The Deep Blue aerosol project: Aerosol retrievals from AVHRR

Data product user guide

Contents

1 Introduction

This document is designed to provide information of relevant to users of the AVHRR 'Deep Blue' aerosol data set. For the initial version (V001), we have processed several years of data from the AVHRR sensors aboard the NOAA11, NOAA14, and NOAA18 satellites; specifically, 1989-1990, 1995-1999, and 2006-2011 respectively. The platforms and years considered may be expanded in the future.

Note: this AVHRR data set should be considered a demonstration research-level data product, still under validation, and comments/suggestions are welcome.

The primary data products are orbit-level (level 2, L2) and daily/monthly gridded aggregates (level 3, L3) of aerosol optical thickness (AOT), also known as aerosol optical depth (AOD), at reference wavelengths of 550 nm and at AVHRR band 1 (~630 nm). Other geophysical quantities and related ancillary information, such as pixel-level quality assurance (QA) flags, are also provided.

All data files are in NetCDF4 format and include metadata compliant with the Climate and Forecast (CF) conventions version 1.6. Note that NetCDF4 files can also generally be accessed with libraries for HDF5.

1.1 Revision history

This document has been revised as follows. The date of the document is indicated in the filename, and the latest version is available online from https://portal.nccs.nasa.gov/datashare/AVHRRDeepBlue

- 03 April 2017 Initial version for V001 release
- 12 April 2017 Minor textual updates

1.2 Algorithm background

The Deep Blue AOD retrieval algorithm family consists of two main algorithms: Deep Blue (DB) over land, and the Satellite Ocean Aerosol Retrieval (SOAR) over ocean. DB itself is further split, having two different methods by which surface reflectance is estimated, dependent upon surface brightness/type. These algorithms have heritage in similar algorithms applied previously to generate AOD records from sensors such as SeaWiFS, MODIS, and VIIRS.

More information on the DB and SOAR algorithms as applied to AVHRR is provided in the following papers:

- N. C. Hsu, J. Lee, A. M. Sayer, N. Carletta, S.-H. Chen, C. J. Tucker, and S.-C. Tsay, Retrieving Global Aerosol Properties over Land and Ocean from AVHRR, *submitted to J. Geophys. Res. Atmos.*
- M. Sayer, N. C. Hsu, J. Lee, N. Carletta, S.-H. Chen, and A. Smirnov, Evaluation of NASA Deep Blue/SOAR aerosol retrieval algorithms applied to AVHRR measurements, *submitted to J. Geophys. Res. Atmos.*

Additional information, and links to other relevant papers, can be found at https://deepblue.gsfc.nasa.gov. Several are also mentioned in the References section at the end of this document.

1.3 Quality flags and data use recommendations

Quality assurance (QA) flags in the level 2 products, sometimes also called confidence flags (CF), are used to identify is there is a suspected problem with an individual retrieval. Examples of this include scenes suspected of cloud contamination or where the retrieval solution was not able to find a good match to the input measurements.

These QA flags take integer values from 1 (worst) to 3 (best). QA=0 indicates no retrieval was performed for a pixel. For scientific applications, **we recommend using only data with a QA flag value of 2 of 3**. Retrievals with QA=1 are in general only useful for imagery where data coverage is more important than accuracy, or identifying regions where retrievals are persistently difficult to perform. Level 3 (daily/monthly aggregates) are only generated using retrievals with a QA flag of 2 or 3.

Location and meanings of the QA flags within the files are described in Section 3. In addition, note that the files contain several “best estimate” data sets which are pre-filtered to remove QA=1 retrievals. These are recommended for most users.

1.4 Contact information and citation for data use

If you have general questions or comments regarding our data products, please email them to Dr. Jaehwa Lee or Dr. Andrew Sayer (jaehwa.lee@nasa.gov and andrew.sayer@nasa.gov). More information is also available on the Deep Blue aerosol project website, https://deepblue.gsfc.nasa.gov.

