Auxiliary Material 1. ‘l2_site_stats.xls’ Comparison between MODIS Level 2 retrieved AOD and AERONET, by site, season, satellite, Microsoft Excel.

This table shows statistical comparisons between MODIS Level 2 AOD and AERONET, as described in the paper.

Sorting fields included are as follows, from left to right:

- Region;
- Site name;
- Satellite (Terra|Aqua|Both);
- Correction applied to data:
    - ‘nocorr’ : no correction applied;
    - ‘albcorr’ : albedo correction applied (using matched MODIS MCD43C3 data)
    - ‘albregcorr’ : albedo correction and regional slope correction applied;
- Filtering applied to data:
    - ‘f0’ : all valid data included;
    - ‘fvg’ : all data with MODIS QA == ‘Very Good’ included;
    - ‘fqa’ : ‘fvg’ + scattering angle + snow filter (matched, extended, see Section 3.3)
    - ‘fqa_nocloud’ : ‘fqa’ + retrievals with MODIS-detected cloud excluded;
    - ‘weak’ : ‘fqa_nocloud’ + matched MODIS MCD43C3 albedo filters (‘weak’ thresholds);
    - ‘strong’ : ‘fqa_nocloud’ + matched MODIS MCD43C3 albedo filters (‘strong’ thresholds);
- Season:
    - ‘djf’ : December/January/February
    - ‘mam’ : March/April/May
    - ‘jja’ : June/July/August
    - ‘son’ : September/October/November
    - ‘all’ : entire year

Statistics presented are as follows, from left to right:

- Distribution of MODIS AOD values:
  - -N($\tau_M < 0.2$);
  - -N($0.2 < \tau_M < 0.6$);
  - -N($0.6 < \tau_M < 1.4$);
  - -N($\tau_M > 1.4$);
- Mean bias MODIS-AERONET AOD;
- Regression slope, calculated using $0.2 < \tau_M < 1.4$, using Equation (2).
- Regression coefficient r^2 for $0.2 < \tau_M < 1.4$.
- Regression slope, calculated using $\tau_M > 1.4$, using Equation (2).
- Regression coefficient r^2 for $\tau_M > 1.4$.
- Percentage BELOW/WITHIN/ABOVE compliance (+/-0.05 + 0.20τ) for:
  - $\tau_M < 0.2$;
  - $0.2 < \tau_M < 0.6$;
  - $0.6 < \tau_M < 1.4$;
  - $\tau_M > 1.4$;
  - all τ_M;
- RMS difference MODIS-AERONET;
- RMS difference MODIS-AERONET for $\tau_A < 0.2$;
- Slope and intercept of diagnostic error model (RMS difference vs τ_A);
  - calculated with $0.2 < \tau_M < 1.4$;
  - calculated with $\tau_M > 1.4$;
- Slope and intercept of prognostic error model (RMS difference vs τ_M);
  - calculated with $0.2 < \tau_M < 1.4$;
  - calculated with $\tau_M > 1.4$;

Note that in some cases data volumes are insufficient to make the regression calculations.

Auxiliary Material 2. 'l2_site_pages' Comparison between MODIS Level 2 retrieved AOD and AERONET, by site and satellite, PDF.

NOTE: Included with Auxiliary Materials to this paper are examples from one site, Cuiaba-Miranda in Brazil. Site pages for all sites are available from the authors upon request.

Each of these PDF files shows various comparisons between MODIS and AERONET data for a specific AERONET site. All statistics and data are derived from the 2005-2008 data set of matched MODIS and AERONET AOD data, matched within tolerances of +/- 30 minutes and +/-30 kilometers.

The bold-face letter in the top left corner indicates data from Terra ('T') or Aqua ('A') MODIS.

The top line gives the geographic coordinates and name of each AERONET site.

The second line at the top gives the number of points in the matched dataset, the mean AERONET AOD (interpolated to 0.55 μm), and the mean fine-mode fraction ('eta'), derived from AERONET data using the method of O'Neill et al. (2003).

The three thin time series plots give auxiliary information for each point in the matched dataset. From top to bottom, plots show:

- Fraction of snow from MODIS MCD43C3 albedo product;
- Fine-mode fraction of each AERONET retrieval;
- Ratio of 0.65 μm and 2.1 μm albedo from MODIS MCD43C3 product.

