ALGORITHM FOR REMOTE SENSING OF TROPOSPHERIC AEROSOL FROM MODIS: Collection 005: Revision 2; Feb 2009

Product ID: MOD04/MYD04

MODIS Science Team:
Lorraine A. Remer[1], Didier Tanré[2] and Yoram J. Kaufman[1]
Additional Points of contact: R. Levy[3] and S. Mattoo[3]

[1] NASA Goddard Space Flight Center, Code 613.2, Greenbelt, MD 20771, USA
[2] Laboratoire d'Optique Atmospherique, Université de Sciences et Techniques de Lille, Villeneuve d'Ascq, France, on frequent visits to NASA/GSFC
[3] Science Systems and Applications, Inc. c/o NASA/GSFC, Greenbelt, MD 20771, USA.

This document updates the 1996 MOD04 Algorithm Theoretical Basis Document (ATBD-96), describing the algorithms for simultaneous remote sensing of aerosol properties over land and ocean from spectral reflectance observed by EOS-MODIS (both Terra and Aqua). The Collection 5 (C005) algorithm retrieves the aerosol optical depth (AOD or τ–proportional to the aerosol total loading in the vertical column) and proxies for the size distribution (such as Fine-Weighting (FW or η)) of the ambient (undisturbed) aerosol, over most of the globe (oceans and the moist parts of the continents) on a daily basis. These aerosol products are primarily intended for radiative budget and climate applications, but are expected to be relevant for hydrological, oceanographic and air quality applications. The combined ocean/land algorithm takes advantage of the MODIS wide spectral range and high spatial resolution with daily global coverage (e.g., 500 m at 0.47 to 2.12 μm with 250 m at 0.66 and 0.86 μm and 1 km at 1.38 μm). These unique MODIS characteristics allow excellent cloud rejection while maintaining high statistics of cloud free pixels. The wide spectral range permits resolution of aerosol size distribution. The land algorithm described here is a complete overhaul from that described in the ATBD-96. The C005 over-land algorithm continues to derive only over sufficiently dark surfaces, but has new surface reflectance parameterizations, aerosol optical models, and assumptions relating to how the 2.12 μm channel relates to surface and surface properties. The C005 ocean algorithm is generally described by the ATBD-96, but we add here a chapter describing the changes to it. Co-located sunphotometer validation is provided for both land and ocean products. The gas (ozone, water vapor and carbon dioxide) absorption corrections to the reflectance are described here. Finally, we include a section describing the structure of the operational product files (i.e. MOD04/MYD04-C005).

TABLE OF CONTENTS

1. Introduction 4
2. Background information 7
   2.1. Characteristics of the MODIS instrument 7
   2.2. Inputs to the aerosol retrieval 8
   2.3. Production environment 8
   2.4. Processing prior to aerosol retrieval 9
   2.5. Quality Assurance, Level 2 'Combined' and Level 3 products 9
   2.6. Strategy for evaluation and validation of MODIS products 10
3. Algorithm description: ocean 12
   3.1. Strategy 12
   3.2. Aerosol models (modes) and lookup tables 14
   3.3. Selection of pixels: cloud, glint and sediment masking 17
   3.4. Retrieval algorithm 18
   3.5. Sensitivity study 19
   3.6. Retrieved ocean products 24
4. Algorithm description: land 26
   4.1. Strategy 26
   4.2. Formulation of the land algorithm 27
   4.3. Aerosol optical models and lookup tables 27
   4.4. VISvs2.12 surface reflectance assumptions 36
   4.5. Retrieval Algorithm 45
   Selection of "dark pixels" 46
   Correcting the LUT for elevation 48
   Procedure A: Inversion for dark surfaces 49
   Procedure B: Alternative Retrieval for Brighter surfaces 50
   Derivation of Fine Mode τ, Mass Concentration and other secondary parameters 50
   Negative (and very low) optical depth retrievals 51
   4.6. Sensitivity Study 51
   4.7. Retrieved Land Products 57
5. Provisional validation 60
   5.1. Visual comparison of V5.2 versus V5.1 products 60
   5.2. Statistics of V5.2 vs V5.1 products 62
   5.3. Comparison of V5.2 and V5.1 over-land products with AERONET 64
   5.4. Comparison of ocean products with AERONET 67
6. File Format, Notes, and Recommendations to users 71
   6.1. File format and content 71
   6.2. Notes on global sampling 72
   6.3. Recommendations for use of data 73
   6.3.1. Total τ 73
   6.3.2. Size parameters 73
   6.3.3. Other derived and diagnostic parameters 74
7. Summary 75
8. Global Evaluation of C005 products 77
9. References 78
A1: Gas Correction of L1B reflectance 82

A2: Masking over land and ocean 84
A2.2 Masking over ocean 84
A2.2 Masking over land 86
A3: Table of Run Time QA Flags of M?D04 Level 2 Aerosol Products 89
A4: Calculation of Mass Concentration 95

1. Introduction

The 1996 version of the MODIS-aerosol Algorithm Theoretical Document (ATBD-96) introduced the rationale for performing aerosol remote sensing from MODIS on a global scale. Aerosols were known to impact the radiative budget, climate change, hydrological processes, and the global carbon, nitrogen and sulfur cycles. To even begin to understand the wide-ranging effects of aerosol, it was considered necessary to explain aerosol characteristics with high spatial and temporal resolution. The polar-orbiting MODerate resolution Imaging Spectrometer (MODIS-Salmonson et al 1989) with its high spatial resolution (up to 250m at nadir), wide swath (2330 km, so that everywhere on the globe is observed at least once daily), and large spectral range (36 channels between 0.412 to 14.2 μm) was expected to be the key for monitoring global aerosol properties.

The use of the MODIS aerosol products has far exceeded nearly everyone's imagination. Since launch of MODIS aboard the Earth Observing System's (EOS) Terra (originally called EOS-AM1) in 1999, and aboard EOS-Aqua in 2002, MODIS data and specifically aerosol data have been used for dozens of applications and referred to in hundreds publications. Figure 1 shows how the MODIS aerosol publications have grown nearly exponentially (as demonstrated by Yoram Kaufman when searching for "MODIS and aerosol" on the ISI citation web site). Not only have MODIS aerosol products been used to answer scientific questions about radiation and climate (e.g. IPCC, 2001; Yu et al., 2005), they are being used for applications not previously intended. Some examples include monitoring surface air quality for health (e.g. Chu et al., 2003, Al-Saadi et al., 2005) and estimating iron nutrients (from dust) deposited into the ocean (Gao et al., 2000).

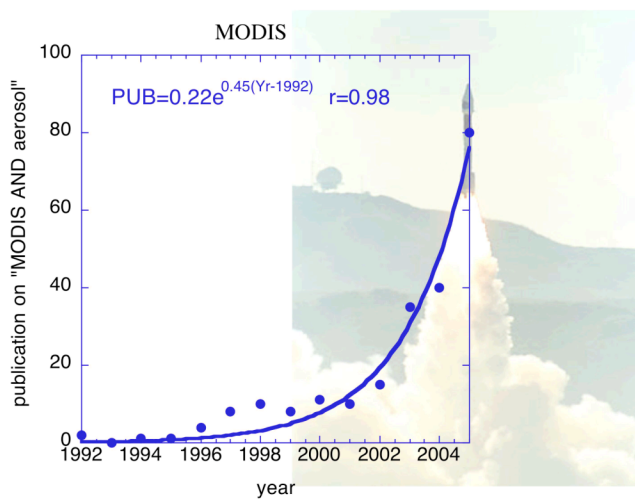

Figure 1: The image shows the exponential rate of publication from the ISI citation web site on "MODIS and aerosol" starting around 1992. (In the background is a photograph of the Terra launch on December 18, 1999). Operational algorithms and data centers that make the data available quickly to the broad international science community, give rise to an exponential publication rate preparing for and using MODIS aerosol data (Image produced by Y. Kaufman (climate.gsfc.nasa.gov).

The fundamental aerosol products from MODIS include total spectral 'aerosol optical depth' (AOD or τ) and 'Fine aerosol Weighting' (FW or η). In the literature, the concept of τ is sometimes known as 'aerosol optical thickness' (AOT), but AOD is preferred for this document. The concept of FW is also variously defined in the literature. In this document, η refers to the fractional contribution of fine (small sized) aerosol to the total τ, and is reported at a particular wavelength (0.55 μm).

The MODIS aerosol algorithm is comprised of two independent algorithms, one for deriving aerosols over land and the second for aerosols over ocean. Both algorithms were conceived and developed before Terra launch and described in depth in Kaufman, et al. (1997b), Tanré, et al. (1997) and ATBD-96. Until now, the theoretical basis of the algorithms has not changed from inception, although some of the mechanics and details of the algorithms have evolved. MODIS data is organized by collections. A collection consists of products that were generated by similar, but not necessarily the same, versions of the algorithm. ATBD-96 describes the pre-launch algorithm. Collection 003 (C003) provided the first globally 'validated' products over ocean (Remer et al., 2002) and over land (Chu et al., 2002) by comparing MODIS with ground based sunphotometer data (Ichoku et al., 2002) of AERONET (Holben et al., 1998). C003 data also was used for regional validation exercises (e.g. Levy et al., 2003; Ichoku et al., 2002). For example the Ichoku et al., (2002) study proved that single scattering albedo (SSA or ω_0) was in fact lower than previously assumed for southern African biomass burning. It was studies like these that led to updates to the algorithm and re-processing older data. The algorithms and products of Collection 004 (C004) were described by Remer et al., (2005), and provided the benchmark for development of the C005 algorithms. A complete history of changes to the operational algorithm over the course of the MODIS mission can be found at http://modis-atmos.gsfc.nasa.gov/MOD04_L2/history.

C004 was a complete dataset from both Terra and Aqua, spanning from Terra first light in February 2000 through the first few months of 2006. Remer et al. (2005) provided global validation over both land and ocean for 2000 through 2003, finding that MODIS was comparable to ground truth (e.g. AERONET) within certain expected uncertainties. The expected uncertainty for τ was found to be:

$$\Delta\tau = \pm 0.03 \pm 0.05\tau \qquad \textbf{(1)}$$

over ocean and

$$\Delta\tau = \pm 0.05 \pm 0.15\tau \qquad \textbf{(2)}$$

over land. The expected errors were designated for the 0.55 μm MODIS channel, but were found to apply in other channels as well. However, while in general the MODIS retrievals met expected accuracy, under certain conditions they did not. Non-spherical dust over the ocean was known to lead to errors in retrieving spectral τ (e.g. Levy et al., 2003). Naïvely assuming constant surface reflectance assumptions was found to overestimate τ in clean conditions along the U.S. East Coast (Levy et al., 2005) and errors (both positive and negative) in other regions (e.g. Remer et al., 2005). The available choices of aerosol optical models over land led to either positive or negative bias for large τ in many regions (such as negative bias over the U.S. East Coast – Levy et al., 2005).

The aerosol fine weighting (η) product over land and ocean was also been separately compared with sunphotometer data, on a multiple of scales (e.g. Kleidman et al., 2005; Chu et al, 2005; Anderson et al., 2005). The expected uncertainty for η as not defined, but it seemed to have valid physical meaning (i.e. correlation with other measurements of fine aerosol weighting).

After the validation efforts of C004 products, the combined algorithm went through a number of revisions. The latest major update included the new snow mask of Li et al., (2005), and a general 'clean up' of confusing output products. This latest minor update was known internally as 'V5.1', but never became operational. Instead, our understanding of aerosol properties and MODIS processing had progressed to the point where we considered to derive new philosophy of deriving aerosol over land.

The result of that work led to the formulation of Version 5.2 ('V5.2'), which included a complete overhaul of the aerosol retrieval over land. Over ocean, V5.2 was revised as well, to include new assumptions about coarse aerosol properties. The combined V5.2 algorithm is being used to create the products of C005. In this document, we detail the combined C005 algorithm (V5.2) and show how it with the last available C004 algorithm (V5.1). Even though V5.1 never became operational, we use it in this document (Section 5) to provide a fair comparison (i.e. same boundary conditions and same radiance data) with V5.2. At the time of this document revision (Feb 2009), MODIS-Terra is being processed by V5.2.5, which includes some bug fixes and other minor updates. MODIS-Aqua is running with V5.3.7, which includes new products from the Deep Blue retrieval (Hsu et al., 2006). The Deep Blue algorithm is not detailed in this ATBD. As evaluation and validation of operational C005 products is performed, the results will be appended as Section 7 of this document.

2. Background information

2.1. Characteristics of the MODIS instrument

The MODerate resolution Imaging Spectrometer (MODIS) instrument flies on the Earth Observation System's (EOS) Terra and Aqua satellites. Both satellites are polar-orbiting, with Terra on a descending orbit (southward) over the equator about 10:30 local sun time, and Aqua on an ascending orbit (northward) over the equator about 13:30 local sun time. From a vantage about 700 km above the surface and a ±55° view scan, each MODIS views the earth with a swath about 2330 km, thereby observing nearly the entire globe on a daily basis, and repeat orbits every 16 days. Each scan is 10 km along track. MODIS performs measurements in the solar to thermal infrared spectrum region from 0.41 to 14.235 μm (Salomonson et al., 1989). Detailed specifications and components can be found at http://modis.gsfc.nasa.gov. The aerosol retrieval makes use of seven wavelength bands (listed in Table 1), and a number of other bands to help with cloud and other screening procedures. Included in Table 1 are estimates of the central wavelength in each band (obtained by integration of the channel-averaged response functions). To keep in line with common references in the aerosol literature, MODIS channels 1, 2, 3, 4, 5, 6 and 7 are known in this document as the 0.66, 0.86, 0.47, 0.55, 1.24, 1.64 and 2.12 μm channels, respectively.

TABLE 1: CHARACTERISTICS OF MODIS CHANNELS USED IN THE AEROSOL RETRIEVAL

Band #	Bandwidth (μm)	Weighted Central Wavelength (μm)	Resolution (m)	NeΔρ (x10^{-4})	Max ρ	Required SNR	Rayleigh optical depth
1	0.620 - 0.670	0.646	250	3.39	1.38	128	0.0520
2	0.841 - 0.876	0.855	250	3.99	0.92	201	0.0165
3	0.459 - 0.479	0.466	500	2.35	0.96	243	0.1948
4	0.545 - 0.565	0.553	500	2.11	0.86	228	0.0963
5	1.230 – 1.250	1.243	500	3.12	0.47	74	0.0037
6	1.628 – 1.652	1.632	500	3.63	0.94	275	0.0012
7	2.105 – 2.155	2.119	500	3.06	0.75	110	0.0004

Notes: Band #26 (1.38 μm channel) is used for cirrus correction; NeΔρ corresponds to the sun at zenith (θ = 0°)

The MODIS algorithm uses of the spectral 'reflectance', ρ_λ, defined as a function of the measured spectral radiance, L_λ, the solar zenith angle (θ_0), and the solar irradiance F_0 in the wavelength band λ:

$$\rho_\lambda = L_\lambda \frac{\pi}{F_{0,\lambda}\cos(\theta_0)}. \qquad (3)$$

To be useful for aerosol retrieval, the MODIS instrument must be spectrally stable and sufficiently sensitive. The spectral stability for each instrument is better than 2 nm (0.002 μm). The 'Noise Equivalent Differential Spectral Radiance' (NeΔL) is a property of the instrument. 'Signal to Noise Ratio' (SNR) is defined as the ratio of the 'typical scene radiance' (L^{ts}) and the NeΔL. The NeΔL and the SNR specifications are given in Table 1. To be understood in the framework of aerosol remote sensing, the definition of SNR should be based on the expected aerosol signal. Therefore, a 'Noise Equivalent Differential optical depth (NeΔτ) is defined, where:

$$Ne\Delta\tau = \pi Ne\Delta\rho \frac{4\cos(\theta_0)\cos(\theta_v)}{\omega_0 P(\Theta)} \qquad (4)$$

where θ_0 and θ_v are the solar and view zenith angles, ω_0 is the aerosol single scattering albedo (ω_0), $P(\Theta)$ is the aerosol phase function as a function of scattering angle, and $Ne\Delta\rho$ ('Noise Equivalent Differential Spectral Reflectance') is related to $Ne\Delta L$ analogous to equation (3). The least sensitivity to aerosol scattering optical depth (largest noise) is expected when both sun and satellite are at nadir views ($\theta_0 = \theta_v = 0.0$), the phase function is a minimum ($\Theta \sim 120°$) and the channel used is the least sensitive (channel 7, at 2.12 μm). With a typical phase function value of 0.08 at 120°, a typical aerosol has $Ne\Delta\tau \sim 1.5x10^{-2}$. The 2.12 μm channel is also where the 'typical scene τ' is (τ^{ts}) is 0.01 or less. Therefore the SNR ratio defined by the ratio of $\tau^{ts}/Ne\Delta\tau$ is about 0.66. This means that single 500 m pixels are insufficiently sensitive to characterize aerosol.

However, if individual pixels are aggregated to larger areas, say to a grid of 10x10 km^2 (20 x 20 500 m pixels), then the noise is reduced by a factor of 20. Instead of 0.66, the SNR becomes 13. Since SNR>10, we decide to use 10x10 km^2 boxes as the default retrieval size.

2.2. Inputs to the aerosol retrieval

The MODIS orbit is separated into 5-minute chunks called 'granules'. Each granule is about 2030 km (about 203 scans of 10 km) along the orbital path. Each scan line has a swath about 2330 km, and at nominal (nadir) 1 km resolution, is covered by 1354 pixels. Note, that due to spherical geometry, the size of each pixel increases from 1km at nadir to nearly 2km at the swath edges. Each granule is 1354 by 2030 pixels in this '1 km' resolution. Only data from MODIS daytime orbits are considered for retrieval.

Both the land and ocean aerosol algorithms rely on calibrated, geolocated reflectances (known as 'Level 1B' or 'L1B') provided by the MODIS Characterization Support Team (MCST). These are identified as products MOD02 and MOD03 for Terra and MYD02 and MYD03 for Aqua (MCST 2000; MCST 2002). Hereby, either 'MOD' or 'MYD' will be denoted by 'M?D'. The algorithm actually uses L1B reflectances at three resolutions (M?D021KM, M?D02HKM and M?D02QKM for 1km, 500m and 250m resolution channels, respectively). Ignatov et al. (2005) provides a good discussion of these reflectances and possible errors associated with them. In addition, the MODIS algorithm uses two processed products known as 'Level 2' (L2). These include the M?D35 Wisconsin cloud mask product (Ackerman et al. 1998) and the M?D07 atmospheric profile product. Finally, the algorithm expects ancillary data from NCEP (National Center for Environmental Prediction) analyses, including the (closest to granule time) GDAS1 1°x1° 6 hourly meteorological analysis and the TOVS (before 2005) or the TOAST (after 2005) 1°x1° daily ozone analysis. Although the algorithm inputs both M?D07 and the NCEP data, it can run successfully without these supplements by using climatology for first guess water vapor and ozone profiles.

2.3. Production environment

All MODIS L1B and L2 atmospheric products are written in 'Hierarchical Data Format' (HDF - http://hdf.ncsa.uiuc.edu), with each parameter stored as a 'Scientific Data Set' (SDS). In order to be shared across multiple computing environments, HDF files must be accessed through HDF library subroutine and function calls. In addition, operational processing employs the 'Science Data

Processing ToolKit' (SDPTK) and the 'MODIS Applications Programming Interface' (M-API), which employ the HDF libraries. The collective system environment is known as the MODIS-Toolkit.

The algorithm to create M?D04 data is written primarily in Fortran 77, and includes subroutines and functions that interface with the MODIS-Toolkit. Finally, there are additional static files, including the aerosol lookup tables, required for aerosol retrieval. More information about the computer nitty-gritty is found at http://modis-atmos.gsfc.nasa.gov/MOD04_L2/production.html. All operational processing is performed by the MODIS Data Processing System (MODAPS) located at NASA-GSFC. All data are archived and freely available from MODAPS web sites.

2.4. Processing prior to aerosol retrieval

The aerosol algorithm reads in the required L1B, L2 and ancillary data into memory. Specifically, it reads one scan line at a time, where each scan is made up of ten 1km pixels along track. The 1354 swath pixels are also collected into ten pixel boxes, so that there are 135 '10km' boxes in a swath. Each of these boxes is separately considered for aerosol retrieval. Note that each 10 km box contains 10 x 10 = 100 '1km' pixels and 20 x 20 = 400 '500m' pixels. Again note that these sizes refer to the nadir view. At the scan edges the *number* of pixels in each box remains the same, but the *area* encompassed in each box will be double the area encompassed at nadir.

Reflectances in all seven MODIS-aerosol channels, plus the 1.38 μm channel are corrected for water vapor, ozone and carbon dioxide (Appendix 1). In addition to the cloud mask, the M?D35 product also identifies whether a pixel is a 'land' pixel or a 'water' pixel. If all pixels in the 10 km box are considered water, the algorithm proceeds with the over-ocean retrieval. However, if any pixel is considered land, then it proceeds with the over-land algorithm. This helps to minimize problems introduced by underwater reflectance in shallow water near the coasts.

2.5. Quality Assurance, Level 2 'Combined' and Level 3 products

During aerosol retrieval on a particular 10 km box, the algorithm may proceed normally, proceed with non-fatal errors, or quit because of a fatal error. The 'quality assurance' (QA) of the retrieved products is assigned based on the behavior of the algorithm. Individual QA flags are assigned particular values when any errors (fatal or non-fatal) are encountered. When stored within the M?D04 HDF aerosol product files, the QA flags are composed of data 'bits' that can be decoded to determine these errors. For example, one QA data flag warns the user to any 'water' pixels within the box, even when the land retrieval is still performed.

Whether ocean or land aerosol retrieval was performed, the products are assigned a QA 'confidence' flag (QAC) that represents the aggregate of all the individual QA flags. This QAC flag indicates to a user how the particular retrieval should be considered. It also is used to derive downstream products, either by its use as a filter for expected quantitative value of the retrieval, or to provide weighting for aggregating/averaging computations. This QAC value ranges from 3 to 0, where 3 means 'good' quality and 0 means 'bad' quality. Presumably, there is more quantitative value to a 'good' quality

retrieval, rather than one that is 'fair', 'poor' or 'bad'. Appendix 2 describes the individual QA flags and how they are used to assign QAC.

The QAC flag is used to decide which land or ocean τ values go into a combined land and ocean product, known here as the *joint* L2 product. There are in fact two joint τ products. The unconstrained product, known as the 'Image_Optical_Depth_Land_And_Ocean', has no QAC threshold and is intended for evaluating plume locations, and creating attractive imagery with fewer holes. The other, known as the 'Optical_Depth_Land_And_Ocean', is constrained by QAC, in order to filter out lower confidence data with presumably less quantitative value. In the start of C005 processing, the constrained joint SDS required QAC≥1 over land (QAC=1,2 or 3), and QAC ≥ 0 (all QAC) over ocean. However, based on later evaluation of the C005 data stream, the QAC filter should be tightened, to require QAC=3 (only) over land and QAC≥1 over ocean. There is also a single joint η product, known as 'Optical_Depth_Ratio_Small_Land_And_Ocean', that combines land and ocean with no QAC requirement.

In addition to providing a filter on the joint L2 aerosol products, the QAC flag helps to derive average values on regular grids. All MODIS-atmosphere products, including the M?D04 product are averaged globally, on a 1° x 1° degree grid, on daily, weekly, and monthly time scales. These gridded products are known as the Level 3 (L3) products. Two averages are produced; the simple average of all L2 within the box, and the *confidence* weighted average, which utilizes the QAC. The QAC flag is used for weighting the 10 km products onto the 1° grid. Those retrievals with QAC = 3 are assigned higher weights than those with QAC = 2 or QAC = 1. Retrievals with QAC=0 are not included in the 1° averages.

2.6. Strategy for evaluation and validation of MODIS products

The first step in determining the quality of any product is to ensure that the products look as expected. Do images of a parameter have discontinuities, gaps, or any other features that indicate logical errors or other problems? Are diagnostic parameters useful and correct? Are the mean values and the histograms of a retrieved or derived parameter reasonable, at different temporal and spatial scales? Many of these tests are subjective, but they are important for solving problems with the algorithm or the assumptions.

For some products we hope to perform *validation*. This means that a product is comparable to some "ground-truth" measurement within some measure of *expected uncertainty*. The primary means of MODIS validation is by showing that a product matches equivalent observations from AERONET or other accurate systems. At least 2/3 (or one standard deviation) of global matched pairs (e.g. MODIS versus AERONET) must be contained within the envelope of expected uncertainty to be considered validated. Using this strategy, Remer et al., (2005), validated that the C004 τ products (i.e. V4.2 and before) were matched within the envelopes of Eqs. (1) and (2) over ocean and land, respectively.

Prior to operational C005 production, the processing code was tested on set of 6300 MODIS granules, known as the testbed. This testbed included granules from many days, seasons, locations and both satellites, so they presumably would be a sufficient sample of a global dataset. However, since that our tests were performed on L1B data processed prior to C005 production (e.g. C004 radiance calibration data), the results should be considered as *provisional*. The provisional validation is described in

Section 5. The operational C005 data will be evaluated and validated as a continual work in progress. Its results will be appended as Section 7.

3. Algorithm description: ocean

3.1. Strategy

The theory and strategy of the C005 aerosol retrieval over the ocean (C005-O) is aptly described in the ATBD-96 (also in Tanré et al., 1997, Levy et al., 2003; Remer et al., 2005) and remained the same for C005. The mechanics of the ocean algorithm are illustrated in Figure 2. The algorithm is based on a 'look-up table' (LUT) approach, i.e., radiative transfer calculations are pre-computed for a set of aerosol and surface parameters and compared with the observed radiation field. The algorithm assumes that one fine and one coarse lognormal aerosol modes can be combined with proper weightings to represent the ambient aerosol properties over the target. Spectral reflectance from the LUT is compared with MODIS-measured spectral reflectance to find the 'best' (least-squares) fit. This best fit, or an 'average' of a set of the best fits is the solution to the inversion. Although the core inversion remains similar to the process described in Tanré, et al. (1997), the masking of clouds and sediments, the special handling of heavy dust including dust retrievals over glint, and revisions of the look-up table are new. The changes from previous over-ocean algorithms are described below.

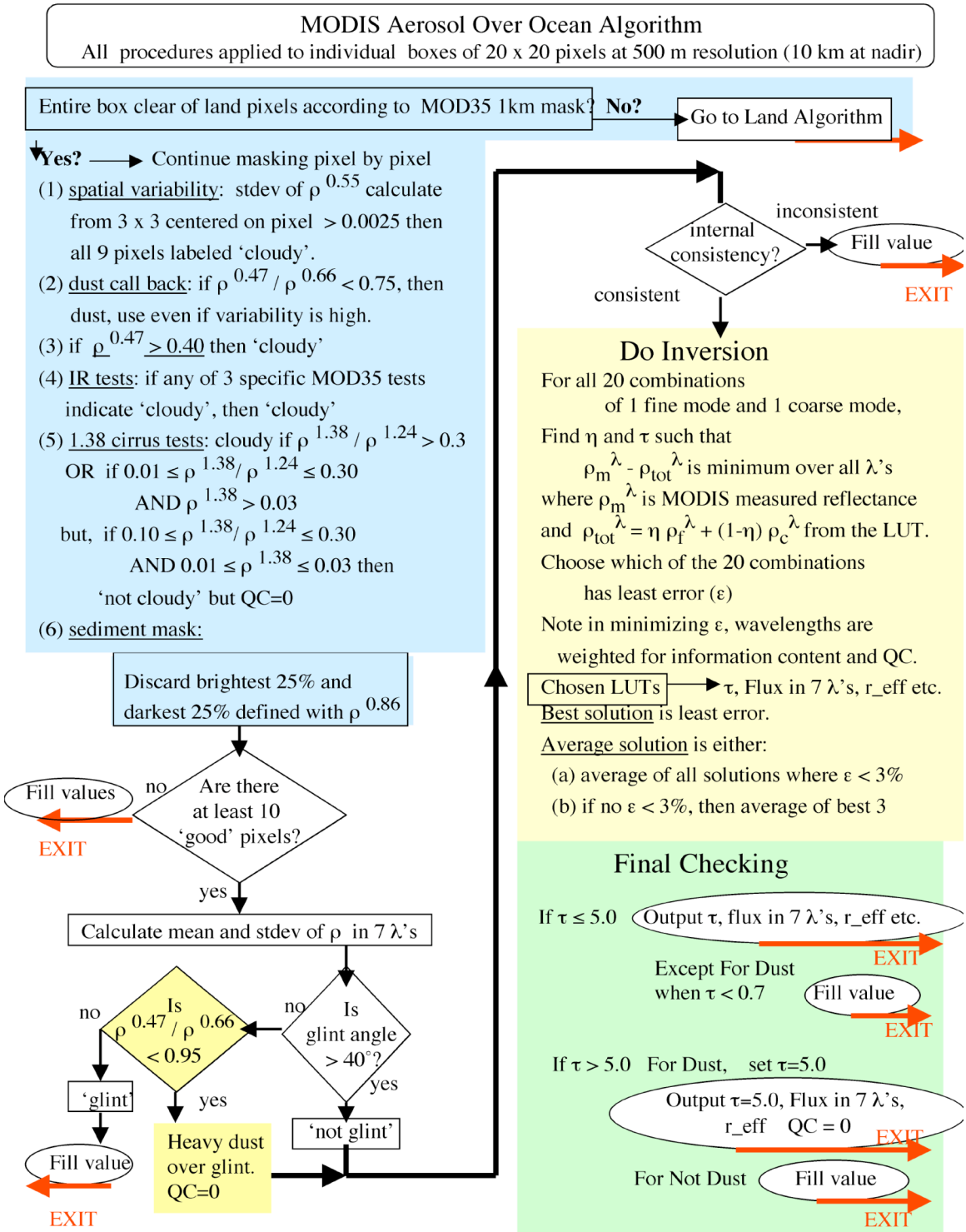

Figure 2: Flowchart of the over-ocean aerosol retrieval algorithm (from Remer et al., 2005).

3.2. Aerosol models (modes) and lookup tables

The climatology used to create the C005 aerosols models was derived mainly from data gleaned from sun/sky photometers (e.g. AERONET; Holben et al., 1998), and from analysis of errors in the products from previous versions of the MODIS algorithm. In the C005-over ocean algorithm (C005-O), there are four fine modes and five coarse modes described in Table 2 (A-D). Coarse modes 7, 8 and 9 have refractive indices different from the modes in the C004-over ocean (C004-O) algorithm (Remer et al., 2005). The change was made based on AERONET almucantur retrievals (e.g. Dubovik et al., 2000) of marine aerosol optical properties. Figure 3 shows how η is strongly reduced to more realistic values in dusty situations, but in smoke or pollution dominated situations the values are still high. Retrievals of τ (AOT) in both situations are nearly unchanged.

TABLE 2A: REFRACTIVE INDICES, NUMBER MEDIAN, STANDARD DEVIATION AND EFFECTIVE RADIUS FOR THE AEROSOL MODES USED IN THE MODIS LOOKUP TABLE FOR THE OCEAN ALGORITHM. MODELS 1-4 ARE FINE MODES AND MODELS 5-9 ARE COARSE MODES.

F	λ=0.47-->0.86µm	λ =1.24µm	λ =1.64µm	λ =2.12µm	r_g	σ	r_{eff}	Comments
1	1.45-0.0035i	1.45-0.0035i	1.43-0.01i	1.40-0.005i	0.07	0.40	0.10	Water Soluble
2	1.45-0.0035i	1.45-0.0035i	1.43-0.01i	1.40-0.005i	0.06	0.60	0.15	Water Soluble
3	1.40-0.0020i	1.40-0.0020i	1.39-0.005i	1.36-0.003i	0.08	0.60	0.20	Water Soluble with humidity
4	1.40-0.0020i	1.40-0.0020i	1.39-0.005i	1.36-0.003i	0.10	0.60	0.25	Water Soluble with humidity

C	λ =0.47-->0.86µm	λ =1.24µm	λ =1.64µm	λ =2.12µm	r_g	σ	r_{eff}	Comments
5	1.35-0.001i	1.35-0.001i	1.35-0.001i	1.35-0.001i	0.40	0.60	0.98	Wet sea salt type
6	1.35-0.001i	1.35-0.001i	1.35-0.001i	1.35-0.001i	0.60	0.60	1.48	Wet sea salt type
7	1.35-0.001i	1.35-0.001i	1.35-0.001i	1.35-0.001i	0.80	0.60	1.98	Wet sea salt type
8	1.53-0.003i (0.47) 1.53-0.001i (0.55) 1.53-0.000i (0.66) 1.53-0.000i (0.86)	1.46-0.000i	1.46-0.001i	1.46-0.000i	0.60	0.60	1.48	Dust-like type
9	1.53-0.003i (0.47) 1.53-0.001i (0.55) 1.53-0.000i (0.66) 1.53-0.000i (0.86)	1.46-0.000i	1.46-0.001i	1.46-0.000i	0.50	0.80	2.50	Dust-like type

TABLE 2B: SPECTRAL EXTINCTION COEFFICIENTS FOR THE AEROSOL MODES USED IN THE MODIS LOOKUP TABLE FOR THE OCEAN ALGORITHM. MODELS 1-4 ARE FINE MODES AND MODELS 5-9 ARE COARSE MODES.

λ (μm) / Mode	0.466	0.553	0.645	0.855	1.24	1.64	2.12
1	1.43E-10	9.33E-11	6.15E-11	2.66E-11	7.91E-12	4.30E-12	1.48E-12
2	3.03E-10	2.33E-10	1.78E-10	9.95E-11	3.93E-11	1.88E-11	6.99E-12
3	6.78E-10	5.45E-10	4.34E-10	2.63E-10	1.15E-10	5.67E-11	2.28E-11
4	1.33E-09	1.12E-09	9.36E-10	6.15E-10	3.01E-10	1.57E-10	6.68E-11
5	2.69E-08	2.78E-08	2.84E-08	2.85E-08	2.55E-08	2.12E-08	1.63E-08
6	5.57E-08	5.76E-08	5.95E-08	6.29E-08	6.44E-08	6.09E-08	5.33E-08
7	9.50E-08	9.72E-08	9.97E-08	1.06E-07	1.13E-07	1.15E-07	1.09E-07
8	5.57E-08	9.72E-08	5.70E-08	6.05E-08	6.60E-08	6.63E-08	6.26E-08
9	6.42E-08	6.54E-08	6.66E-08	6.92E-08	7.31E-08	7.43E-08	7.36E-08

TABLE 2C: SPECTRAL SINGLE SCATTERING ALBEDOS FOR THE AEROSOL MODES USED IN THE MODIS LOOKUP TABLE FOR THE OCEAN ALGORITHM. MODELS 1-4 ARE FINE MODES AND MODELS 5-9 ARE COARSE MODES.

λ (μm) / Mode	0.466	0.553	0.645	0.855	1.24	1.64	2.12
1	0.9735	0.9683	0.9616	0.9406	0.8786	0.539	0.4968
2	0.9782	0.9772	0.9757	0.9704	0.9554	0.8158	0.8209
3	0.9865	0.9864	0.9859	0.9838	0.9775	0.9211	0.9156
4	0.9861	0.9865	0.9865	0.9855	0.9819	0.9401	0.9404
5	0.9781	0.982	0.9847	0.9886	0.9914	0.9923	0.9925
6	0.9661	0.9716	0.976	0.9825	0.9882	0.9906	0.9919
7	0.955	0.9619	0.9673	0.9759	0.9842	0.988	0.9904
8	0.9013	0.9674	1	1	1	0.9901	1
9	0.8669	0.953	1	1	1	0.9835	1

TABLE 2D: SPECTRAL ASYMMETRY PARAMETERS FOR THE AEROSOL MODES USED IN THE MODIS LOOKUP TABLE FOR THE OCEAN ALGORITHM. MODELS 1-4 ARE FINE MODES AND MODELS 5-9 ARE COARSE MODES.

