Quality Issues in Multi-sensor Aerosol Level 3 Satellite Data

Gregory Leptoukh, NASA GSFC

Funded in part by NASA ESTO and ACCESS programs

Overview

- Why Level 3 data and what is missing in L3?
- Quality needs: fitness for purpose
- Level 3 quality aspects
- Biases: sampling and processing-related
- Perspectives of Data Quality: Pixel vs. Product
- What is Level 3 validation?
- What needs to be done? A framework for consistent assessment and quantification of Level 3 data quality

Why use Level 3 products?

- Satellite Level 2 data are difficult to work with:
  - Complex formats
  - Complicated projection (swath)
  - Data volume
  - Number of files, etc., etc.
- Level 3 products are widely used by modelers, application users, climate change scientists
- Level 3 data are easy to use … but how good are these data for various purposes?

Challenge: to answer a typical data user question:
Which product is better for me?

Why now? What has changed?

- Growing attention to Climate change
- More models needs validation
- Revolutionary progress in data systems → dealing with data from many different sensors finally has become a reality.

Only now, a systematic approach to remote sensing quality is on the table.

Data quality needs: fitness for purpose

- Measuring Climate Change:
  - Model validation - gridded contiguous data with uncertainties in grid cells
  - Long-term time series – bias assessment is the must , e.g., sensor degradation, orbit and spatial sampling change (e.g., changing cloud cover over tropical oceans due to El-Nino)
- Studying phenomena using multi-sensor data:
  - Consistently processed and presented data with quality information
- Realizing Societal Benefits through Applications:
  - Near-Real Time for transport and event monitoring - in some cases, coverage and timeliness might be more important that accuracy
  - Pollution monitoring (e.g., air quality exceedance levels) – accuracy
- Educational (users generally not well-versed in the intricacies of quality; just taking all the data as usable can impair educational lessons) – only the best products

Why is it so complicated for Level 3?

Historical reasons

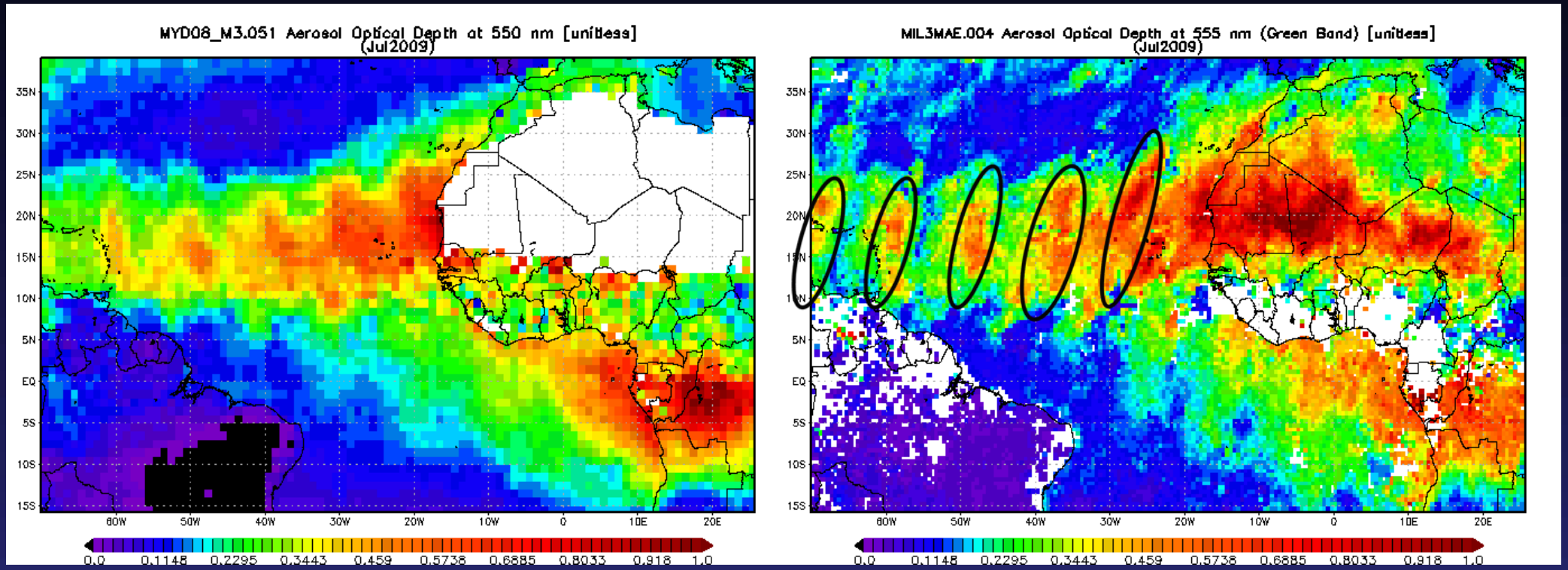
- Usually, Science Teams are tasked to produce & validate Level 2 data
- Usability of L3 data usually is not a high priority for Science Teams
- Level 3 products are treated mostly as just imagery, to assess gross features and variability of geophysical parameters
- L3 data are constructed differently for different instruments
- L2 uncertainty usually not propagated to L3
- The L3 “validation”, in most cases, is done by either comparing with point data or consistency checking with L3 data from other sensors or models, or just declaring it “validated” if L2 data are
- No consistent efforts to characterize & quantify L3 uncertainties across sensors besides some individual efforts