The main time series plot shows matched MODIS and AERONET data from 2005-2008. Symbols on this plot are as follows:

- BLACK CROSSES CONNECTED BY THIN BLACK LINES are AERONET retrieved AOD;
- GRAY Xs are MODIS AOD data with less than 'very good' QA flags;
- MAGENTA Xs are MODIS AOD data with 'very good' QA and partial cloud detected by MODIS;
- GREEN Xs are MODIS AOD with 'very good' QA and no MODIS-detected cloud;
- BLUE Xs are MODIS AOD which pass all QA filters as well as the albedo filter (matched data, 'weak' thresholds) as described in the paper.
- All MODIS data shown in the time series plots are UNCORRECTED.

Scatter plots below the main time series directly compare MODIS and AERONET retrieved AOD with various permutations of filtering and correction. The complete descriptions are as follows, from left to right:

- RAW = All valid MODIS AOD, no corrections applied
- BASE = QA filters applied as described in paper, including exclusion of partially cloudy retrievals;
- FILT = 'QA' + snow/albedo filters (matched data, 'weak' albedo thresholds) applied;
- NEW = 'FILT' + albedo correction and regional slope correction applied.

The tables at the bottom show statistical comparisons corresponding to the data sets in the scatter plots. Statistics presented are as follows for the upper table, from left to right:

- Number of retrievals in the matched dataset;
- Mean MODIS AOD;
- Fraction of MODIS AOD > 0.2;
- Fraction of MODIS AOD > 1.0;
- Mean bias MODIS-AERONET AOD;
- RMS difference MODIS-AERONET;
- Percentage BELOW/WITHIN/ABOVE the compliance limit (+/-0.05 + 0.20τ);
- Regression slope, calculated using $0.2 < \tau_M < 1.4$, using Equation (2).
- Regression coefficient r^2.

The lower tables show parameters and results from a regression model of the AOD error. Statistics presented are as follows, from left to right:

- RMS difference MODIS-AERONET for $\tau_A < 0.2$;
- Slope and intercept of diagnostic error model (RMS difference vs τ_A);
- Slope and intercept of prognostic error model (RMS difference vs τ_M);
- Estimated uncertainty at various AOD values, using the prognostic error model.

Note that in some cases data volumes are insufficient to make the regression calculation for the prognostic error model.

Auxiliary Material 3. 'l3_region_stats.xls' Comparison between MODIS Level 3 gridded AOD and AERONET, by region, satellite, Microsoft Excel.

This table shows statistical comparisons between MODIS Level 3 gridded AOD and AERONET, as described in the paper.

Sorting fields included are as follows, from left to right:

- Region;
- Satellite (Terra|Aqua);
- Correction applied to data:
    - 'nocorr' : no correction applied;
    - 'albcorr' : albedo correction applied (using matched MODIS MCD43C3 data)
    - 'albregcorr' : albedo correction and regional slope correction applied;
    - 'albregclim' : climatological albedo correction and regional slope correction applied;
- Filtering applied to data:
    - 'f0' : all valid data included;
    - 'context' : context filtering applied (see Section 6.2);
    - 'fqa' : QA=='Very Good' + scattering angle filter
    - 'fqa_nocloud' : 'fqa' + retrievals with MODIS-detected cloud excluded;
    - 'fmatchedweak' : 'fqa_nocloud' + matched MODIS MCD43C3 albedo filters ('weak' thresholds) + matched MODIS MCD43C3 snow filter;
    - 'fmatchedweak_nocloud' : 'fmatchedweak' + partially cloudy excluded;
    - 'fmatchedstrong' : 'fqa_nocloud' + matched MODIS MCD43C3 albedo filters ('strong' thresholds) + matched MODIS MCD43C3 snow filter;
    - 'fmatchedstrong_nocloud' : 'fmatchedstrong' + partially cloudy excluded;
    - 'fclimweak' : 'fqa_nocloud' + climatological MODIS MCD43C3 albedo filters ('weak' thresholds) + climatological MODIS MCD43C3 snow filter;
    - 'fclimweak_nocloud' : 'fclimweak' + partially cloudy excluded;
    - 'fclimstrong' : 'fqa_nocloud' + climatological MODIS MCD43C3 albedo filters ('strong' thresholds) + climatological MODIS MCD43C3 snow filter;
    - 'fclimstrong_nocloud' : 'fclimstrong' + partially cloudy excluded;

Scenarios presented in the paper correspond to these combinations:

- 'RAW' == 'nocorr' + 'f0'
- 'BASE' == 'nocorr' + 'fqa_nocloud'
- 'NEW' == 'albregcorr' + 'fmatchedweak_nocloud'
- 'CLIM' == 'albregclim' + 'fclimweak_nocloud'
- 'STRONG' == 'albregcorr' + 'fmatchedstrong_nocloud'

Statistics presented are as follows, from left to right:

- Number of retrievals in the matched dataset;
- Distribution of MODIS AOD values:
  - -N($\tau_M < 0.2$);
  - -N($0.2 < \tau_M < 0.6$);
  - -N($0.6 < \tau_M < 1.4$);
  - -N($\tau_M > 1.4$);
- Regression slope, calculated using $0.2 < \tau_M < 1.4$, using Equation (2).
- Regression coefficient r^2 for $0.2 < \tau_M < 1.4$.
- Percentage BELOW/WITHIN/ABOVE compliance (+/-0.05 + 0.20τ) for:
  - $\tau_M < 0.2$;
  - $0.2 < \tau_M < 0.6$;
  - $0.6 < \tau_M < 1.4$;
  - $\tau_M > 1.4$;
  - all τ_M;
- RMS difference MODIS-AERONET for $\tau_A < 0.2$;
- Slope and intercept of diagnostic error model (RMS difference vs τ_A);
  - calculated with $0.2 < \tau_M < 1.4$;
- Slope and intercept of prognostic error model (RMS difference vs τ_M);
  - calculated with $0.2 < \tau_M < 1.4$;

Note that in some cases data volumes are insufficient to make the regression calculations.

Auxiliary Material 4. 'l3_region_pages' Comparison between MODIS Level 3 gridded AOD and AERONET, by region and satellite, PDF.

Each of these PDF files shows various comparisons between gridded MODIS and AERONET data for a specific geographic region. All statistics and data are derived from the matched gridded MODIS and AERONET AOD data from 2005-2008

The bold-face letter in the top left corner indicates data from Terra ('T') or Aqua ('A') MODIS.

The top line gives the approximate geographic coordinates and name of each region.

The second line at the top gives the number of points in the matched dataset, the mean AERONET AOD (interpolated to 0.55 μm), and the mean fine-mode fraction ('eta'), derived from AERONET data using the method of O'Neill et al. (2003).

The map shows the are of the region shaded in gray, with symbols indicating the location of each AERONET site with data in the matched dataset. Site symbols are sized according to the total number of matching points, as processed with the RAW scenario (BLUE CIRCLES). Concentric circles indicate the reduction in matched data volume associated with the filtering for the BASE (TAN CIRCLES), NEW (GREEN CIRCLES), and STRONG (RED CIRCLES) scenarios.

Scatter plots below the map directly compare MODIS and AERONET retrieved AOD under the different scenarios. The scenario descriptions are as follows, from left to right:

- RAW = All valid MODIS AOD, no corrections applied
- BASE = QA filters applied as described in paper, including exclusion of partially cloudy retrievals;
- NEW = QA filters + matched snow filters + matched albedo filters ('weak' thresholds) + matched albedo correction + regional slope correction;
- CLIM = QA filters + climatological snow filters + climatological albedo filters ('weak' thresholds) + climatological albedo correction +regional slope correction;

The tables at the bottom show statistical comparisons corresponding to the data sets in the scatter plots. Statistics presented are as follows for the upper table, from left to right:

- Number of retrievals in the matched dataset;
- Mean MODIS AOD;
- Fraction of MODIS AOD > 0.2;
- Fraction of MODIS AOD > 1.0;
- Mean bias MODIS-AERONET AOD;
- RMS difference MODIS-AERONET;
- Percentage BELOW/WITHIN/ABOVE the compliance limit (+/-0.05 + 0.20τ);
- Regression slope, calculated using $0.2 < \tau_M < 1.4$, using Equation (2).
- Regression coefficient r^2.

The second table is identical to the first, but calculated using only points from the matched data set with $\tau_A > 0.2$.

The bottom table shows parameters and results from a regression model of the AOD error. Statistics presented are as follows, from left to right:

- RMS difference MODIS-AERONET for $\tau_A < 0.2$;
- Slope and intercept of diagnostic error model (RMS difference vs τ_A);
- Slope and intercept of prognostic error model (RMS difference vs τ_M);
- Estimated uncertainty at various AOD values, using the prognostic error model.

These numbers represent the estimated error values used in the data assimilation system.

Note that in some cases data volumes are insufficient to make the regression calculation for the prognostic error model.