λ (μm) / Mode	0.466	0.553	0.645	0.855	1.24	1.64	2.12
1	0.5755	0.5117	0.4478	0.3221	0.1773	0.1048	0.0622
2	0.6832	0.6606	0.6357	0.5756	0.4677	0.3685	0.2635
3	0.7354	0.7183	0.6991	0.651	0.559	0.4715	0.3711
4	0.7513	0.7398	0.726	0.6903	0.6179	0.5451	0.4566
5	0.7852	0.7865	0.7891	0.7945	0.7951	0.7865	0.769
6	0.7947	0.7885	0.7857	0.7868	0.794	0.7963	0.7922
7	0.8102	0.8005	0.7931	0.7858	0.7884	0.7937	0.7963
8	0.7534	0.72	0.6979	0.6795	0.7129	0.72	0.719
9	0.7801	0.7462	0.7352	0.7065	0.722	0.7222	0.7151

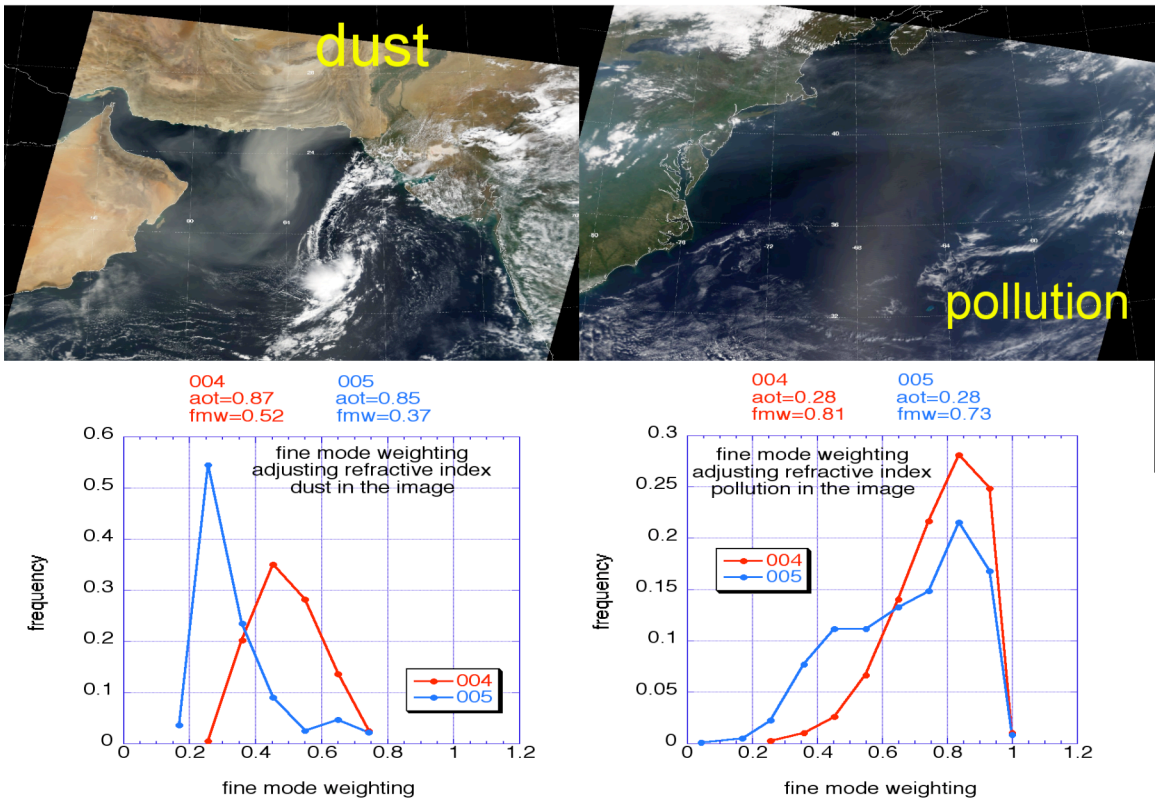

Figure 3: In dusty situations the Collection 004 (red) data reported a too high proportion of fine mode aerosols with average Fine Weighting (η or η) of ~0.5. In smoke or pollution aerosols it retrieved roughly the correct fmw of ~0.8. Using the new refractive indices in the three coarse modes, the Collection 005 (blue) retrievals report less fine mode particles in the dust (η reduces to ~0.3), but there is still a high frequency of predominately fine mode particles retrieved in the pollution episode. Note that the mean τ (AOT) remains the same in the two collections.

Like all previous versions, the C005-O LUT was created with the radiative transfer code developed by Ahmad and Fraser (1981). The spectral reflectance at the satellite level is assumed to be from a combination of radiation from the surface and the atmosphere. The ocean surface calculation includes sun glint reflection off the surface waves (Cox and Munk, 1954), reflection by whitecaps (Koepke, 1984) and Lambertian reflectance from underwater scattering (sediments, chlorophyll, etc). The surface wind speed is assumed at 6.0 m/s. Zero water leaving radiance is assumed for all compared wavelengths, except for at 0.55 µm, where a fixed reflectance of 0.005 is used. The atmospheric contribution includes multiple scattering by gas and aerosol, as well as reflection of the atmosphere by the sea surface. Thus, spectral reflectance was computed for each of the nine aerosol models described in the Table 2. Six values of τ, τ^a, (normalized to 0.55 µm) were considered for each mode, ranging from a pure molecular (Rayleigh) atmosphere ($\tau^a = 0.0$) to a highly turbid atmosphere ($\tau^a = 3.0$), with intermediate values of 0.2, 0.5, 1.0 and 2.0. For each model and aerosol optical depth at 0.55 µm, the associated aerosol optical depths were stored for the other six wavelengths, including the blue (at 0.47 µm). Computations are performed for combinations of 9 solar zenith angles (6°, 12°, 24°, 36°, 48°, 54°, 60°, 66° and 72°), 16 satellite view zenith angles (0° to 72°, increments of 6°) and 16 relative sun/satellite azimuth angles (0° to 180°, increments of 12°) for a total of 2304 angular combinations.

3.3. Selection of pixels: cloud, glint and sediment masking

The masking of clouds and sediments and the selection of pixels are described in Remer et al., (2005). Prior to ocean retrieval, the reflectance is organized into the nominal (at nadir) 10km boxes of 20 by 20 pixels (at nominal 500m resolution) and corrected for gas (water, carbon dioxide and ozone) absorption (see Appendix 1). This requires degrading the resolution of the 250 m channels ($\rho_{0.66}$ and $\rho_{0.86}$). Note that if land is encountered in any of the 400 pixels, the entire box is left for the land algorithm.

The algorithm has the arduous task of separating 'good' pixels from 'cloudy' pixels. The standard M?D35 cloud mask includes using the brightness in the visible channels to identify clouds. This procedure will mistake heavy aerosol as 'cloudy', and miss retrieving important aerosol events over ocean. On the other hand, relying on IR-tests alone permits low altitude, warm clouds to escape and be misidentified as 'clear', introducing cloud contamination in the aerosol products. Thus, our cloud mask over ocean combines spatial variability tests (e.g. Martins et al., 2002) along with tests of brightness in visible and infrared channels. Appendix 1 describes the cloud mask over ocean. Underwater sediments have proved to be a problem in shallow water (near coastlines) as the sediments can easily have land-like surface properties. Thus, the sediment mask of Li et al. (2003) is used in addition to the cloud mask.

The algorithm sorts the remaining pixels that have evaded all the cloud masks and the sediment mask according to their $\rho_{0.86}$ value, discards the darkest and brightest 25%, and thereby leaves the middle 50% of the data. The filter is used to eliminate residual cloud contamination, cloud shadows, or other unusual extreme conditions in the box. Because the ocean cloud mask and the ocean surface are expected to be less problematic than their counterparts over land, the filter is less restrictive than the one used in the land retrieval. Of the 400 pixels in the original box, at least 10 must remain from the masking and filtering. Otherwise, no retrieval is attempted and fill values are given for the aerosol products in that10 km box. If there are at least 10 good pixels in the 0.86µm channel and at least 30 good pixels in the remaining 5 channels, the mean reflectance and standard deviation are calculated for the remaining 'good' pixels at the six pertinent wavelengths.

Ocean Glint and Internal Consistency:

The ocean algorithm was designed to retrieve only over dark ocean, (i.e. away from glint). There is a special case when we retrieve over glint, and that is described below. The algorithm calculates the glint angle, which denotes the angle of reflection, compared with the specular reflection angle. The glint angle is defined as

$$\Theta_{glint}=\cos^{-1}((\cos\theta_s\cos\theta_v)+(\sin\theta_s\sin\theta_v\cos\phi)) \qquad (5)$$

where θ_s, θ_v, and ϕ are the solar zenith, the satellite zenith and the relative azimuth angles (between the sun and satellite), respectively. Note that Fresnel reflection corresponds to $\Theta_{glint} = 0$. If $\Theta_{glint} > 40^\circ$, we can avoid glint contamination and proceed with the retrieval. The algorithm performs several consistency checks of the spectral reflectances. Depending on the outcome of these consistency checks, the algorithm may either declare the reflectances to be beyond the range necessary for a successful inversion and exit the procedure, or continue onto the inversion after assigning quality flags (Quality Assurance – QA) to each wavelength.

3.4. Retrieval algorithm

The inversion procedure described in ATBD-96, Levy et al., (2003) and Remer et al., (2005) remains the same for the C005-O algorithm. Following Tanré et al. (1996), we know that the MODIS-measured spectral radiance (0.55 – 2.13 μm) contains almost three pieces of independent information about the aerosol loading and size properties. With some assumptions, the algorithm can derive three parameters: the τ at one wavelength ($\tau_{0.55}^{tot}$), the 'reflectance weighting parameter' (the over-ocean definition of Fine Weighting - η) at one wavelength ($\eta_{0.55}$) and the 'effective radius' (r_e), which is the ratio of the 3rd and 2nd moments of the aerosol size distribution. The effective radius is represented by choosing a single 'fine' (f) and single 'coarse' (c) aerosol mode for combining with the η parameter. The inversion is based on the look-up table (LUT) of four fine modes and five coarse modes (Table 2), Remember that although the LUT is defined in terms of a single wavelength of optical thickness, the parameters of each of the single mode models define a unique spectral dependence for that model, which is applied to the retrieved value of $\tau_{0.55}^{tot}$ to determine optical thickness at other wavelengths.

The retrieval requires a single fine mode and a single coarse mode. The trick, however, is to determine which of the (4 x 5 =) twenty combinations of fine and coarse modes and their relative optical contributions that best mimics the MODIS-observed spectral reflectance. The reflectance from each mode is combined using η as the weighting parameter,

$$\rho_\lambda^{LUT}(\tau_{0.55}^{tot}) = \eta\rho_\lambda^{f}(\tau_{0.55}^{tot}) + (1-\eta)\rho_\lambda^{c}(\tau_{0.55}^{tot}) \qquad (6)$$

where $\rho_\lambda^{LUT}(\tau_{0.55}^{tot})$ is a weighted average reflectance of an atmosphere with a pure fine mode 'f' and optical thickness $\tau_{0.55}^{tot}$ and the reflectance of an atmosphere with a pure coarse mode 'c' also with the same $\tau_{0.55}^{tot}$. For each of the twenty combinations of one fine mode and one coarse mode, the inversion finds the pair of $\tau_{0.55}^{tot}$ and $\eta_{0.55}$ that minimizes the 'fitting error' (ε) defined as

$$\varepsilon = \sqrt{\frac{\sum_{\lambda=1}^{6} N_\lambda \left(\frac{\rho_\lambda^{m} - \rho_\lambda^{LUT}}{\rho_\lambda^{m} - \rho_\lambda^{ray} + 0.01} \right)^2}{\sum_{\lambda=1}^{6} N_\lambda}} \qquad (7)$$

where N_λ is the sum of good pixels at wavelength λ, ρ_λ^{m} is the measured MODIS reflectance at wavelength λ, ρ_λ^{ray} is the reflectance contributed by Rayleigh scattering, and ρ_λ^{LUT} is calculated from the combination of modes in the look-up table and defined by Equation (6). The 0.01 is to prevent a division by zero for the longer wavelengths under clean conditions (Tanré et al. 1997). Note the inclusion of the Rayleigh reflectance scattering in Equation (7), as compared with the formula in C004-O (Remer et al., 2005). The inversion requires $\rho_{0.87}^{LUT}$ to exactly fit the MODIS observations at that wavelength and then finds the best fits to the other five wavelengths via Equation (7). The 0.87 μm channel was chosen to be the primary wavelength because it is expected to be less affected by variability in water leaving radiances than the shorter wavelengths, yet still exhibit a strong aerosol signal, even for aerosols dominated by the fine mode. By emphasizing accuracy in this channel variability in chlorophyll will have negligible effect on the optical thickness retrieval and minimal effect on $\eta_{\lambda=0.55}$.

The twenty solutions are then sorted according to values of ε. The 'best' solution is the combination of modes with accompanying $\tau_{0.55}^{tot}$ and $\eta_{0.55}$ that minimizes ε. The solution may not be unique. The 'average' solution is the average of all solutions with ε< 3% or if no solution has ε < 3%, then the average of the 3 best solutions. Once the solutions are found, then the chosen combination of modes is the de facto derived aerosol model and a variety of parameters can be inferred from the chosen size distribution including spectral optical depth, effective radius, spectral flux, mass concentration, etc.

Final Checking.

Before the final results are output, additional consistency checks are employed. In general, if the retrieved τ at 0.55 μm is greater than –0.01 and less than 5, then the results are output. Negative optical depths are given lower quality (QA) values. There are exceptions and further checking for heavy dust retrievals made over the glint. The final QA-confidence flag may be adjusted during this final checking phase.

Special case: Heavy dust over glint.

If $\Theta_{glint} \leq 40^{\circ}$ then we check for heavy dust in the glint. We use a similar technique as before during the masking operations when we noted that heavy dust has a distinctive spectral signature because of light absorption at blue wavelengths. In the situation of identifying heavy dust over glint we designate all values of $\rho_{0.47}/\rho_{0.66} < 0.95$ to be heavy dust. If heavy dust is identified in the glint, the algorithm continues with the retrieval, although it sets QAC=0. This permits the retrieval, but prohibits the values from being included in the Quality Weighted Level 3 statistics (Remer et al., 2005). If heavy dust is not identified in the glint, then the algorithm writes fill values to the aerosol product arrays and exits the procedure.

3.5. Sensitivity study

Tanré et al., (1997) and ATBD-96, describe a rigorous sensitivity study that tests the ability of the inversion procedure to retrieve the 'correct' values of aerosol properties. Since the Tanré et al., (1997) study uses the pre-launch aerosol models (11 models), we expect that there would be no significant difference if repeated for the nine models in the C005-O inversion described here. Here, we will attempt to summarize the major results of their sensitivity experiments, noting that the some of the details are obsolete for this version of the algorithm.

In the first experiment, Tanré et al., (1997) fed exact values from the lookup table ('Input in the LUT') into the inversion algorithm, attempting to determine the accuracy in which the algorithm recreated the simulated aerosol properties. They considered two small modes 'S_A' and 'S_B', as well as one large mode 'L_A' out of their choice of lookup table modes. For each of these modes, they considered four optical depths at 0.55 μm (τ = 0.2, 0.5, 1.0 and 2.0) and three choices of η mixing, (η = 1.0, 0.0 and 0.41). One set of solar/surface/sensor scattering geometry was used. In each case, the authors attempted to retrieve total τ ($\tau_{0.55}^{tot}$), η ($\eta_{0.55}$), the effective radius (R_{eff}) and the asymmetry parameter (g).

Figures 4 (quadrant 'a') and 5 (quadrant 'a' for each A-D) represent 'Input in the LUT'. Figure 4 is for the retrieved τ, whereas Figure 5 plots retrievals of size parameters. Figure 5D is a blowup of 5C, showing $R_{eff} < 0.40$. In each figure, the 'best' solution is plotted as the black dots, 'average solution is plotted with red 'X', whereas the standard deviation of the 'average' solution is plotted as the error bars. In all cases, the 'best' solutions exactly match the input scenarios, even the size parameters. The 'average' solutions, however, show instability compared to the input aerosol properties, especially for the size parameters. Large variations are observed in the 'average' solutions (crosses) for double modes, i.e., when η=0.0 and 1.0. Because of the lack of uniqueness in the relationship between physical and optical properties (Tanré et al., 1996), large fluctuations in the retrieved physical properties may occur, and as a result the algorithm has to compensate them by selecting a wrong ratio η or the effective radius. It is interesting to note that the standard deviation (shown as error bars) is a good measure of the quality of the retrieval; when it is small, the 'average' and 'best' solution are similar and quite close to the true values. The point is that although the exact characteristics of the different modes are difficult to assess, the algorithm is generally sensitive to the relative *size* of the aerosol particles.

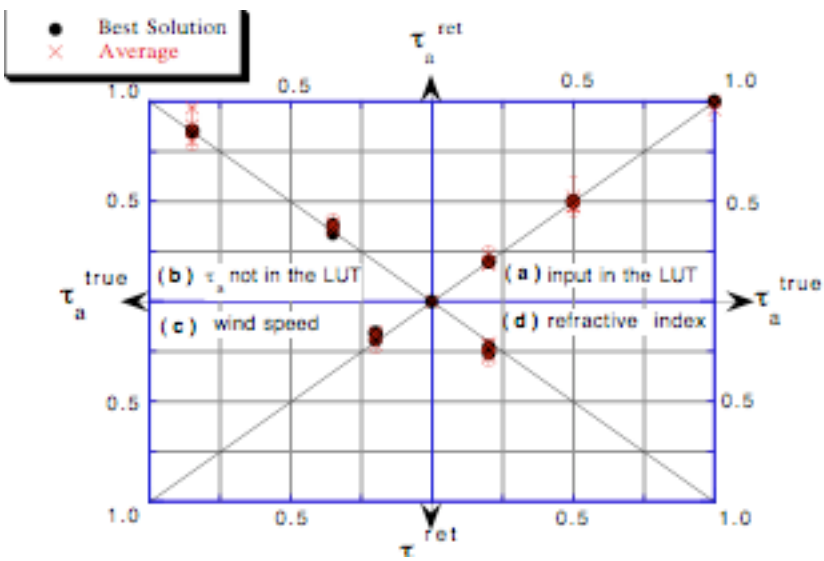

Figure 4: Scatter diagram of the optical depth, the x-axis correspond to the input, the y-axis to the retrieved values. Each quarter is devoted to a specific sensitivity study. Quarter (a) corresponds to an inversion where all the input are included in the LUT. Quarter (b) corresponds to an inversion with values of the optical thickness not included in the LUT. Quarter (c) corresponds to a different wind speed. Quarter (d) corresponds to a different refractive index. The black dots correspond to the 'best' solutions and the crosses to the 'average' solutions (see the text).

The next experiment was an attempt to simulate aerosol conditions that were not already in the LUT, such as $\tau_{0.55}^{tot}$ = 0.35 or 0.85. To simulate these scenarios, they performed additional (offline) RT calculations to simulate the TOA reflectance in these conditions. These spectral reflectances were then fed into the inversion code to determine the quality of retrieval. All of the quadrants 'a' in Figures 4-5 represent the conditions of inputs 'Not in the LUT' and show that the algorithm performs about as well as for inputs 'in the LUT'.

Quadrants 'c' and 'd' of each subfigure in Figures 4 and 5 show how the retrieval performs in cases of where the wind speed or refractive index is different from that in the LUT. Again the details are described in Tanré et al., (1997).

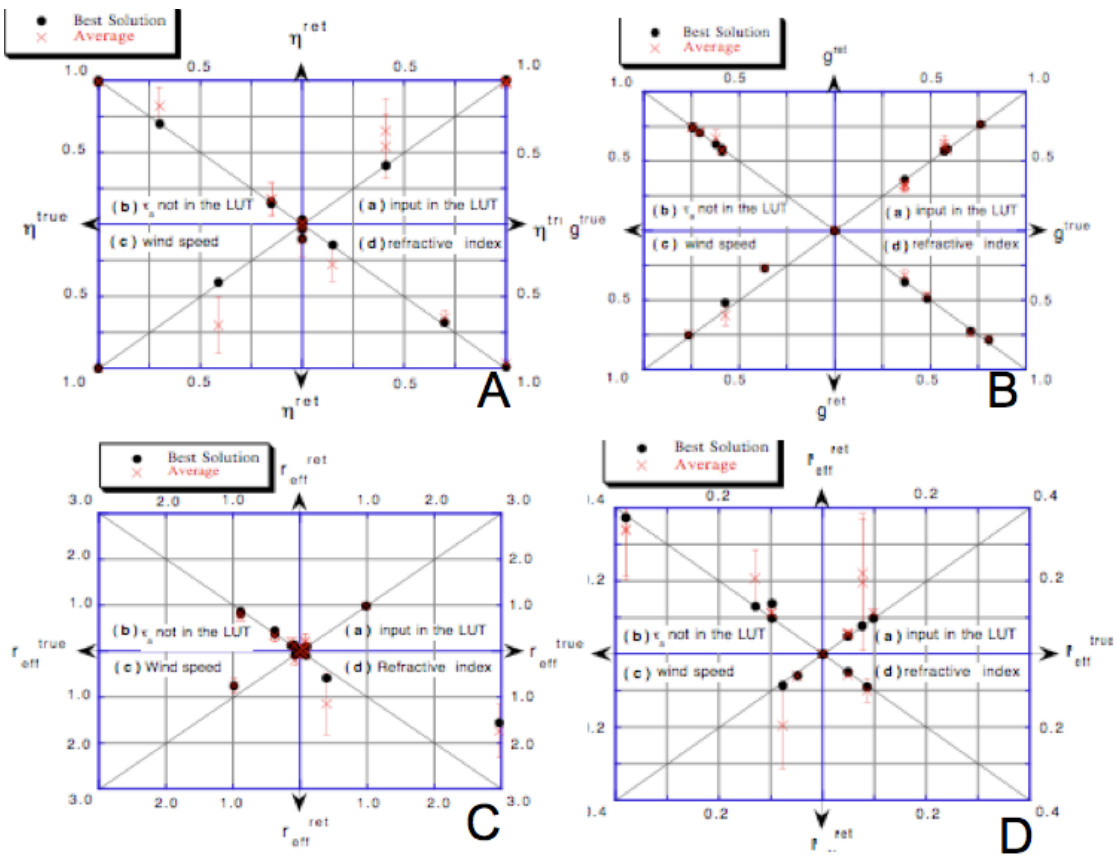

Figure 5: Same as Fig 4 but for (A) ratio η, (B) asymmetry parameter g, (C) Effective Radius (R_{eff}) and for (D) blowup of Effective Radius (R_{eff}) < 0.40. Error bars correspond to the standard deviation of the 'average' solution.

In the next set of experiments, Tanré et al., (1997) consider sources of error on the aerosol retrieval, such as sensor calibration, contamination by glint, or false estimate of the water-leaving radiance. To simulate these effects, they added the error separately to the measurements ρ_j^m for each channel (j) in the following way.

(a) 'calibration error': $\rho_j^m => \rho_j^m$ (1.-Rnd_j) where Rnd_j is random error scaled between ±0.01. It represents a random spectral calibration error of maximum of 1%.

(b) 'Glint error': $\rho_j^m => \rho_j^m$+0.01. This considers that the glint effect may not be completely avoided or predicted, which adds a constant value to the reflectance in all channels.

(c) 'Type 1 surface error': $\rho_j^m => \rho_j^m$+ Rnd_j, where Rnd_j is random error between ±0.002. It represents, for instance, possible errors in the water leaving radiance.

(d) 'Type 2 surface error': $\rho_j^m => \rho_j^m$+ (0.005/λ_j), where where λj is the center wavelength (in μm) of channel j. For doing so, the reflectance is increased by approximately 0.01 in 0.55 μm and 0.0025 in 2.12 μm channels, representing systematic errors in the spectral dependence of the reflectances, like uncertainties resulting from the foam spectral dependence.

Figures 6 and 7 (A-D) are analogous to Figures 4 and 5 (A-D), except each quadrant in each subfigure represents the errors described by (a)-(d). For randomly distributed errors (calibration and type 1 surface errors – quadrants (a) and (c)) Figure 6 shows that there is no systematic effect on the retrieval of τ and that the impact is almost negligible in most of the cases. Surface errors due to the glint or Type 2 surface errors (i.e., non-random errors) lead to an overestimate of the optical thickness. That is because additional surface contribution is translated into a larger atmospheric contribution, and in consequence results in a larger optical thickness. This effect is more important for small optical thickness, as the relative contribution of the surface is larger. As for the primary derived size parameter η, Figure 7A clearly shows that it is difficult it them accurately in the presence of these errors. The dispersion of η is quite large for both 'best' and 'average' solutions, so that the retrieved values are considered only as an estimate of the relative contributions of the modes. However, the derived retrievals of R_{eff} and g tend to be much better, especially for the 'best' solution. Let us note that the glint effect is the most destructive error for retrieving size paramters; it may result in 100% error but there is no systematic bias. Better values of size parameters can be retrieved far from the specular reflection. It was from this sensitivity study that glint angles > 40° would not be considered for quality aerosol retrieval.

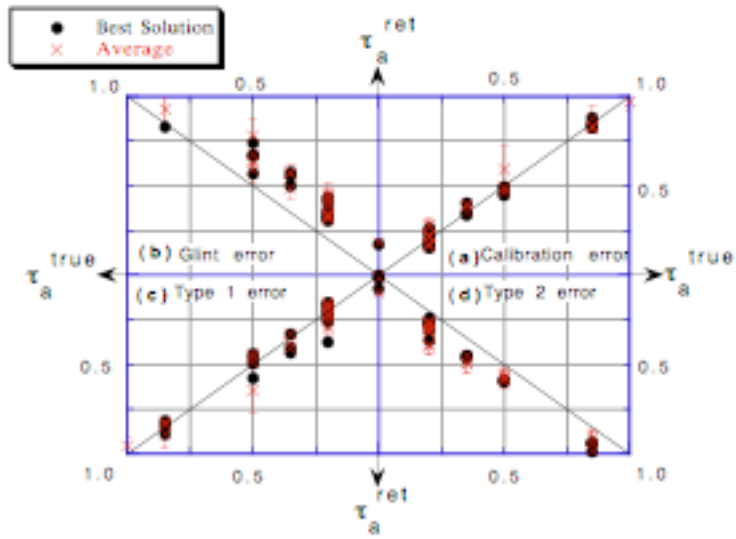

Figure 6: Scatter diagram of the optical depth, with additional assumed 'errors'. The x-axes correspond to the input, the y-axes to the retrieved values. Each quarter is devoted to a specific source of errors. Quarter (a) corresponds to calibration errors. Quarter (b) corresponds to glint error. Quarter (c) corresponds to Type 1 surface error. Quarter (d) corresponds to Type 2 surface error. The black dots correspond to the 'best' model and the crosses to the 'average' solution (see the text). Errors bars correspond to the standard deviation of the 'average' solution.

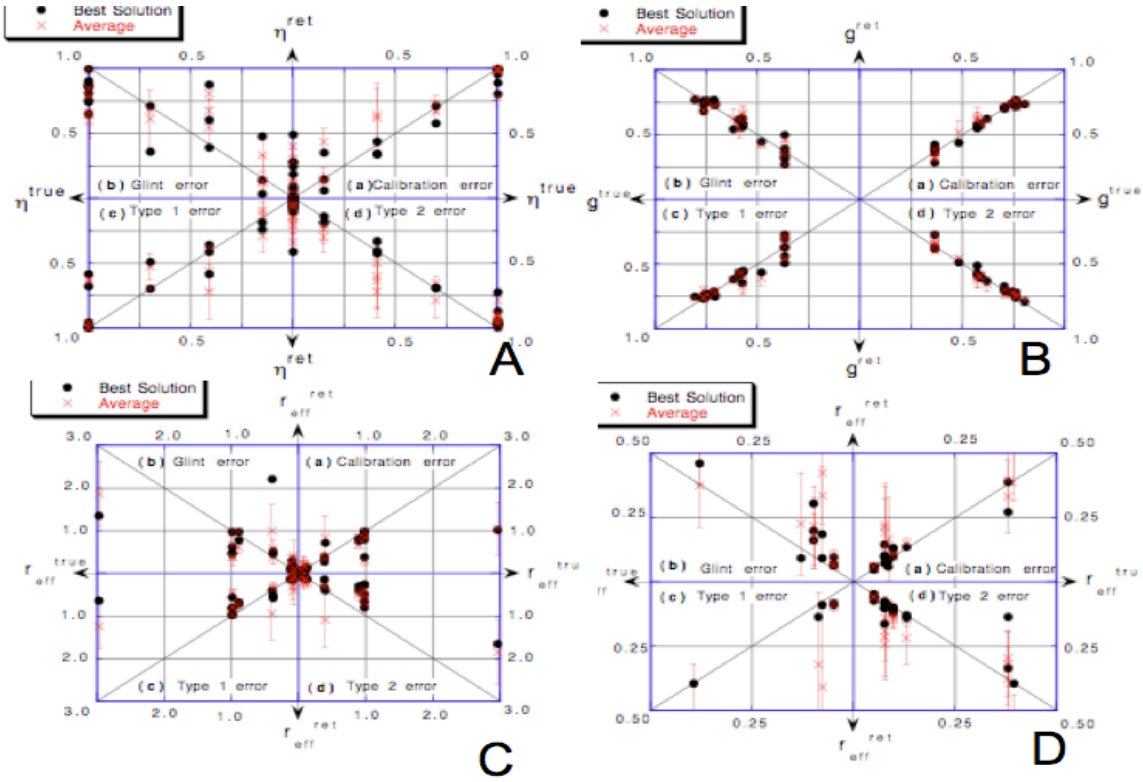

Figure 7: Same as Fig 6 but for (A) ratio η, (B) asymmetry parameter g, (C) Effective Radius (R_{eff}) and for (D) blowup of Effective Radius (R_{eff}) < 0.40. Error bars correspond to the standard deviation of the 'average' solution.

3.6. Retrieved ocean products

As discussed earlier, the primary retrieved products of the ocean algorithm are the total τ at 0.55 μm ($\tau_{0.55}^{tot}$), the Fine (Mode) Weighting (η or η) and which Fine (f) and which Coarse (c) modes were used in the retrieval. The fitting error (ε) of the simulated spectral reflectance is also retrieved. Both the 'best' and 'average' solutions are reported. From these primary products a number of other parameters can be easily derived. Examples include the effective radius (r_e) of the combined size distribution, the spectral total, fine and coarse τs (τ_λ^{tot}, τ_λ^{f} and τ_λ^{c}), and a measure of the columnar aerosol mass concentration (M_C). Table 3 lists the products retrieved and derived during the ocean retrieval algorithm. Details about the 'Quality_Assurance_Ocean' (QA) SDS are given in the Appendix. Note that a parameter's type does not indicate whether a parameter should be used in a quantitative way. Each parameter must be independently validated via comparison with ground-truth measurements, where *validation* means that the product can be shown to be comparable with a set of ground-truth observations on a global scale, within some defined expected uncertainty. Prior to operational C005 processing, a testbed of 6300 MODIS granules were used to evaluate a few of the derived parameters. This *provisional* evaluation is described in Section 5. This ATBD document will be updated as the MODIS products are evaluated and reach validation status (Section 7).

TABLE 3: CONTENTS OF MODIS C005 AEROSOL LEVEL 2 FILE (MOD04/MYD04): OCEAN PRODUCTS

Name of Product (SDS)	Dimesions: 3rd Dimension	Type of product
Effective_Optical_Depth_Average_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Retrieved Primary
Effective_Optical_Depth_Best_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Retrieved Primary
Optical_Depth_Ratio_Small_Ocean_0_55micron	X,Y,2: average, best	Retrieved Primary
Solution_Index_Ocean_Small	X,Y,2: average, best	Retrieved Primary
Solution_Index_Ocean_Large	X,Y,2: average, best	Retrieved Primary
Least_Squares_Error_Ocean	X,Y,2: average, best	Retrieved Diagnostic
Effective_Radius_Ocean	X,Y,2: average, best	Derived
Optical_Depth_Small_Best_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Optical_Depth_Small_Average_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Optical_Depth_Large_Best_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Optical_Depth_Large_Average_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Mass_Concentration_Ocean	X,Y,2: average, best	Derived
Cloud_Condensation_Nuclei_Ocean	X,Y,2: average, best	Derived
Asymmetry_Factor_Best_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Asymmetry_Factor_Average_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Backscattering_Ratio_Best_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Backscattering_Ratio_Average_Ocean	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Derived
Angstrom_Exponent_1_Ocean (0.55/0.86 micron)	X,Y,2: average, best	Derived
Angstrom_Exponent_2_Ocean (0.86/2.1 micron)	X,Y,2: average, best	Derived
Cloud_Condensation_Nuclei_Ocean	X,Y,2: average, best	Derived
Optical_Depth_by_models_ocean	X,Y,9: 9 models	Derived
Cloud_Fraction_Ocean	X,Y:	Diagnostic
Number_Pixels_Used_Ocean	X,Y:	Diagnostic
Mean_Reflectance_Ocean	X,Y:	Diagnostic
STD_Reflectance_Ocean	X,Y:	Diagnostic
Aerosol_Cldmask_Byproducts_Ocean	X,Y:	Diagnostic
Quality_Assurance_Ocean	X,Y,5 bytes	Diagnostic
Optical_Depth_Land_And_Ocean	X,Y: 0.55 μm	Joint (QA≥0) **
Image_Optical_Depth_Land_And_Ocean	X,Y: 0.55 μm	Joint (QA≥0)
Optical_Depth_Ratio_Small_Land_And_Ocean	X,Y: 0.55 μm	Joint (QA≥0)

X = 135; Y = 203. If there is a 3rd dimension of the SDS, then the indices of it are given. The "Retrieved" parameters are the solution to the inversion, whereas "Derived" parameters follow from the choice of solution. "Diagnostic" parameters aid in understanding of the directly Retrieved or Derived products. "Experimental" products are unrelated to the inversion but may have future applications. "Joint" products are the combined land and ocean products, with associated QAC constraint (for over ocean) in parentheses. **Based on evaluation of operational C005-O data, the QAC for quantitative studies should be limited to QAC≥1 only.

Some of the ocean products are combined with products from land (discussed in the next section) as the *Joint* products. For τ, two joint products are reported, the 'Optical_Depth_Land_And_Ocean' and the 'Image_Optical_Depth_Land_And_Ocean'. The first product is supposed to have more quantitative meaning, so is constrained by QAC. In practice, however, the two joint products are identical over ocean, meaning that the values of 'Effective_Optical_Depth_Average_Ocean' (at 0.55 μm) are written into both SDSs, regardless of QAC. Subsequent validation exercises (Section 7) suggest limiting the more quantitative joint product to ocean data having QAC≥1, which will be implemented in Feb 2009. The 'Optical_Depth_Ratio_Small_Ocean' product is copied (regardless of QAC) into 'Optical_Depth_Ratio_Small_Land_And_Ocean'..

4. Algorithm description: land

The upward reflectance (normalized solar radiance), at the top of the atmosphere (TOA), is a function of successive orders of radiation interactions, within the coupled surface-atmosphere system. The TOA angular (θ_0,θ, and ϕ = solar zenith, view zenith and relative azimuth angles) spectral reflectance ($\rho_\lambda(\theta_0,\theta,\phi)$) at a wavelength λ results from: scattering of radiation within the atmosphere without interaction with the surface (known as the 'atmospheric path reflectance'), the reflection of radiation off the surface that is directly transmitted to the TOA (the 'surface function'), and the reflection of radiation from outside the sensor's field of view (the 'environment function'). The environment function is neglected so that to a good approximation:

$$\rho_\lambda^*(\theta_0,\theta,\phi) = \rho_\lambda^a(\theta_0,\theta,\phi) + \frac{T_\lambda(\theta_0)T_\lambda(\theta)\rho_\lambda^s(\theta_0,\theta,\phi)}{1 - s_\lambda \rho_\lambda^s(\theta_0,\theta,\phi)} \qquad (8)$$

where ρ^a_λ is the atmospheric 'path reflectance', $T_\lambda(\theta_0)$ and $T_\lambda(\theta)$ is the downward and upward atmospheric transmissions (reciprocity implied), s_λ is the atmospheric backscattering ratios, and ρ^s_λ is the angular spectral surface reflectance.

Except for the surface reflectance, each term on the right hand side of Equation (8) is a function of the aerosol type and loading (τ). Assuming that a small set of aerosol types and loadings can describe the range of global aerosol, we can derive a lookup table that contains pre-computed simulations of these aerosol conditions. The goal of the algorithm is to use the lookup table to determine the conditions that best mimic the MODIS-observed spectral reflectance ρ^m_λ, and retrieve the associated aerosol properties (including τ and η). The difficulty lies in making the most appropriate assumptions about both the surface and atmospheric contributions.

4.1. Strategy

Compared to the C004-L family of algorithms, the C005-L algorithm (V5.2 and beyond) is a complete overhaul (Levy et al., 2007a, 2007b). Whereas C004-L essentially retrieved aerosol properties independently in two visible channels (0.47 and 0.66 μm) retrieval, C005-L retrieves aerosol properties in three channels simultaneously (the two visible channels, plus the 2.12 μm channel). The C004-L algorithms assumed that aerosol was transparent in the 2.12 μm channel, and that surface reflectance in the visible channels was a constant ratio of the observed (equal to surface) reflectance at 2.12 μm. C005-L assumes that the 2.12 μm channel contains information about coarse mode aerosol as well as the surface reflectance. The surface reflectance in the visible is still a function of the surface reflectance at 2.12 μm, but is also a function of the scattering angle and the "greenness" of the surface in the mid-IR spectrum (NDVI-like parameter based on 1.24 and 2.12 μm).