Addressing Level 3 data "quality"

- Terminology: Quality, Uncertainty, Bias, Error budget, etc.
- Quality aspects (examples):
  - Completeness:
    - *Spatial* (MODIS covers more than MISR)
    - *Temporal* (Terra mission has been longer in space than Aqua)
    - *Observing Condition* (MODIS cannot measure over sun glint while MISR can)
  - Consistency:
    - *Spatial* (e.g., not changing over sea-land boundary)
    - *Temporal* (e.g., trends, discontinuities and anomalies)
    - *Observing Condition* (e.g., exhibit variations in retrieved measurements due to the viewing conditions, such as viewing geometry or cloud fraction)
  - Representativeness:
    - Neither pixel count nor standard deviation fully express how representative the grid cell value is
    - Example from R. Kahn: for global, ~ 1˚ x 1˚ AOD, in general, MISR data need to be aggregated to ~ 3-month sampling to converge with MODIS

Spatial and temporal sampling – how to quantify to make it useful for modelers?

MODIS Aqua AOD July 2009 **MISR Terra AOD July 2009**



- Completeness: MODIS dark target algorithm does not work for deserts
- Representativeness: monthly aggregation is not enough for MISR and even MODIS
- Spatial sampling patterns are different for MODIS Aqua and MISR Terra: "pulsating" areas over ocean are oriented differently due to different direction of orbiting during day-time measurement → *Cognitive bias*

Is L3 quality different from L2 quality?

- If L2 errors are known, the corresponding L3 error can be computed, in principle
- Processing from L2→L3 daily → L3 monthly may reduce random noise but can also exacerbate systematic bias and introduce additional sampling bias
- However, at best, standard deviations (mostly reflecting variability within a grid box), and sometimes pixel counts and quality histograms are provided
- Convolution of natural variability with sensor/retrieval uncertainty and bias – need to understand their relative contribution to differences between data
- This does not address sampling bias