Deep Blue data are free to the public to use. If you make use of our data in a publication or report, we request that you cite the relevant paper(s) for the specific data set(s) used. In the case of AVHRR, these are:

- N. C. Hsu, J. Lee, A. M. Sayer, N. Carletta, S.-H. Chen, C. J. Tucker, and S.-C. Tsay, Retrieving Global Aerosol Properties over Land and Ocean from AVHRR, *submitted to J. Geophys. Res. Atmos.*
- M. Sayer, N. C. Hsu, J. Lee, N. Carletta, S.-H. Chen, and A. Smirnov, Evaluation of NASA Deep Blue/SOAR aerosol retrieval algorithms applied to AVHRR measurements, *submitted to J. Geophys. Res. Atmos.*

Note: this AVHRR data set should be considered a demonstration research-level data product, still under validation, and comments/suggestions are welcome.

If a significant portion of our data is used in your publication, offers of co-authorship are also appreciated. In this case, please contact Dr. N. Christina Hsu (christina.hsu@nasa.gov), the PI of the Deep Blue aerosol project.

In addition to citation, the following text can be used in an Acknowledgements or Data Availability section of a paper:

We thank the Deep Blue science team (https://deepblue.gsfc.nasa.gov) for the AVHRR Deep Blue aerosol data record.

2 Data organization

Level 2 (L2) files are available as 5-minute 'granules' along the orbit track. Level 3 (L3) files are available as gridded aggregates in daily and monthly timesteps.

2.1 File naming convention

For level 2 files, a sample filename is as follows:

DBAER_avhrrnoaa18_D20090225T1241_C20170320132046000000_V001.nc

This is interpreted as follows:

- **DBAER** indicates the data product type (Deep Blue aerosol).
- **avhrrnoaa18** indicates the instrument and platform (here, AVHRR aboard NOAA18).
- **D20090225** indicates the date (YYYYMMDD) at which the granule begins.
- **T1241** indicates the time (HHMM UTC) at which the granule begins. Granules are 5 minutes long.
- **C20170320132046000000** indicates the date and time (UTC) at which the file was created.
- **V001** indicates the processing version of the file (here, version 1 of the AVHRR Deep Blue data release).
- **nc** indicates a NetCDF4 file.

For level 3 daily files, a sample filename is as follows:

DBAER_avhrrnoaa18_daily_1deg_D20060101_V001.nc

This is interpreted as follows:

- **DBAER** indicates the data product type (Deep Blue aerosol).
- **avhrrnoaa18** indicates the instrument and platform (here, AVHRR aboard NOAA18).
- **daily** indicates that this is a daily L3 file.
- **1deg** indicates that the spatial resolution of the file is 1 degree.
- **D20060101** indicates the date (YYYYMMDD) that the file covers.
- **V001** indicates the processing version of the file (here, version 1 of the AVHRR Deep Blue data release).
- **nc** indicates a NetCDF4 file.

For level 3 monthly files, a sample filename is as follows:

DBAER_avhrrnoaa18_monthly_1deg_D200601_V001.nc

This is interpreted as follows:

- **DBAER** indicates the data product type (Deep Blue aerosol).
- **avhrrnoaa18** indicates the instrument and platform (here, AVHRR aboard NOAA18).
- **monthly** indicates that this is a monthly L3 file.
- **1deg** indicates that the spatial resolution of the file is 1 degree.
- **D200601** indicates the date (YYYYMM) that the file covers.

- **V001** indicates the processing version of the file (here, version 1 of the AVHRR Deep Blue data release).
- **nc** indicates a NetCDF4 file.

2.2 File format and structure

Each data file for both L2 and L3 data is in NetCDF format, compliant with climate and forecast (CF) conventions version 1.6. Each file contains multiple Scientific Data Sets (SDS), listed in Section 3.

2.3 L2 production and filtering

Each L2 file contains data from a 5-minute portion of a single AVHRR swath. We produce L2 data at a resolution of 2x2 native AVHRR Global Area Coverage (GAC) pixels. These 2x2 aggregates are often referred to as 'cells' to distinguish from native instrument 'pixels'. These L2 products can be thought to have an approximate size around 8.8 km x 8.8 km at the sub-satellite point, and larger away from this, due to a combination of the sensor scanning geometry and Earth's curvature. All scientific data are filtered for clouds, sea ice, and surface snow or ice.