Like the ocean algorithm (C005-O), the C005 land (C005-L) algorithm is an inversion, but takes only three (nearly) independent observations of spectral reflectance (0.47, 0.66 and 2.1 μm) to retrieve three (nearly) independent pieces of information. These include total τ at 0.55 μm ($\tau_{0.55}$), Fine (model) Weighting at 0.55 μm (η or $\eta_{0.55}$), and the surface reflectance at 2.1μm ($\rho^s_{2.12}$). Like the ocean

algorithm, C005-L is based on look-up table (LUT) approach, i.e., radiative transfer calculations are pre-computed for a set of aerosol and surface parameters and compared with the observed radiation field. The algorithm assumes that one fine-dominated aerosol model and one coarse-dominated aerosol model (each may be comprised of multiple lognormal modes) can be combined with proper weightings to represent the ambient aerosol properties over the target. Spectral reflectance from the LUT is compared with MODIS-measured spectral reflectance to find the best match. This best fit is the solution to the inversion.

In addition to the new philosophy of over-land retrieval for C005, the C005-L algorithm makes use of new climatology of aerosol optical properties, vector radiative transfer for creating the lookup tables, and new logic for creation of the Quality Assurance (QA) flags.

4.2. Formulation of the land algorithm

For formulating C005-L, we used a large collection of AERONET L2A (quality assured Level 2 data) sunphotometer data (http://aeronet.gsfc.nasa.gov) co-located with the MODIS C004 data (Ichoku et al. 2002. From the AERONET database, we used a combination of direct 'sun' measurements of τ in four or more wavelengths (at least 0.44, 0.67, 0.87 and 1.02 μm) and indirect 'sky' measurements of almucantur radiance that were inverted into aerosol optical properties and size distributions. Sun measurements are made approximately every 15 minutes, whereas almucantur sky measurements are performed about every hour. Recently, O'Neill et al., (2003) developed an algorithm to invert sun-measured spectral τ to yield estimates of Fine aerosol Weighting (η). Over 15,000 pairs of MODIS and AERONET sun data, at over 200 global sites, have been co-located in time via the technique of Ichoku et al., (2002). A valid MODIS/AERONET match is considered when there at least five (out of a possible 25) MODIS retrievals (10 km x 10 km resolution) within the box, and at least two (out of a possible five) AERONET observations within an hour. About 136,000 individual AERONET sky retrievals were used to develop the C005-L global/seasonal aerosol climatology.

4.3. Aerosol optical models and lookup tables

A number of studies (e.g. Chu et al., 2002, Remer et al., 2005, Ichoku et al., 2002, Levy et al., 2005) have demonstrated that MODIS/AERONET regression of τ over land resulted in slope less than one; meaning that MODIS tended to under-retrieve optical depth for large τ. This indicated that the aerosol models used in C004-L were not truly representative of the optical conditions viewed by MODIS. For example, over the east coast of the United States during summertime, Levy et al. (2005) showed MODIS retrievals would be improved by switching to the GSFC urban/industrial aerosol model derived from AERONET data (Dubovik et al., 2002). Omar et al., (2005) performed a "cluster analysis" of AERONET data and found that six aerosol models (composed of desert dust, biomass burning, background/rural polluted continental, marine, and dirty pollution, respectively) sufficiently represented the entire AERONET dataset. The models varied mainly by their ω_0 and size distribution (asymmetry parameter, g). Two models were representative of very clean conditions (marine and background/rural). One of the models (dust) was coarse-dominated, analogous to the MODIS coarse dust model, and three were fine models having different ω_0 (biomass burning, polluted continental, and dirty pollution), that were analogous to the C004-L set of fine models. Because the MODIS over-land retrieval employs only three channels (and suffers from surface and other contaminations), it is not able to select among choices of fine aerosol model. Therefore, the aerosol retrieval algorithm must

assign the fine aerosol model a priori of the retrieval. This section describes how cluster analysis was used to determine a set of global aerosol models, and how they were assigned as a function of location and season. More information is in Levy et al., (2007a).

For our subjective cluster analysis, we used all AERONET Level 2 (L2A) data that were processed as of February 2005, encompassing both spherical and spheroid retrievals. We discriminated the retrievals by the minimum quality parameters suggested by the AERONET team, including: τ at 0.44 μm greater than 0.4, solar zenith angle greater than 45°, 21 symmetric left/right azimuth angles, and radiance retrieval error less than 4%. The resulting data set was comprised of 13,496 spherical retrievals and 5128 spheroid retrievals at over 100 sites. In order to differentiate between aerosol types, we separated the AERONET data set into ten discrete bins of τ. Each bin, then, was used separately to differentiate aerosol types. Presumably, this would help to identify expected dynamic properties (function of τ) of each aerosol type (e.g. Remer et al., (1998)). In contrast to Omar et al. (2005), we desired to pursue not necessarily the most statistically significant clustering, but rather to identify three distinct fine-dominated models useful for MODIS. With this goal of fine model identification in mind, we clustered with respect to only two optical parameters: ω_0 at 0.67 μm and the asymmetry parameter *g* at 0.44 μm. Presumably the ω_0 would differentiate between non-absorbing aerosols (such as urban/industrial pollution – (Remer et al., 1998; Dubovik et al., 2002)) and much more absorbing aerosols (such as savanna burning smoke – (Ichoku et al., 2003; Dubovik et al., 2002)), and *g* at 0.44 μm would help differentiate between relative size (via the phase function) of primarily fine aerosols. We numbered the clusters so that in each τ bin, there is always a 'low', 'medium' and 'high' cluster, and upon re-combination across τ bins, that they would yield dynamical information. As for the coarse aerosol model, we found that a single cluster described the spheroid-based almucantur inversions (Dubovik et al., 2006). Since the sites contributing to spheroid data were primarily those known to be in dust regions, we assumed that the spheroid model represented coarse ('dust') aerosol.

For each AERONET site, and for each season, we determined the percentage of the retrievals attributed to each cluster. Figure 8 (a-d) displays pie-plots at each site, as a function of season. To remove poor statistics, we show pie plots only at sites having at least 10 observations (per season) during the history of AERONET. Green pie segments represent the non-absorbing (ω_0~0.95) model (presumably urban/industrial aerosol), blue segments are the moderately absorbing (ω_0~0.90) model (presumably background, forest smoke and developing world aerosol), and red segments designate the highly absorbing (ω_0~0.85) model (presumably savanna/grassland smoke aerosol). At most sites and most seasons, the aerosol type is as expected. Non-absorbing aerosol (green) dominates the U.S. East Coast and far western Europe, whereas highly absorbing aerosol (red) dominates the savannas of South America and Africa. Most other sites are dominated by moderately absorbing aerosol (blue) or are a mix of all clusters. There are some exceptions to expectation, however. Surprisingly, Southeast Asia seems to be primarily non-absorbing aerosols, as opposed to the absorbing aerosol assumed in C004-L. Recent studies (e.g. Eck et al., 2005) confirm that the urbanized areas of Southeast Asia are primarily non-absorbing. A few sites in Western Europe have large fractions of absorbing aerosol, yet the reason is not known.

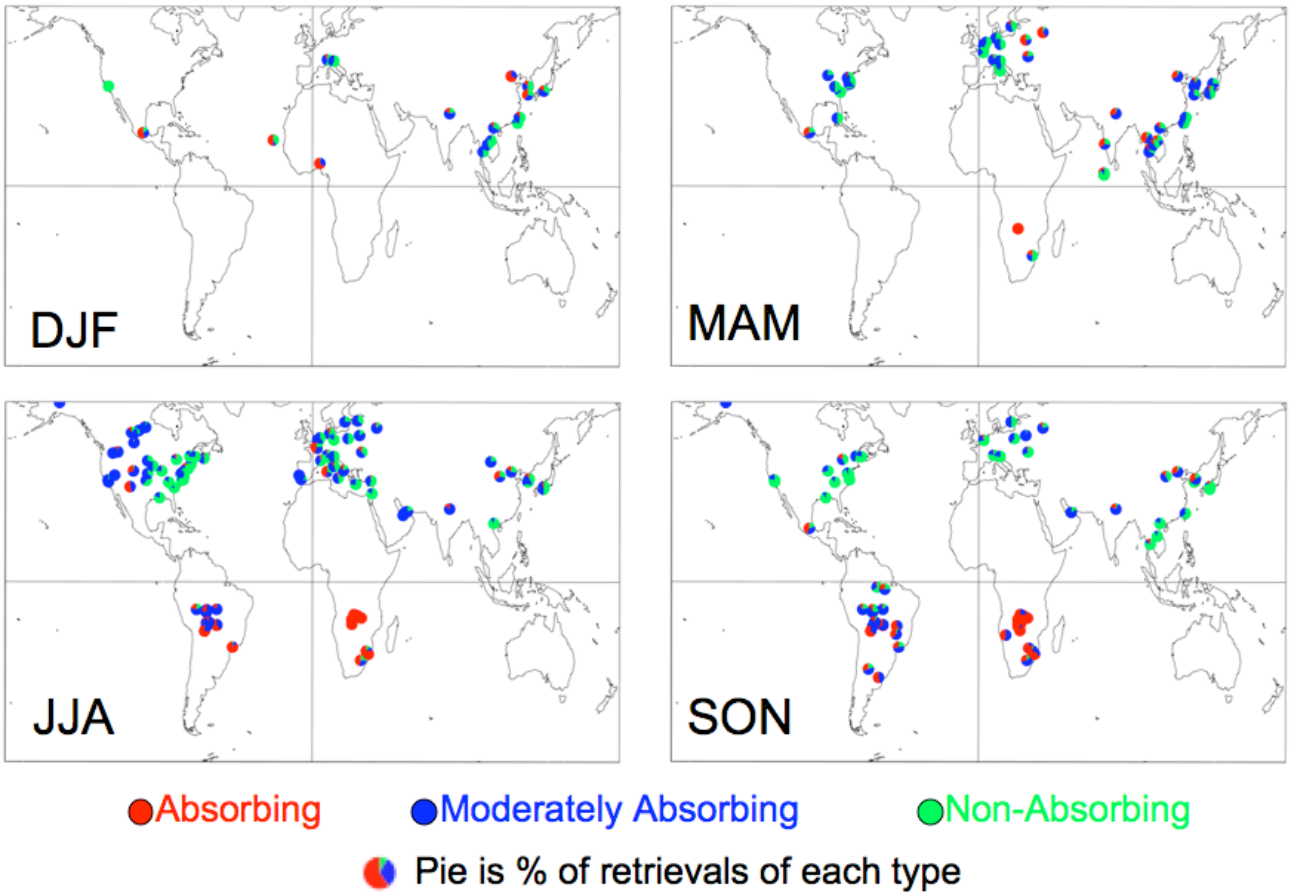

Figure 8: Percentage (pie charts) of spherical aerosol model type retrieved at each AERONET site per season. Colors represent absorbing ($\omega_0 \sim 0.85$), moderately absorbing ($\omega_0 \sim 0.90$) and non-absorbing ($\omega_0 \sim 0.95$), respectively.

Keeping in mind our goal of dividing the world into plausible aerosol types, we decided that each site should have an assumed aerosol type attached to it. The Moderately absorbing aerosol model was set as the default, and would be overwritten only if clear dominance of one of the other two aerosol types was observed. If either the non-absorbing or the absorbing aerosol occupied more than 40% of the pie, and the other occupied less than 20%, then the site was designated as the dominant aerosol type. For example, GSFC (Longitude = -77; Latitude = 37.93) during the summer months (JJA) recorded 87% non-absorbing and 13% moderately absorbing, meaning it would be designated as non-absorbing. Figure 9 (a-d) displays the designated aerosol types at each site. As in Figure 8, green represents non-absorbing, blue represents moderately absorbing and red designates absorbing aerosol types. Most site designations seem reasonable and were expected from our experience. North America during the summer (JJA) is split between non-absorbing and moderately absorbing aerosol types, much the same way (approximately -100° longitude) as was prescribed for C-004. Southern Africa during the winter season (DJF) is solidly designated as absorbing aerosol (e.g. Ichoku et al., 2003). Western Europe is evenly split between non-absorbing and moderately absorbing (except for two absorbing sites), meaning that a subjective decision is needed here. To follow C004-L, the non-absorbing aerosol model was chosen for Western Europe.

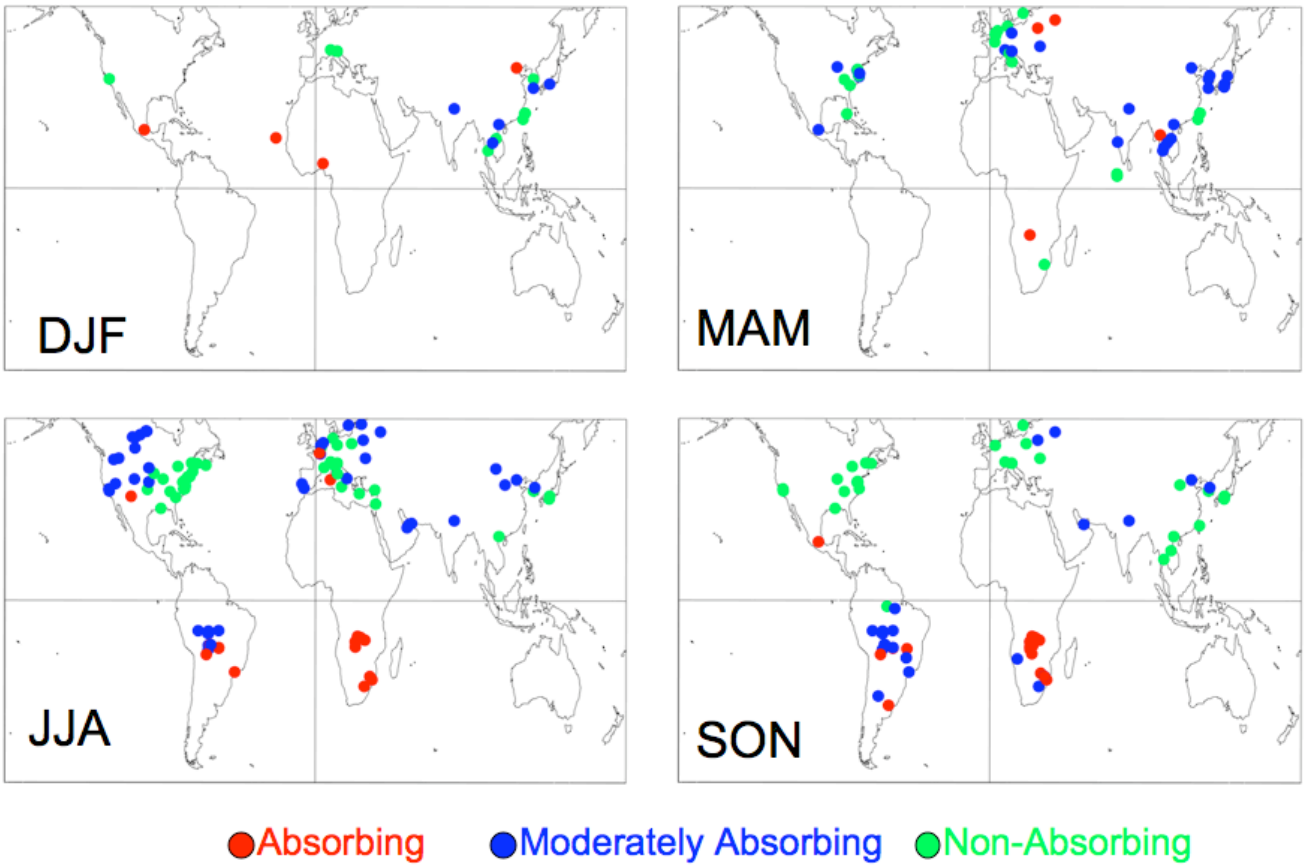

Figure 9: Final spherical aerosol model type designated at each AERONET site per season. Colors represent absorbing ($\omega_0 \sim 0.85$), moderately absorbing ($\omega_0 \sim 0.90$) and non-absorbing ($\omega_0 \sim 0.95$), respectively.

Figure 10 plots the final decision for designating aerosol types around the globe, as a function of season. Note that where possible the shapes correspond with the clustering. At some regions, however, some subjectivity was needed to connect areas. For example, even though insufficient data exists for Africa north of the equator, the known surface types and seasonal cycles suggest that heavy absorbing aerosol would be produced during the biomass burning season. Red designates regions where the absorbing aerosol is chosen, whereas green represents non-absorbing aerosol. The moderately absorbing ($\omega_0 \sim 0.90$) model is assumed everywhere else. These images were mapped onto a 1° longitude x 1° latitude grid, such that a fine aerosol type is assumed for each grid point, globally. This global map approach, that is not hardwired into the processing code, will allow for easy alterations as new information becomes available.

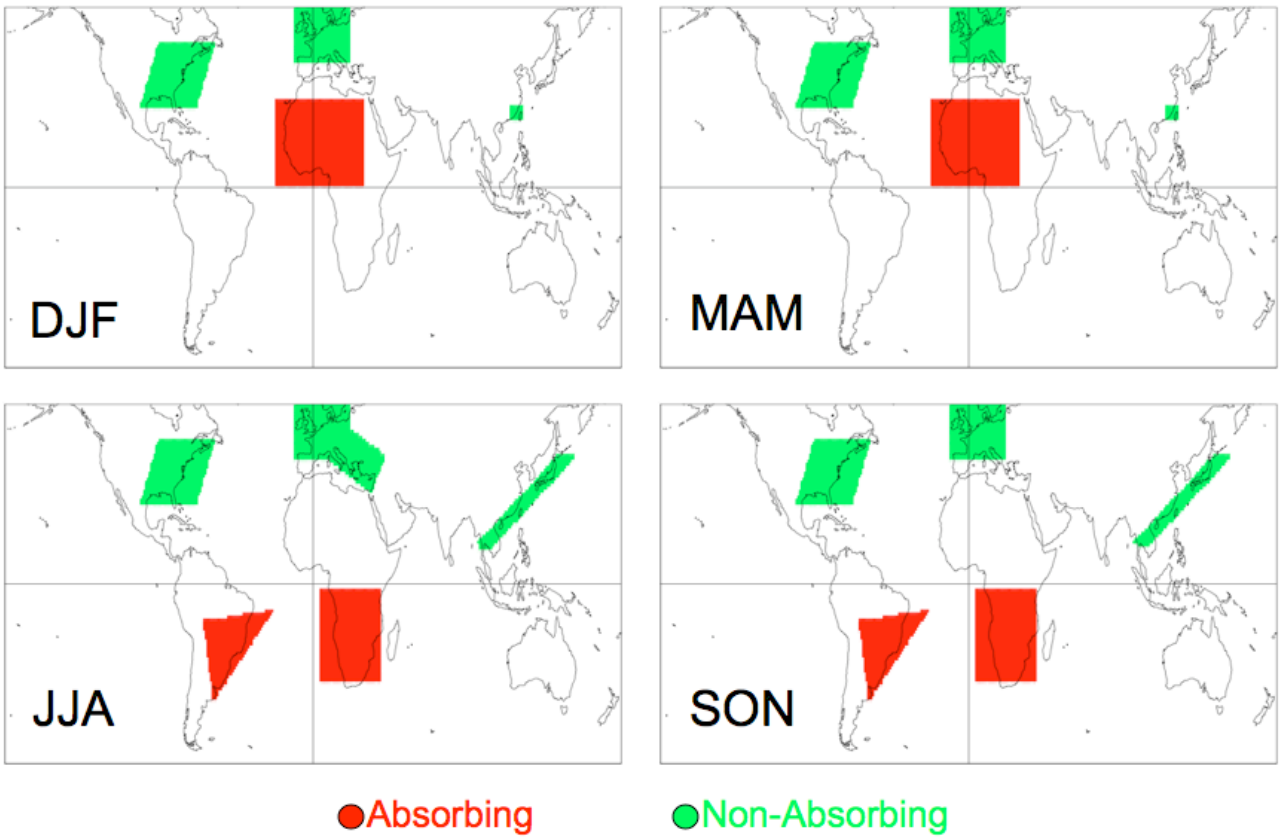

Figure 10: Final spherical aerosol model type designated at 1° x 1° gridbox per season. Red and green represent absorbing (ω_0 ~ 0.85) or non-absorbing (ω_0 ~ 0.95) models, respectively. Moderately absorbing (ω_0 ~ 0.90) is assumed everywhere else.

Table 4 displays the optical properties and size distributions for the three spherical (moderately absorbing, absorbing and non-absorbing) fine-dominated models and the one spheroid coarse aerosol (dust) model, and the Continental model. Figure 11 shows the size distributions for the four AERONET-derived models. Note the dynamic nature (function of τ) of the size properties of the fine models, especially the non-absorbing model. Figure 12 plots the final phase function at 0.55 μm for each model as well as spectral dependence of three parameters (τ, ω_0 and g), all for $\tau_{0.55} = 0.5$.

TABLE 4: OPTICAL PROPERTIES OF THE AEROSOL MODELS USED FOR THE C005-L LOOKUP TABLE

Model	**Mode**	**r_v (μm)**	**σ**	**V_0 (μm³/μm²)**	**Refractive Index: k**	**ω_0 / g (0.47/0.55/0.66/2.1μM) for $\tau_{0.55}$ = 0.5**
Continental						0.90/0.89/0.88/0.67 0.64/0.63/0.63/0.79
	Soluble	0.176	1.09	3.05	1.53-0.005i; 0.47 μm 1.53-0.006i; 0.55 μm 1.53-0.006i; 0.66 μm 1.42-0.01i; 2.12 μm	
	Dust	17.6	1.09	7.364	1.53-0.008i; 0.47 μm 1.53-0.008i; 0.55 μm 1.53-0.008i; 0.66 μm 1.22-0.009i; 2.12 μm	
	Soot	0.050	0.693	0.105	1.75-0.45i; 0.47 μm 1.75-0.44i; 0.55 μm 1.75-0.43i; 0.66 μm 1.81-0.50i; 2.12 μm	
Moderately absorbing / Developing						0.93/0.92/0.91/0.87 0.68/0.65/0.61/0.68
	Accum	$0.0203\tau + 0.145$	$0.1365\tau + 0.3738$	$0.1642\ \tau^{0.7747}$	$1.43 - (-0.002\tau+0.008)i$	
	Coarse	$0.3364\tau + 3.101$	$0.098\tau + 0.7292$	$0.1482\ \tau^{0.6846}$	$1.43 - (-0.002\tau+0.008)i$	
Non-absorb/ Urban-Ind						0.95/0.95/0.94/0.90 0.71/0.68/0.65/0.64
	Accum	$0.0434\tau + 0.1604$	$0.1529\tau + 0.3642$	$0.1718\ \tau^{0.8213}$	$1.42 - (-0.0015\tau+0.007)i$	
	Coarse	$0.1411\tau + 3.3252$	$0.1638\tau + 0.7595$	$0.0934\ \tau^{0.6394}$	$1.42 - (-0.0015\tau+0.007)i$	
Absorbing/ Heavy Smoke						0.88/0.87/0.85/0.70 0.64/0.60/0.56/0.64
	Accum	$0.0096\tau + 0.1335$	$0.0794\tau + 0.3834$	$0.1748\ \tau^{0.8914}$	$1.51 - 0.02i$	
	Coarse	$0.9489\tau + 3.4479$	$0.0409\tau + 0.7433$	$0.1043\ \tau^{0.6824}$	$1.51 - 0.02i$	
Spheroid/ Dust						0.94/0.95/0.96/0.98 0.71/0.70/0.69/0.71
	Accum	$0.1416\ \tau^{-0.0519}$	$0.7561\ \tau^{0.148}$	$0.0871\ \tau^{1.026}$	$1.48\tau^{-0.021} - (0.0025\ \tau^{0.132})i$; 0.47 μm $1.48\tau^{-0.021} - 0.002i$; 0.55 μm $1.48\tau^{-0.021} - (0.0018\ \tau^{-0.08})i$; 0.66 μm $1.46\tau^{-0.040} - (0.0018\ \tau^{-0.30})i$; 2.12 μm	
	Coarse	2.2	$0.554\ \tau^{-0.0519}$	$0.6786\ \tau^{1.0569}$	$1.48\tau^{-0.021} - (0.0025\ \tau^{0.132})i$; 0.47 μm $1.48\tau^{-0.021} - 0.002i$; 0.55 μm $1.48\tau^{-0.021} - (0.0018\ \tau^{-0.08})i$; 0.66 μm $1.46\tau^{-0.040} - (0.0018\ \tau^{-0.30})i$; 2.12 μm	

Listed for each model are the individual lognormal modes, and the final ω_0 at different wavelengths. Listed for each mode are the mean radius r_v, standard deviation σ of the volume distribution, and total volume of the mode, V_0. The complex refractive index is assumed for all wavelengths (0.47, 0.55. 0.66 and 2.1 μm), unless otherwise noted. The Absorbing and Moderately absorbing model parameters (r_v, σ and k) are defined for $\tau \leq 2.0$; for $\tau > 2.0$, we assume $\tau = 2.0$. Likewise, the Non-absorbing and Spheroid model parameters are defined for $\tau \leq 1.0$. V_0 (for all models) is defined for all τ.

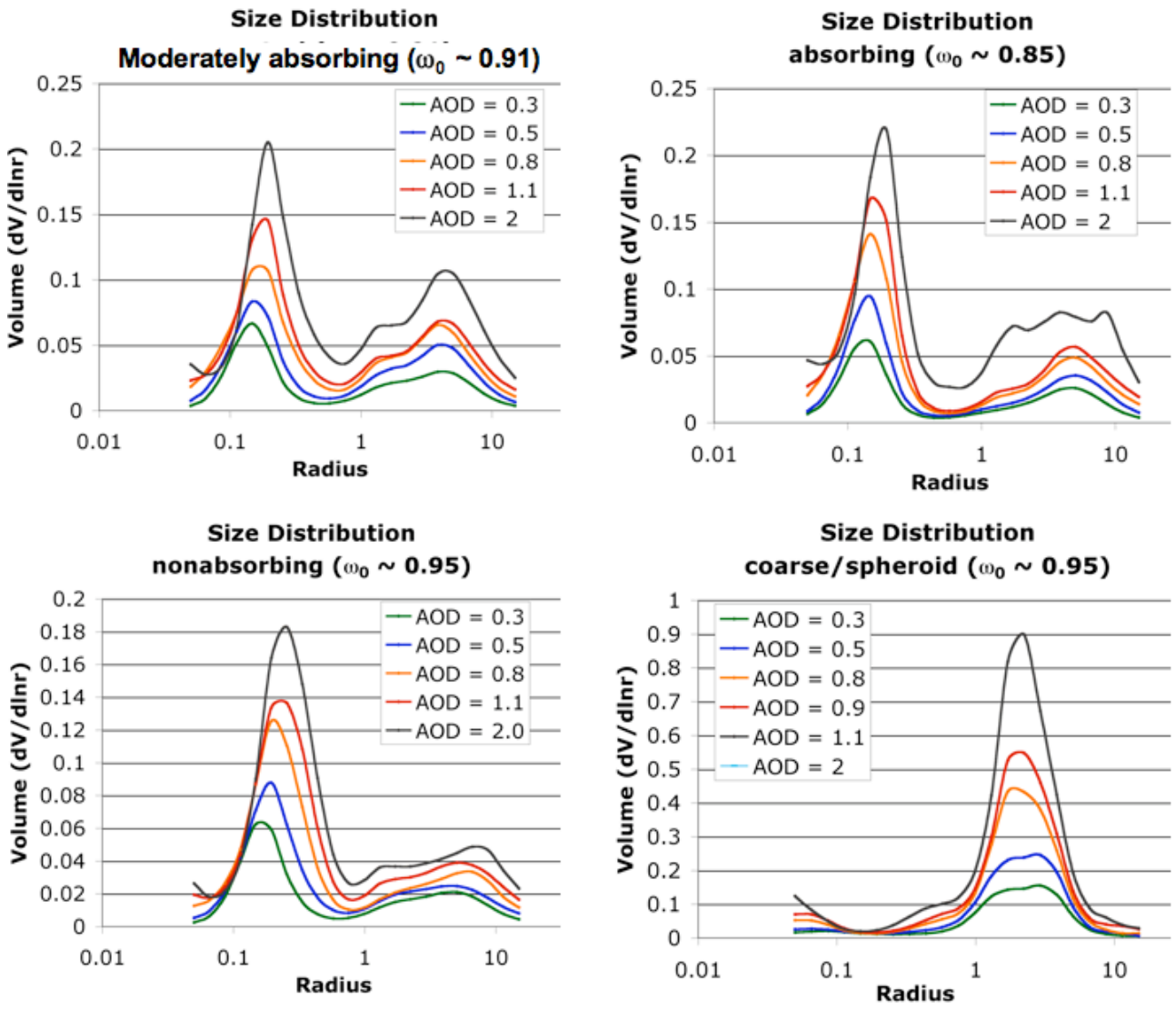

Figure 11: Aerosol size distribution as a function of optical depth for the three spherical (moderately absorbing, absorbing and non-absorbing) and spheroid (dust) models identified by clustering of AERONET.

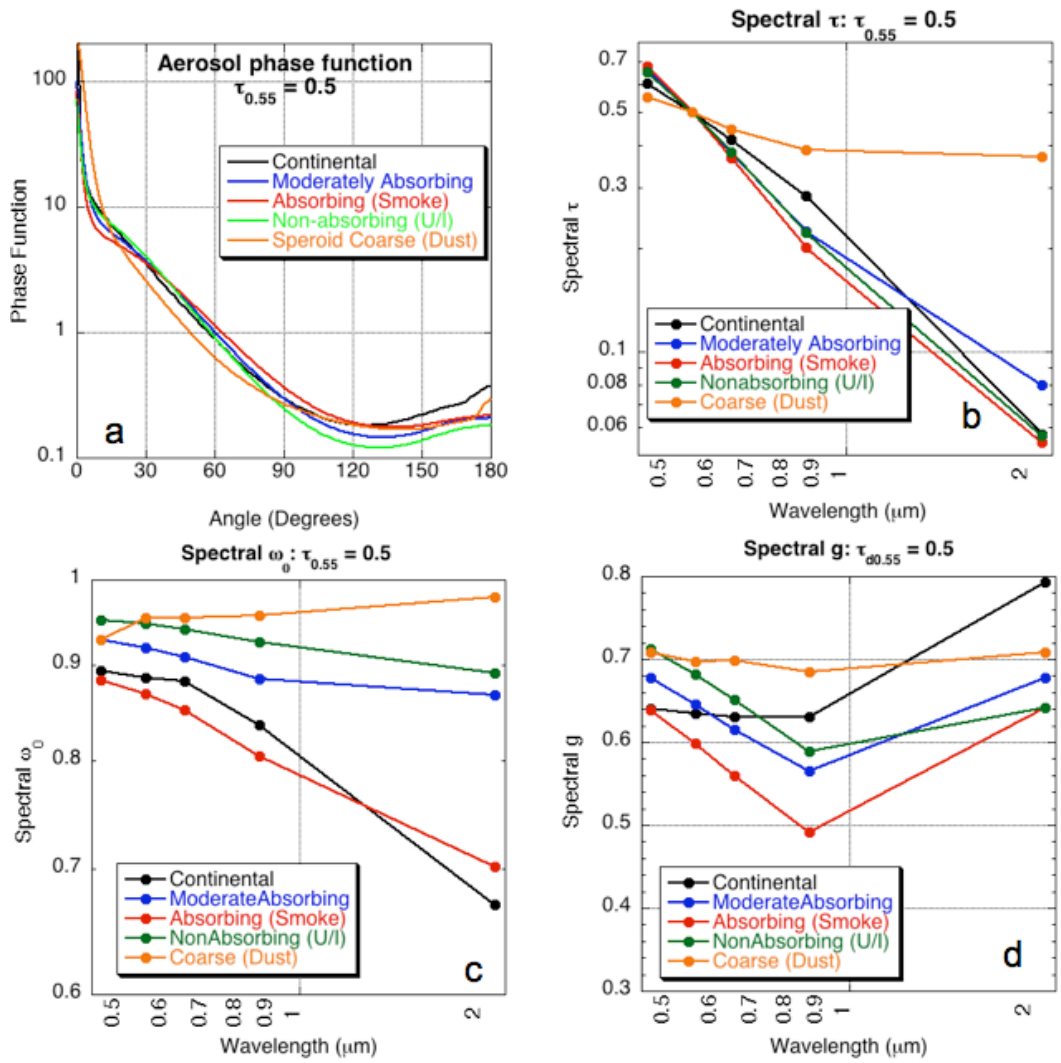

Figure 12: Plot of optical properties for the 5 aerosol models of the C005-L LUT, for τ = 0.5. a) phase function at 0.55 μm (as a function of angle) b) optical depth spectral dependence, c) single scattering albedo spectral dependence and d) asymmetry parameter spectral dependence.

The C004-L MODIS lookup table (LUT) contained simulated aerosol reflectance in two channels (0.47 and 0.66 μm), calculated using the non-polarized (scalar) SPD radiative transfer (RT) code (Dave et al., 1970). Levy et al., (2004) demonstrated that under some geometries, neglecting polarization could lead to significant errors in top of atmosphere reflectance, further leading to significant errors in τ retrieval. Figures 13a and 13b are taken from Levy et al., (2004), plotting errors in 0.47μm reflectance and associated errors in τ retrieval at the eight sun/surface/satellite geometries given in Table 5. As some errors are large (up to 0.01 for reflectance and 0.1 for τ), it was decided to use a vector RT code for creating the C005-L LUT. Fig. 14 shows the differences between the scalar versions of Dave and RT3 (Evans and Stephens, 1991), when simulating the Continental aerosol model (see Table 5). Plotted are the differences in 0.466 μm reflectance for the eight scattering geometries of Table 5. Under most geometries and optical depths, differences between the two RT codes are less than 0.001

(which is about 1%). Note that as in Levy et al., (2004), the aerosol scattering phase function elements (inputs to RT3) were calculated by the MIEV Mie code (Wiscombe et al., 1980).

TABLE 5: SOLAR/SURFACE/SATELLITE GEOMETRY FOR EIGHT EXAMPLES

Reference	Solar Zenith	View Zenith	Relative Azimuth	Scattering Angle
A	12.00	6.97	60.00	163.40
B	12.00	52.84	60.00	120.53
C	12.00	6.97	120.00	169.59
D	12.00	52.84	120.00	132.35
E	36.00	6.97	60.00	140.12
F	36.00	52.84	60.00	104.74
G	36.00	6.97	120.00	147.00
H	36.00	52.84	120.00	136.29

All units are degrees

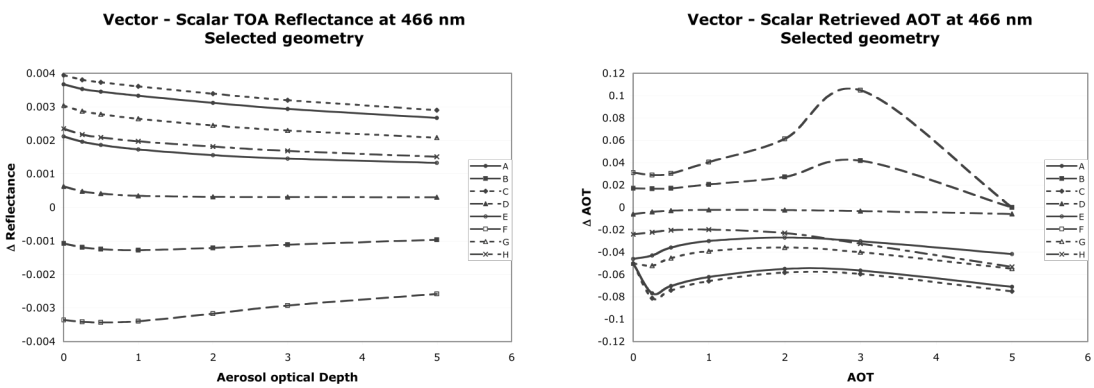

Figure 13: Difference between vector- and scalar- derived reflectance (a) and retrieved τ (b) at 0.466 μm, for eight example sun/surface/satellite geometries as a function of the input τ.

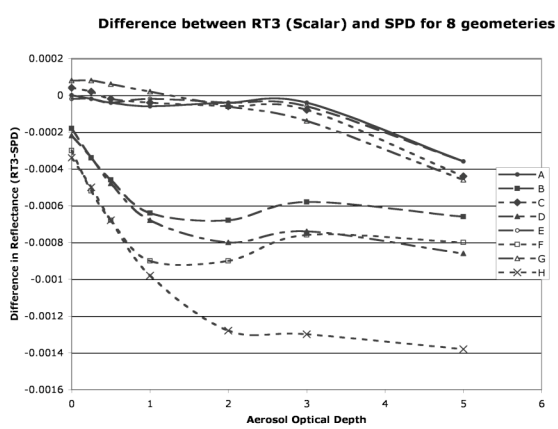

Figure 14: Difference between MIEV/RT3 combination and SPD derived reflectance at 0.466 μm, for eight example sun/surface/satellite geometries as a function of the input τ.