Differences in L3 from different sensors due to processing

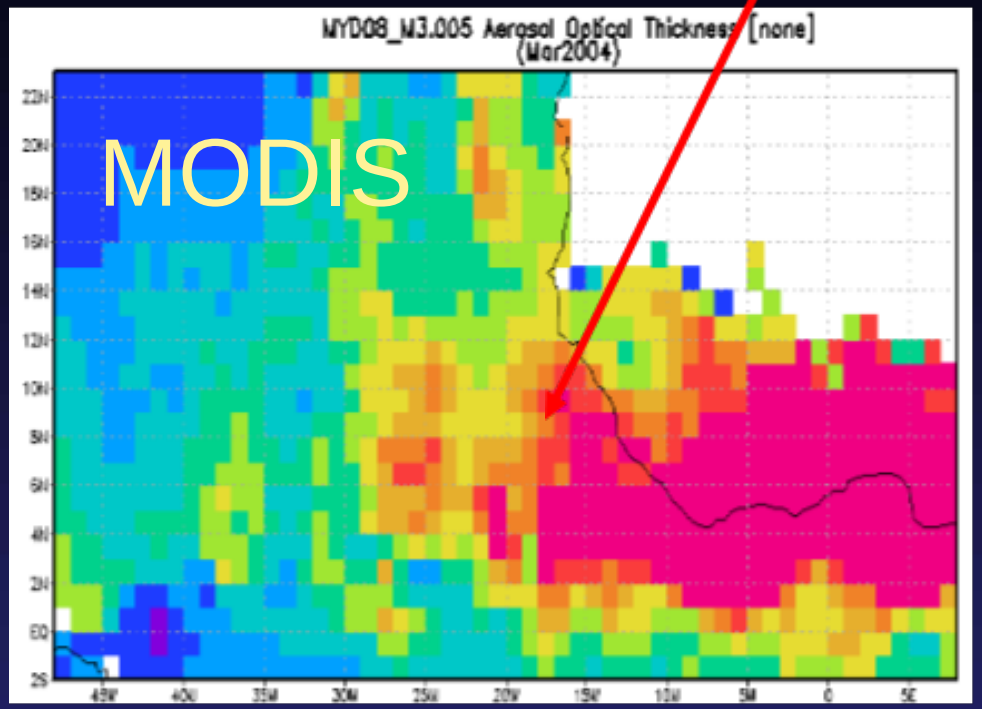
- Spatial and temporal binning (L2→L3 daily) leads to *Aggregation bias*:
  - Measurements (L2 pixels) from one or more orbits can go into a single grid cell → different within-grid variability
  - Different weighting: pixel counts, quality
  - Thresholds used, i.e., > 5 pixels
- Data aggregation (L3D → L3monthly → regional → global):
  - Weighting by pixel counts or quality
  - Thresholds used, i.e., > 2 days

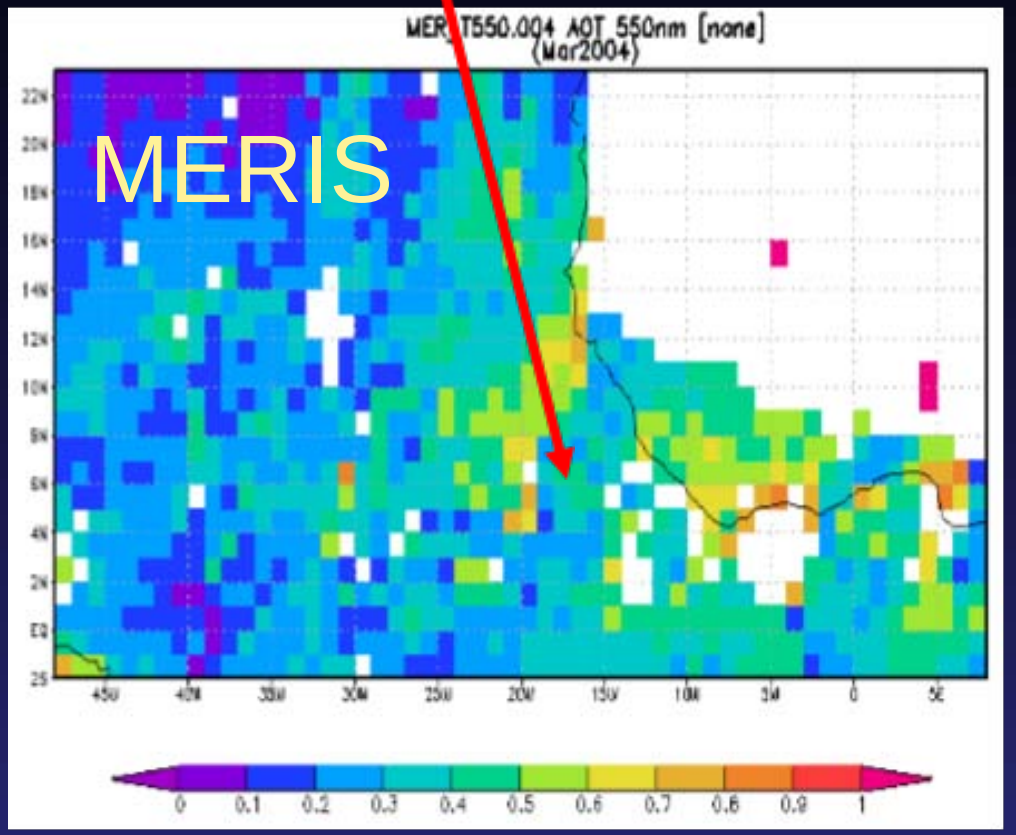
While these algorithms have been documented in ATBD, reports and papers, the typical data user is not immediately aware of how a given portion of the data has been processed, and what is the resulting impact