Data fields which remove suspected poor-quality (QA=1) retrievals contain _best_estimate in the SDS name. These are provided for user convenience, so most users can use these without the need for further filtering of data, and it is expected that most users will use these for analysis. The content of these is the same as the non-best-estimate equivalents, except that QA=1 cells are populated with fill values.

Data fields that contain both land and ocean data (referred to as 'combined' in the SDS names) are provided to simplify the user experience. Each cell in the L2 data is designated as land if 3 or 4 of the pixels (determined from our land/water mask) in that 2x2 pixel cell are over land, and was water otherwise. The values of land cells are derived using only data from land pixels, and vice versa for water cells. Thus, such 'combined' data fields are simply composites of the individual land and ocean data fields since there is no overlap between them.

2.4 Daily L3 production and filtering

Daily level 3 (gridded) files at 1 degree resolution are produced from the L2 data. In most cases, each data field represents the arithmetic mean of all cells whose latitude

and longitude places it within the bounds of each grid element, although other quantities including medians, standard deviations, and (for some variables) AOD-weighted means are also provided.

Only retrievals from best_estimate data sets (i.e. QA-filtered) are used in the aggregation. Furthermore, only cells measured on the day of interest are included in this calculation. At least 5 such retrievals are required for a given grid cell to be valid on a given day.

2.5 Monthly L3 production and filtering

We also produce a monthly L3 gridded product based on the daily L3 gridded data and at the same 1 degree resolution. Statistics for the monthly product are based on the arithmetic mean values from the daily L3 gridded products (e.g. mean, median, standard deviation of daily means).

To remove poorly-sampled grid elements, at least 3 valid days of data in the month are required for the monthly grid element to be populated.

SDS names in the monthly L3 products are the same as in the daily L3 products.

3 Data contents

A summary of some of this information about file contents can also be found on the Deep Blue aerosol project website at https://deepblue.gsfc.nasa.gov/data .

3.1 Dimensions

Two dimensions are defined within the L2 files, valid for all 2D SDS:

- nxtrack, the number of L2 cells in the across-track (roughly East-West) direction. This is typically 205.
- natrack, the number of L2 cells in the along-track (roughly North-South) direction. This is typically 300.

Two dimensions are defined within the L3 files, valid for all 2D SDS:

- nlons, the number of elements in the longitudinal direction. This is 360 for 1 degree resolution.
- nlats, the number of elements in the latitudinal direction. This is 180 for 1 degree resolution.

3.2 Global attributes

The global attributes in the table below are present in both the L2 and L3 files.

Attribute name	Description	Data type
Conventions	Metadata conventions followed (e.g CF-1.6)	string
description	What data set the file is	string
institution	Institution of the original data producers	string
instr_long_name	Long, descriptive name of measuring instrument	string
instr_short_name	Abbreviated name of measuring instrument	string
long_name	Long, descriptive name for data set	string
platform_long_name	Descriptive name for platform on which instrument flew	string
platform_short_name	Abbreviated name of platform containing instrument	string
prod_type	Type of data product	string
reference	Published or online information about the data	string
short_name	Short name for data set	string
src	Source instrument type providing measurements	string
string	Data level	string
suffix	File format suffix	string
version	Data product version	string
contact person address	Physical address of contact person	string
contact_person_email	Email address for contact person	string
contact_person_name	Name of point of contact regarding the data	string
contact_person_role	Role of contact person in project	string
data_set_language	Language the data set is in	string
file_format	File format of the data (NetCDF4)	string
keywords	Descriptive words or phrases related to the data	string
related_url	Related website for more information on the data	string

3.2 Data field attributes

The attributes in the table below are present for each SDS in both L2 and L3 (daily and monthly) files, where applicable.

Attribute name	Description	Data type
long_name	Long, descriptive name of data field	string
units	Units of the data field	string
_FillValue	Value assigned to missing/invalid data	same type as data field

Note that scale_factor and add_offset attributes are not defined for the geophysical data in these data products, as they take values of 1 and 0 respectively.