As described above, the fine aerosol models are assumed to be spherical particles. We used the combination of MIEV (Wiscombe et al., 1980) and RT3 (Evans and Stephens, 1991), described above and by Colarco et al., (2003). For the spheroids of the coarse aerosol model, Mie theory is not sufficient. We used instead, a version of the T-matrix code described in Dubovik et al., (2002, 2005), to calculate the scattering properties of the model. Not only is this a necessary approximation for

integrating a spheroid size distribution, it is consistent with the calculations used in fitting the original almucantur radiance in the first place. In summary, then, a combination of the T-matrix and RT3 codes is used for the coarse (dust) model LUT. Assumed central wavelengths and Rayleigh optical depths are shown in Table 1.

The C005-L LUT contains pre-computed optical properties of aerosol at four discrete wavelengths (0.466, 0.553, 0.644 and 2.119 μm, representing MODIS channels 3, 4, 1 and 7, respectively) for several values of aerosol total loadings, and for a variety of geometry. For discrete optical depths (described by the τ or τ at 0.55 μm) each spherical aerosol model (Continental, Moderately Absorbing, Absorbing and Non-absorbing) and non-spherical model (Dust), scattering/extinction properties of aerosol size distributions are calculated by either MIEV or the Dubovik T-matrix code. Assuming a Rayleigh atmosphere and realistic layering of the aerosol, the Legendre moments of the combined Rayleigh/aerosol are computed for each layer of a US Standard Atmosphere (U.S. Government, 1976). These moments are fed into RT3 to simulate TOA reflectance and total fluxes.

The parameters of Equation (8) were calculated for seven aerosol loadings ($\tau_{0.55}$ = 0.0, 0.25, 0.5, 1.0, 2.0 3.0, and 5.0). TOA reflectance was calculated for 9 solar zenith angles (θ_0 = 0.0, 6.0, 12.0, 24.0, 36.0, 48.0, 54.0, 60.0 and 66.0), 16 sensor zenith angles (θ = 0.0 to 65.0, approximate increments of 6.0), and 16 relative azimuth angles (ϕ = 0.0 to 180.0 increments of 12.0). All of these parameters are calculated assuming a surface reflectance of zero. The approximate increments of θ arise from the use of the Lobatto quadrature function in RT3; allowing for values that resemble C004-L LUT without having to perform extra interpolation.

When surface reflectance is present, the second term in Equation (8) is nonzero. The flux is a function only of the atmosphere, however, the atmospheric backscattering term, s, and the transmission term, T, are functions of both the atmosphere and the surface. Therefore, RT3 is run two additional times with distinct positive values of surface reflectance.

$$s = (1/\rho_1^s)(1-(F_d T\rho_1^s/(\rho^* - \rho^a)))$$

and (9a,b)

$$s = (1/\rho_2^s)(1-(F_d T\rho_2^s/(\rho^* - \rho^a)))$$

Here, we chose values of 0.1 and 0.25 for our surface reflectance. $\rho^s{}_1$ and $\rho^s{}_2$. These two equations can be solved for the two unknowns, s and T. The values of F_d, s, and T are saved into the LUT, for each τ index, wavelength and aerosol model.

Other parameters contained in the LUT include the scattering and extinction coefficients Q and variables describing the physical properties (lognormal size parameters r_g and σ, and complex refractive indices, n_r+in_k) of the aerosol models. We also compute a Mass concentration coefficient, M_C, in units of (μg per cm^2) that is a function of optical thickness and Q. The derivation of M_C and some typical values are explained in Appendix 3.

4.4. VISvs2.12 surface reflectance assumptions

When performing atmospheric retrievals from MODIS or any other satellite, the major challenge is separating the total observed reflectance into atmospheric and surface contributions, and then defining the aerosol contribution. Over the ocean, the surface is nearly black at red wavelengths and longer, so that assuming negligible surface reflectance in these channels is a good approximation. Over land, however, the surface reflectance in the visible and SWIR is far from zero and varies over surface type. As the land surface and the atmospheric signals are comparable, errors of 0.01 in assumed surface reflectance will lead to errors on the order of 0.1 in τ retrieval. Errors in multiple wavelengths can lead to poor retrievals of spectral τ, which in turn would be useless for estimating size parameters.

Kaufman and colleagues (e.g. Kaufman et al., 1997b) observed that over vegetated and dark soiled surfaces, the surface reflectance in some visible wavelengths correlated with the surface reflectance in the SWIR. Parallel simulations by vegetation canopy models, showed that the physical reason for the correlation was the combination of absorption of visible light by chlorophyll and infrared radiation by liquid water in healthy vegetation (Kaufman et al., 2002). These relationships were such that the surface reflectance values in the visible (blue and red channels) were nearly fixed ratios of that in the SWIR (Kaufman et al., 1997c). As applied within C004-L (and all previous versions), surface reflectance in the blue, 0.47µm channel 3 and the red, 0.66 µm channel 1 channels were assumed to be one-quarter and one-half, respectively, of the surface reflectance in the mid-SWIR 2.1µm channel 7 channel (Kaufman et al., 1997b). We note these as the '0.47vs2.12' and '0.66vs2.12' ratios, respectively, and collectively as 'VISvs2.12'

However, correlation of C004-L (and prior) MODIS-derived τ to AERONET sunphotometer data (Chu et al., 2002; Remer et al., 2005) showed positive offset of about 0.1, likely meaning that the surface reflectance was under-estimated. From data observed during the CLAMS experiment of 2001, Levy et al., (2005) demonstrated that higher values of VISvs2.12 surface ratios (e.g. 0.33 and 0.65 for the blue and red, respectively) improved the continuity of the MODIS over-land and over-ocean aerosol products along the coastline of the DelMarVa Peninsula. The MODIS/AERONET τ regression over near-coastal sites was also improved. However, at locations far from the coastline, the CLAMS VISvs2.12 ratios tended toward over-correction of the surface reflectance and retrievals of τ less than zero. It is also known that earth's surface is not Lambertian, and that some surface types exhibit strong bi-directional reflectance functions (BRDF). Gatebe et al., (2001) flew the Cloud Absorption Radiometer (CAR) low over vegetated surfaces and found that the VISvs2.12 surface ratios varied as a function of angle, and often greatly differed from the one-quarter and one-half ratios assumed by in C004-L. Remer et al., (2001) also noted that the VISvs2.12 surface ratios varied as a function of scattering geometry. In fact, under certain geometry, these VISvs2.12 surface ratios broke down completely. Some of this is related to BRDF effects (e.g. Lypastin et al., 2001).

To understand how VISvs2.12 surface reflectance relationships vary by location, season and angle, we performed atmospheric correction on the co-located C004-L-MODIS/AERONET data. Atmospheric correction (Kaufman and Sendra, 1988) attempts to calculate the optical properties of the surface, by theoretically subtracting the effects of the atmosphere from the satellite-observed radiation field. The atmospherically corrected surface reflectance ρ^s_λ is calculated by re-arranging Equation (9). In order to minimize errors, arising from multiple scattering, we limited our exercise to conditions of τ in the green less than 0.2. Out of the original 15,000 co-located MODIS/AERONET points (described in section 2), there are over 10,000 collocations with low τ. The archive contains "gas absorption corrected" MODIS-Level 2 observed reflectance (see Appendix) and AERONET-observed spectral τ,

column water vapor depth. For each MODIS parameter, four statistical parameters are reported, including: 'pval' (value for the central pixel-closest to the sunphotometer), 'npix' (number of valid retrievals within a 5 x 5 box; ≤ 25), 'mean' and 'sdev' (mean and standard deviation within the box). For each AERONET parameter, the analogous statistics are: 'pval' (value for the AERONET retrieval closest in time to MODIS overpass), 'nval' (number of valid retrievals within one hour of overpass; ≤ 5), 'mean' and 'sdev'. For the atmospheric correction, we used the *pval* values of MODIS spectral reflectance, and the *mean* values of AERONET τ and water vapor. The molecular properties of the atmosphere are assumed those of the U.S. standard atmosphere. The sea level Rayleigh optical depth (ROD) values are assumed for each MODIS spectral channel, and scaled according to the elevation/air pressure of the sunphotometer.

The relation between the satellite-measured reflectance and the surface reflectance is a complicated function of the atmospheric effects of scattering and absorption by the aerosol. Previous atmospheric correction exercises often assumed some form of the Continental aerosol model (e.g. Vermote et al., 1997), to describe both the scattering and absorption properties. While this model may provide reasonable simulations in channels near to 0.55 μm (such as 0.47 and 0.66 μm), it cannot be expected to provide accurate simulations at 2.12 μm, even for low τ. For example, for $\tau_{0.55} = 0.2$, $\tau_{2.12}$ ranges from 0.03 to 0.16, depending on which of our aerosol types are assumed. Thus, assuming the wrong aerosol size in the correction procedure will lead to errors in estimating 2.12 μm surface reflectance.

The LUT spectral reflectance is interpolated for geometry and AERONET measured τ, thus estimating TOA reflectance over a black surface (path reflectance). From the AERONET-derived Ångstrom exponent, we can decide whether to assume a fine model or a coarse model. Since ω_0 is not known, the moderately absorbing/developing world aerosol type ($\omega_0 \sim 0.9$) was chosen to represent fine-dominated aerosol. When $\alpha < 0.6$ (400 cases), the correction procedure assumed the coarse-dominated model. Co-locations where $0.6 < \alpha < 1.6$ (about 6000 cases) were not used due to uncertainties of aerosol mixing.

The atmospheric correction resulted in two datasets: surface reflectance at three wavelengths (0.47, 0.66, 2.12 μm) for each of the two regimes (fine and coarse-dominated). Separate comparison of 0.66μm versus 2.12μm and 0.47μm versus 2.12μm, for each regime indicated that their regressions differed by less than 10% (both slope and y-offset values), suggesting to combine the two surface reflectance datasets into one.

Figure 15a plots the entire set of atmospherically corrected visible surface reflectance (in the blue $\rho^s_{0.47}$ and the red $\rho^s_{0.66}$) versus that in the mid-SWIR ($\rho^s_{2.12}$) and their regression lines. While not plotted, also considered were the regressions if they were forced through zero, thereby assuming that zero SWIR reflectance is zero reflectance over the entire spectrum (which would be equivalent to deriving simple ratios). Correlation (R) values are 0.93 for the red, but only about 0.75 for the blue. In the blue, forcing a regression through zero is quite different than that not constrained. If forced through zero, the slope tends toward about 0.36, whereas including the offset (about +0.011) yields a slope closer to the assumed one-quarter (0.258). In the red, whether including offset or not, the slope is about 0.55. Thus in a mean sense, atmospheric correction of MODIS data yields VISvs2.12 surface reflectance relationships that differ from the assumed C004-L VISvs2.12 ratios. Figure 15b shows that fitting blue to red (0.47vs0.66) has higher correlation and less scatter than 0.47vs2.1 directly, suggesting better relationship between the two visible channels. There is less difference between fitting through zero and

not, such that a straight blue/red ratio is about 0.54, and the full regression has slope = 0.508 and offset = 0.008. Therefore, instead of the 0.47 μm and 0.66 μm surface reflectance being calculated separately from 2.12 μm, we calculate the 0.66 μm surface reflectance from that in 2.12 μm, followed by calculating 0.47 μm from the 0.66 μm, i.e.

$$\begin{aligned} \rho^s_{0.66} &= f(\rho^s_{2.12}) \\ \rho^s_{0.47} &= g(\rho^s_{0.66}) \end{aligned}, \qquad (10)$$

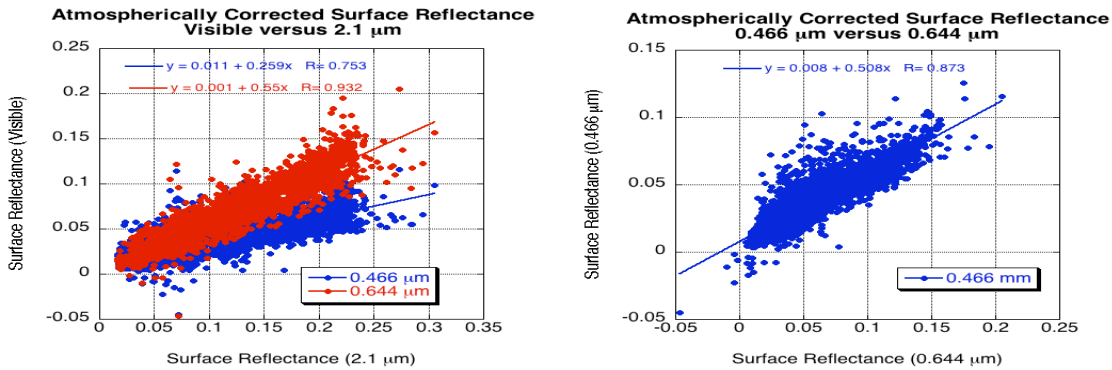

Figure 15: Atmospherically corrected surface reflectance in the visible (0.47 and 0.66 μm channels) compared with that in the 2.12 μm SWIR channel (a), and the 0.47μm compared with that in the 0.66 μm channel (b).

As noted in Figures 15a and 15b the VISvs2.12 surface reflectance regressions display large scatter. For example, if surface reflectance is 0.15 at 2.12 μm, applying the regressed relationships of 0.66vs2.12 and 0.47vs0.66 results in estimates of surface reflectance of 0.083±0.03 at 0.66 μm and 0.050±0.03 at 0.47 μm. Obviously, this could result in very large errors in retrieved τ, on the order of 0.3 or more. Therefore, to reduce the scatter we look for dependencies on other parameters to refine the relationships.

The works of Gatebe et al. (2002) and Remer et al. (2001) suggests that the VISvs2.12 surface reflectance relationships are angle dependent. Therefore, we tested which type of angle (solar zenith angle, sensor zenith angle, glint angle or scattering angle) most affected the VISvs2.12 surface reflectance relationship. The highest correlation of the VISvs2.12 surface variability was found to be with scattering angle Θ, defined as:

$$\Theta = \cos^{-1}(-\cos\theta_0 \cos\theta + \sin\theta_0 \sin\theta \cos\phi) \qquad (11)$$

where θ_0, θ and ϕ are the solar zenith, sensor view zenith and relative azimuth angles, respectively. Fig ure 16 (a) displays the median values of surface reflectance as a function of scattering angle, and shows a definite relationship at 2.12 μm, less at 0.66 μm, and nearly none at 0.47 μm. Since Figures 15a and 15b showed that both a slope and y-offset was necessary to regress VIS to 2.12 μm surface reflectance, we look for scattering angle dependence on both parameters. Fig 16 (b-d) plots the slope, y-offset and correlation of the surface reflectance relationships, as a function of scattering angle. The 0.66vs2.12 regression slope shows dependence on scattering angle, whereas the 0.47vs0.66 regression slope shows nearly none. The regressed y-intercept shows strong dependence on scattering angle for

both relationships. Especially interesting is that the 0.66vs2.12 y-offset goes from positive to negative with increasing scattering angle, with a value of zero near Θ=135°.

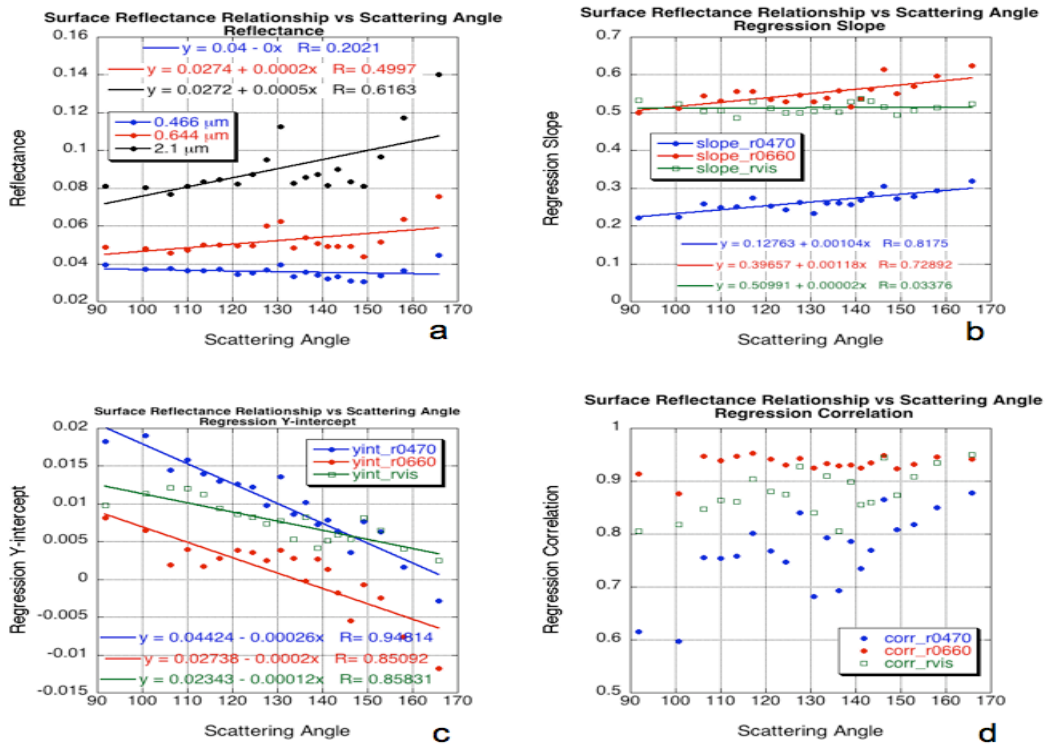

Figure 16: VISvs2.12 surface reflectance relationships as a function of scattering angle. The data were sorted according to scattering angle and put into 20 groups of equal size (about 230 points for each scattering angle bin). On all subplots, each point is plotted for the median value of scattering angle in the bin. Part (a) plots median values of reflectance at each channel as a function of the scattering angle. Linear regression was calculated for the 230 points in each group. The slope of the regression (for each angle bin) is plotted in (b), the y-intercept is plotted in (c) and the regression correlation is plotted in (d). Note for (b), (c) and (d) that 0.47 µm vs 2.12 µm (r0470) is plotted in blue, 0.66 µm vs 2.12 µm (r0660) is plotted in red and 0.47 vs 0.66 µm (rvis) is plotted in green.

Because AERONET sites are located in different surface type regimes, it could be expected that the VISvs2.12 surface relationships will vary based on surface type and/or season. Using the International Geosphere/Biosphere Programme's (IGBP) scene map of USGS surface types and formatted for MODIS validation (http://edcdaac.usgs.gov/modis/mod12c1v4.asp), we determined the scene type of the MODIS/AERONET validation box. We then separated urban from non-urban surfaces, and grouped into season (winter or summer) and general location (mid-latitude or tropical). Figures 17 and 18 display some of the surface reflectance relationships as a function of different regions and locations. Generally, "greener" surfaces (midlatitude summer sites both urban and nonurban) have higher red to SWIR ratios (red/SWIR>0.55) than winter sites (red/SWIR<0.55). Many of the AERONET sites in the tropics are in savanna or grassland regions, where the landscape is not as green, and hence the red to SWIR ratios are also lower. As for the blue to red channel surface reflectance relationships, except for the urban sites during summer (blue/red ratio ~ 0.766), the relationships around the globe are relatively consistent (blue/red ~ 0.52).

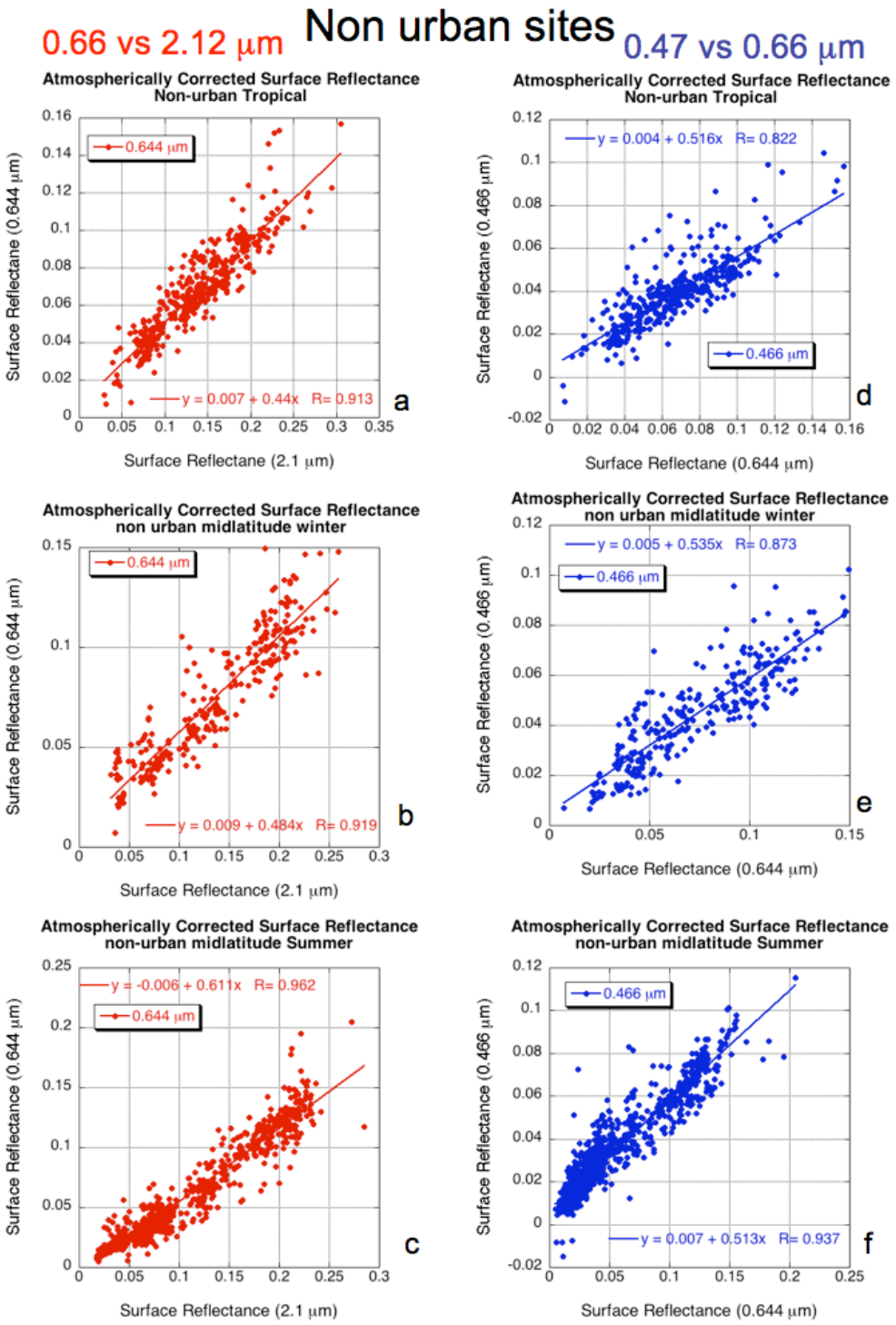

Figure 17: Surface Reflectance relationships for non-urban sites. The left three subfigures (a-c) are for Visible versus 2.12 µm channels, whereas the right three subfigures (d-f) are for 0.47 µm versus 0.66 µm channels. From the top to bottom, subfigures (a) and (d) are for tropical sites, (b) and (e) are for midlatitude sites in winter, and (c) and (f) are for midlatitude sites during summer.

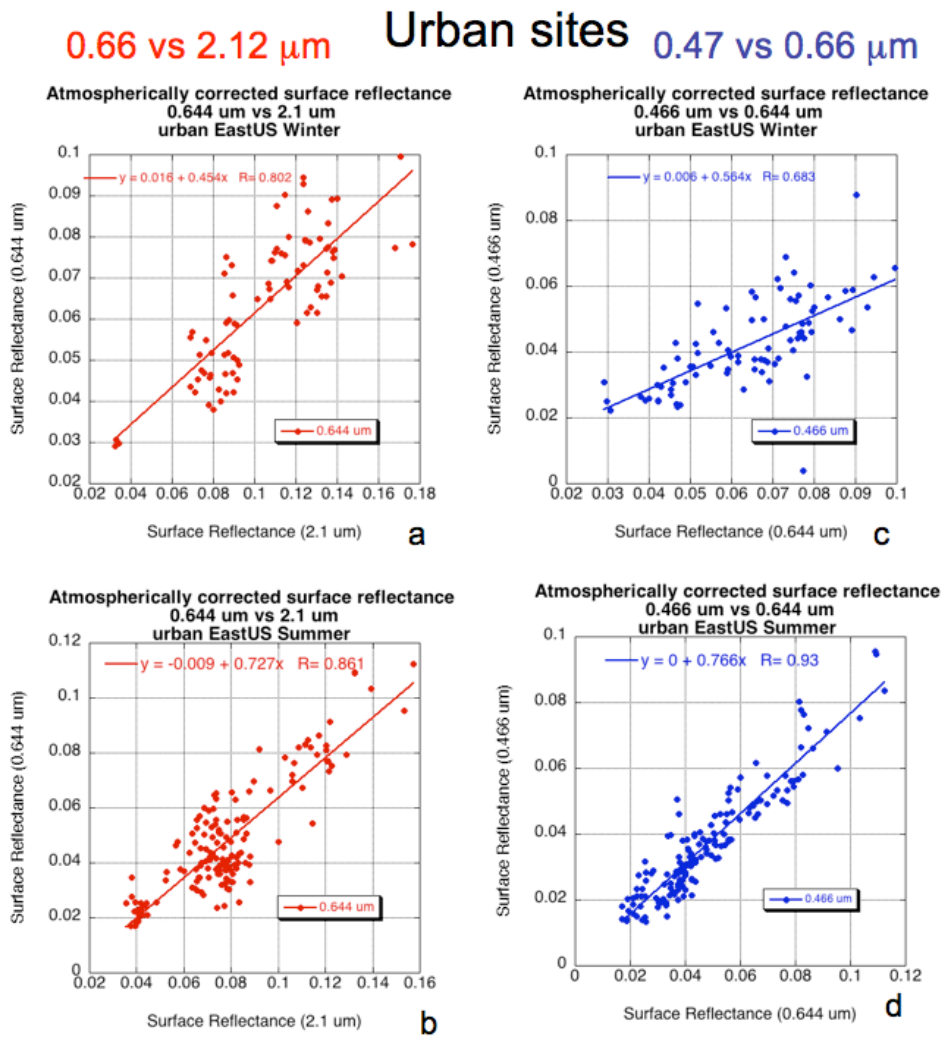

Figure 18: Surface Reflectance relationships for urban sites along the US East Coast. The left two subfigures (a-b) are for 0.66 µm versus 2.12 µm channels, whereas the right two subfigures (c-d) are for 0.47 µm versus 0.66 µm channels. From the top to bottom, subfigures (a) and (c) are for the sites in winter, and (b) and (d) are for the sites during summer.

Except for urban areas, most surfaces seem to have VISvs2.12 surface reflectance relationships that vary as a function of their "greenness." Can we relate the surface reflectance relationships to a vegetation index (VI)? The Normalized Difference Vegetation Index (NDVI), defined as a function of the red (0.66 µm – channel 1) and near-IR (0.86 µm – channel 2), can be heavily influenced by aerosol (Tucker et al., 1979). We attempted to work with other VIs (such as described by Karnieli et al., 2000) that have different sensitivity to atmospheric (aerosol) conditions, and found the most promising to be the $NDVI_{SWIR}$, defined as:

$$NDVI_{SWIR} = (\rho^m_{1.24} - \rho^m_{2.12})/(\rho^m_{1.24} + \rho^m_{2.12}) \qquad (12)$$

where $\rho_{1.24}$ and $\rho_{2.12}$ are the MODIS-measured reflectances of the 1.24 µm channel (MODIS channel 5) and the 2.1 µm channel (channel 7), which are much less influenced by aerosol (except for heavy aerosol or dusts). This index is also known as $NDVI_{MIR}$ (Mid-InfraRed). In aerosol free conditions

$NDVI_{SWIR}$ is highly correlated with regular *NDVI*. A value of $NDVI_{SWIR} > 0.6$ is active vegetation, whereas $NDVI_{SWIR} < 0.2$ is representative of dormant or sparse vegetation. Figure 19 plots the relationship of 0.66 μm channel and 2.12 μm channel (atmospherically corrected) surface reflectance relationship, for nonurban sites, as a function of low, medium and high values of $NDVI_{SWIR}$. Clearly, as the $NDVI_{SWIR}$ increases, the ratio between 0.66μm and 2.12μm surface reflectance increases, and we will use this relationship in the final VISvs2.12 surface reflectance parameterization. Since the 0.47vs0.66 relationship does not strongly vary as a function of $NDVI_{SWIR}$, we assume it to be constant.

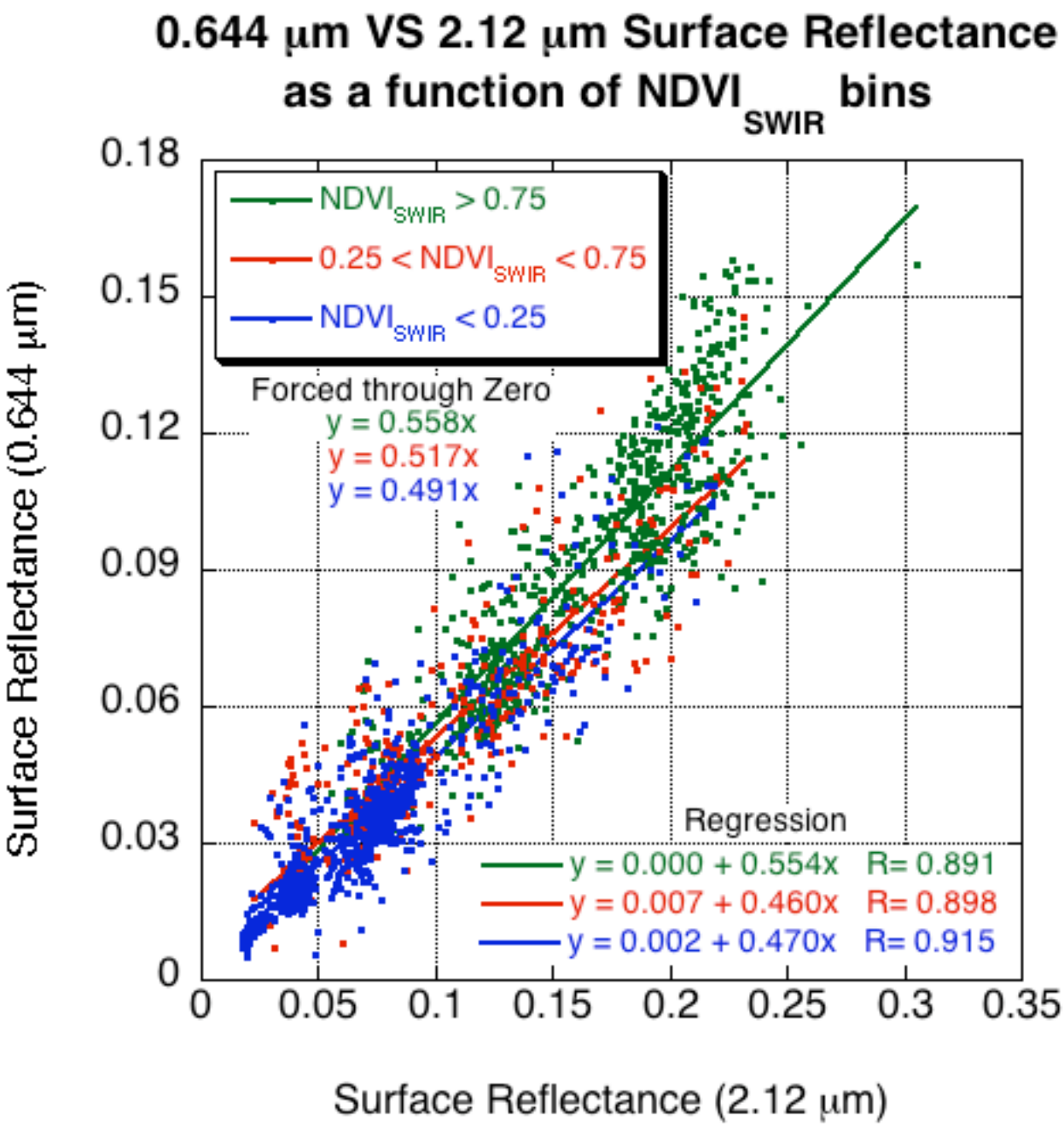

Figure 19: 0.66 μm versus 2.12 μm surface reflectance as a function of bins of $NDVI_{SWIR}$ values (low, medium and

high). Both standard regression and "forced through zero" are plotted.

Results of the global atmospheric correction exercise imply that not only do the VISvs2.12 surface relationships differ from the C004-L VISvs2.12 ratios, they also have a strong dependence on both geometry and surface type. The C005-L VISvs2.12 surface reflectance relationship is parameterized as a function of both $NDVI_{SWIR}$ and scattering angle Θ, such that Equation (10) can be expanded:

$$\rho^s_{0.66} = f(\rho^s_{2.12}) = \rho^s_{2.12} * slope_{0.66/2.12} + yint_{0.66/2.12}$$

and (13)

$$\rho^s_{0.47} = g(\rho^s_{0.66}) = \rho^s_{0.66} * slope_{0.47/0.66} + yint_{0.47/0.66}$$

where

$$\begin{aligned} slope_{0.66/2.12} &= slope^{NDVI_{SWIR}}_{0.66/2.12} + 0.002\Theta - 0.27, \\ yint_{0.66/2.12} &= -0.00025\Theta + 0.033, \\ slope_{0.47/0.66} &= 0.49, \text{and} \\ yint_{0.47/0.66} &= 0.005 \end{aligned} \tag{14}$$

where in turn

$$\begin{aligned} slope^{NDVI_{SWIR}}_{0.66/2.12} &= 0.48; NDVI_{SWIR} < 0.25, \\ slope^{NDVI_{SWIR}}_{0.66/2.12} &= 0.58; NDVI_{SWIR} > 0.75 \\ slope^{NDVI_{SWIR}}_{0.66/2.12} &= 0.48 + 0.2(NDVI_{SWIR} - 0.25); 0.25 \le NDVI_{SWIR} \le 0.75 \end{aligned} \tag{15}$$

Note that while the above parameterization was based on the results of Figures 15, 16 and 19, the coefficients are not identical to those in the figures. The atmospheric corrected data set is the broadest and most comprehensive representation of global surface reflectance relationships, still it is limited to AERONET site locations, which are in turn are most concentrated in certain geographical regions. Trial and error was used to modify the basic results from the AERONET-based atmospheric correction, to give more realistic MODIS retrievals globally, (especially in places were few or no AERONET sites are located).

4.5. Retrieval Algorithm

A major limitation of C004-L was that aerosol is assumed transparent in the 2.12 μm SWIR channel. The surface reflectance in 2.1 μm was assumed to be exactly the value of the observed TOA reflectance in that channel. Under a dust aerosol regime, aerosol transparently is an extremely poor assumption. Even in a fine aerosol dominated regime, τ is not zero. For example, for our 'moderately absorbing/developing world' ($\omega_0 \sim 0.9$) aerosol, $\tau_{0.55}$ of 1.0 corresponds to $\tau_{2.12}$ of 0.114. For a given angle (say $\theta_0 = 36°$, $\theta = 36°$, and $\phi = 72°$) assuming $\tau_{2.12} = 0.0$ instead leads to error in 2.12 μm path reflectance of about 0.012. Via the VISvs2.12 reflectance relationship, the reflectance error at 0.66 μm would be on the order of 0.006, leading to ~ 0.06 error in retrieved τ. As a percentage of the actual τ, the error is not very large. However, combined with errors at 0.47 μm, the resulting incorrect Ångstrom exponent leads to error in estimating η.

In the spirit of the MODIS aerosol over ocean algorithm (Tanré et al., 1997; Levy et al., 2003; Remer et al., 2005), we developed a multi-channel reflectance inversion for retrieving aerosol properties over land, known as the "second-generation algorithm" (Levy et al., 2007b). Analogous to the ocean

algorithm's combination of fine and coarse aerosol *modes*, the land algorithm attempts to combine fine-dominated and coarse-dominated aerosol *models* (each composed of multiple modes) to match with the observed spectral reflectance. The 2.1 μm channel is assumed to contain both surface and aerosol information, and the visible surface reflectance is a function of the parameterized VISvs2.12 surface reflectance relationships. Simultaneously inverting the aerosol and surface information in the three channels (0.47 μm, 0.66 μm and 2.12 μm) yields three parameters: τ ($\tau_{0.55}$), the η ($\eta_{0.55}$) and the surface reflectance ($\rho^s_{2.12}$).

