998 TRMM/VIRS (CERES-SSF4) τ data before and after making Rayleigh optical thickness, surface diffuse reflectance, and imaginary part of the aerosol refractive index corrections and adjustments (no corrections were made to VIRS calibration and to Rayleigh optical thickness for the 1.61-μm channel). The errors after adjustments in 0.63-μm channel are not optimized since adjustments were taken directly from the AEROBS analyses. Surface diffuse reflectance before and after adjustment for the 1.61-μm channel is 0.000 and 0.002, respectively, while the imaginary part of the aerosol refractive index is adjusted from 0.000 to 0.015.

Correction status	Channel (μm)	Systematic errors: Minimum (τ = 0.00)	Systematic errors: Mean (τ = 0.16 at λ_1) (τ = 0.07 at λ_2)	Systematic errors: Maximum (τ = 1.00)	Random error ($\pm$)
Before correction	λ_1 = 0.63	+0.09	+0.06	−0.07	0.06
	λ_2 = 1.61	+0.03	+0.02	−0.07	0.04
After correction	λ_1 = 0.63	+0.03	+0.04	+0.08	0.08
	λ_2 = 1.61	+0.004	+0.005	+0.01	0.04

or effective radius) from two (or three for *NOAA-16*) channels, used dependently, rather than independently. This makes the retrieval sensitive to regional differences in aerosol, particularly with respect to particle size. The validation procedure presented here will be applied to the results from the third-generation algorithm to evaluate and document its performance. We also hope the standardized validation procedure presented here can be used by other research groups for their algorithm validation, which will make algorithm intercomparison much more meaningful in the future. The archived match-up dataset is available for algorithm intercomparison upon request from the authors.

Acknowledgments. We express our deep gratitude to the AERONET project leader Dr. Brent Holben at NASA Goddard Space Flight Center and to the principal investigator for Dry Tortugas site, Dr. Kenn Voss at University of Miami, for supplying 1998 AERONET sun-photometer observations; Mr. David Donahue at Science and Technology Corporation, for supplying AVHRR–AEROBS 1998 data; and Ms. Linda Hunt at NASA Langley, for supplying TRMM/VIRS (CERES-SSF4) 1998 data. Discussions with Dr. Alexander Ignatov at NOAA/NESDIS also were of great benefit. Two anonymous reviewers' comments were very constructive for improvements to the manuscript. This work was funded by NASA Langley through Contract L90987C managed by Drs. Bruce Wielicki and Bruce Barkstrom, CERES science team managers; and by NOAA/NESDIS through a grant from the ocean remote sensing program managed by Dr. Lee Dantzler.

REFERENCES

Beyer, H. W., Ed.,1991: Probability and statistics. *CRC Standard Mathematical Tables and Formulae,* 29th ed., CRC Press, 470–527.

Charlson, R. J., S. E. Schwartz, J. M. Hales, R. D. Cess, J. A. Coakley Jr., J. E. Hansen, and D. J. Hofmann, 1992: Climate forcing by anthropogenic aerosols. *Science,* **255,** 423–430.

Chu, D. A., Y. J. Kaufman, L. A. Remer, and B. N. Holben, 1998: Remote sensing of smoke from MODIS airborne simulator during the SCAR-B experiment. *J. Geophys. Res.,* **103,** 31 979–31 987.

——, ——, D. Tanre, and B. N. Holben, 1999: Smoke optical properties derived from airborne measurements in Amazonia. *Proc. ALPS'99 Conf*. Meribel, France, CNES, WK1-p8.

Dubovik, O., and M. King, 2000: A flexible inversion algorithm for retrieval of aerosol optical properties from sun and sky radiance measurements. *J. Geophys. Res.,* **105,** 20 673–20 696.

——, A. Smirnov, B. N. Holben, M. D. King, Y. J. Kaufman, T. E. Eck, and I. Slutsker, 2000: Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) sun and sky radiance measurements. *J. Geophys. Res.,* **105,** 9791–9806.

Durkee, P. A., and Coauthors, 2000: Regional aerosol optical depth characteristics from satellite observations: ACE-1, TARFOX and ACE-2 results. *Tellus,* **52B,** 484–497.

Goloub, P. D., D. Tanre, J. L. Deuze, M. Herman, A. Marchand, and F. M. Breon, 1999: Validation of the first algorithm applied for deriving the aerosol properties over the ocean using POLDER/ADEOS measurements, *IEEE Trans. Geosci. Remote Sens.,* **37,** 1586–1596.

Hansen, J. E., and A. A. Lacis, 1990: Sun and dust versus greenhouse gases: An assessment of their relative roles in global climate change. *Nature,* **346,** 713–719.

——, M. Sato, R. Ruedy, A. Lacis, and V. Oinas, 2000: Global warming in the 21st century: An alternate scenario. *Proc. Natl. Acad. Sci.,* **97,** 9875–9880.

Higurashi, A., and T. Nakjima, 1999: Development of a two-channel aerosol retrieval algorithm on a global scale using NOAA AVHRR. *J. Atmos. Sci.,* **56,** 924–941.

——, ——, B. N. Holben, A. Smirnov, R. Frouin, and B. Chatenet, 2000: A study of global aerosol optical climatology with two-channel AVHRR remote sensing. *J. Climate,* **13,** 2011–2027.