3.3 SDS names and descriptions

3.3.1 Level 2 data fields

SDS name	Description	Units
latitude	Latitude of the center of the cell	degrees_north
longitude	Longitude of the center of the cell	degrees_east
solar_zenith_angle	Angle between the pixel zenith and the Sun	degrees
viewing_zenith_angle	Angle between the cell zenith and the instrument	degrees
relative_azimuth_angle	Difference between the satellite and solar azimuth angles NOTE: Follows the Gordon convention and is measured relative to South.	degrees
scattering_angle	Angle between the direction of incident radiation and the direction into which it is scattered	degrees
hour	Hour of day (UTC) that measurement was made	hour (UTC)
minute	Minute after hour of day that measurement was made	minute
aerosol_optical_thickness_confidence_flag_land	Confidence flags (QA flags) associated with all aerosol optical thickness over land data fields. 0 = No retrieval 1 = Poor 2 = Good 3 = Best	none
aerosol_optical_thickness_confidence_flag_ocean	Confidence flags (QA flags) associated with all aerosol optical thickness over water data fields. 0 = No retrieval 1 = Poor 2 = Good 3 = Best	none
deep_blue_algorithm_flag	Flag identifying which algorithm was used to populate cells in the combined land and ocean SDS. 0= Deep Blue data base method (land) 1= Deep Blue vegetated method (land) 2= Deep Blue mixed (land) 3= SOAR (ocean)	none
aerosol_optical_thickness_550_land_ocean_best_estimate	AOT at 550 nm over land and ocean, filtered to remove QA=1 points	none
aerosol_optical_thickness_550_land_ocean	AOT at 550 nm over land and ocean	none
aerosol_optical_thickness_630_land_ocean_best_estimate	AOT at AVHRR band 1 (near 630 nm) over land and ocean, filtered to remove QA=1 points	none
aerosol_optical_thickness_630_land_ocean	AOT at AVHRR band 1 (near 630 nm) over land and ocean	none
aerosol_optical_thickness_550_land	AOT at 550 nm over land	none
aerosol_optical_thickness_630_land	AOT at AVHRR band 1 (near 630 nm)	none
aerosol_optical_thickness_550_land_best_estimate	AOT at 550 nm over land, filtered to remove QA=1 points	none
aerosol_optical_thickness_630_land_best_estimate	AOT at AVHRR band 1 (near 630 nm) over land, filtered to remove QA=1 points	none

aerosol_optical_thickness_550_ocean_best_estimate	AOT at 550 nm over ocean, filtered to remove QA=1 points	none
aerosol_optical_thickness_630_ocean_best_estimate	AOT at AVHRR band 1 (near 630 nm) over ocean, filtered to remove QA=1 points	none
aerosol_optical_thickness_830_ocean_best_estimate	AOT at AVHRR band 2 (near 830 nm) over ocean, filtered to remove QA=1 points	none
aerosol_optical_thickness_550_ocean	AOT at 550 nm ocean	none
aerosol_optical_thickness_630_ocean	AOT at AVHRR band 1 (near 630 nm) over ocean	none
aerosol_optical_thickness_830_ocean	AOT at AVHRR band 2 (near 830 nm) over ocean	none
fine_mode_fraction_550_ocean	Fine mode fraction of AOT at 550 nm over ocean	none
fine_mode_fraction_550_ocean_best_estimate	Fine mode fraction of AOT at 550 nm over ocean, filtered to remove QA=1 points	none
angstrom_exponent_ocean_best_estimate	Ångström exponent (between 550 nm and band 2) over ocean, filtered to remove QA=1 points	none
angstrom_exponent_ocean	Ångström exponent (between 550 nm and band 2) over ocean	none
model_flag_ocean	Aerosol optical model of best-fit solution over ocean: 0 = no retrieval 1 = dust 2 = fine-mode dominated 3 = marine, standard 4 = marine, high FMF	none