We rewrite Equation 8, but note that the calculated spectral total reflectance ρ^*_λ at the top of the atmosphere is the weighted sum (η) of the spectral reflectance from a combination of fine and coarse – dominated aerosol models, i.e.

$$\rho^*_\lambda = \eta\rho^{*f}_\lambda + (1-\eta)\rho^{*c}_\lambda \qquad (16)$$

where ρ^{*f}_λ and ρ^{*c}_λ are each composites of surface reflectance ρ^s_λ and atmospheric path reflectance of the separate aerosol models. That is:

$$\rho^{*f}_\lambda = \rho^{af}_\lambda + F^f_{d\lambda}T^f_\lambda\rho^s_\lambda/(1-s^f_\lambda\rho^s_\lambda)$$
$$\text{and} \qquad (17)$$
$$\rho^{*c}_\lambda = \rho^{ac}_\lambda + F^c_{d\lambda}T^c_\lambda\rho^s_\lambda/(1-s^c_\lambda\rho^s_\lambda)$$

where ρ^{af}_λ and ρ^{ac}_λ are the fine and coarse model atmospheric path reflectance, $F^f_{d\lambda}$ and $F^c_{d\lambda}$ are normalized downward fluxes for zero surface reflectance, T^f_λ and T^c_λ represent upward total transmission into the satellite field of view, and s^f_λ and s^c_λ are atmospheric backscattering ratios. Note the angular and τ dependence of some of the terms: $\rho^a=\rho^a(\tau, \theta_0,\theta,\phi)$, $F=F(\tau,\theta_0)$, $T=T(\tau,\theta)$, $s = s(\tau)$ and $\rho^s=\rho^s(\theta_0,\theta,\phi)$. Whereas the other terms are a function of the aerosol and are contained within the lookup tables, the surface reflectance is independent. However, we know the VISvs2.12 surface reflectance relationships.

Due to the limited set of aerosol optical properties in the lookup table, the equations may not have exact solutions, and solutions may not be unique. Therefore, we find the aerosol solution most closely resembling the set of MODIS measured reflectance. In order to reduce the possibility of non-unique retrievals we only allow discrete values of η. During the retrieval, the algorithm tests whether certain criteria are met for consistency and valid retrieval steps. Results of these are encoded into a product called the 'Quality_Assurance_Land'. Upon completion, the retrieval is assigned a final Quality Assurance confidence (QAC) value that ranges from 0 (bad quality) to 3 (good quality). Details of the QA and QAC are given in the Appendix. In the following subsections, we describe the mechanics of the inversion algorithm in more detail.

Selection of "dark pixels"

Figure 20 illustrates the main steps of the C005-L land algorithm. The relevant Level 1 B (L1B) data include calibrated spectral reflectance in eight wavelength bands at a variety of spatial resolutions, as well as the associated geo-location information. The spectral data include the 0.66 and 0.86 μm channels (MODIS channels 1 and 2 at 250 m resolution), the 0.47, 0.55, 1.2, 1.6 and 2.1 μm channels

(channels 3, 4, 5, 6 and 7 at 500 m), and the 1.38 µm channel (channel 26 at 1 km). The geo-location data are at 1 km and include angles (θ,θ,ϕ, and Θ), latitude, longitude, elevation and date. The L1B reflectance values are corrected for water vapor, ozone, and carbon dioxide (described in Appendix I) before proceeding.

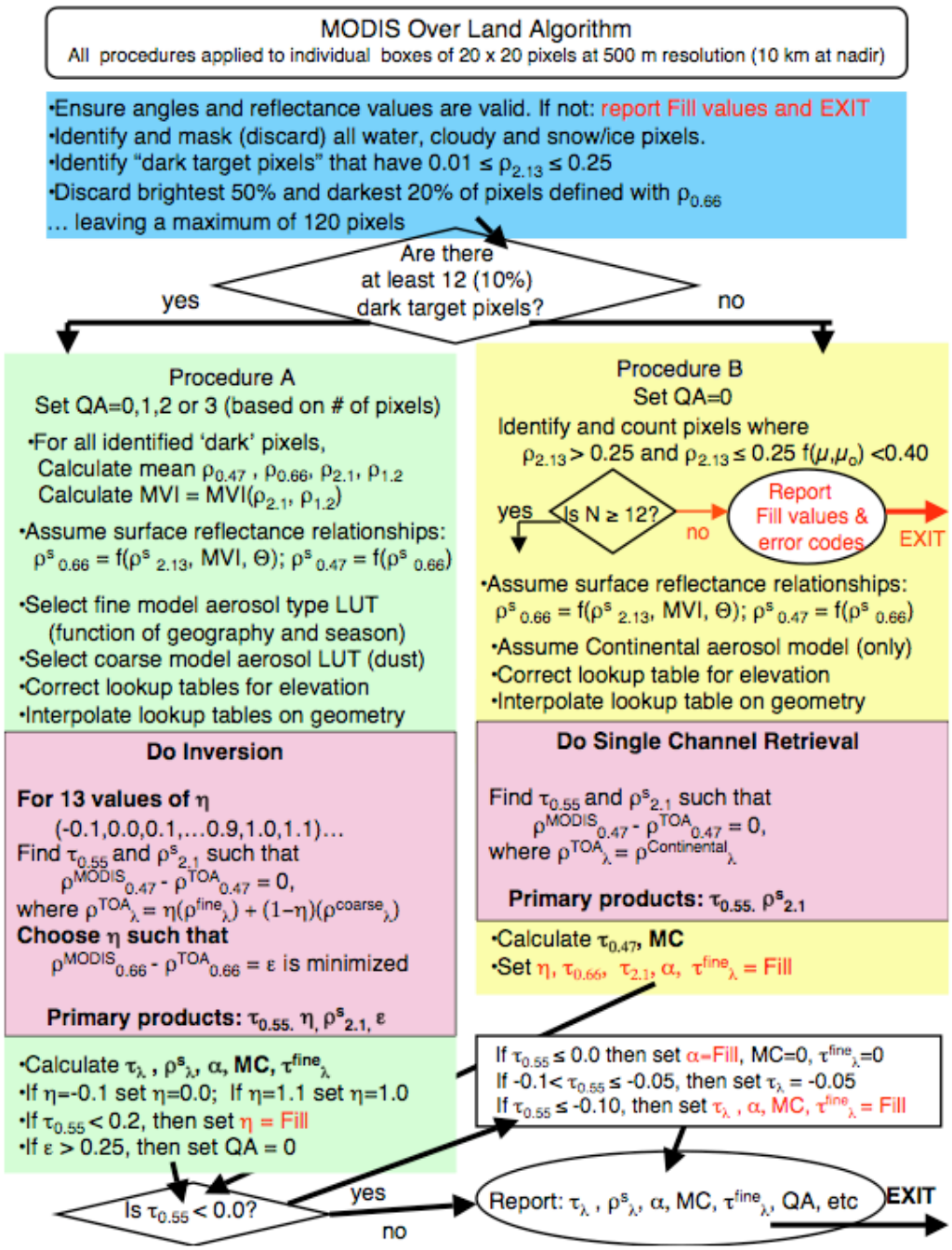

Figure 20: Flowchart illustrating the derivation of aerosol over land for C005-L.

The first step is to organize the measured reflectance into nominal (at nadir) 10 km by 10 km boxes (corresponding to 20 by 20, or 40 by 40 pixels, depending on the channel). The 400 pixels in the box are evaluated pixel by pixel to identify whether the pixel is suitable for aerosol retrieval. Clouds (Martins et al., 2002), snow/ice (Li et al., 2004) and inland water bodies (via NDVI tests) are considered not suitable and are discarded. Details of masking are also described in Appendix I.

The non-masked pixels are checked for their brightness. Pixels having 2.12 μm measured reflectance between 0.01 and 0.25 are grouped and sorted. The brightest 50% and darkest 20% are discarded, in order to reduce cloud and surface contamination and scale towards darker targets. If there are at least 12 pixels remaining (10% of 30% of the original 400), then the reflectance in each channel is averaged, yielding the 'MODIS-measured' spectral reflectance $\rho^m_{0.47}$, $\rho^m_{0.66}$, $\rho^m_{2.12}$, and $\rho^m_{1.24}$. These reflectance values are used for 'Procedure A'. If less then 12 pixels remain, then 'Procedure B' (described later) is followed and QAC is lowered to zero.

Correcting the LUT for elevation

A major change from C004-L concerns how the algorithm corrects for elevated surface targets. The sea-level Rayleigh optical depth (*ROD*, $\tau_{R,\lambda}$) at a wavelength λ (in μm) can be approximated over the visible range (e.g. Dutton et al., 1994; Bodhaine et al., 1998) by:

$$\tau_{R,\lambda} = 0.00877\lambda^{-4.05} \qquad (18)$$

When not at sea level (pressure = 1013 mb), the ROD is a function of pressure (or height, *z*) so that it can be approximated by:

$$\tau_{R,\lambda}(z = Z) = \tau_{R,\lambda}(z = 0)\exp(\frac{-Z}{8.5}) \qquad (19)$$

where Z is the height (in kilometers) of the surface target and 8.5 km is the exponential 'scale height' of the atmosphere. The difference between ROD at z=0 and z=Z is $\Delta\tau_{R,\lambda}$.

In C004-L, the algorithm (too) simply corrected the retrieved τ product by adding the optical depth that was neglected by assuming sea level for the retrieval, (i.e. $\tau_\lambda(z = Z) = \tau_\lambda(z = 0) + \Delta\tau_{R,\lambda}$). However, this correction can give poor results because of the large differences between molecular and aerosol phase functions.

Instead, the C005-L algorithm makes use of the procedure described in Fraser et al., (1989). The algorithm adjusts the lookup table to simulate different ROD by adjusting the wavelength. Substitution of Equation (18) into equation (19) yields

$$\lambda(z = Z) = \lambda(z = 0)\exp(\frac{Z}{34}). \qquad (20)$$

For example, at Z = 0.4 km, λ increases by about 1.2%. For the blue 0.47 μm channel, (centered at 0.466 μm) this means that $\tau_{R,\lambda}(z=0)=0.194$, $\tau_{R,\lambda}(z=0.4)=0.185$ and $\lambda(z=0.4)=0.471$ μm. In other words, the algorithm simulates an elevated surface by adjusting the blue channel's wavelength to 0.471 μm. Assuming that gases and aerosols are optically well mixed in altitude, the algorithm substitutes for the parameter values of the 0.47 μm LUT by interpolating (linearly as functions of log wavelength and log parameter) between the 0.47 μm (0.466 μm) and the 0.55 μm (0.553 μm) entries. Similar interpolations are performed for the other channels (for example, 0.55 μm would be adjusted to 0.559 μm). For the 0.4 km case, this means that lower values of TOA atmospheric path reflectance and higher values of transmission are chosen to represent a given aerosol model's optical contribution. However, also note that since the 0.55 μm channel has also been adjusted, the associated values of the τ indices have been adjusted accordingly.

Whereas most global land surfaces are at sea level or above, a few locations are below sea level (Z < 0). In these cases, the algorithm is allowed to extrapolate below 0.466 μm. Since the extrapolation is at most for a hundred meters or so, this is not expected to introduce large errors, and these cases can still be retrieved. Note also that due to the extremely low ROD in the 2.12 μm channel, little is gained by adjusting this channel.

Procedure A: Inversion for dark surfaces

If following Procedure A (for dark surfaces), the QA confidence (QAC) is initially set to a value between 0 ('bad quality') and 3 ('good quality'), depending on the number of dark pixels remaining. In Procedure A, the algorithm assigns the fine aerosol model, based on the location and time (Figure 10). From the lookup table, ρ^a, F, T and s (for the fine model and coarse model separately) are interpolated for angle, resulting in six values for each parameter, corresponding to aerosol loading (indexed by τ at 0.55 μm).

The 2.12 μm path reflectance is a non-negligible function of the τ, so that the surface reflectance is therefore also a function of the τ. For discrete values of η between -0.1 and 1.1 (intervals of 0.1), the algorithm attempts to find the τ at 0.55 μm and the surface reflectance at 2.12 μm that exactly matches the MODIS measured reflectance at 0.47 μm. There will be some error, ε, at 0.66 μm. The solution is the one where the error at 0.66 μm is minimized. In other words,

$$\begin{aligned} \rho^m_{0.47} - \rho^*_{0.47} &= 0 \\ \rho^m_{0.66} - \rho^*_{0.66} &= \varepsilon \\ \rho^m_{2.12} - \rho^*_{2.12} &= 0 \end{aligned} \qquad \text{(21abc)}$$

where

$$\rho^*_{2.12} = \eta(\rho^{fa}_{2.12} + F^f_{d\,2.12}T^f_{2.12}\rho^f_{2.12}/(1-s^f_{2.12}\rho^s_{2.12})) + (1-\eta)(\rho^{ca}_{2.12} + F^c_{d\,2.12}T^c_{2.12}\rho^s_{2.12}/(1-s^c_{2.12}\rho^s_{2.12}))$$

$$\rho^*_{0.66} = \eta(\rho^{fa}_{0.66} + F^f_{d0.66}T^f_{0.66}f(\rho^s_{2.12})/(1-s^f_{0.66}f(\rho^s_{2.12}))) + (1-\eta)(\rho^{ca}_{0.66} + F^c_{d\,0.66}T^c_{0.66}f(\rho^s_{2.12})/(1-s^c_{0.66}f(\rho^s_{2.12}))) \quad \text{(22a, b, c),}$$

and

$$\rho^*_{0.47} = \eta(\rho^{fa}_{0.47} + F^f_{d0.47}T^f_{0.47}g(\rho^s_{0.66})/(1-s^f_{0.47}g(\rho^s_{0.66}))) + (1-\eta)(\rho^{ca}_{0.47} + F^c_{d\,0.47}T^c_{0.47}g(\rho^s_{0.66})/(1-s^c_{0.47}g(\rho^s_{0.66}))),$$

where in turn, $\rho^a=\rho^a(\tau)$, $F=F(\tau)$, $T=T(\tau)$, $s = s(\tau)$ are functions of τ indices in the lookup table, and $f(\rho^s_{2.12})$, $g(\rho^s_{0.66})$ are described by Equations (13-15). Note that non-physical values of η are tried (1.1

and -0.1) to allow for the possibility of inappropriate assumptions in either aerosol models or surface reflectance. Again, the primary products are τ ($\tau_{0.55}$), η ($\eta_{0.55}$), and the surface reflectance ($\rho^s_{2.12}$). The error ε is also noted.

Procedure B: Alternative Retrieval for Brighter surfaces

The derivation of aerosol properties is possible when the 2.12 μm reflectance is brighter than 0.25, but is expected to be less accurate (Remer et al., 2005), due to increasing errors in the VISvs2.12 relationship. However, if Procedure A was not possible, and if there are at least 12 cloud-screened, non-water pixels, satisfying

$$0.25 < \rho^m_{2.12} < 0.25G < 0.40 \qquad (23)$$

where

$$G = 0.5((1/\mu) + (1/\sqrt{\mu_0})) \ , \qquad (24)$$

then Procedure B is attempted. In this case, QAC is automatically set to 0 ('bad quality').

Procedure B is analogous to 'Path B' described in Remer et al., (2005). Like in C004-L, the Continental aerosol model is assumed. Unlike C004-L, the VISvs2.12 surface reflectance assumptions are those described by Equations (13-15) and the Continental aerosol properties are indexed to 0.55 μm. In other words, it uses equations (16-17), except with the first term only (i.e. η = 1.0). The primary products for Procedure B are τ ($\tau_{0.55}$) and the surface reflectance ($\rho^s_{2.12}$). The 'land fitting error' ε is also saved.

Derivation of Fine Mode τ, Mass Concentration and other secondary parameters

Following the derivation of primary products by Procedure A ($\tau_{0.55}$, $\eta_{0.55}$ and $\rho^s_{2.12}$), a number of secondary products can also be calculated. These include the fine and coarse model optical depths $\tau^f_{0.55}$ and $\tau^c_{0.55}$:

$$\tau^f_{0.55} = \tau_{0.55}\eta_{0.55} \quad \text{and} \quad \tau^c_{0.55} = \tau_{0.55}(1-\eta_{0.55}) \qquad (25)$$

the 'mass concentration', *M:*

$$M = M^f_c \tau^f_{0.55} + M^c_c \tau^c_{0.55} \qquad (26)$$

the spectral total, fine and coarse model optical thicknesses τ_λ, τ^f_λ and τ^c_λ:

$$\tau_\lambda = \tau^f_\lambda + \tau^c_\lambda$$

where (27)

$$\tau^f_\lambda = \tau^f_{0.55}(Q^f_\lambda / Q^f_{0.55}) \quad \text{and} \quad \tau^c_\lambda = \tau^c_{0.55}(Q^c_\lambda / Q^c_{0.55})$$

the Ångstrom Exponent α:

$$\alpha = \ln(\tau_{0.47}/\tau_{0.66})/\ln(0.466/0.644) \qquad (28)$$

and the spectral surface reflectance ρ^s_λ, computed by re-arranging Equations (13-15). M^f_c and M^c_c are mass concentration coefficients for the fine and coarse models, whereas Q^f_λ and Q^c_λ represent model extinction coefficients at wavelength, λ. See Appendix 3 for derivation of the extinction coefficients. If the products are inconsistent, then the QAC value initially assigned to the pixel is changed to 0 ('bad quality'), and only a subset of the products is reported.

If Procedure B was followed, the only secondary products calculated are M and $\tau_{0.47}$, and the QAC is set to 0. The other products are undefined.

Negative (and very low) optical depth retrievals

A major philosophical change from C004-L to C005-L is that negative τ retrievals are allowed. Given that there is both positive and negative noise in the MODIS observations, and that surface reflectance and aerosol properties may be under or over-estimated depending on the retrieval conditions, it is statistically useful to allow retrieval of negative τ. In fact it is necessary for creating an unbiased dataset from any instrument. Without negative retrievals the τ dataset is biased by definition.

The trick is to determine the cutoff between a retrieved τ that is "zero, plus or minus," and a retrieved τ that is truly wrong. Since we assume that MODIS should retrieve between the expected error defined by Equation (2) ($\pm0.05\pm0.15\tau$) for very clean conditions when $\tau\sim0$ there is essentially no difference between a retrieval of -0.05, 0 or +0.05. All negative values -0.05 to 0 are reported with QAC rated 'good' (QAC=3). Retrievals in the range -0.10 to -0.05 are reported as -0.05 and the QAC value is lowered. Retrievals less than -0.10 are regarded as 'out of range' and are not reported. Other products that are retrieved or derived (such as the η or Ångstrom Exponent) are set to zero or reported as not defined when the retrieved τ is negative.

In case low τ is retrieved ($\tau < 0.2$), the η is too unstable to be retrieved with any accuracy. Therefore, η is reported as un-defined even though other parameters (such as Ångstrom exponent and Fine τ) may be reported.

4.6. Sensitivity Study

Following the lead of Tanré et al (1997), we have tested the sensitivity of Procedure A by applying it for the following exercises: (1) simulation of conditions that are included within the LUT, (2) simulations where one of the parameters (i.e. τ) is not included within the LUT, and (3) simulations for conditions that include one or more errors.

Whereas the study of Tanré et al, (1997) tested the algorithm on a single geometrical combination, we performed the study in (1) by simulating the 720 reasonable geometrical combinations in the LUT ($0°\leq\phi\leq180°$, $\theta\leq60°$, $\theta_0\leq48°$). We assumed the "fine" aerosol model to be the moderately absorbing ($\omega_0\sim0.9$) aerosol model and that the "coarse" model was our Spheroid (dust) model. For each combination of geometry, and for each MODIS channel, we extracted the fine and coarse mode values of atmospheric path reflectance ρ^a_λ, backscattering ratio s_λ, downward flux F_d and transmission T_λ. We assumed that the 2.1 μm surface reflectance $\rho^s_{2.12}$ = 0.15, and the C004-L VISvs2.12 surface

reflectance ratios (i.e, $\rho^s_{0.66}$= 0.5 $\rho^s_{2.12}$ and $\rho^s_{0.47}$= 0.5$\rho^s_{0.66}$). Using Equations (16-17), we simulated TOA reflectance ρ^*_λ for 5 discrete values of the η (η = 0.0, 0.25, 0.5, 0.75 and 1.0). Therefore, for each value of τ in the LUT, there are 720 x 5 = 3600 attempts to retrieve that τ.

Figure 21 plots the mean (open circles) and standard deviation (error bars) of retrieved τ, as a function of the input τ. For smaller τ (τ ≤ 1), the τ was retrieved nearly perfectly. As τ increases, however, computational instabilities lead to a less exact solution. Still, though, the retrieved τ is certainly within 10%. It is interesting that the algorithm tends to over-estimate τ. Figure 22 shows attempted retrievals of η when the τ is held constant at τ = 0.5 (for all 720 geometrical combinations). η of 0.0 and 1.0 can be considered to be exactly within the LUT, and are retrieved nearly perfectly.

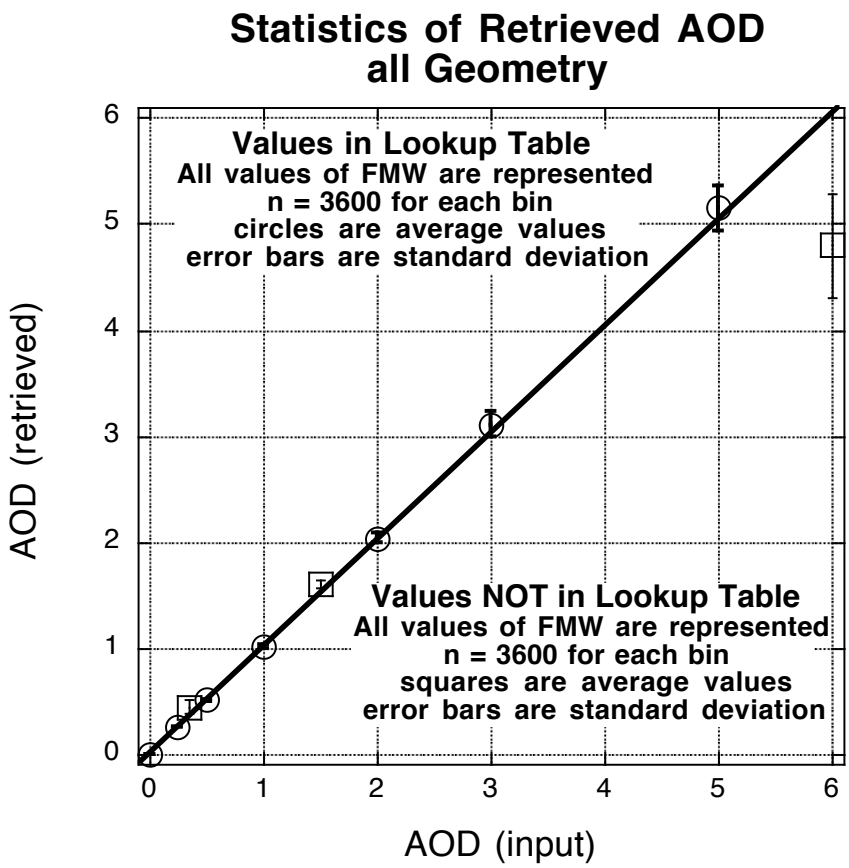

Figure 21: Scatter diagram of the retrieved τ (AOD) versus the input τ. The open shapes correspond to the average of the solutions for each inputted τ, where there are 3600 combinations of geometry and input η (FMW in the figure). The error bars are the standard deviation of the 3600 inputs. The open circles represent cases where the inputted τ are included in the C005-L lookup table, whereas the open squares represent inputted τ that are not in the LUT.

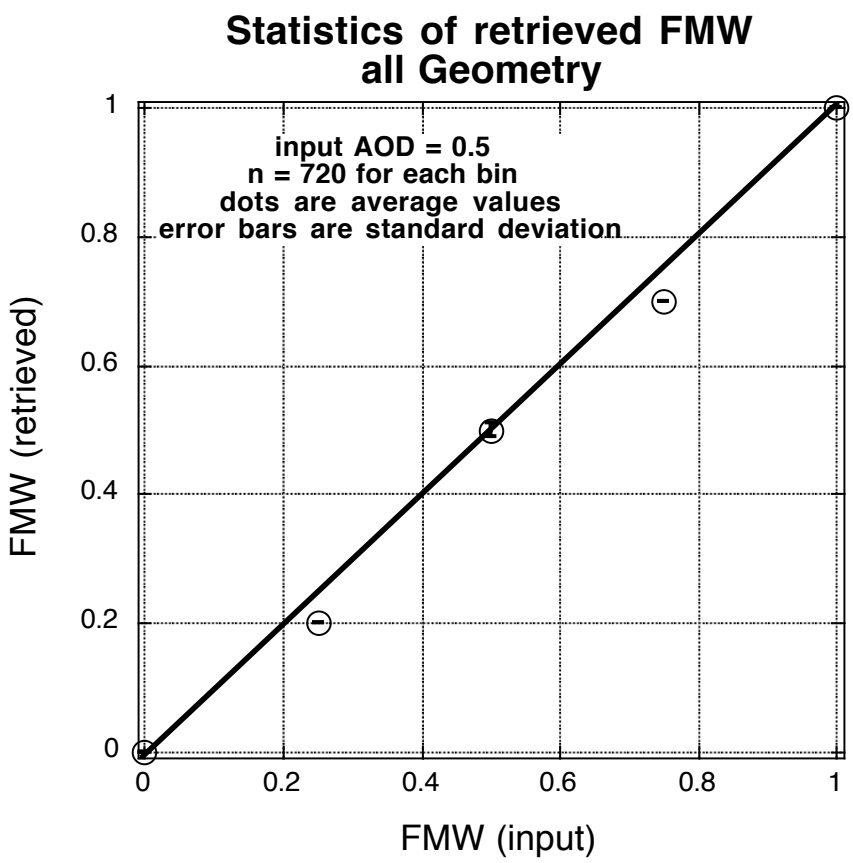

Figure 22: Scatter diagram of the retrieved η (FMW) versus the input η. The open shapes correspond to the average of the solutions for each inputted η, where there are 720 combinations of geometry and input τ (AOD). The error bars are the standard deviation of the 720 inputs. Only the η values of 0.0 and 1.0 would be represented exactly in the LUT, as all coarse and all fine model, respectively. Note the cases for which input η=0.25 and η=0.75, where the algorithm is only allowed to determine η in 0.10 increments.

Figures 23 and 24 provide another way of assessing the retrieved MODIS products. Fig 23 plots retrieved τ, surface reflectance and fitting error as a function of either air mass (top) and scattering angle (bottom), given that the input conditions are $\tau_{0.55}$=0.5, η=0.5 and $\rho^{s}_{2.12}$=0.15. In this case, we plotted all of the 720 geometrical combinations in the LUT. It is interesting that retrieval never exactly matches the input reflectances, although the errors are very small (less than 0.1%). What is even more interesting is how the retrieval uses an under-estimated surface reflectance to balance the over-estimated optical depth. Fortunately, though, most errors are small, and are well within any expected error bars. Figure 24 is similar, but for η=0.25, and plotted only for the air mass dependence.

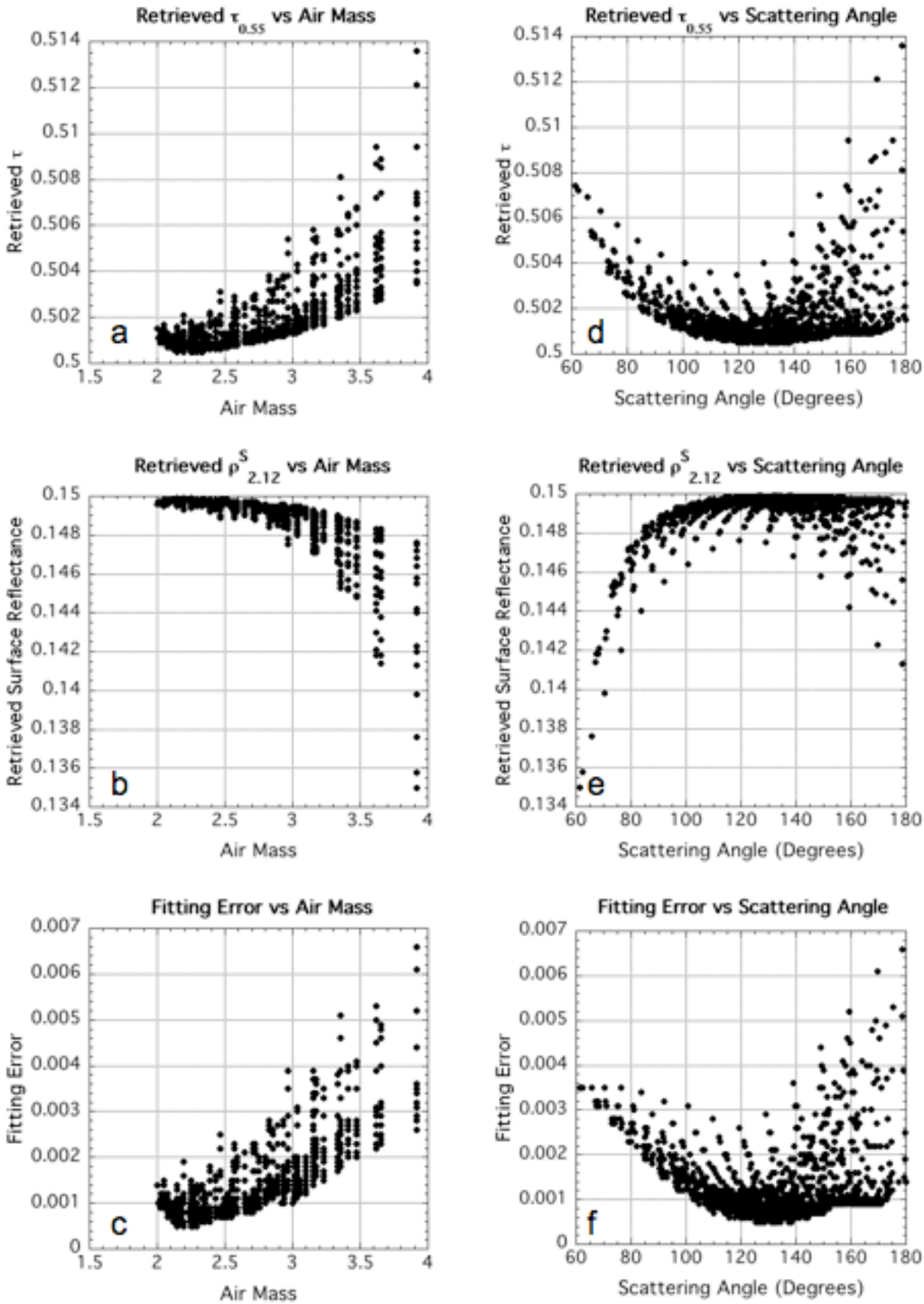
Retrieved $\tau_{0.55}$ vs Air Mass
Retrieved τ
Air Mass
a
Retrieved $\tau_{0.55}$ vs Scattering Angle
Retrieved τ
Scattering Angle (Degrees)
d
Retrieved $\rho^S_{2.12}$ vs Air Mass
Retrieved Surface Reflectance
Air Mass
b
Retrieved $\rho^S_{2.12}$ vs Scattering Angle
Retrieved Surface Reflectance
Scattering Angle (Degrees)
e
Fitting Error vs Air Mass
Fitting Error
Air Mass
c
Fitting Error vs Scattering Angle
Fitting Error
Scattering Angle (Degrees)
f

Figure 23: Retrieved MODIS products as a function of Air Mass (a-c) and Scattering Angle (d-f) for given atmospheric conditions (τ=0.5, η=0.5 and $\rho^s_{2.12}$=0.15) and 720 LUT geometrical combinations. The retrieved τ is plotted in (a) and (d), the 2.12 surface reflectance in (b) and (e) and the fitting error is plotted in (c) and (f). Note that in all cases, the η value of 0.5 was retrieved exactly.

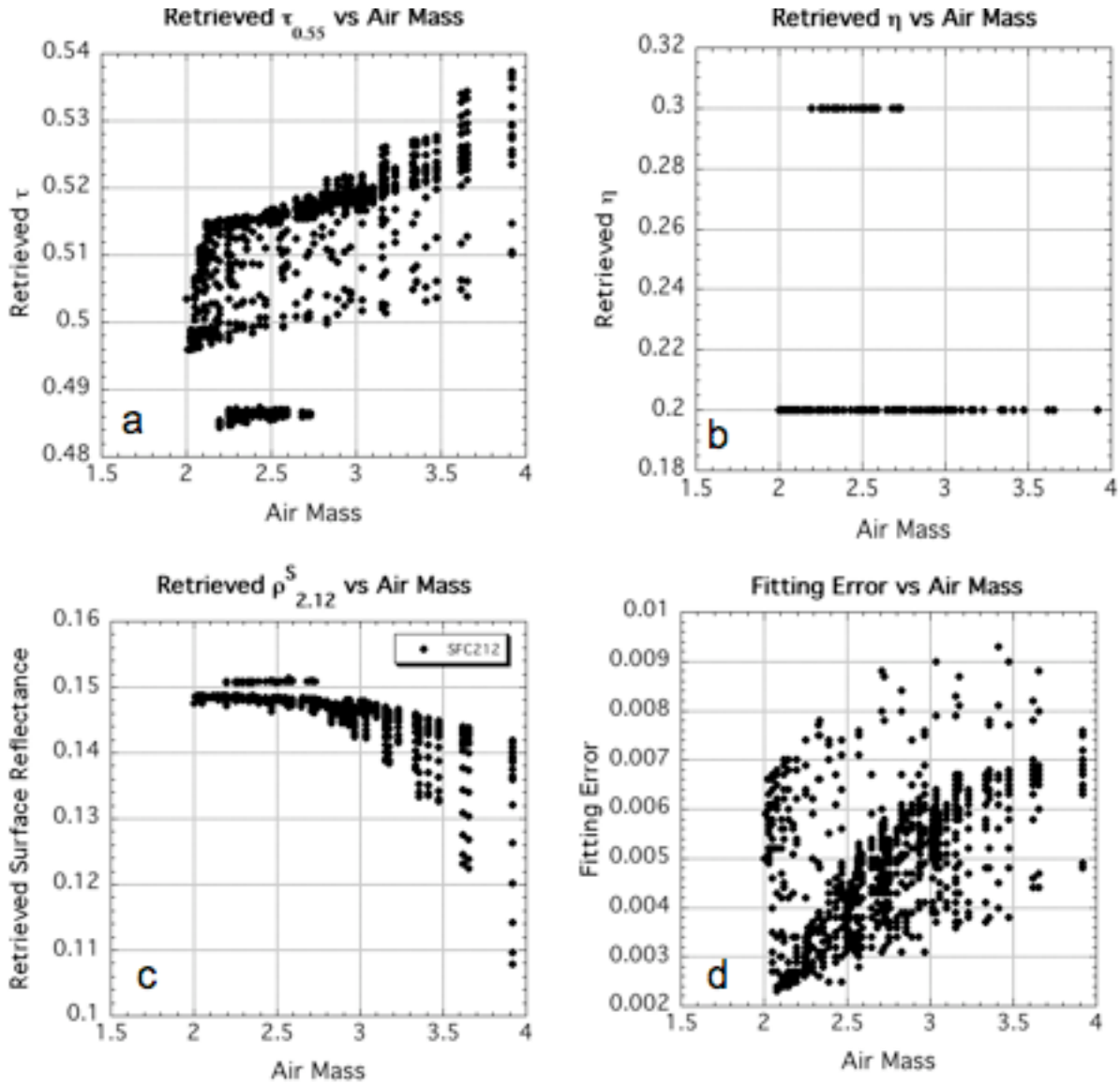

Figure 24: Retrieved MODIS products as a function of Air Mass for given atmospheric conditions (τ=0.5, η=0.25 and $\rho^s_{2.12}$=0.15) and 720 LUT geometrical combinations. The retrieved τ is plotted in (a), η in (b) the 2.12 μm surface reflectance in (c) and the fitting error is plotted in (d). Errors are much larger (up to 1%), but τ is still well within expected error.

(2) We used the same combination of radiative transfer codes (MIEV + RT3) used for the LUT to simulate additional values of aerosol loading ($\tau_{0.55}$ = 0.35, 1.5 and 6.0) to create an "extended" LUT.

As in exercise (1) we simulated the same 720 geometrical combinations as in the C005-L LUT and the five values of η.

Also plotted in Fig 21 are the average values (open squares) and standard deviation (error bars) of the retrieved τ for conditions of this extended LUT. On average the retrieval is very close to the expected value, however, the standard deviation over all geometry is larger than for τ in the normal LUT. A notable exception is the attempt at retrieving $\tau_{0.55}$ = 6.0, where the algorithm does a poor job of extrapolating. In the operational algorithm, we constrain the maximum possible τ to be 5.0. As for retrieving η values not included in the LUT, Figure 22 demonstrates that the algorithm is successful. The η=0.5 retrieval is well behaved. The attempt at resolving either η=0.25 or η=0.75 leads to retrieving η=0.20 and η=0.70. Although it is impossible for an exact retrieval, due to the algorithm choosing between 0.1 intervals, it is interesting that no retrievals of η=0.30 or η = 0.80 are produced.