Holben, B. N., and Coauthors, 1998: AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.,* **66,** 1–16.

Husar, B. R., J. M. Prospero, and L. L. Stowe, 1997: Characterization of tropospheric aerosols over the oceans with the NOAA advanced very high resolution radiometer optical thickness operational product. *J. Geophys. Res.,* **102,** 16 889–16 909.

Ignatov, A. M., and L. L. Stowe, 2000: Physical basis, premises, and self-consistency checks of aerosol retrievals from TRMM/VIRS. *J. Appl. Meteor.,* **39,** 2259–2277.

——, and ——, 2002a: Aerosol retrievals from individual AVHRR channels. Part I: Retrieval algorithm and transition from Dave to 6S radiative transfer model. *J. Atmos. Sci.,* **59,** 313–334.

——, and ——, 2002b: Aerosol retrievals from individual AVHRR channels. Part II: Quality control, probability distribution functions, information content, and consistency checks of retrievals. *J. Atmos. Sci.,* **59,** 335–362.

——, ——, and G. Korotaev 1995a: validation of the NOAA/NESDIS satellite aerosol product over the North Atlantic in 1989. *J. Geophys. Res.,* **100,** 5123–5132.

——, ——, R. Singh, D. Kabanov, and I. Dergileva, 1995b: Validation of NOAA/AVHRR aerosol retrievals using sun-photometer measurements from RV Akademik Vernadsky in 1991. *Adv. Space Res.,* **16** (10), 95–98.

——, ——, and ——, 1998: Sensitivity study of the Ångström exponent derived from AVHRR over the oceans. *Adv. Space Res.,* **21,** 439–442.

IPCC, 1996: *Climate Change 1995. The Science of Climate Change.* Cambridge University Press, 572 pp.

Kaufman, Y., J. Tucker, and I. Fung, 1990: Remote sensing of biomass burning in the Tropics. *J. Geophys. Res.,* **95,** 9895–9909.

——, and Coauthors, 1997: Passive remote sensing of tropospheric aerosol and atmospheric correction for the aerosol effect. *J. Geophys. Res.,* **102,** 16 815–16 830.

King, M. D., Y. J. Kaufman, D. Tanre, and T. Nakajima, 1999: Remote sensing of tropospheric aerosols from space: Past, present, and future. *Bull. Amer. Meteor. Soc.,* **80,** 2229–2259.

Lacis, A. A., and M. I. Mishchenko, 1995: Climate forcing, climate sensitivity, and climate response: A radiative modeling perspective on atmospheric aerosols. *Aerosol Forcing of Climate,* R. Charlson and J. Heintzenberg, Eds., Wiley and Sons, 11–42.

Mishchenko, M. I., I. V. Geogdzhayev, B. Cairns, W. B. Rossow, and A. Lacis, 1999: Aerosol retrievals over the oceans by use of channels 1 and 2 AVHRR data: Sensitivity analysis and preliminary results. *Appl. Opt.,* **38,** 7325–7341.

Nakajima, T., and A. Higurashi, 1997: AVHRR remote sensing of aerosol optical properties in the Persian Gulf region, summer 1991. *J. Geophys. Res.,* **102,** 16 935–16 946.

——, G. Tonna, R. Rao, P. Boi, Y. Kaufman, and B. Holben, 1996: Use of sky brightness measurements from ground for remote sensing of particulate polydispersions. *Appl. Opt.,* **35,** 2672–2686.

Ostle, B., and R. W. Mensing, 1988: *Statistics in Research.* Iowa State University Press, 596 pp.

Price, S., 1983: *Atlas of Hawaii.* University of Hawaii Press, 596 pp.

Schmid, B., and Coauthors, 1999: Comparison of aerosol optical depth from solar radiometers during fall 1997 ARM intensive observation period. *Geophys. Res. Lett.*, **26,** 2725–2728.

Smirnov, A., B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker, 2000: Cloud screening and quality control algorithms for the AERONET data base. *Remote Sens. Environ.*, **73,** 337–349.

——, ——, Y. J. Kaufman, O. Dubovik, T. F. Eck, I. Slutsker, C. Pietras, and R. Halthore, 2002: Optical properties of atmospheric aerosol in maritime environments. *J. Atmos. Sci.*, **59,** 501–524.

Stowe, L. L., and H. Jacobowitz, 1997: The AVHRR Pathfinder Atmosphere (PATMOS) Project. Preprints, *Ninth Atmospheric Radiation Symposium*, Long Beach, CA, Amer. Meteor. Soc., 196–197.

——, A. M. Ignatov, and R. R. Singh, 1997: Development, validation, and potential enhancements to the second-generation operational aerosol product at the National Environmental Satellite, Data, and Information Service of the National Oceanic and Atmospheric Administration. *J. Geophys. Res.*, **102,** 16 923–16 934.

Tanre, D., M. Herman, and Y. J. Kaufman, 1996: Information on the aerosol size distribution contained in the solar reflected spectral radiances. *J. Geophys. Res.*, **101,** 19 043–19 060.

——, Y. J. Kaufman, M. Herman, and S. Matoo, 1997: Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *J. Geophys. Res.*, **102,** 16 971–16 988.