3.3.2 Level 3 data fields

SDS name	**Description**	**Units**
latitude	Latitude of the center of the grid element	degrees_north
longitude	Longitude of the center of the grid element	degrees_east
aerosol_optical_thickness_550_land_ocean_mean	Mean of QA-filtered AOT at 550 nm over land and ocean	none
aerosol_optical_thickness_630_land_ocean_mean	Mean of QA-filtered AOT at AVHRR band 1 (near 630 nm) over land and ocean	none
aerosol_optical_thickness_550_land_ocean_median	Median of QA-filtered AOT at 550 nm over land and ocean	none
aerosol_optical_thickness_630_land_ocean_median	Median of QA-filtered AOT at AVHRR band 1 (near 630 nm) over land and ocean	none
aerosol_optical_thickness_550_land_ocean_stddev	Standard deviation of QA-filtered AOT at 550 nm over land and ocean	none
aerosol_optical_thickness_630_land_ocean_stddev	Standard deviation of QA-filtered AOT at AVHRR band 1 (near 630 nm) over land and ocean	none
land_ocean_count	Number of individual retrievals (for daily files) or days (for monthly files) contributing to statistics for combined land and ocean SDS	none
aerosol_optical_thickness_550_land_mean	Mean of QA-filtered AOT at 550 nm over land	none
aerosol_optical_thickness_630_land_mean	Mean of QA-filtered AOT at AVHRR band 1 (near 630 nm) over land	none

aerosol_optical_thickness_550_land_median	Median of QA-filtered AOT at 550 nm over land	none
aerosol_optical_thickness_630_land_median	Median of QA-filtered AOT at AVHRR band 1 (near 630 nm) over land	none
aerosol_optical_thickness_550_land_stddev	Standard deviation of QA-filtered AOT at 550 nm over land	none
aerosol_optical_thickness_630_land_stddev	Standard deviation of QA-filtered AOT at AVHRR band 1 (near 630 nm) over land	none
land_count	Number of individual retrievals (for daily files) or days (for monthly files) contributing to statistics for land SDS	none
aerosol_optical_thickness_550_ocean_mean	Mean of QA-filtered AOT at 550 nm over ocean	none
aerosol_optical_thickness_630_ocean_mean	Mean of QA-filtered AOT at AVHRR band 1 (near 630 nm) over ocean	none
aerosol_optical_thickness_830_ocean_mean	Mean of QA-filtered AOT at AVHRR band 2 (near 830 nm) over ocean	none
fine_mode_fraction_550_ocean_mean	Mean of QA-filtered fine mode fraction of AOT at 550 nm over ocean	none
angstrom_exponent_ocean_mean	Mean of QA-filtered Ångström exponent (between 550 nm and band 2) over ocean	none
aerosol_optical_thickness_550_ocean_median	Median of QA-filtered AOT at 550 nm over ocean	none
aerosol_optical_thickness_630_ocean_median	Median of QA-filtered AOT at AVHRR band 1 (near 630 nm) over ocean	none
aerosol_optical_thickness_830_ocean_median	Median of QA-filtered AOT at AVHRR band 2 (near 830 nm) over ocean	none
fine_mode_fraction_550_ocean_median	Median of QA-filtered fine mode fraction of AOT at 550 nm over ocean	none
angstrom_exponent_ocean_median	Median of QA-filtered Ångström exponent (between 550 nm and band 2) over ocean	none
aerosol_optical_thickness_550_ocean_stddev	Standard deviation of QA-filtered AOT at 550 nm over ocean	none
aerosol_optical_thickness_630_ocean_stddev	Standard deviation of QA-filtered AOT at AVHRR band 1 (near 630 nm) over ocean	none
aerosol_optical_thickness_830_ocean_stddev	Standard deviation of QA-filtered AOT at AVHRR band 2 (near 830 nm) over ocean	none
fine_mode_fraction_550_ocean_stddev	Standard deviation of QA-filtered fine mode fraction of AOT at 550 nm over ocean	none
angstrom_exponent_ocean_stddev	Standard deviation of QA-filtered Ångström exponent (between 550 nm and band 2) over ocean	none
fmf_ocean_wtdmean	Weighted mean (by AOT at 550 nm) of QA-filtered fine mode fraction of AOT at 550 nm over ocean	none
ang_ocean_wtdmean	Weighted mean (by AOT at 550 nm) of QA-filtered Ångström exponent (between 550 nm and band 2) over ocean	none
ocean_count	Number of individual retrievals (for daily files) or days (for monthly files) contributing to statistics for ocean SDS	none

4 Reading the data

Data usability is very important to us. We have selected CF-compliant NetCDF4 to maximize the usability and accessibility of our data now and into the future. If you have trouble reading our data, or have suggestions on how to make it more useful, please contact us.