(3) This exercise studied the impact of different types of errors that could creep into the retrieval process. Potential errors include (but are not limited to) random, systematic or spectrally dependent errors that arise from issues like sensor calibration, assuming the wrong aerosol model at a given location, coarse input topography mapping, or wrong estimates of the VISvs2.12 surface reflectance relationships. These errors are expressed by adding random or systematic errors in the measurements of one or more spectral channels, geometrical conditions or other input boundary conditions. Table 6 lists some prescribed errors, whereas Table 7 shows results when attempting to retrieve conditions of $\tau_{0.55}$=0.5, η=0.5 and $\rho^{s}_{2.12}$=0.15, for the eight sample geometries described in Table 5. Under most conditions, introducing minor calibration or random errors does not kill the retrieval. Even when introducing geo-location errors (angle or assumed aerosol model), the retrieved τ is within expected error bars. As expected, introducing multiple errors to the retrieval leads to poorer retrievals. However, the combination of all sensitivity tests shows that the C005-L algorithm is generally stable.

TABLE 6: LIST OF PRESCRIBED ERRORS FOR C005-LAND SENSITIVITY STUDY

Reference	Error Name	Description
1	LUTinput	LUT input: Use the LUT with no prescribed errors
2	ModError	Aerosol model error: We tried to retrieve with the Non-absorbing fine model LUT
3	RndError	Random Error: All channels have random reflectance error of up to ±0.002
4	SfcError	Surface Error: 10% error in assumed 0.66/2.12 surface reflectance relationship
5	CalError	Calibration Error: All channels have random error of up to ±1%
6	ElvError	Elevation Error: Elevation is 1km instead of assumed sea level
7	GeoError	Geometry Error: All angles have random error of up to ±5 degrees
8	AllError	Combination of 2,3,4,5,6 and 7.

TABLE 7: RESULTS OF SENSITIVITY STUDY USING PRESCRIBED ERRORS

Geometry Error Name	LUTinput	RndError	CalError	GeoError	ModError	ElvError	SfcError	AllError
A	0.501	0.4786	0.5242	0.5143	0.5015	0.6068	0.5402	0.6963
B	0.501	0.4887	0.5242	0.4977	0.4993	0.6035	0.5422	0.6677
C	0.501	0.5227	0.5227	0.4657	0.4835	0.5104	0.4955	0.4809
D	0.5011	0.5104	0.4995	0.4761	0.5014	0.5228	0.498	0.4892
E	0.5008	0.4754	0.502	0.4893	0.4866	0.5211	0.4877	0.5737
F	0.501	0.5135	0.5029	0.4922	0.5035	0.531	0.488	0.5536
G	0.5014	0.4973	0.5199	0.4698	0.4811	0.5097	0.488	0.427
H	0.5016	0.4961	0.5001	0.4744	0.5198	0.5299	0.4939	0.5106

A: Retrieved τ at 0.55 μm (expected τ=0.5)

Product Name	Error LUTinput	RndError	CalError	GeoError	ModError	ElvError	SfcError	AllError
τ	0.0011(+)	0.0159	0.0162	0.0215	0.0123	0.0561(+)	0.0221	0.1006
η	0.0000	0.0000	0.0707	0.1000	0.0707	0.4243 (+)	0.1323 (+)	0.4912 (+)
ρ	0.0004 (-)	0.0008	0.0022	0.0025	0.0031 (-)	0.0067	0.0020 (+)	0.0074 (+)
ε	0.0010	0.0021	0.0037	0.0028	0.0020	0.0025	0.0035	0.0052

B: MSE of retrieved τ,η,ρ^{σ} and ε (expected τ=0.5, η=0.5, ρ^{σ}=0.15 and ε=0.0). Entries designated with (+) mean that the product was over-estimated for all 8 geometries, whereas those with a (-) means it was under-estimated for all geometries.

4.7. Retrieved Land Products

Examples of the three primary aerosol products ($\tau_{0.55}$, η and $\rho^{s}_{2.12}$) are shown in Fig. 25B-D, along with a color composite (RGB image) of the L1B reflectances (0.47, 0.55 and 0.66 μm channels) in Figure 25A. This Terra image was taken over the Eastern U.S. on May 4, 2001, and is the same granule that was used in King et al., (2003). Like the earlier image, it shows a plume of aerosol, seemingly transported from the Ohio Valley, through Maryland, and into the Atlantic. Aerosol optical depths in the plume are high, $\tau_{0.55} \sim 1.0$. The aerosol is also dominated by fine particles as seen in the η image (plotted where $\tau_{0.55} > 0.2$). The land/ocean continuity is very good, and the τ continuity is discussed later. AERONET observations in Baltimore (MD_Science_Center) show τ ~ 1.0 and Ångstrom exponent ~ 2.0, also indicating fine-dominated heavy aerosol.

An example of a new derived over-land product in C005 is the Fine τ, which is simply the product of η and τ at 0.55 μm. This product is expected to useful in estimating the anthropogenic contribution to the total τ because generally smoke and urban pollution are fine mode dominated (e.g. Kaufman et al., 2005).

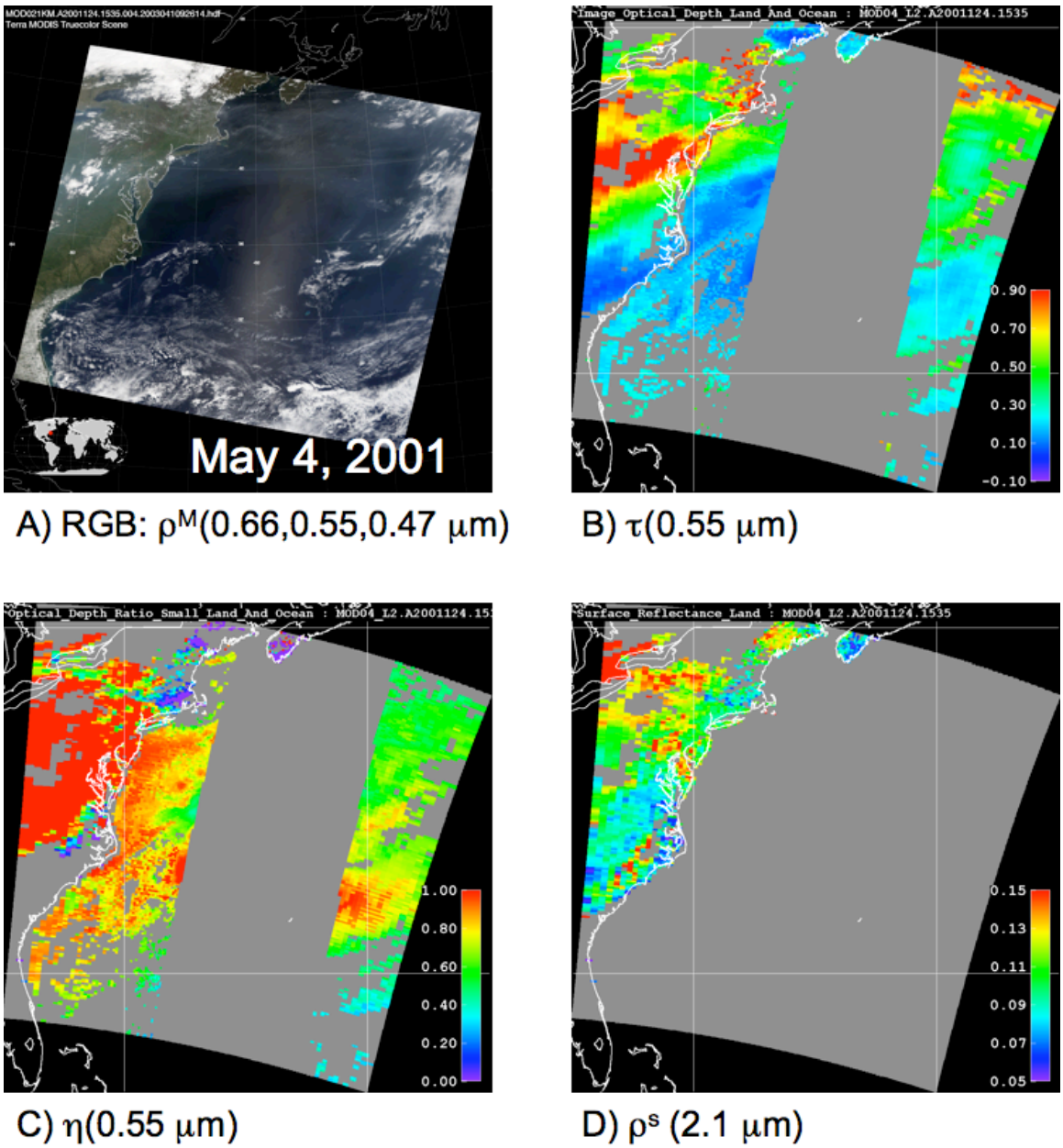

Figure 25: Retrieved aerosol and surface properties over the Eastern U.S. on May 4, 2001. This figure can be compared with that plotted in King et al., (2003). Panel A) is a 'true-color' composite image of three visible channels, showing haze over the mid-Atlantic. Panels B) and C) show retrieved τ and η, showing that the heavy aerosol ($\tau \sim 1.0$) is dominated by fine particles. The transport of the aerosol into the Atlantic is well represented with good agreement between land and ocean. In fact the continuity of τ seems to be improved since earlier versions of the aerosol algorithm. Note that over-land η is not reported when $\tau < 0.2$. Panel D) shows the retrieved surface reflectance during the retrieval.

Table 8 lists the aerosol over land products that are within each "M?D04" Level 2 granule. For each product, the table lists its name within the file, its dimension, and its "type." All products are at least two-dimensional (nominally 135 x 204 at 10 km x 10 km resolution), and many have three dimensions. If there is a third dimension, the channels (usually wavelengths) are listed. A parameter's type may be

Retrieved, Derived, Diagnostic, Experimental, or Joint Land and Ocean. A *Retrieved* parameter is one that is a solution to the inversion (Procedure A). *Derived* parameters are computed based on products directly retrieved. For example, the Ångstrom Exponent is derived based on the spectral AOD that characterizes the retrieved solution. Products that are *Diagnostic* include QA parameters and those parameters that were calculated during intermediate steps. These diagnostic parameters can be used to understand how the retrieval worked. Products denoted *Experimental* are superfluous to the main inversion, but are useful in other applications. They are described in an Appendix. Note that a parameter's type does not signify that the parameter should be used in a quantitative way. Each parameter should be independently validated through comparison with ground-truth. Some provisional validation is described in section 5, however more rigorous evaluation of each parameter are subjects of future study, and will be appended as needed in Section 7.

Finally, the *Joint* products are those that are composites of over-land and over-ocean aerosol retrievals. For example, Image_Optical_Depth_Land_And_Ocean includes all 0.55 μm τ data from 'Corrected_Optical_Depth_Land' including those with QAC=0. This product provides a full picture of the aerosol distribution, even if some of the retrievals are more qualitative in nature than the validated quality assured data. The 'Optical_Depth_Land_And_Ocean' product is designed to include the more quantitative data, and is filtered for higher QAC. Upon the implementation of C005 in early 2006, it included all retrievals where QAC > 0 (e.g. 1, 2 or 3). However, based on evaluation of the operational C005 data stream, the recommended QAC filter is for QAC=3 only to ensure the most quantitative data.

TABLE 8: CONTENTS OF MODIS C005 AEROSOL LEVEL 2 FILE (MOD04/MYD04): LAND PRODUCTS

Name of Product (SDS)	Dimesions: 3rd Dimension	Type of product
Corrected_Optical_Depth_Land	X,Y,3: 0.47, 0.55, 0.66 μm	Retrieved Primary
Corrected_Optical_Depth_Land_wav2p1	X,Y,1: 2.12 μm	Retrieved Primary
Optical_Depth_Ratio_Small_Land	X,Y: (for 0.55 μm)	Retrieved Primary
Surface_Reflectance_Land	X,Y,3: 0.47, 0.66, 2.12 μm	Retrieved Primary
Fitting_Error_Land	X,Y: (at 0.66 μm)	Retrieved By-Product
Quality_Assurance_Land	X,Y,5: 5 bytes	Diagnostic
Aerosol_Type_Land	X,Y:	Diagnostic
Angstrom_Exponent_Land	X,Y: (for 0.66/0.47 μm)	Derived
Mass_Concentration_Land	X,Y:	Derived
Optical_Depth_Small_Land	X,Y,4: 0.47,0.55,0.66,2.12 μm	Derived
Mean_Reflectance_Land	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Diagnostic
STD_Reflectance_Land	X,Y,7: 0.47,0.55,0.66,0.86,1.2,1.6,2.12μm	Diagnostic
Cloud_Fraction_Land	X,Y:	Diagnostic
Number_Pixels_Used_Land	X,Y:	Diagnostic
Path_Radiance_Land	X,Y,2: 0.47, 0.66 μm	Experimental
Error_Path_Radiance_Land	X,Y,2: 0.47, 0.66 μm	Experimental
Critical_Reflectance_Land	X,Y,2: 0.47, 0.66 μm	Experimental
Error_Crit_Reflectance_land	X,Y,2: 0.47, 0.66 μm	Experimental
Error_Critical_Reflectance_Land	X,Y,2: 0.47, 0.66 μm	Experimental
Quality_Weight_Path_Radiance_Land	X,Y,2: 0.47, 0.66 μm	Experimental
Quality_Weight_Crit_Reflectance_Land	X,Y,2: 0.47, 0.66 μm	Experimental
Optical_Depth_Land_And_Ocean	X,Y: 0.55 μm	Joint (QAC>0)**
Image_Optical_Depth_Land_And_Ocean	X,Y: 0.55 μm	Joint (QAC≥0)
Optical_Depth_Ratio_Small_Land_And_Ocean	X,Y: 0.55 μm	Joint (QAC≥0)

X = 135; Y = 203. If there is a 3rd dimension of the SDS, then the indices of it are given. The "Retrieved" parameters are the solution to the inversion, whereas "Derived" parameters are computed from the choice of solution. "Diagnostic" parameters are derived during the retrieval process. "Experimental" products are unrelated to the inversion but may have future applications. "Joint" products are the combined land and ocean products, with associated QAC constraint (for over land) in parentheses. **Based on evaluation of operational C005-L data, the QAC for quantitative studies should be limited to QAC=3.

5. Provisional validation

As described in Section 2.6, the retrieved products must look reasonable, and must be evaluated with respect to ground-truth data. Before being run operationally (by MODAPS), the C005 algorithm (as V5.2) was tested on a *testbed* of MODIS granules, and compared to results derived by C004 (as V5.1) using the same input data, and masking procedures. The testbed consisted of 6300 granules, including one full month (August 2001), fifteen entire days (listed in Table 9) and about 141 individual granules. These granules included observations from both Terra and Aqua, and were seasonally and yearly representative of the MODIS time series. This section reports on the V5.2 vs V5.1 comparison.

TABLE 9: DESCRIPTION OF DATA USED IN V5.2 PROVISIONAL VALIDATION

Date of MODIS Observations	Terra/Aqua	Why interesting?
August 2001 (full month: 4138 granules)	Terra and Aqua	
7 July 2002 (full day: 132 granules)	Aqua	Quebec Smoke in NE US
8 July 2002 (full day: 136 granules)	Aqua	Quebec Smoke in NE US
6 Mar 2004 (full day: 132 granules)	Aqua	Asian Dust
7 Mar 2004 (full day: 138 granules)	Aqua	Asian Dust
Eight days in 2003 (full days: 1070 granules)	Aqua	Yearly Cycle
14 Nov 2005 (full day: 138 granules)	Terra	Low τ globally
22 Apr 2001 (full day: 136 granules)	Terra	ACE-Asia
26 Jun 2002 (full day: 138 granules)	Terra	Summer time haze
Test_bed_Aqua: (39 granules)	Aqua	Test bed of interesting Aqua data
Test_bed_Terra: (102 granules)	Terra	Test bed of interesting Terra data

Total granules = 6299

5.1. Visual comparison of V5.2 versus V5.1 products

Figs 26 and 27 plot retrieved τ at 0.55 μm from both V5.1 and V5.2, over small areas of two MODIS granules. In both figures, V5.1 (OLD) is presented in (a), whereas V5.2 (NEW) is shown in (b). Fig 26 shows a region in the western U.S. from 30 Sep 2003, whereas, Fig 27 displays the U.S. mid-Atlantic from 4 May 2001 (the same granule is in King et al., 2003). The west coast images (Fig 21) show that the V5.2 aerosol retrieval adds more valid retrievals over very low τ areas (coastal Oregon and northern California). V5.2 reports these areas as having near zero or slightly negative τ, where V5.1 would have reported fill values (errors). In areas farther from the coastline, V5.2 tends to clean up contamination presumably caused by clouds, elevation, and inhomogeneous surface properties, and produces a much more reasonable picture of τ. The east coast retrievals (Fig 27) show subtle changes in retrieved τ from V5.1 to V5.2. The land/ocean τ seems a bit more continuous over southeastern Virginia and northeastern North Carolina. That is at the expense of retrieving negative τ values over Maine, and generally lowering the τ over other areas (for example, the common borders of New York, Pennsylvania and New Jersey).

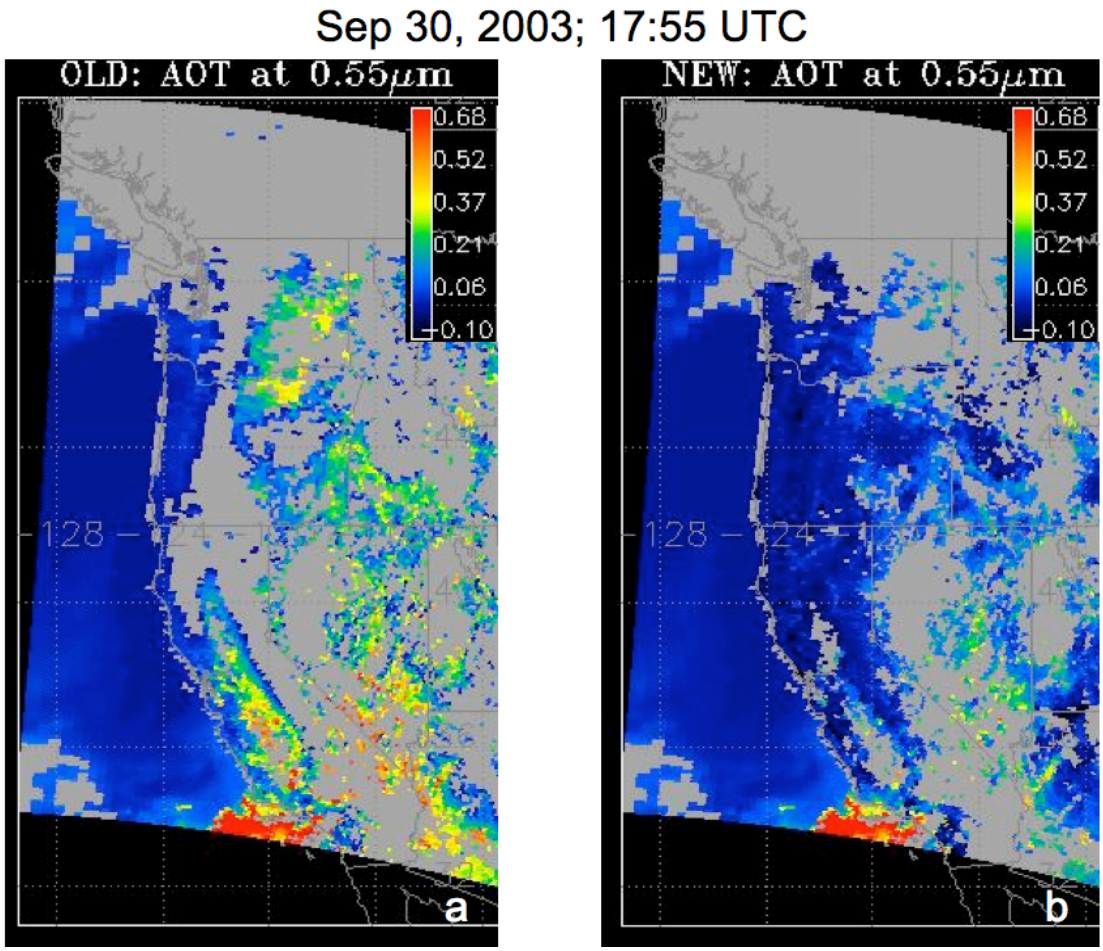

Figure 26: Retrieved τ at 0.55 μm for Old V5.1 (a) and New V5.2 (b) over California for 30 September 2003. The color scale is the same for both plots. Note the increase in the retrieval spatial coverage and reduction in surface contamination for V5.2.

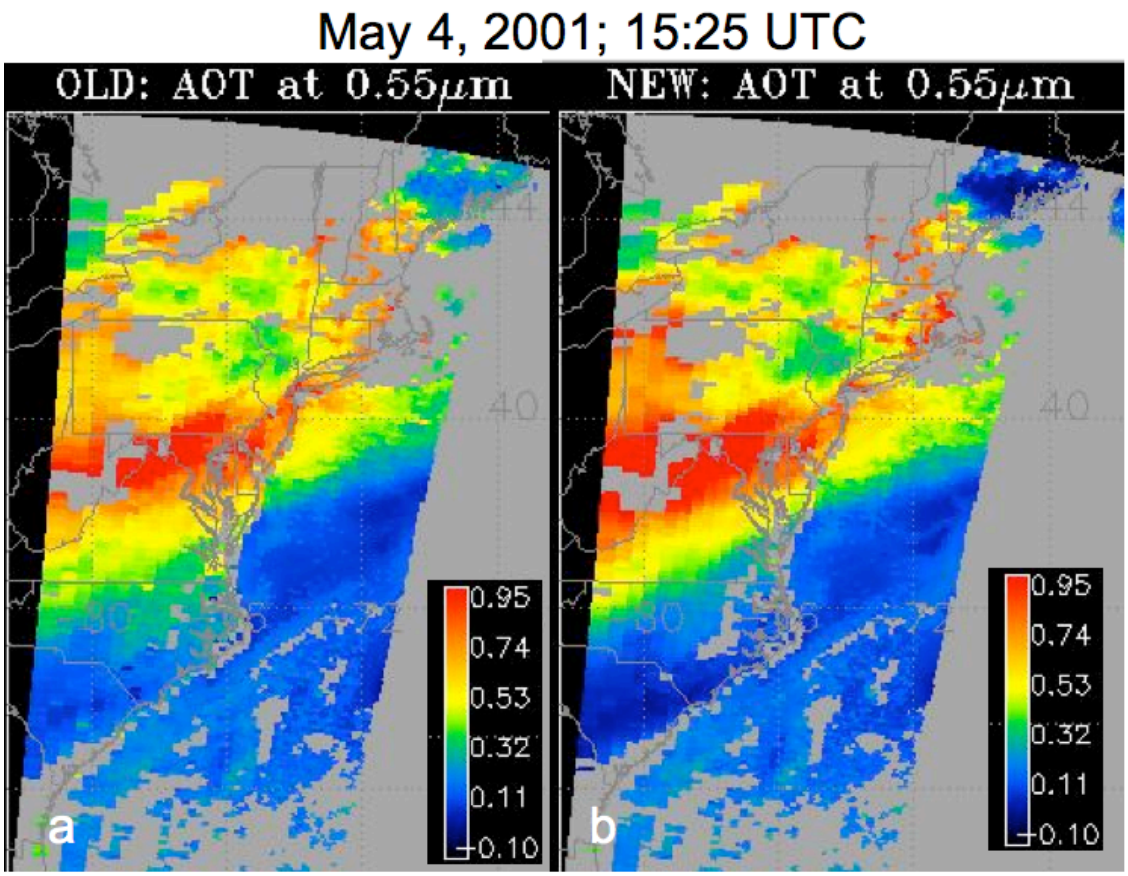

Figure 27: Retrieved τ (AOT) at 0.55 μm for Old V5.1 (a) and New V5.2 (b) over the Mid-Atlantic U.S. for 4 May 2001. The color scale is the same for both plots. Note the slight improvement of land/ocean continuity in V5.2.

5.2. Statistics of V5.2 vs V5.1 products

Of most interest to the climate community will be the changes in the statistics of the aerosol products. These include the global mean values and the distribution (histogram) of the values. For the set of MODIS granules listed in Table 9 (about 6300 granules of both Terra and Aqua), the mean 0.55 μm τ over land is reduced from 0.28 to 0.21. This is a significant reduction that should be compared with model estimates. Over ocean, the mean 0.55 μm τ remains constant, about 0.14.

Figs 28 and 29 plot the histograms of retrieved τ and η over land and ocean at 0.55 μm from both V5.1 and V5.2. These histograms include the 141 individual Terra and Aqua granules (the "test_bed") and twelve days of global data (Table 9 except for the August 2001 Terra data). The use of global data is especially important for determining how the retrieval behaves in regions not selected for algorithm development.

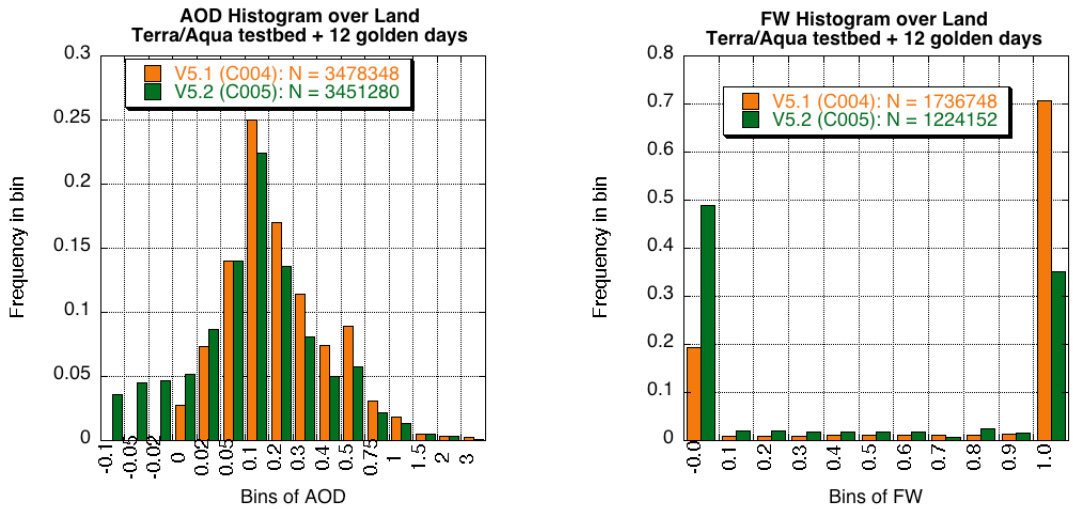

Figure 28: Histograms of retrieved τ (a) and η (b) over land, from V5.2 (C005) in green, compared to V5.1 (C004) in orange. The data include the 141 granules of the Terra and Aqua "test_bed" as well as twelve complete days. The value of each bin refers to the minimum value of the bin (the max value would be the value of the next bin). Note that the general lognormal nature of the retrievals is preserved, except now there are some negative values. Note fewer retrievals of η in C005 due to the constraint that τ > 0.2.

An obvious change in the V5.2 over-land product is that small magnitude negative τ (Fig 28a) retrievals are valid. About 10-11% of the total τ retrievals are now retrieved as below zero, of which only about 3% are below -0.05. This result indicates that V5.2 has reasonable ability to detect very clean conditions within the expected error of ±0.05. Over land, the fraction of retrieved medium to medium high τ ($0.2 < \tau < 0.75$) is reduced, while the fraction of high τ ($\tau > 0.75$) remains nearly constant. The η product over land (Fig 28b) continues to show binary retrievals (either zero or one), although the product is more evenly balanced between one and zero. Note the retrievals are not completely analogous, due to the constraint that $\tau > 0.2$ in V5.2.

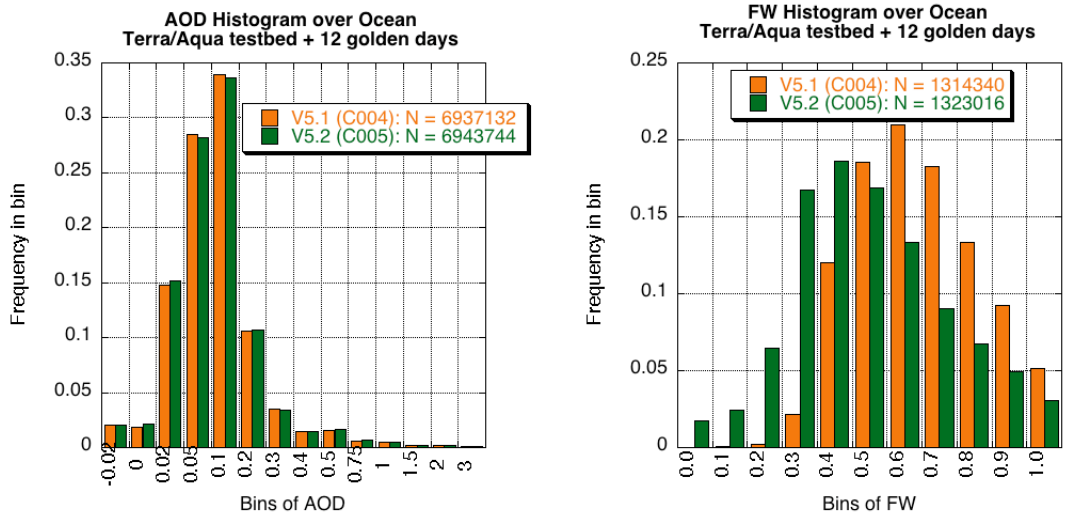

Figure 29: Histograms of retrieved τ (a) and η (b) over ocean, from V5.2 (C005) in green, compared to V5.1 (C004) in orange. The data include the 141 granules of the Terra and Aqua "test_bed" as well as twelve complete days. The value of each bin refers to the minimum value of the bin (the max value would be the value of the next bin). Note that the retrieved τ from V5.1 to V5.2 is nearly unchanged, whereas in general the values of η are shifted by about -0.2.

Over ocean, there is negligible change in the total τ histogram (Figure 29a). However, the retrieved values of η (Figure 29b) have been lowered by about 0.2 in most cases.

5.3. Comparison of V5.2 and V5.1 over-land products with AERONET

Our primary means of this provisional validation is comparison with ground-based sunphotometer measurements, specifically, those of AERONET (Holben, et al. 1998). In 'sun' mode, the AERONET instruments measure spectral τ, τ_λ, to within ~0.01 in the MODIS visible and near-IR wavelength regions (Eck et al., 1999) and can be used to derive Fine Weighting (η) by the spectral deconvolution method of O'Neill et al., (2003). The AERONET measured τ is easily interpolated to the exact MODIS wavelengths (for example 0.55 μm) by quadratic interpolation in log reflectance/log τ space. The AERONET 'sun-measured' definition of η differs from either of the MODIS (land or ocean) definitions, but should be correlated with either. The methodology of comparing temporally varying AERONET data with spatially varying MODIS data is described in (Ichoku et al., 2002). In the following validation, we use AERONET Level 2.0 data (cloud screened and quality assured for instrument calibration) when available (Smirnov, et al. 2000). Although North America and Europe provide the most stations in the data base (http://aeronet.gsfc.nasa.gov), all continents (except Antarctica), all oceans and all aerosol types are represented. Remer et al., (2005) provide a comprehensive validation of τ from MODIS C004, whereas Kleidman et al., (2005) provide comparisons of the η product.

Figures 30-31 plot the comparisons of both the V5.1 and V5.2 over-land aerosol products (on 6300 globally/seasonally representative granules) with analogous AERONET data, via the spatio-temporal

co-location method of Ichoku et al., (2002). Figure 30 plots the retrieved MODIS τ against AERONET τ, both at 0.55 μm. In this plot and others, the data were sorted according to AERONET aerosol optical thickness, and grouped into twenty bins of equal number of points. In Figure 30, there are 62 points per bin. Plotted are the mean and standard deviation (error bars) of the points in each bin. At higher optical thickness where the data become sparser, fewer points are used in the average, as indicated. The regression equation and correlation given at the top of each plot were calculated from the full (un-binned) scatter plots. The solid black line is the 1:1 line, and the dashed lines denote the expected uncertainty (±0.05 ± 0.15τ). The dashed lines should encompass one-standard deviation (66%) of the aerosol retrievals for the MODIS products to be considered 'validated' by the sunphotometer measurements. Figure 30 shows that in V5.1, only about 57% of the points were within error bars, whereas V5.2 shows more than 67%. The regression equation has improved, from "y=0.097+0.91x" to "y=0.029+1.01x." Correlation R is also improved, from R=0.847 to R=0.894. It should be noted that slight differences in the number of points arise due to different selection of valid dark pixels and allowance of below zero τ retrievals.

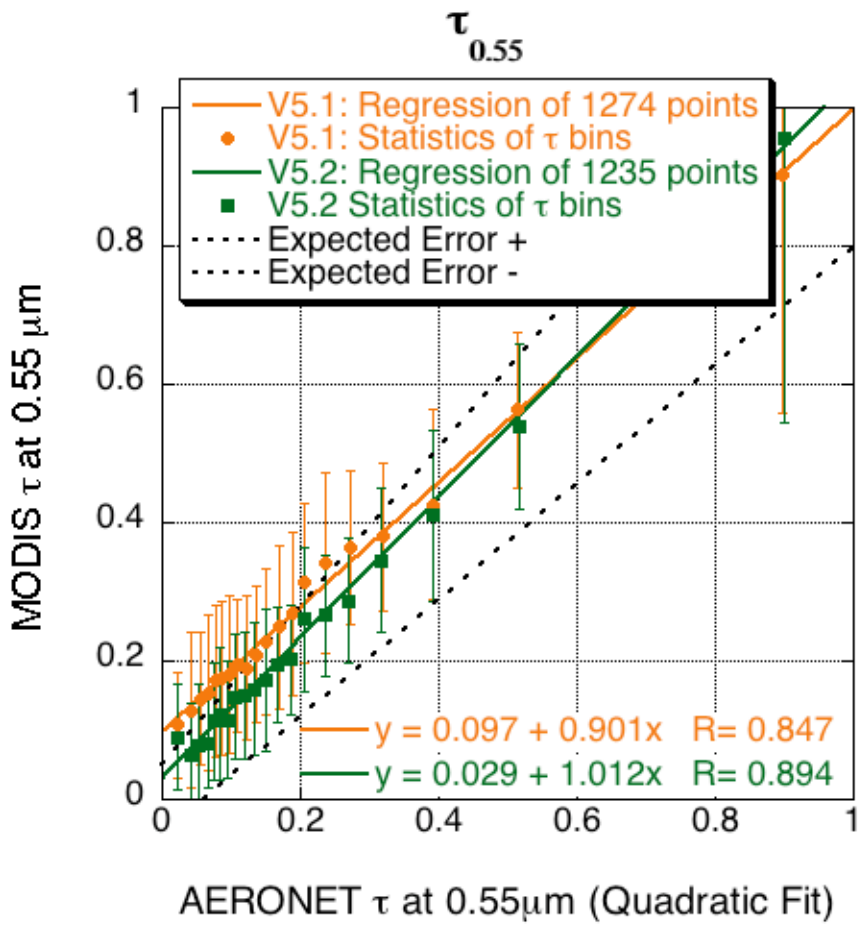

Figure 30: MODIS τ over land retrieved at 0.55 μm, compared with AERONET τ interpolated to 0.55 μm. The solid shapes and error bars represent the mean and standard deviation of the MODIS retrievals, in 20 bins of AERONET-derived τ. Both the retrievals from V5.1 (orange) and V5.2 (green) are shown. The regressions (solid lines) are for the cloud of all points (not shown). The expected errors for MODIS (±0.05 ±0.15τ) are also shown (dashed lines).

Figure 31A plots the retrieved MODIS η (fine model weighting or non-dust weighting) against AERONET η, where the AERONET η is that defined by O'Neill et al., (2003). Note, that the MODIS η and the AERONET η are not exactly the same parameter. The MODIS value is the percentage of τ attributed to fine mode dominated model, which includes fine and coarse modes. The AERONET value is simply the single fine mode. We would expect these quantities to be correlated, but not necessarily the same. The improvement to the MODIS η product is mainly its correlation to AERONET. Note that due to the V5.2 η product defined only when τ > 0.2, a fair comparison necessitates that the constraint is placed on the V5.1 points as well.