Case 1: MODIS vs. MERIS

Same parameter

Same space & time



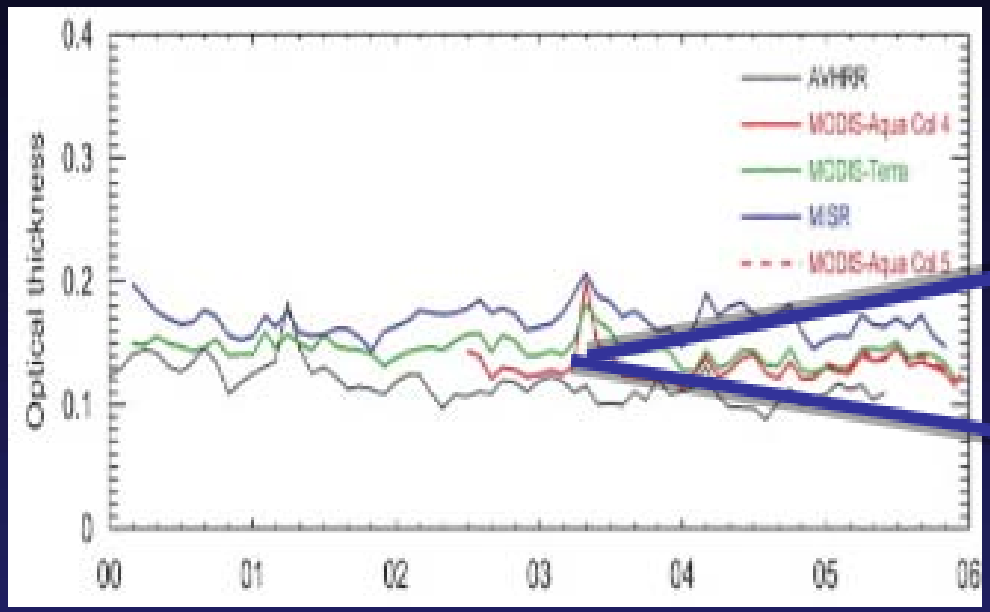


Different results – why?

A threshold used in MERIS processing effectively excludes high aerosol values. *Note: MERIS was designed primarily as an ocean-color instrument, so aerosols are "obstacles" not signal.*