More information on NetCDF, including tools to access files in this format, can be found at https://www.unidata.ucar.edu/software/netcdf/ . For quick browsing of the contents of individual files, the Panoply tool (http://www.giss.nasa.gov/tools/panoply/) provides a quick and easy interface. NetCDF libraries are also available in a variety of higher-level programming languages, such as IDL, Python, C/C++, and FORTRAN.

The Deep Blue website includes a page with more information about the content and format of various Deep Blue data products, including a brief tutorial on the use of Panoply to map aerosol optical depth, at https://deepblue.gsfc.nasa.gov/data .

5 Where to download the data

The data set is currently available through the NASA Center for Climate Simulation (NCCS) data portal:

- http access is at https://portal.nccs.nasa.gov/datashare/AVHRRDeepBlue
- ftp access is at ftp://dataportal.nccs.nasa.gov or ftp://nccs.nasa.gov with username *AVHRRDeepBlue* and no password

Access is free and does not require registration. If you have difficulties using the NCCS portal, please use the contact information on that webpage for support.

Acknowledgements

Data hosting resources were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center.

References

Scientific references about Deep Blue/SOAR algorithm development and validation can be found on the Deep Blue website at https://deepblue.gsfc.nasa.gov/publications . Some key references dealing with calibration, algorithm, and validation are additionally listed below:

- Hsu, N. C., S. -C. Tsay, M. D. King, and J. R. Herman (2004), Aerosol properties over bright-reflecting source regions, *IEEE Trans. Geosci. Remote Sens.*, 42(3), 557–569, doi:10.1109/TGRS.2004.824067

- Hsu, N. C., S.-C. Tsay, M. D. King, and J. R. Herman (2006), Deep Blue retrievals of Asian aerosol properties during ACE-Asia, *IEEE Trans. Geosci. Remote Sens.*, 44, 3180–3195, doi:10.1109/TGRS.2006.879540

- Hsu, N. C., M.-J. Jeong, C. Bettenhausen, A. M. Sayer, R. Hansell, C. S. Seftor, J. Huang, and S.-C. Tsay (2013), Enhanced Deep Blue aerosol retrieval algorithm: The second generation, *J. Geophys. Res. Atmos.*, 118, 9296–9315, doi:10.1002/jgrd.50712

- Jeong, M.-J., N. C. Hsu, E. J. Kwiatkowska, B. A. Franz, G. Meister, and C. Salustro (2011), Impacts of Cross-Platform Vicarious Calibration on the Deep Blue Aerosol Retrievals for Moderate Resolution Imaging Spectroradiometer Aboard Terra, *IEEE Trans. Geosci. Remote Sens.*, 49 (12), 4877-4888, doi:10.1109/TGRS.2011.2153205

- Sayer, A. M., N. C. Hsu, C. Bettenhausen, Z. Ahmad, B. N. Holben, A. Smirnov, G. E. Thomas, and J. Zhang (2012), SeaWiFS Ocean Aerosol Retrieval (SOAR): Algorithm, validation, and comparison with other data sets, *J. Geophys. Res.*, 117, D03206, doi:10.1029/2011JD016599

- Sayer, A. M., N. C. Hsu, C. Bettenhausen, and M.-J. Jeong (2013), Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data, *J. Geophys. Res. Atmos.,* 118, 7864–7872, doi:10.1002/jgrd.50600

- Sayer, A. M., L. A. Munchak, N. C. Hsu, R. C. Levy, C. Bettenhausen, and M.-J. Jeong (2014), MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and “merged” data sets, and usage recommendations, *J. Geophys. Res. Atmos.,* 119, 13,965–13,989, doi:10.1002/2014JD022453

- Sayer, A. M., N. C. Hsu, C. Bettenhausen, M.-J. Jeong, and G. Meister (2015), Effect of MODIS Terra radiometric calibration improvements on Collection 6 Deep Blue aerosol products: Validation and Terra/Aqua consistency, *J. Geophys. Res. Atmos.,* 120, doi:10.1002/2015JD023878