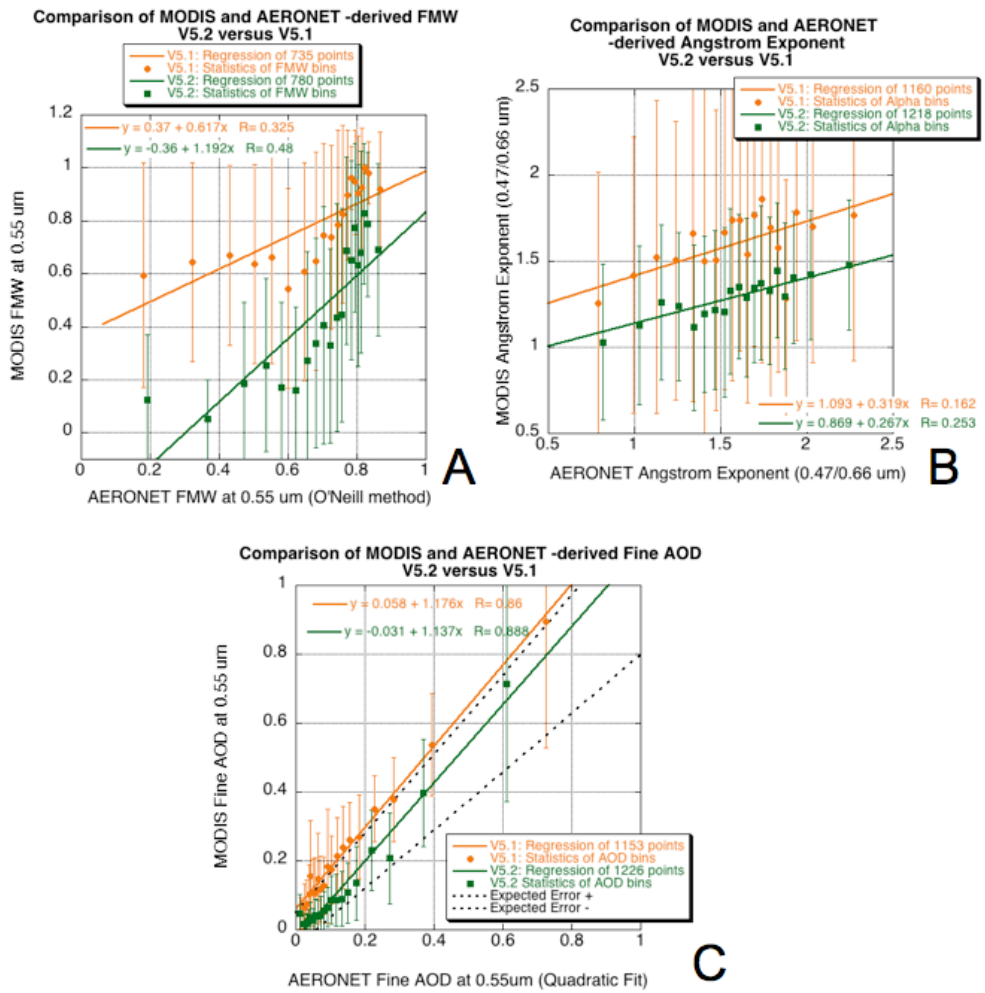

Figure 31: MODIS aerosol size retrievals compared with AERONET derived products. The solid shapes and error bars represent the mean and standard deviation of the MODIS retrievals, in 20 bins of AERONET-derived product. Both the retrievals from V5.1 (orange) and V5.2 (green) are shown. The regressions (solid lines) are for the cloud of all points (not shown). A) η over land retrieved at 0.55 μm, compared with AERONET η retrieved by the O'Neill method. Note that due to the V5.2 η product defined only when $\tau > 0.2$, a fair comparison necessitates that the constraint is placed on the V5.1 points as well. B) MODIS Ångstrom Exponent (0.466/0.644 μm) over land with AERONET Ångstrom Exponent interpolated to the same wavelengths. C) MODIS Fine model τ over land retrieved at 0.55 μm, compared with AERONET Fine mode τ interpolated to 0.55 μm by the O'Neill method. The expected errors for MODIS ($\pm 0.05 \pm 0.15\tau$) are also shown (dashed lines).

Figs 31B and 31C show comparisons for derived size information over land, including the the Ångstrom Exponent (defined by 0.47 and 0.66 μm) and Fine τ (i.e. τ x η), respectively. The Ångstrom exponent is defined the same way in both MODIS and AERONET, but has little improvement from V5.1 to V5.2. There is slightly better but still poor correlation with the AERONET measured quantities. For the Fine τ, the correlation and slopes are nearly unchanged, however, the offset goes from +0.051 to -0.031. The result is that if the same expected errors are defined (i.e. Equation (2)), then nearly two-thirds of all MODIS Fine τ is within expected errors. Note again that the difference in the number of points is due to different selection of dark pixels and treatment of negative τ retrievals. Again, the MODIS and AERONET quantities are not exactly the same because there is difference in the definitions of η. It is clear, however, that MODIS still derives too much fine-dominated aerosol over land.

5.4. Comparison of ocean products with AERONET

Over ocean, the expected τ error bars are much smaller ($\pm 0.03 \pm 0.05\tau$) than over land. This is because the (non glint) ocean contribution to spectral reflectance can be characterized with a great deal of authority. Figure 32 compares MODIS τ at 0.55 μm compared to the AERONET τ (quadratic fit to 0.55 μm). The changes made to the refractive indices in three coarse mode models are the only differences between V5.1 ('C004') and V5.2 (C005). We do not expect these changes to affect the τ retrieval. The difference between the V5.1 and V5.2 τ is indeed insignificant. In either version the correlation of MODIS/AERONET is extremely high ($R > 0.96$), although there tends to be a slight positive bias (slope > 1.06). Two-thirds of the points do lie within the expected error bars, thus validating the MODIS τ. Note there are far fewer points over ocean than land due to less AERONET coverage.

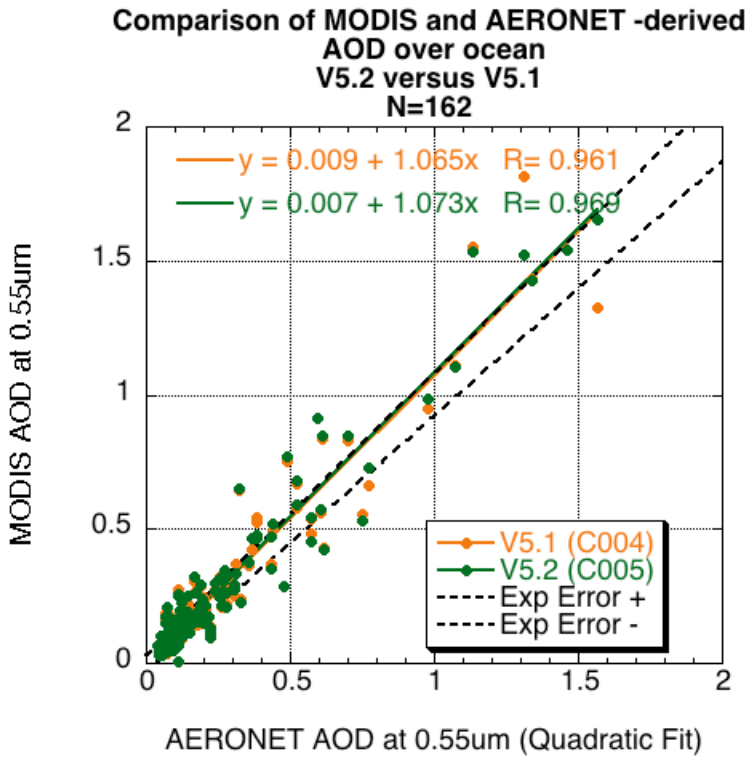

Figure 32:MODIS τ over ocean retrieved at 0.55 μm, compared with AERONET τ interpolated to 0.55 μm. Both the retrievals from V5.1 (orange) and V5.2 (green) are shown. The regressions (solid lines) are shown. The expected errors for MODIS (±0.03 ±0.05τ) are also shown (dashed lines).

We expect the changes made to the ocean algorithm to primarily affect the retrieved size parameters. Kleidman et al., (2005) noted a large positive bias (>0.2) in the C004-O MODIS η (Fine mode Weighting). It is expected that the aerosol size retrievals in C005 should note larger particles and a smaller η than those in C004, especially in dust aerosol. Figs 33 (A-C) are the analogous ocean plots to the Figs 31 (A-C) land. Fig 33A shows V5.2 and V5.1 MODIS η data from our test datasets, compared with AERONET η, when the τ at 0.55 μm is greater than 0.15. V5.2 η is lower by up to 0.15, especially in coarse-mode dominated (presumably dust) conditions. In fact, when performing the same C005 algorithm on the Kleidman et al., (2005) data (dominated by dust), the positive bias is nearly removed. Ångstrom exponent (Fig 33B) is also plotted for τ > 0.15, and confirms that C005 derives bigger particles over the ocean.

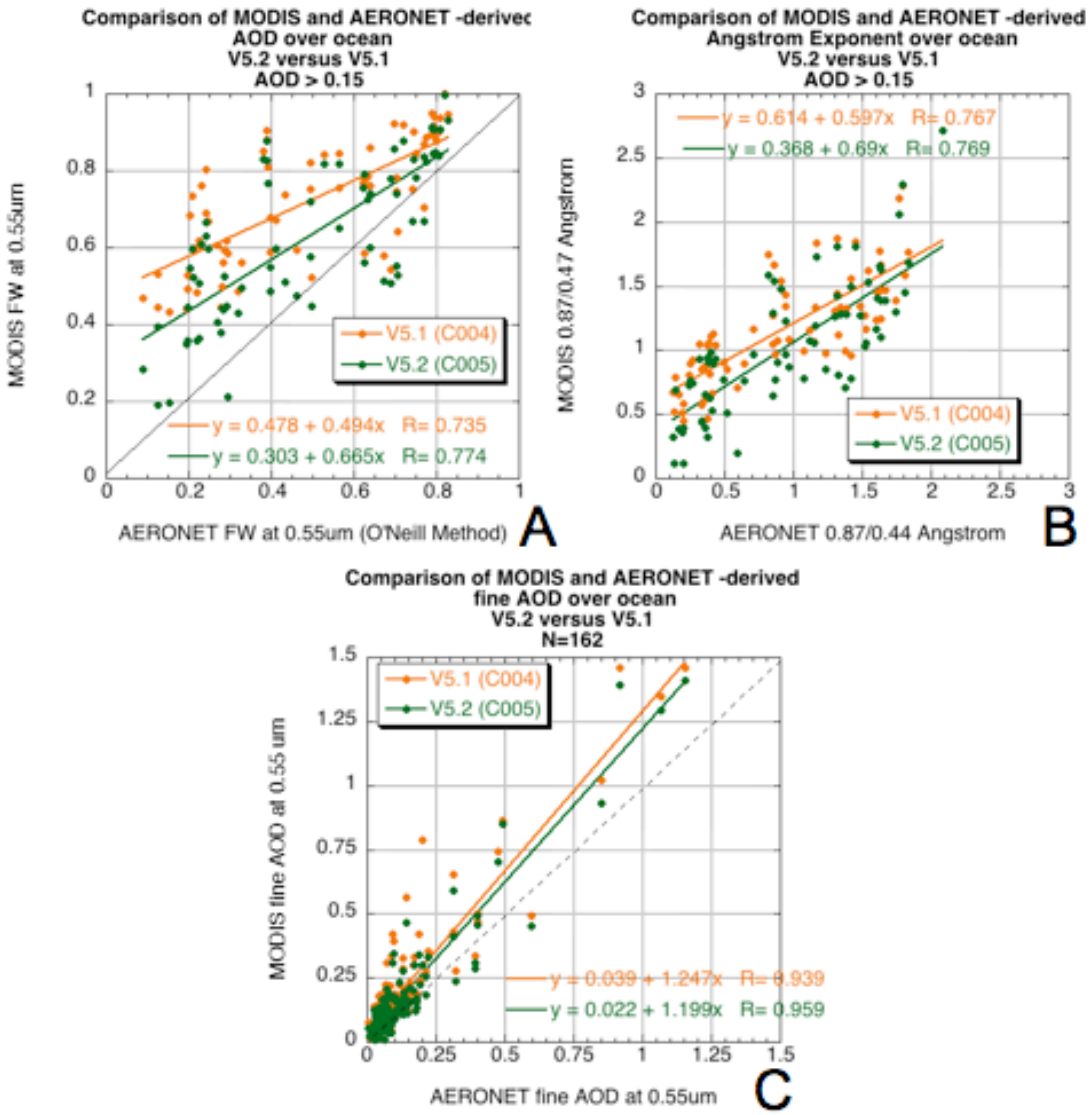

Figure 33: MODIS aerosol size retrievals over ocean compared with AERONET derived products. Both the retrievals from V5.1 (orange) and V5.2 (green) are shown. The regressions (solid lines) are shown. A) η over ocean retrieved at 0.55 μm, compared with AERONET η retrieved by the O'Neill method, for when τ at 0.55 μm > 0.15. B) MODIS Ångstrom Exponent (0.466/0.644 μm) over ocean with AERONET Ångstrom Exponent interpolated to the same wavelengths, also for when τ at 0.55 μm > 0.15. C) MODIS Fine model τ over ocean retrieved at 0.55 μm, compared with AERONET Fine mode τ interpolated to 0.55 μm by the O'Neill method. The one-to-one line is also shown (dashed line).

Finally, Fig 33C plots the fine τ (η x total τ) over the ocean. In general, the retrievals of V5.2 compare slightly better with AERONET derived fine τ, but the change is probably insignificant. There is a strong positive bias of both V5.1 and V5.2, which is surprising given that the definitions of η are much closer to AERONET for ocean rather than for land. Although the changes to the coarse mode refractive indices in the ocean LUT make a significant improvement to the retrieval of aerosol size parameters over ocean, not all issues have been solved. Non-sphericity effect described by Levy et al., (2003) may

continue to pose problems. It should be noted that non-spherical models (spheroids) were attempted for V5.2-ocean, but their effects on the retrieval did not yield the desired results.

6. File Format, Notes, and Recommendations to users

6.1. File format and content

The MODIS Level 2 aerosol product files (MOD04_L2 for Terra and MYD04_L2 for Aqua, known commonly as M?D04) are stored in Hierarchical Data Format (HDF). HDF is a multi-object file format for sharing scientific data in multi-platform distributed environments. HDF files can be accessed through HDF library subroutine and function calls, which can be downloaded from the HDF web site (http://www.hdfgroup.org). Proprietary software, such IDL (http://www.ittvis.com/ProductServices/IDL.aspx) and MATLAB (http://www.mathworks.com/) have built in routines to process IDL files. Also, free tools, including HDFLook (http://www-loa.univ-lille1.fr/Hdflook/) are available, can be found on the MODIS atmosphere web site (http://modis-atmos.gsfc.nasa.gov/tools.html).

All C005 M?D04 data files (and corresponding L1B and geolocation data files used to create them) are available on-line (at no charge) through the user-friendly Level 1 and Atmosphere Archive and Distribution System (LAADS; http://ladsweb.nascom.nasa.gov/data/search.html). M?D04 HDF file names have this naming convention:

M?D04_L2.AYYYYDDD.HHMM.CCC.YEARDAYHRMNSC.hdf

where ? is O or Y for Terra or Aqua, YYYY, DDD and HHMM are four digit year, three digit Julian day, and time of day in UTC, CCC is the collection (005 for C005), and YEARDAYHRMNSC represents when the file was processed.

The M?D04 files include 53 parameters, each stored as a Scientific Data Set (SDS) within the HDF file. In addition to the SDSs listed in previous sections of this document (in Table 8 for land and Table 3 for ocean), the HDF file includes SDSs for location (Latitude, Longitude), geometry (Solar and Sensor zenith and azimuth angles). These and other details can be found within the MODIS-atmosphere web site (http://modis-atmos.gsfc.nasa.gov/MOD04_L2/format.html), and may be updated as needed.

The M?D04 product contains data that has a spatial resolution (pixel size) of 10 x 10 kilometers (at nadir), with larger size closer to swath edges. Each M?D04 product file represents a five-minute time interval of L1B observations (a granule), such that the output grid is 135 pixels in width (cross track) and 203 pixels in length (along track). Every tenth granule is 135 by 204 pixels. The geolocation (location and geometry) of each output pixel is computed from aggregation of 10x10 boxes of L1B 1 km input, taking the average of the 4 central (5,5), (5,6), (6,5), (6,6) 1-km L1B input pixels. Again, more details can be found within the MODIS-atmosphere web site (http://modis-atmos.gsfc.nasa.gov/MOD04_L2/grids.html), and may be updated as needed.

6.2. Notes on global sampling

Although MODIS views a swath approximately 2300 km wide, there are gaps near the equator and multiple viewings of the poles. Near swath edges, view angles exceed 55˚, such that individual pixels can represent three times the surface area of nadir views. Thus parts of the globe may be over or under sampled by each MODIS instrument each day.

During a given month, each instruments' M?D04 products represent MODIS observations on 1.1×10^8 pixels (at 10 x 10 km). Except for the Deep Blue algorithm (not included in this version of the ATBD), the MODIS aerosol algorithms rely on observing dark surface targets, meaning that glint (over ocean), deserts (over land), ice, snow and clouds are avoided, and aerosol properties are not retrieved over them. Thus out of the monthly M?D04 coverage, only ~15% can be used for aerosol retrieval. These 'valid' pixels are comprised of approximately 75% ocean and 25% land retrievals, but varies monthly. High confidence data (QAC=3) account for approximately 60% and 25% of the valid data over land and ocean, respectively, while at least marginal confidence data (QAC≥1) accounts for about 85% and 98%, respectively. The algorithm over ocean has stricter criteria for high confidence than that over land.

The fraction of retrieval coverage versus MODIS-observed coverage varies by month. In addition, due to changing surface and cloud conditions, the QAC (quality) distribution varies by month. Fig. 34 shows how the number of retrievals (pixel counts, or P_k, black circles) and their QAC (plotted as QC, square symbols) varies by month of 2003, for Terra over land and ocean separately. Clearly, both the quality and quantity of the over-land retrievals has a strong seasonal dependence, with strong maxima in both during the northern summer. There is much less seasonal dependence over ocean. Interestingly, over the ocean, the maximum in PC is not associated with the maximum in confidence. Note the relative dominance of QAC=3 over land, but QAC=1 over ocean, due the stricter QA criteria over each surface.

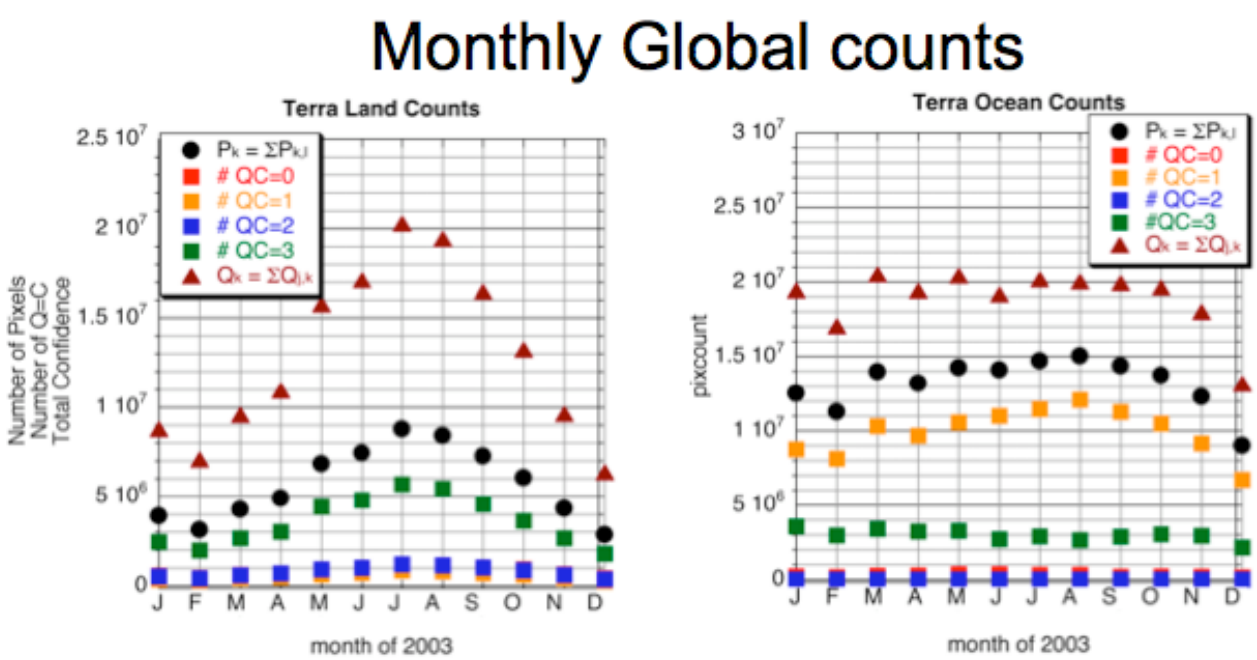

Figure 34: Time series (Terra, 2003) of monthly total counts over A) land and B) ocean. For each panel, black circles represent the monthly total pixel counts (P_k), the squares are the total pixel counts that are a given QC value, and the red triangles are the total confidence for that month (i.e., $Q_k = \Sigma_i Q_{i,k}$). Note the different seasonal dependencies of the curves.

6.3. Recommendations for use of data

6.3.1. Total τ

According to the provisional validation (Section 5), the C005 MODIS retrievals of total τ can be validated as compared to AERONET. The expected uncertainties for τ over land and ocean are expected to be similar to those defined for C004 (Eqs. 1 and 2), however many of the biases over land will be reduced. While the provisional validation did not explicitly mention that the data were separated by QAC, all data had minimum QAC=1. Fig 34 shows that where QAC≥1, the most common values are QAC=1 over ocean, and QAC=3 over land.

Up through V5.2.8 of the MODIS dark target algorithms (current as of Feb. 2009), the quantitative joint τ product (i.e., 'Optical_Depth_Land_And_Ocean') considered minimum QAC=0 over ocean and QAC=1 over land. We do not recommend using that SDS for quantitative studies, instead, we suggest that to ensure sufficient quality for quantitative use, the user should work with the separately derived τ SDSs ('Effective_Optical_Depth_Average_Ocean' and 'Corrected_Optical_Depth_Land') and filter by QAC (QAC≥1 over ocean and QAC=3 over land). The use of these SDSs, coupled with the QA SDSs will allow the user the most control over data quality. In future processing, the quantitative joint product may be revised to reflect our new expectation. On the other hand, the joint qualitative SDS 'Image_Optical_Depth_Land_And_Ocean' reports all valid (non-fill value) AOD, over land and ocean, regardless of QAC value. This product can be used to trace aerosol plumes and make aerosol maps, but should not be used for quantitative study.

It also should be repeated that over land, the algorithm permits retrieval of small negative τ values in order to avoid arbitrary biases in long-term statistics. MODIS does not have enough sensitivity over land to retrieve aerosol to better than ±0.05, meaning that in very clean conditions the algorithm cannot determine $\tau = 0.00$ from 0.05 or -0.05. It is up to a user to decide what to do with these values, whether to force them to zero or some small positive value (for use within models). However, these retrievals of negative τ are considered to be good retrievals, and are assigned QAC=3, unless other conditions apply. Over ocean, however, retrievals of negative τ are NOT permitted, so that they are automatically forced to zero, and the QAC lowered to zero.

6.3.2. Size parameters

The over land algorithm reports size parameters if the retrieved total AOD is valid, and QAC levels are sufficient. Values for η are reported when retrieved τ>0.2 and QAC≥1. Ångstrom exponent is reported when τ>0.0, but QAC may be 0. Values for τ_f are reported for any value of τ as long as QAC≥1. Over

ocean, η is reported, as long as τ is valid and the observed reflectance (at 0.86 μm) is above some threshold value. Ångstrom exponent is reported.

The provisional validation of Section 5 indicated that the size parameters from MODIS were qualitatively comparable to AERONET, especially when constrained to higher aerosol loadings (τ>0.2 over land and τ>0.15 over ocean). Over land, there was some minor improvement between C004 and C005, for derivations of both η and α. The regression fit lines (slope and offset) moved towards the one to one line. However, the low correlations (R<0.5 for η and R<0.25 for α) still mean that the algorithm over land lacks information about aerosol size. Over ocean, the wide spectral range and dark surfaces leads to better discrimination (R>0.75) of the size parameters. For either surface, MODIS seems to have some skill at retrieving τ_f (R>0.88). Therefore, there is a good possibility that τ_f could become a validated parameter, even over land.

At this point a user should not expect quantitative information about either η and α over land, even if the total aerosol loading is large. There may be some regions, where MODIS is well tuned to expected aerosol types and surface conditions, that the size parameters may be useful. However, this is the exception, and would require further study. Over ocean, MODIS has skill at retrieving size information globally, however, there may be biases in certain regions/seasons that must be characterized.

The MODIS aerosol team will be continually evaluating the size parameter retrievals of C005, and updates will be published in the literature and appended to the end of this document.

6.3.3. Other derived and diagnostic parameters

At this point, the MODIS aerosol team has not evaluated the reliability of other parameters, including certain derived parameters (e.g. Mass Concentration) and diagnostic parameters (cloud fraction). The user is cautioned that there is no expected quantitative value to these parameters.

7. Summary

In this document, we have introduced a new algorithm (C005), intended for deriving aerosol properties from MODIS observed spectral reflectance. In fact, this algorithm is comprised of two separate algorithms, one for deriving aerosol properties over ocean, the other over land. The quality assurance (QA) plan has been overhauled and is now more useful to users within the aerosol community. Operational C005 processing began in April 2006.

The core inversion of the C005 over-ocean algorithm (C005-O) remains similar to the pre-launch MODIS algorithm formulated by Tanré et al., (1997) and ATBD-96, as well as that described for C004 over ocean (C004-O) by Remer et al., (2005). However, updates have been made to the assumed fine and coarse aerosol modes, cloud and other masking, and to the descriptions of the retrieved products. Specifically, the refractive indices of the coarse modes have been changed since the C004-O algorithm.

The algorithm for deriving C005 aerosol over-land (C005-L) has been completely overhauled from the pre-launch algorithm formulated by Kaufman et al., (1997) and ATBD-96 and the C004-L algorithm described by Remer et al., (2005). We have updated a number of assumptions, including the representative global aerosol optical models, the VISvs2.12 surface reflectance parameterization, and the statistical implications of deriving below zero aerosol optical thickness. We also have converted the algorithm over land from an independent two-channel retrieval to a simultaneous three-channel inversion, in order to make use of aerosol information contained in the SWIR (2.12 μm) channel.

The C005 algorithm was tested (as V5.2), both for its theoretical ability to derive realistic aerosol properties, and on a test bed of 6300 MODIS granules. The results were compared to those produced by a version of C004 algorithm (V5.1), and evaluated at co-located AERONET sites. In its provisional validation, V5.2 performed better than V5.1, especially over land. Over land, retrievals τ met the expected accuracy levels ($\pm0.05\pm0.15\tau$) defined by Remer et al., (2005). MODIS/AERONET τ regression had an equation of: $y = 1.01x + 0.03$, $R = 0.90$, with nearly 67% within error bars for a test dataset of 6300 granules. Over ocean, the expected accuracy of MODIS retrieved τ are more stringent ($\pm0.03\pm0.05\tau$), but except for a slightly high bias (regression: $y = 1.07x + 0.01$, $R = 0.96$), the algorithm is performing very well. For the test data set, mean τ over land was reduced from 0.28 in V5.1 to 0.21 in V5.2, whereas over ocean has stayed constant at about 0.135.

Retrievals of aerosol size distribution of the C005 algorithms (e.g. V5.2) were shown to be generally better than those in V5.1, yet no claim is made to their quantitative value and their validation. Retrievals of Fine Model Weighting (η) over land show slight improvement in their correlation to AERONET values, while retrievals of spectral Ångstrom Exponent show little or no improvement. In spite the algorithm overhaul, there is still very little information content for retrieving aerosol size over land and the results should not be considered to be quantitative. The C005 algorithm, however, has shown usefulness, however, in deriving Fine τ ($\tau \times \eta$), which is a product that can be related to the anthropogenic contribution to the total τ (e.g. Kaufman et al., 2005). Over ocean, the changes in coarse mode refractive index have yielded the desired results of lowering the MODIS retrieved η. The high bias of ocean η seen by Kleidman et al., (2005) has been reduced. The retrievals of Ångstrom Exponent have high correlation with AERONET, and their bias has been reduced as well. This means

that C005-O will indicate larger particles than C004-O. Finally, the over-ocean retrievals of fine τ are biased high compared to AERONET values. This may be because there are problems of non-sphericity and under representation of absorption effects that are yet to be understood.

The provisional validation suggests that the operational C005 retrieval will be capable of deriving τ over both land and ocean within expected uncertainty. However, we note that the comparison was performed for products derived using L1B radiance data from C004. The C005 radiance data, including those representing observations prior to 2006, may be revised due to recalibration. We also point out that the provisional validation was performed using data that was filtered using the QA definitions for C004, so that it was easy to compare "new" versus "old." QAC$\geq$1 was required for both over land and over ocean. Any official C005 validation should be filtered using the C005 definitions for QA. Thus, it is necessary to perform global validation only after C005 has been processed for some time.

8. Global Evaluation of C005 products

TBD

9. References

Ahmad, Z. and R. S. Fraser (1981). "An Iterative Radiative Transfer Code For Ocean-Atmosphere Systems." Journal of the Atmospheric Sciences **39**: 656-665.

Al-Saadi, J., J. Szykman, R. B. Pierce, C. Kittaka, D. Neil, D. A. Chu, L. Remer, L. Gumley, E. Prins, L. Weinstock, C. MacDonald, R. Wayland, F. Dimmick, and J. Fishman, 2005: Improving National Air Quality Forecasts with Satellite Aerosol Observations. Bull. Am. Met. Soc., 86 (9), 1249-1261.

Anderson, T.L., Y. Wu, D.A. Chu, B.Schmid, J. Redemann and O. Dubovik, 2006: Testing the MODIS satellite retrieval of aerosol fine-mode fraction, J.Geophys.Res., 110,D18204, doi: 10.1029/2005JD005978.

Bodhaine, B. A., N. B. Wood, et al. (1999). "On Rayleigh optical depth calculations." Journal of Atmospheric and Oceanic Technology **16**(11): 1854-1861.

Chin, M., P. Ginoux, et al. (2002). "Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements." Journal of the Atmospheric Sciences **59**(3): 461-483.

Chu, D. A., Y. J. Kaufman, et al. (2002). "Validation of MODIS aerosol optical depth retrieval over land." Geophysical Research Letters **29**(12): art. no.-1617.

Chu, D. A., L. A. Remer, Y. J. Kaufman, B. Schmid, J. Redemann, K. Knobelspiesse, J. D. Chern, J. Livingston, P. B. Russell, X. Xiong, and W. Ridgway, 2005: Evaluation of aerosol properties over ocean from Moderate Resolution Imaging Spectroradiometer (MODIS) during ACE-Asia. J. Geophys. Res., 110 (D07308), doi: 10.1029/2004/JD005208. (Abstract) (Full Text (PDF))

Chu, D. A., Y. J. Kaufman, G. Zibordi, J. D. Chern, J. Mao, C. Li, and B. N. Holben, 2003: Global monitoring of air pollution over land from EOS- Terra MODIS. J. Geophys. Res, 108 (D21), 4661, doi: 10.1029/2002JD003179.

Chylek, P., B. Henderson, et al. (2003). "Satellite based retrieval of aerosol optical thickness: The effect of sun and satellite geometry." Geophysical Research Letters **30**(11).

Dave, J. V. (1970). "Intensity and Polarization of Radiation Emerging from a Plane-Parallel Atmosphere Containing Monodispersed Aerosols." Applied Optics **9**(12): 2673-&.

Dubovik, O., B. Holben, et al. (2002). "Variability of absorption and optical properties of key aerosol types observed in worldwide locations." Journal of the Atmospheric Sciences **59**(3): 590-608.

Dubovik, O., B. N. Holben, et al. (2002). "Non-spherical aerosol retrieval method employing light scattering by spheroids." Geophysical Research Letters **29**(10): art. no.-1415.

Dubovik, O. and M. D. King (2000). "A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements." Journal of Geophysical Research-Atmospheres **105**(D16): 20673-20696.

Eck, T. F., B. N. Holben, et al. (2005). "Columnar aerosol optical properties at AERONET sites in central eastern Asia and aerosol transport to the tropical mid-Pacific." Journal of Geophysical Research-Atmospheres **110**(D6).

Eck, T. F., B. N. Holben, et al. (1999). "Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols." Journal of Geophysical Research-Atmospheres **104**(D24): 31333-31349.

Evans, K.F. and G. L. Stephens, 1991, A New Polarized Atmospheric Radiative Transfer Model, *J.*

Quant. Spectrosc. Radiat. Transfer, **46**(5):413-423.
Fraser, R. H., Ferrare, R. A., Kaufman, Y. J., Mattoo, S. (1989). Algorithm for Atmospheric Corrections of Aircraft and Satellite Imagery. NASA Technical Memorandum 100751. Greenbelt, MD USA, NASA Goddard Space Flight Center.
Gatebe, C. K., M. D. King, et al. (2001). "Sensitivity of off-nadir zenith angles to correlation between visible and near-infrared reflectance for use in remote sensing of aerosol over land." Ieee Transactions on Geoscience and Remote Sensing **39**(4): 805-819.
Gao, B.-C., Y.J. Kaufman, D. Tanré and R.-R. Li, 2002: Distinguishing tropospheric aerosols from thin cirrus clouds for improved aerosol retrievals using the ratio of 1.38-µm and 1.24-µm channels. Geophys. Res. Lett., 29, 1890, doi:10.1029/2002GL015475.
Gao, Y., Y. J. Kaufman, D. Tanré, and P. G. Falkowski, 2000: Seasonal distribution of aeolian iron fluxes to the global ocean. Geophys. Res. Lett., 28, 29-33.
Holben, B. N., T. F. Eck, et al. (1998). "AERONET - A federated instrument network and data archive for aerosol characterization." Remote Sensing of Environment **66**(1): 1-16.
Holben, B. N., D. Tanré, et al. (2001). "An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET." Journal of Geophysical Research-Atmospheres **106**(D11): 12067-12097.
Hubanks, P. A. (2005). MODIS Atmosphere QA Plan for Collection 005. Greenbelt, MD USA, NASA Goddard Space Flight Center**:** 57.
Ichoku, C., L. A. Remer, et al. (2003). "MODIS observation of aerosols and estimation of aerosol radiative forcing over southern Africa during SAFARI 2000." Journal of Geophysical Research-Atmospheres **108**(D13): art. no.-8499.
Ichoku, C., D. A. Chu, et al. (2002). "A spatio-temporal approach for global validation and analysis of MODIS aerosol products." Geophysical Research Letters **29**(12): art. no.-1616.
Ignatov, A., P. Minnis, N. Loeb, B. Wielicki, W. Miller, S. Sun-Mack, D. Tanré, L. Remer, I. Laszlo, and E. Geier, 2005: Two MODIS Aerosol Products over Ocean on the Terra and Aqua CERES SSF Datasets,. J. Atm. Sci., 62, 1008-1031.
Intergovernmental Panel on Climate Change (IPCC), 2001. Climate Change 2001: The Scientific Basis, J. T. Houghton, Y. Ding, D.J. Griggs, M. Noguer, P. J. van der Linden and D. Xiaosu (Eds.), Cambridge University Press, UK. pp 944.
Kaufman, Y. J., O. Boucher, et al. (2005). "Aerosol anthropogenic component estimated from satellite data." Geophysical Research Letters **32**(17).
Kaufman, Y. J., N. Gobron, et al. (2002). "Relationship between surface reflectance in the visible and mid-IR used in MODIS aerosol algorithm - theory." Geophysical Research Letters **29**(23): art. no.-2116.
Kaufman, Y. J., D. Tanré, et al. (1997). "Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer." Journal of Geophysical Research-Atmospheres **102**(D14): 17051-17067.
Kaufman, Y. J., A. E. Wald, et al. (1997). "The MODIS 2.1-mu m channel - Correlation with visible reflectance for use in remote sensing of aerosol." Ieee Transactions On Geoscience and Remote Sensing **35**(5): 1286-1298.
King, M. D., Y. J. Kaufman, et al. (1992). "Remote-Sensing of Cloud, Aerosol, and Water-Vapor Properties From the Moderate Resolution Imaging Spectrometer (Modis)." Ieee Transactions On Geoscience and Remote Sensing **30**(1): 2-27.
King, M. D., W. P. Menzel, Y. J. Kaufman, D. Tanre, B.-C. Gao, S. Platnick, S. A. Ackerman, L. A. Remer, R. Pincus, and P. A. Hubanks, 2003: Cloud and aerosol properties, precipitable water,

and profiles of temperature and humidity from MODIS. IEEE Trans. Geosci. Remote Sens., 41, 442-458.

Kleidman, R. G., N. T. O'Neill, et al. (2005). "Comparison of moderate resolution Imaging spectroradiometer (MODIS) and aerosol robotic network (AERONET) remote-sensing retrievals of aerosol fine mode fraction over ocean." Journal of Geophysical Research-Atmospheres **110**(D22).