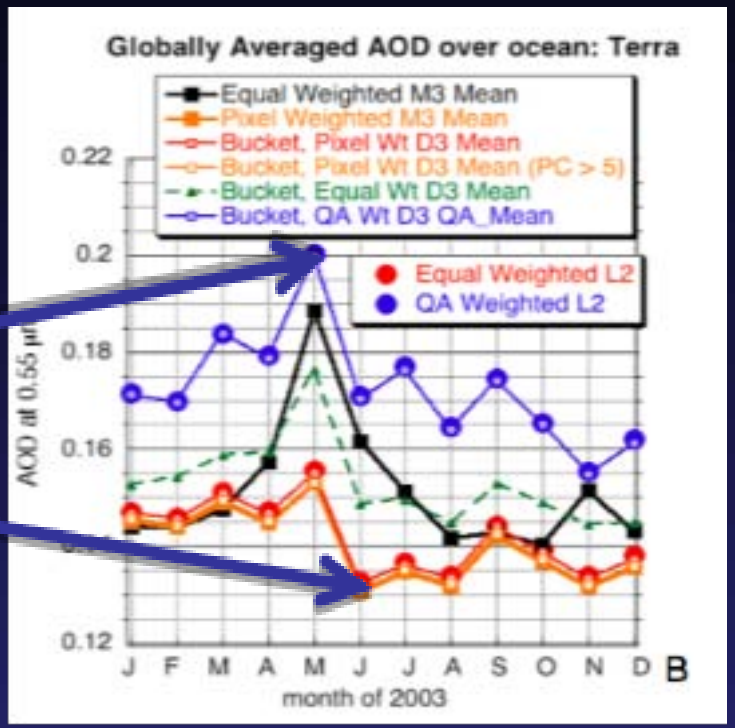
Case 2: Aggregation

AOD difference between sensors



Mishchenko et al., 2007

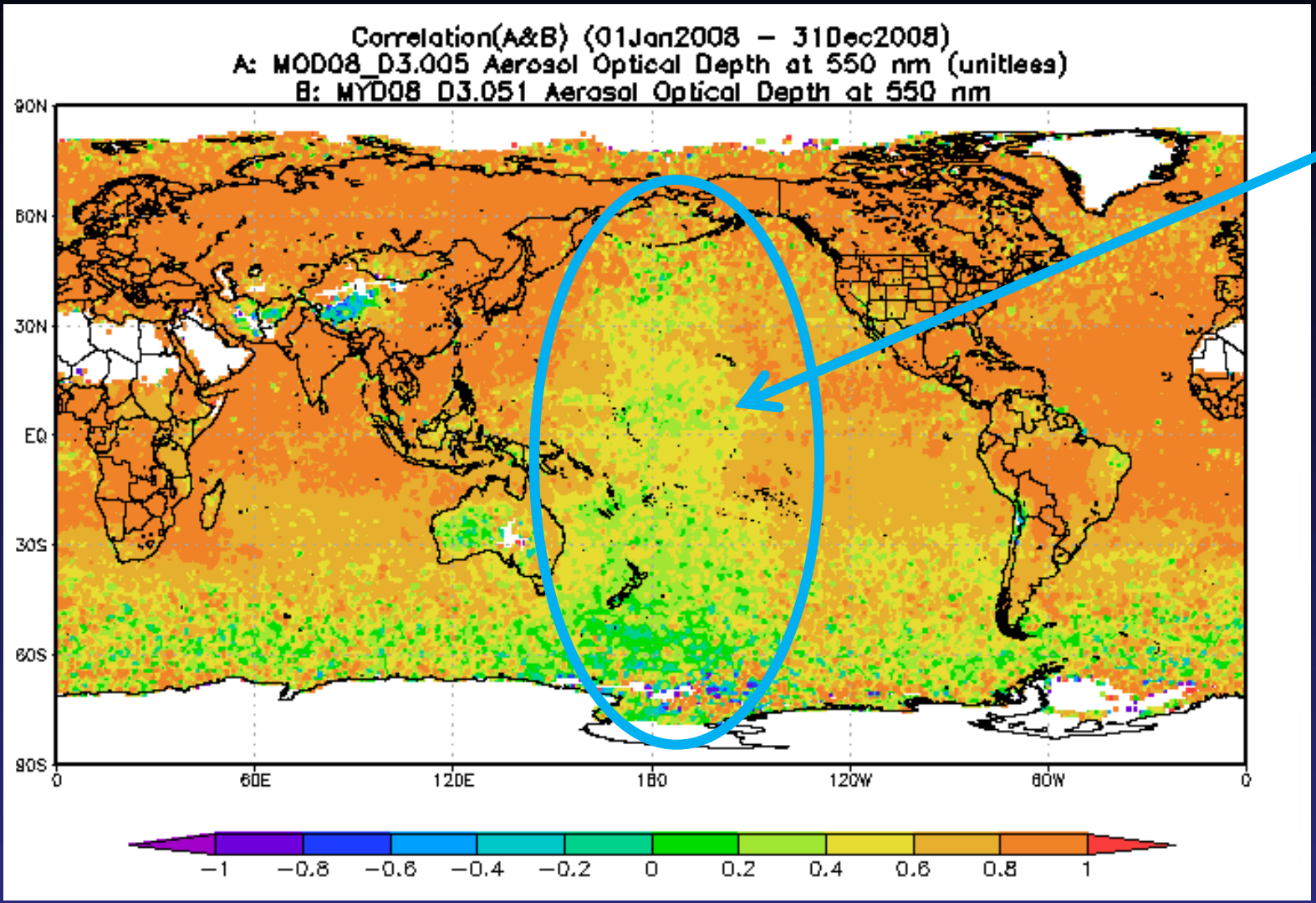
MODIS Terra only AOD: difference between diff. aggregations