Levy, R. C., L. A. Remer, et al. (2004). "Effects of neglecting polarization on the MODIS aerosol retrieval over land." Ieee Transactions on Geoscience and Remote Sensing **42**(11): 2576-2583.

Levy, R. C., L. A. Remer, et al. (2005). "Evaluation of the MODIS aerosol retrievals over ocean and land during CLAMS." Journal of the Atmospheric Sciences **62**(4): 974-992.

Levy, R. C., L. A. Remer, and O. Dubovik, (2007a) Global aerosol optical properties and application to MODIS aerosol retrieval over land . J. Geophys. Res., **112**, D13210, doi:10.1029/2006JD007815.

Levy, R. C., L. Remer, S. Mattoo, E. Vermote, and Y. J. Kaufman, (2007b). Second-generation algorithm for retrieving aerosol properties over land from MODIS spectral reflectance. J. Geophys. Res., **112**, D13211, doi:10.1029/2006JD007811.

Li, R. R., L. Remer, et al. (2005). "Snow and ice mask for the MODIS aerosol products." Ieee Geoscience and Remote Sensing Letters **2**(3): 306-310.

Lyapustin, A. I. (2001). "Three-dimensional effects in the remote sensing of surface albedo." IEEE Transactions on Geoscience and Remote Sensing **39**(2): 254-263.

Martins, J. V., D. Tanré, et al. (2002). "MODIS Cloud screening for remote sensing of aerosols over oceans using spatial variability." Geophysical Research Letters **29**(12): art. no.-1619.

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N. and Thulasiraman, S.: Spectral discrimination of coarse and fine mode optical depth., J Geophys. Res., 108, doi:10.1029/2002JD002975, 2003.

O'Neill, N. T., Smirnov, A., Holben, B., and Thulasiraman, S. (2005). Spectral Deconvolution algorithm: Technical memo.

Omar, A. H., J. G. Won, et al. (2005). "Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements." Journal of Geophysical Research-Atmospheres **110**(D10).

Remer, L. A. and Y. J. Kaufman (1998). "Dynamic aerosol model: Urban/industrial aerosol." Journal of Geophysical Research-Atmospheres **103**(D12): 13859-13871.

Remer, L. A., Y. J. Kaufman, et al. (2005). "The MODIS aerosol algorithm, products, and validation." Journal of the Atmospheric Sciences **62**(4): 947-973.

Remer, L. A., A. E. Wald, et al. (2001). "Angular and seasonal variation of spectral surface reflectance ratios: Implications for the remote sensing of aerosol over land." Ieee Transactions on Geoscience and Remote Sensing **39**(2): 275-283.

Salomonson, V. V., W. L. Barnes, et al. (1989). "MODIS: Advanced Facility Instrument for Studies of the Earth as a System." IEEE Trans. Geosci. Remote Sensing **27**: 145-153.

Tanré, D., M. Herman, and Y. Kaufman, 1996: Information on the aerosol size distribution contained in the solar reflected spectral radiances. J. Geophys. Res., 101, 19043-19060.

Tanré, D., Y. J. Kaufman, et al. (1997). "Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances." Journal of Geophysical Research-Atmospheres **102**(D14): 16971-16988.

Tucker, C. J.: Red and photographic infrared linear combinations monitoring vegetation, Remote Sensing Environment, 8, 127-150, 1979.

U.S. Government Printing Office, 1976, U.S. Standard Atmosphere, Washington, D.C.

Vermote, E. F., D. Tanré, et al. (1997). "Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview." Ieee Transactions On Geoscience and Remote Sensing **35**(3): 675-686.

Wiscombe, W. J. (1981). "Improved Mie scattering algorithms." Appl. Opt. **19**: 1505-1509.

Yu, H., Y. J. Kaufman, M. Chin, G. Feingold, L. Remer, T. Anderson, Y. Balkanski, N. Bellouin, O. Boucher, S. Christopher, P. DeCola, R. Kahn, D. Koch, N. Loeb, M. S. Reddy, M. Schulz, T. Takemura, and M. Zhou, 2006: A review of measurement-based assessments of aerosol direct radiative effect and forcing. Atmos. Chem . Phys., 6, 613-666.

A1: Gas Correction of L1B reflectance

During the course of MODIS processing, the appropriate *ancillary* data are acquired from National Center for Environmental Prediction (NCEP) analyses. These include the 1° by 1° global meteorological analysis data (created every six hours – format "gdas1.PGrbF00.YYMMDD.HHz") and the 1° by 1° global ozone data (created daily). Before 2005, the daily TOVS data (having format "YYMMDD.grb") represented the global ozone, whereas starting in early 2005, the daily TOAST data (format "TOAST_16.YYMMDD.grb") was used instead. The primary parameters extracted from the NCEP data are total precipitable water vapor (PWAT or w in [cm]), and total column ozone O (in [Dobson units]). In case the NCEP data are missing or invalid, the algorithm instead assumes fixed values of water vapor and ozone optical depths (climatological global averages).

For each pixel in the L1B data, the following gas transmission factors are calculated as a function of wavelength λ, each a function of the air mass factor (G), and some weighting coefficients, K. Table A1 lists the coefficients appropriate for each gas (water vapor, ozone and carbon dioxide). The air mass G is a function of solar and sensor zenith angles, such that

$$G = \frac{1}{\cos(\theta_0)} + \frac{1}{\cos(\theta)}.$$

If NCEP water vapor is valid, the transmission factor $T_\lambda^{H_2O}$ for water vapor is:

$$T_\lambda^{H_2O} = \exp(\exp(K_{1,\lambda}^{H_2O} + K_{2,\lambda}^{H_2O}\ln(Gw) + K_{3,\lambda}^{H_2O}(\ln(Gw))^2)).$$

If the NCEP water vapor is missing, then

$$T_\lambda^{H_2O} = \exp(G\overline{\tau_\lambda^{H_2O}}).$$

Ozone transmission $T_\lambda^{O_3}$ is calculated in a similar way, that is:

$$T_\lambda^{O_3} = \exp(GK_\lambda^{O_3}O),$$

for a valid NCEP value, and

$$T_\lambda^{O_3} = \exp(G\overline{\tau_\lambda^{O_3}}),$$

when it is not.

Carbon dioxide optical depth is assumed fixed globally, so that its transmission is:

$$T_\lambda^{CO_2} = \exp(G\overline{\tau_\lambda^{CO_2}}).$$

Therefore, The total gas transmission factor is a multiplication of all individual gas transmission terms, that is

$$T_\lambda^{gas} = T_\lambda^{CO_2}T_\lambda^{O_3}T_\lambda^{H_2O},$$

and the corrected reflectance is given by

$$\rho_\lambda^m = T_\lambda^{gas}\rho_\lambda^{L1B}$$

TABLE A1: GAS ABSORPTION COEFFICIENTS AND CLIMATOLOGY

Wavelength	$K_{1,\lambda}^{H_2O}$	$K_{2,\lambda}^{H_2O}$	$K_{3,\lambda}^{H_2O}$	$\overline{\tau_{\lambda}^{H_2O}}$	$K_{\lambda}^{O_3}$	$\overline{\tau_{\lambda}^{O_3}}$	$\overline{\tau_{\lambda}^{CO_2}}$
0.47					4.26E-06	2.432E-03	
0.55					1.05E-04	2.957E-02	
0.66	-5.73888	0.925534	-0.0188365	1.543E-02	5.09E-05	2.478E-02	
0.86	-5.32960	0.824260	-0.0277443	1.947E-02			
1.24	-6.39296	0.942186	-0.0131901	1.184E-02			4.196E-04
1.64	-7.76288	0.979707	0.007784	9.367E-03			8.260E-03
2.12	-4.05388	0.872951	-0.0268464	5.705E-02			2.164E-02

Note that the K-coefficients are used when NCEP data are valid, whereas the global average τ-values are used when NCEP data are missing. In case of carbon dioxide, global average τ-value is always assumed

A2: Masking over land and ocean

Masking clouds without masking aerosol events remains one of the most challenging issues faced by the aerosol retrieval algorithms. At Terra launch, the aerosol retrievals relied on the standard cloud mask products available in M?D35. Almost immediately we realized that these products were not going to be adequate. We learned that the 1.38 µm reflectance could be used to identify cirrus in the pixel. This channel is especially sensitive to cirrus clouds because in the absence of particles high in the atmosphere the channel returns reflectances near zero due to the strong water vapor absorption at this wavelength. A new cloud mask based on a spatial variability test, supplemented by cirrus tests using the 1.38 µm channel and a few remaining M?D35 products, was implemented in the over ocean aerosol algorithm (Martins et al., 2002). The mask proved to be very successful, especially after adjustments to the cirrus identification part of the algorithm (Gao et al. 2002). All of Collection 004 data over ocean, both from Terra and Aqua, were produced using this cloud mask.

A separate but similar cloud mask for masking clouds over land was developed later and not implemented until November 2002. The spatial variability cloud mask over land improved the aerosol retrievals, especially when it came to confusing heavy aerosol with cloud. However, isolated, residual cloud contamination in the product remained. For Collection 005, we have made a few adjustments to the technique, but maintained the general philosophy and structure of using spatial variability tests coupled with threshold tests only in the 1.38 µm channel. As it turns out, near zero reflectance at 1.38 µm could either be slightly positive or slightly negative. The C004 aerosol algorithm required incoming reflectances in this channel to be non-negative. C005 relaxes this requirement to allow small magnitude negative values, resulting in the recovery of many aerosol retrievals, especially over land.

The following describes the steps for masking cloud and other unsatisfactory pixels when performing aerosol retrieval over land or ocean. Common to land and ocean are that the L1B reflectance data ρ_λ^{L1B} are corrected for gas absorption (Appendix 1). If any of the pixels within the box are considered 'land' by the M?D35 cloud mask product, then the algorithm proceeds with the land retrieval algorithm. Only if all pixels in the box are 'ocean' is the ocean inversion performed. Separate cloud masks are used over land and ocean. Additional masks over land include snow and ice contamination as well as residual water bodies (such as lakes and swamps). The ocean retrievals require masking for glint and underwater sediments.

A2.2 Masking over ocean

Cloud Masking Over Ocean

If all 400 pixels in the 10km box are considered to be 'water' then the ocean algorithm is followed. After the gas correction described in section A1.1, the algorithm has the arduous task of separating 'good' pixels from 'cloudy' pixels. The standard M?D35 cloud mask includes using the brightness in the visible channels to identify clouds. This procedure will mistake heavy aerosol as 'cloudy', and miss retrieving important aerosol events over ocean. On the other hand, relying on IR-tests alone permits low altitude, warm clouds to escape and be misidentified as 'clear', introducing cloud contamination in

the aerosol products. Thus, our cloud mask over ocean combines spatial variability tests (e.g. Martins et al., 2002) along with tests of brightness in visible and infrared channels.

The algorithm marches through the 10 by 10 box, examining the standard deviation of $\rho_{0.55}$ in every group of 3 by 3 pixels. Any group of 9 pixels with standard deviation greater than 0.0025 is labeled as 'cloudy', and all 9 pixels in the group are discarded (Martins, et al. 2002). The only exception to this rule is for heavy dust, which may at times be as spatially inhomogeneous as clouds. Heavy dust is identified by its absorption at 0.47 µm using the ratio ($\rho_{0.47}/\rho_{0.66}$). This quantifies the difference that our eyes witness naturally. Dust absorbs at blue wavelengths and appears brown. Clouds are spectrally moderately absorbing and appear white to our eyes. If $\rho_{0.47}/\rho_{0.66} < 0.75$, then the central pixel of the group of 9 is identified as 'dust' and will be included in the retrieval even if it is inhomogeneous. This is a conservative threshold that requires very heavy dust in order to avoid clouds. Less restrictive thresholds would permit more dust retrievals, but might accidentally permit cloud contamination.

The spatial variability test separates aerosol from most cloud types, but sometimes fails at the centers of large, thick clouds and also with cirrus, both of which can be spatially smooth. The centers of large, thick clouds are very bright in the visible, and so we identify these clouds when $\rho_{0.47} > 0.40$. This is an extremely high threshold that translates into an aerosol optical thickness greater than 5.0, but only for non-absorbing aerosol. Absorbing aerosol never reaches that high value of reflectance and will pass this cloud test unscathed. Some high values of non-absorbing aerosol may be discarded along with bright clouds, but this confusion is rare. Most heavy aerosol loading, with $\tau > 5.0$, absorbs somewhat at 0.47 µm and fails to reach the 0.40 threshold value, exhibited by very bright white clouds.

Cirrus clouds are identified with a combination of infrared and near-infrared tests. Three infrared tests provided by the standard MODIS cloud mask, M?D35, are examined. These tests are IR cirrus test (byte 2, bit 4), 6.7 µm test (byte 2, bit 8) and Delta IR test (byte 3, bit 3) (Ackerman, et al. 1998). If any one of the 3 indicates cloud, we label the pixel as 'cloudy'. The near-infrared cirrus test is based on the reflectance in the 1.38 µm channel and the ratio $\rho_{1.38} / \rho_{1.24}$ (Gao, et al. 2002). It is applied in the algorithm as a three step process. If $\rho_{1.38} / \rho_{1.24} > 0.3$, then the pixel is 'cloudy'. If $0.10 \leq \rho_{1.38} / \rho_{1.24} \leq 0.30$ and $\rho_{1.38} > 0.03$ and $\rho_{0.66} > 1.5\rho^{ray}_{.0.66}$, then the pixels is also 'cloudy'. However, if $0.10 \leq \rho_{1.38} / \rho_{1.24} \leq 0.30$ and $0.01 \leq \rho_{1.38} \leq 0.03$ and $\rho_{0.66} > 1.5\rho^{ray}_{.0.66}$, then the situation is ambiguous. The algorithm labels the pixel as 'not cloudy' and will include the pixel in the retrieval process, but the quality of the retrieval (Quality Assurance Confidence - QAC) is reduced to 0, 'poor quality'. This permits aerosol retrieval at the orbital level (Level 2), but prohibits the retrieval from contributing to the long-term global aerosol statistics (Level 3). Only retrievals with QAC > 0, contribute to the Level 3 Quality Weighted products. The products and product levels are explained further in Section 3. If the reflectance at 0.66 µm ($\rho_{0.66}$) does not exceed 1.5 times the Rayleigh reflectance in that channel ($\rho^{ray}_{.0.66}$) or the reflectance at 1.38 µm does not exceed 0.01, then the pixel is assumed to be 'not cloudy' with no ambiguity, unless the ratio $\rho_{1.38} / \rho_{1.24}$) exceeds 0.3.

Ocean sediment mask

The final mask applied to the data is the sediment mask, which identifies which ocean scenes are contaminated by river sediments (Li et al, 2002) and discards those pixels. The sediment mask takes advantage of the strong absorption by water at wavelengths longer than 1 µm. The resulting spectral

reflectances over water with suspended sediments thus show elevated values in the visible, but not in the longer wavelengths. This creates a unique spectral signature quite different from clear ocean water and also different from airborne dust.

A2.2 Masking over land

Cloud Masking Over Land

The algorithm generates a cloud mask over land using spatial variability of the 0.47 μm (>0.0025) and 1.38 μm (> 0.003) channel reflectance, as well as the absolute value of the 0.47 μm reflectance (>0.4) and the 1.38 μm reflectance (> 0.025). The combination of the two channels yields information about both visibly bright thick clouds and visibly dim thin cirrus. Note, that the refl_1.38 < 0.025 threshold allows cirrus contamination into the land aerosol retrieval. However, those retrievals will have QAC (Quality Assurance Confidence) set to zero.

The visible channel reflectance spatial variability test for land is similar to that for ocean (as described by Martins et al., (2002)), although the 0.47 μm channel (at 500 m resolution) is used instead. The algorithm calculates the absolute standard deviation of the reflectance of each group of 3 x 3 pixels. If that standard deviation is greater than 0.0025, then the area of the entire 3x3 pixel box is considered a cloud. This means that regardless of the resolution, reflectances in all channels are considered cloud contaminated for that 1.5 km x 1.5 km area, and are masked.

The visible channel reflectance threshold test is also performed on the 0.47 μm channel. These are performed pixel by pixel (500 m resolution). If the reflectance at 0.47 μm is greater than 0.4, the pixel is considered a cloud, and all channels are cloud masked for the 500 m pixel.

The 1.38 μm channel reflectance spatial variability and threshold tests are analogous to those in the visible, however, as the resolution is 1 km, the tests cover a larger area. If the spatial variability test finds that the 3x3 pixel standard deviation is greater than 0.003, then the entire 3 km x 3 km area is cloud masked for all channels. The reflectance threshold for the 1.38 μm channel is 0.01.

The final cloud mask is the union of the four cloud mask tests described above. Figure A2.1 shows how C005-L cloud mask is improved from C004-L. The masked pixels are not considered for the over land retrieval.

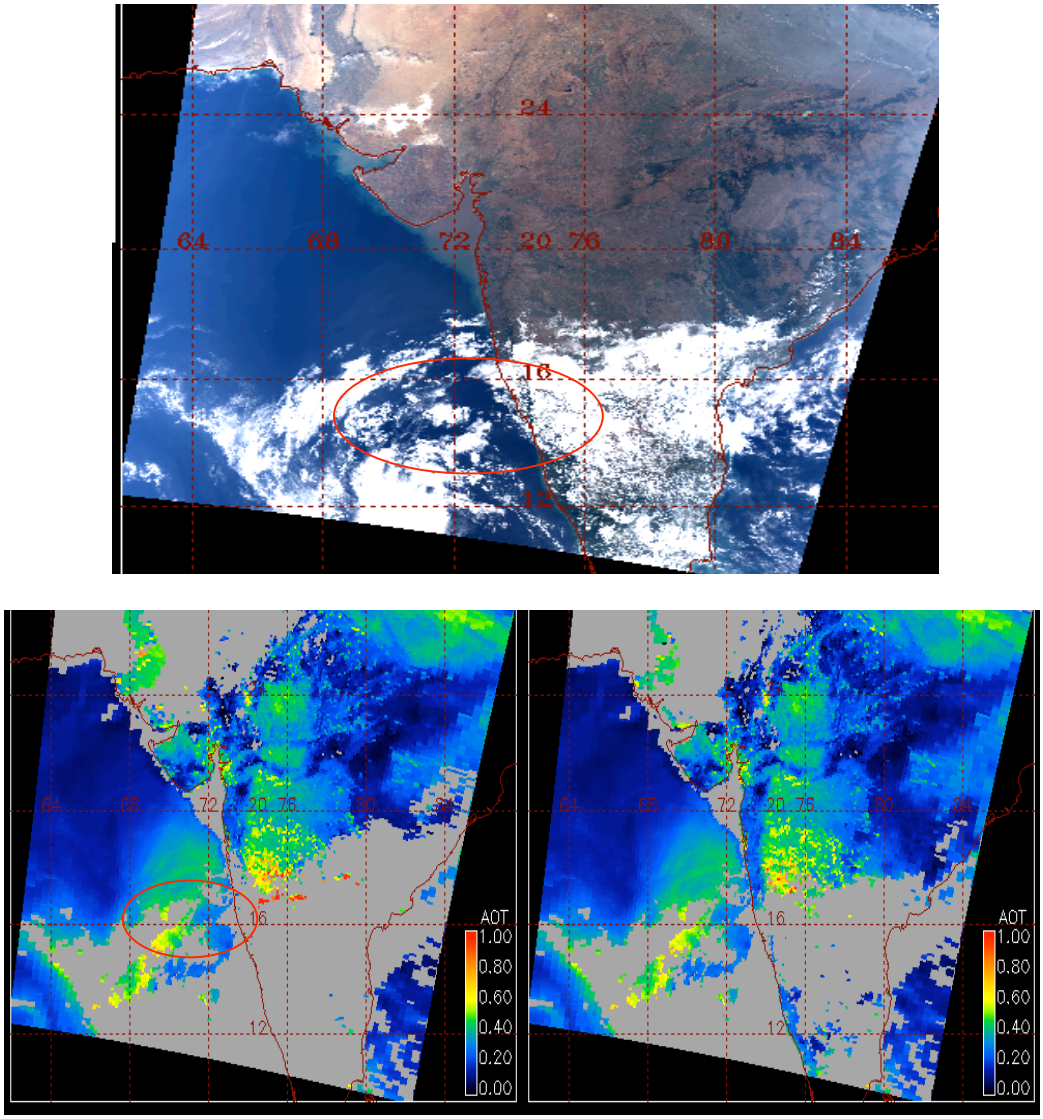

Figure A2.1. Images from MOD04 from year 2000, day 337, time 0555 located over India. The top image is a true color RGB with the red oval highlighting the edge of a cloudy area. The same area is identified by the oval in the lower left image, which shows aerosol optical thickness retrieval available in Collection 004. The bright red spots are artifacts of cloud contamination. These are eliminated in the lower right image, which was produced with Collection 005 software and improved cloud masking logic over land. The increased number of retrievals in the Collection 005 image result from permitting negative values in the 1.38 µm channel.

Snow and inland water masking

A number of other tests are performed at a variety of resolutions to remove contamination by "wet" pixels, including snow fields, swamps, and inland water bodies. These pixels would be expected to have poorly behaved VISvs2.12 surface reflectance relationships. Note that the reflectance data have already been cloud masked and corrected for gas absorption.

The snow/ice mask determined by an NDVI-like ratio of 0.86 μm and 1.24 μm channel reflectances (at 500 m resolution), i.e. $ratio = (\rho^m_{0.86} - \rho^m_{1.24})/(\rho^m_{0.86} + \rho^m_{1.24})$, and the brightness temperature of the 11 μm MODIS thermal band (channel 31 – interpolated to 500 m resolution). If the ratio is greater than 0.1 and the temperature is less than 285K, then the pixel is considered a snow or ice contaminated and masked [Li et al., 2005]. Figure A2.2 shows an example of improvements gained by the snow mask.

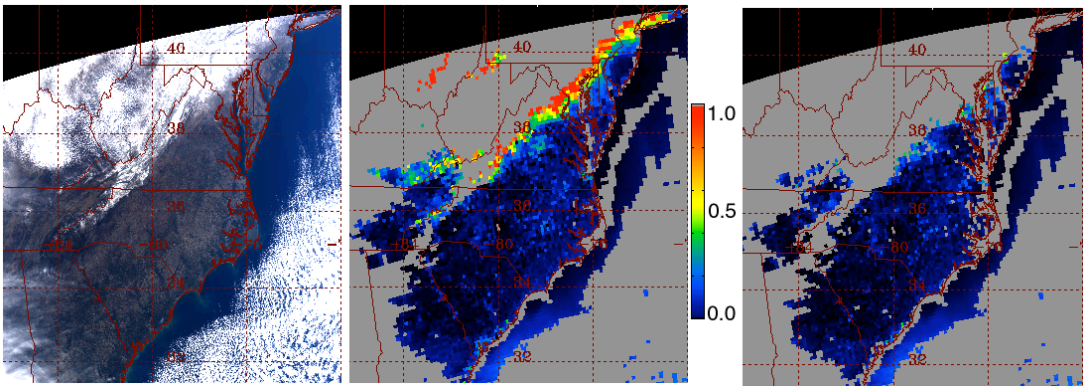

Fig. A2.2. (a) – An Aqua MODIS image over North America on February 8, 2004; (b) the derived aerosol optical depth image using the Collection 004 MODIS aerosol algorithm; and (c) the Collection 005 aerosol optical depth image with improved snow masking.

The inland water mask is determined by computing the traditional NDVI for the 0.66 μm and the 0.86 μm channels 250 m resolution, i.e. $NDVI = (\rho^m_{0.66} - \rho^m_{0.86})/(\rho^m_{0.66} + \rho^m_{0.86})$. If the NDVI value is greater than 0.1 it is considered an inland water body.

A3: Table of Run Time QA Flags of M?D04 Level 2 Aerosol Products

This appendix updates the information for dark target aerosol flags, given in the QA plan for C005 (MODIS Atmosphere QA Plan for Collection 005 (Deep Blue Aerosol Update)) available online at (http://modis-atmos.gsfc.nasa.gov/reference_atbd.php).
The Aerosol (dark target) run time *Quality Assurance* (QA) flags are stored as three Scientific Data Sets (SDSs); the *Cloud_Mask_QA*, *Quality_Assurance_Land*, and *Quality_Assurance_Ocean*. *Cloud_Mask_QA* is a single byte SDS that contains several cloud mask QA flags (from MOD35) recomputed on the 10x10 km aerosol grid. Each of the two aerosol SDSs are five bytes that provide information on the processing (logic) path taken during the aerosol retrieval. The aerosol QA includes product quality flags, retrieval processing flags, and input data resource flags which are designed separately for land and ocean because of the differences of retrieval algorithms. Particular flags may indicate: a) conditions why retrieval was not attempted at all (e.g. input data outside of boundary conditions), b) cases where input data quality may be poor (e.g. large cloud fraction), so that the retrieval is performed with lower confidence, or c) cases where retrieval may have been performed but the results were poor (e.g. results outside of realistic physical conditions).

The Quality Assurance *confidence* (QAC) flags summarize the QA logic, and are referred to in the main text of this ATBD. The QAC flags are the 'Estimated quality flag of aerosol optical thickness at 0.47 μm' for land and the 'Estimated quality of aerosol parameter of average solution' for ocean retrievals.

All Aerosol QA Flag arrays have the following characteristics:

- Spatial resolution: 10 ´ 10 km
- Processing mode: Day time mode only

The following tables describe the byte decoding of the M?D04 Cloud_Mask_QA, Quality_Assurance_Land, and Quality_Assurance_Ocean SDSs. Each flag corresponds to a certain number of bits, and bit values corresponding to results of certain tests. Note that the flags representing the case of valid retrieval but lower confidence is known as "Part I" over land, but "Part II" over ocean. Similarly the flags representing the case of no valid retrieval are known as Part II over land, but Part I over ocean. Under the column "Comments", we list how other flags may be reset (if applicable). For example, if Part I over land receives the value of 8 (less than optimal clear sky pixels) then the QAC bits will be set to 2 (good quality).

Cloud mask QA flags

Flag name	# of bits	Bit value	Description	Comments
Cloud mask summary flag	1	0 1	Undetermined Determined	
Cloud mask quality flag	3	0 1 2 3	0-30% cloudy pixels 30-60% cloudy pixels 60-90% cloudy pixels >90% cloudy pixels	
Snow/Ice flag	1	0 1	Yes No	
Land/Sea flag	3	0 1 2 3	Water (Ocean and Lake) Coastal Desert Land	

Product quality and retrieval processing QA flags over land

Flag name	# of bits	Bit value	Description	Comments
Product quality QA summary flags				
Summary quality flag for aerosol optical thickness at 0.47 μm	1	0 1	Not useful data Useful	"QA usefulness flag" 0) All products are fill values 1) Valid products
Estimated quality flag of aerosol optical thickness at 0.47 μm	3	0 1 2 3 4-7	Poor Marginal Good Very Good Not Used (TBD)	"QA Confidence flag" (QAC)
Summary quality flag for aerosol optical thickness at 0.66 μm	1	0 1	Not useful data Useful	Same value as 0.47 usefulness
Estimated quality flag of aerosol optical thickness at 0.47 μm	3	0 1 2 3 4-7	Poor Marginal Good Very Good Not Used (TBD)	Same value as 0.47 QAC flag
Retrieval processing QA flags - Processing path flags				
Part I: retrieving condition flags when inversion is performed - retrieved value will be output	4	0 1 2 3 4 5 6 7 8 9 10 11	Retrieval performed normally (no issues) Procedure 2 performed (semi-bright surface) Water pixels in 10 x10 box Cirrus present Fitting error > 0.25 -0.1 < Retrieved τ < 0.0 # pixels between 12 & 20 # pixels between 21 & 30 # pixels between 31 & 50 Ångstrom out of bounds Retrieved τ < 0.2 No Retrieval	1) QAC=3 2) QAC=0 3) QAC=0 4) QAC=0 5) QAC=0 6) QAC=3 7) QAC=0 8) QAC=1 9) QAC=2 10) QAC=0 11) QAC=3 12) QAC=0
Part II: retrieving condition flags when inversion is NOT performed – fill values are output	4	0 1 2	No error Solar/sensor geometry out of bounds in LUT Apparent reflectance out of bounds in LUT	QAC=0 QA Useful flag = 0

		3 4 5 6	# pixels < 12 ρ2.1 μm > 0.35 (too bright) Retrieved τ < -0.1 Retrieved τ > 5.0	
Aerosol Type	2	0 1 2 3	All empty	Not currently filled
Thin cirrus or stratospheric aerosol index	2	0 1 2 3	All empty	Not currently filled
Retrieval processing QA flags - Input data resource flags				
Total ozone	2	0 1 2 3	TOVS TOMS Climatology DAO	
Total perceptible water	2	0 1 2 3	NCEP/GDAS MOD05 - NIR Climatology DAO	
Snow cover	2	0 1 2-3	MOD35-cloud mask MOD10-L3 8 day product. TBD	
Spare	6		TBD	

Product quality and retrieval processing QA flags over ocean

Flag name	# of bits	Bit value	Description	Comments
Product quality QA summary flags				
Summary quality flag	1	0 1	Not useful Useful	"QA usefulness flag" 0) products are fill values 1) valid products
Estimated quality of aerosol parameters of "best" solution	3	0 1 2 3 4-7	Poor Marginal Good Very Good Not Used (TBD)	"QA Confidence flag" (QAC)
Summary quality flag	1	0 1	Not useful Useful	"QA usefulness flag" 0) products are fill values 1) valid products
Estimated quality of aerosol parameter of "average" solution	3	0 1 2 3 4-7	Poor Marginal Good Very Good Not Used (TBD)	"QA Confidence flag" (QAC) average solution used for joint product
Retrieval processing QA flags - Processing path flags				
Part I: retrieving condition flags when inversion is NOT performed - fill values are output	4	0 1 2 3 4 5 6 7 8 9 10 11-15	Retrieval is performed Glitter present Cloudy ρ(0.86 μm) too low for retrieving τ Number of valid VIS/SWIR channels (0.55-1.24 μm) is insufficient Number of valid channels < 3 Geometry out of bounds Land pixels in 10 x 10 km box Retrieved τ < -0.01 Retrieved τ > 5.0 No valid reflectance data in any channel TBD	QAC = 0 QA usefulness = 0
Part II: retrieving condition flags when inversion is performed -	4	0 1	Retrieval performed normally Number of pixels within 10×10 km box is < 10%	0) QA=3 1) QA=1

retrieved value will be output		2 3 4 5 6 7 8 9 10 11 12 13 14 15	ρ(0.86 μm) marginal for retrieving τ; the size distribution η = fill value 1.65 μm channel not used 2.13 μm channel not used 2.13 & 1.65 μm not used Large uncertainty in both retrieved τ and aerosol type Large uncertainty in retrieved τ, but aerosol type is stable. The best value of ε is larger than the threshold value (5%) $-0.01 < \tau$ (550 nm) < 0 but to avoid bias in level 3 product Glint angle between 30° and 40° Glint: store only reflectance, SD and Number of pixels used Glint thick dust retrievals Possible cirrus contamination Off glint thick dust retrievals No retrieval performed	2) QA = 0 3) QA=1 4) QA=1 5) QA=0 6) QA=1 7) QA=2 8) QA=1 9) QA=0 10) QA=1 11) QA=0 12) QA=0 13) QA=0 14) QA=2 15) QA=0
Retrieval processing QA flags - Input data resource flags				
Total ozone	2	0 1 2 3	TOVS TOMS Climatology DAO	
Total perceptible water	2	0 1 2 3	NCEP/GDAS MOD05 - NIR Climatology DAO	
Snow cover	2	0 1 2-3	MOD35-cloud mask MOD10-L3 8 day product. TBD	
Spare	2		TBD	

A4: Calculation of Mass Concentration

The following equations lead to derivation of Mass Concentration in units of [μg per cm^2]. In these equations: $dN/d\ln r$ is the number size distribution with r denoting radius (in μm). For a lognormal mode, r_g is the geometric mean radius, σ is $\ln\sigma_g$ representing the standard deviation of the radius, and N_0 is the number of particles per cross section of the atmospheric column (i.e. the amplitude of the lognormal size distribution). In our case, we assume that the distribution is normalized, so that N_0=1.

The number N is related to the volume V and area A distributions by:

$$\frac{dN}{d\ln r} = \frac{3}{4}\pi r^{-3}\frac{dV}{d\ln r} = \pi r^{-2}\frac{dA}{d\ln r}.$$

N_0, V_0, and A_0 are the amplitudes of the corresponding distributions, i.e.

$$V_0 = \int_0^\infty \frac{dV}{d\ln r} d\ln r \quad N_0 = \int_0^\infty \frac{dN}{d\ln r} d\ln r \quad A_0 = \int_0^\infty \frac{dA}{d\ln r} d\ln r.$$

For a single lognormal mode defined by:

$$\frac{dN}{d\ln r} = \frac{1}{r}\frac{N_0}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\ln(r/r_g)^2}{2\sigma^2}\right)$$

and

$$N_0 = V_0\frac{3}{4}\pi r_g^{-3}\exp\left(-\frac{9}{2}\sigma^2\right),$$

the Moments of order k, M^k are defined as

$$M^k = \int_0^\infty r^k \frac{dN}{d\ln r} d\ln r = (r_g)^k \exp(0.5k^2\sigma^2).$$

The effective radius r_{eff} in [μm] is defined by the moments, i.e.

$$r_{eff} = \frac{M^3}{M^2} = \frac{\int_0^\infty r^3 \frac{dN}{d\ln r} d\ln r}{\int_0^\infty r^2 \frac{dN}{d\ln r} d\ln r} = \frac{3}{4}\frac{V_0}{A_0} = r_g \exp(\frac{5}{2}\sigma^2)$$

The extinction coefficient, β_{ext} is related to the extinction efficiency Q_{ext} through the area distribution, and is specific to each mode

$$Q_{ext} = \frac{\beta_{ext}}{\int_0^\infty \pi r^2 \frac{dN}{d\ln r} d\ln r}$$

These parameters are calculated via Mie code (MIEV, [Wiscombe, 1980]). Note that the scattering coefficient β_{sca} and efficiency Q_{sca} are related the same way. The mass extinction coefficient B_{ext} is in

units of [area per mass] and depends on the extinction efficiency and the particle density ρ (assumed to be 1 g per cm^3), such that

$$B_{ext} = \frac{3Q_{ext}}{4\rho r_{eff}}.$$

(Chin et al., 2002). For a single lognormal mode,

$$B_{ext} = \frac{3Q_{ext}}{4\rho r_{eff}} = \frac{3}{4}\frac{Q_{ext}}{\rho}\frac{A_0}{(3/4)V_0} = \frac{Q_{ext}}{\rho}\frac{A_0}{V_0} = \frac{\beta_{ext}}{\rho V_0} = \frac{3\beta_{ext}}{4\rho\pi r_g^3 \exp(4.5\sigma^2)}.$$

However our aerosol models are sums of multiple modes, so that the area and volume distributions must take into account the contributions of each mode. If there are two modes, (i.e. modes 1 and 2), r_{eff} must be calculated this way:

$$r_{eff} = \frac{\int_0^\infty r^3 \frac{(dN_1 + dN_2)}{d\ln r} d\ln r}{\int_0^\infty r^2 \frac{(dN_1 + dN_2)}{d\ln r} d\ln r}$$

Similar modifications are made when calculating Q and thus B.
We can define the Mass Concentration conversion factor, M_c, as the inverse of B, such that

$$M_c = \frac{1}{B}.$$

The columnar mass concentration, M [mass per area] is then defined as

$$M = \tau M_c = \frac{\tau}{B}$$

The final columnar mass concentration product is a weighted combination of the fine and coarse model τs (τ^f and τ^c) and the mass concentration coefficients of each model, i.e.,

$$M = \tau^f M_c^f + \tau^c M_c^c$$

TABLE A4: EXTINCTION PROPERTIES OF THE AEROSOL MODELS USED FOR THE V5.2 OVER-LAND LOOKUP TABLE

Model	ω_0	Q_{ext} **[unitless]**	r_{eff} **[μm]**	β_{ext} **[μm^2]**	B_{ext} **[m^2/g]**	M_c **[μg/cm^2]**
Continental	.8860	0.6210	0.292625	.0010	1.5910	62.8600
Moderately Absorbing / Developing World	.9200	1.0180	0.261287	.0370	2.9220	34.2230
Non-absorbing / Urban-Industrial	.8690	0.9770	0.256210	.0300	3.5330	28.3070
Absorbing / Smoke	.9470	1.1720	0.207507	.0580	3.4310	29.1460
Spheroid / Dust	.9530	1.3390	0.679582	.5450	1.4770	67.6960

Listed for each model are the single scattering albedo, extinction efficiency, effective radius, extinction coefficient, mass extinction coefficient and mass concentration conversion factor. These parameters are defined at 0.55 μm, for $\tau_{0.55} = 0.5$. The particle density is assumed to be 1 g /cm^3.