Levy, Leptoukh, et al., 2009

The AOD difference can be up to 40% due to differences in aggregation

Case 3: DataDay definition



MODIS-Terra vs. MODIS-Aqua: Map of AOD temporal correlation, 2008

MODIS Level 3 dataday definition leads to artifact in correlation

Different kinds of reported and perceived data quality

- Pixel-level Quality (reported): algorithmic guess at usability of data point (some say it reflects the algorithm "happiness")
  - Granule-level Quality: statistical roll-up of Pixel-level Quality
- Product-level Quality (wanted/perceived): how closely the data represent the actual geophysical state
- Record-level Quality: how consistent and reliable the data record is across generations of measurements

Different quality types are often erroneously assumed having the same meaning

Different focus and action at these different levels to ensure Data Quality

General Level 2 Pixel-Level Issues

- How to extrapolate validation knowledge about selected Level 2 pixels to the Level 2 (swath) product?
- How to harmonize terms and methods for pixel-level quality?

AIRS Quality Indicators

0 Best
1 Good
2 Do Not Use

Data Assimilation
Climatic Studies

Purpose

Match up the recommendations?

MODIS Aerosols Confidence Flags

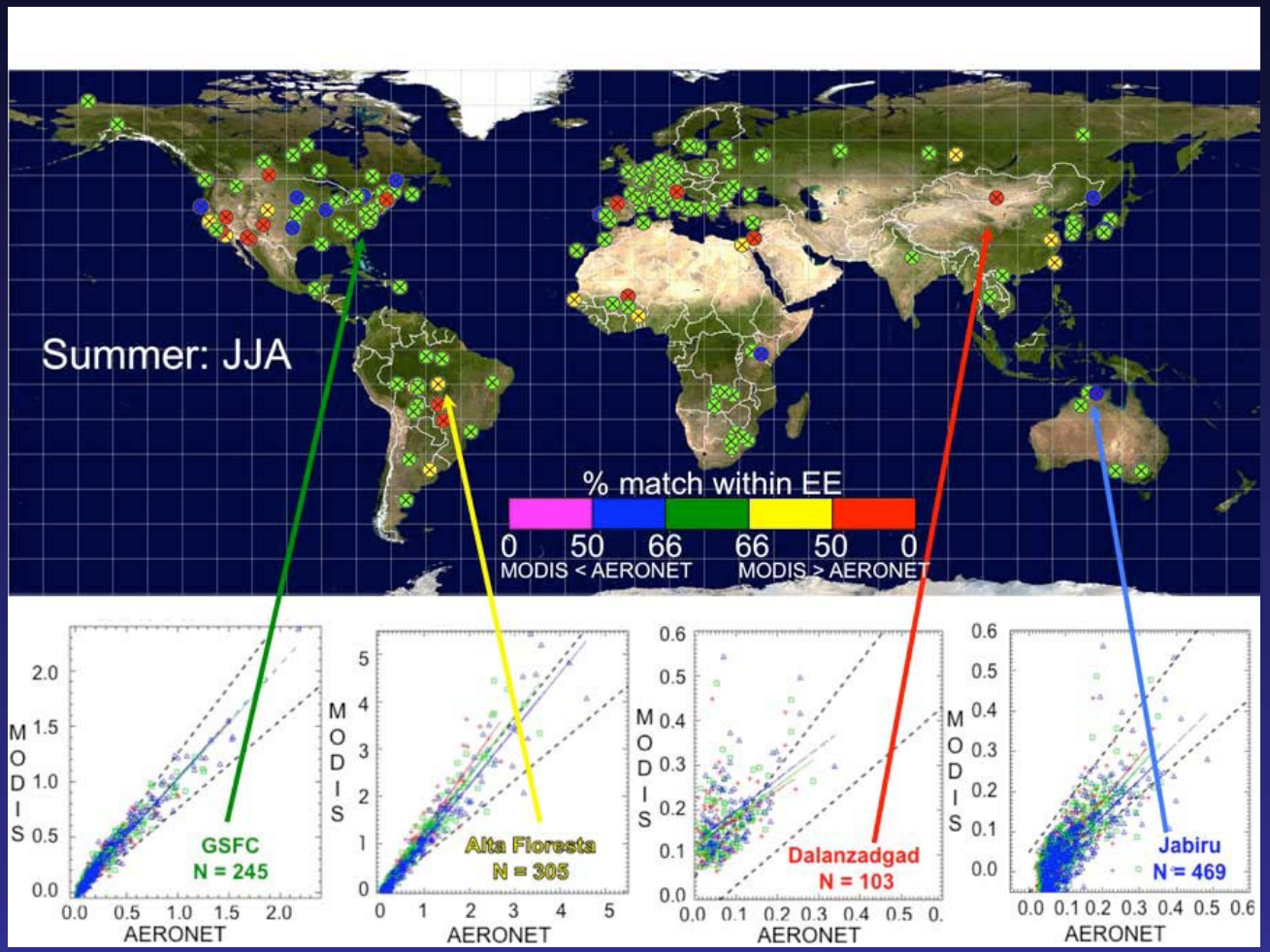
Ocean	Land
3 Very Good	3 Very Good
2 Good	2 Good
1 Marginal	1 Marginal
0 Bad	0 Bad

Use these flags in order to stay within expected error bounds

Ocean	Land
±0.03 ± 0.10 t	±0.05 ± 0.15 t

Data Quality Issues

- Validation of aerosol data show that not all data pixels labeled as “bad” are actually bad if looking at from a bias perspective.
- But many pixels are biased differently due to various reasons



From Levy et al, 2009

Percent of Biased Data in MODIS Aerosols Over Land Increase as Confidence Flag Decreases

*Compliant data are within $\pm 0.05 \pm 0.2\tau_{Aeronet}$

Statistics from Hyer, E., J. Reid, and J. Zhang, 2011, An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, Atmos. Meas. Tech., 4, 379-408, doi:10.5194/amt-4-379-2011.

Factors contributing to uncertainty and bias in L2

- *Physical*: instrument, retrieval algorithm, aerosol spatial and temporal variability, measuring geometry …
- *Input*: ancillary data used by the retrieval algorithm
- *Classification*: erroneous flagging of the data
- *Simulation*: the geophysical model used for the retrieval
- *Sampling*: the averaging within the retrieval footprint

Borrowed from the SST study on error budget

Why can't we just apply L2 quality to L3?

Aggregation to L3 introduces new issues where aerosols co-vary with some observing or environmental conditions:

- *Spatial*: sampling polar areas more than equatorial
- *Temporal*: sampling one time of a day only *(not obvious when looking at L3 maps)*
- *Vertical*: not sensitive to a certain part of the atmosphere thus emphasizing other parts
- *Contextual:* bright surface or clear sky bias
- *Pixel Quality*: filtering or weighting by quality may mask out areas with specific features

Validation of Level 3

- Usual:
  - Level 2: regress against the truth
  - Level 3: aggregate and then regress against the aggregated truth?
- Comparing a mean value in 1 deg grid box with data from stations in the same big area →*representativeness bias*
  - Increasing aggregation: spatial over satellite data and temporal over station data – works well only for large homogenous fields
- Comparing variance in the data with knowledge about atmospheric variability. *Comparison of retrieved maps with climatology can indicate systematic effects*
- Comparison with models (how ironic!) for initial validation

Doesn't look comprehensive…

Current initiatives

- NASA puts more emphasis on data quality
- ESA has requirements for providing quality information within the Climate Change Initiative
- 2010 Guideline for the Generation of Datasets and Products Meeting GCOS Requirements
- CEOS QA for Earth Observations (QA4EO) recommendations for capturing uncertainties (do not go beyond Level 1 or 2)
- QUAlity aware VIsualisation for the Global Earth Observation system of systems (GeoViQua)
- GEWEX panel on aerosols (several incarnations)

What do we recommend?

A framework for consistent assessment, capture and presentation of data quality information

- Establish terminology for Level 3 quality and validation (currently it differs from field to field, group to group)
- Harmonize quality across products
- Consistently aggregate to Level 3 to ensure compatibility between data from different instruments
- Directly address and quantify various bias types at product level
- Extrapolate validation knowledge about L2 product quality to Level 3
- Deliver quality information to users of data in a way they can understand and use it
- Extend QA4EO and other efforts to Level 3 data

So we can answer a typical user question:

Which product is better for my purpose?