



Multi-Sensor Data Synergy Advisor

IN11B-1054

NASA Goddard Earth Sciences (GES) Data & Information Services Center (DISC)

Code 610.2

NASA Goddard Space Flight Center

Greenbelt, Maryland 20771 USA

Gregory.Leptoukh@nasa.gov

G. G. Leptoukh¹, C. Lynnes¹, P. A. Fox⁴, A. I. Prados³, S. Shen², S. Zednik⁴, P. West⁴, D. J. Lary³

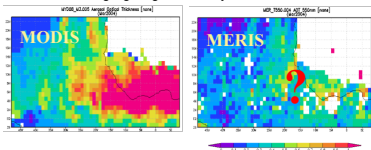
¹GSFC, NASA, Greenbelt, MD, USA; ²GMU, Fairfax, VA, USA; ³UMBC, Greenbelt, MD, USA; ⁴PRI, Troy, NY, USA

Abstract: As our inventory of Earth science data sets grows, more and more scientists try combining information from multiple datasets to get a better assessment of global aerosol spatial and temporal distribution. Advances in data standards and analysis tools make it easier to compare, merge and fuse remote sensing data from multiple sources. However, as the mechanics become easier, the risk of scientifically naive fusion increases. Subtle differences in the data provenance, e.g., sensor characteristics, sampling patterns, processing algorithms (among others), can produce significant systematic differences.

These differences can vary with spatial location, surface type or local time of day of the measurements. Also, systematic differences can arise from differences in cloud screening, calibration and model assumptions, quality screening, and aggregation schemas in the processing algorithms. These differences, if not recognized and accounted for, cast doubt on the validity and usefulness of data intercomparisons, merging and fusion.

The Multi-Sensor Data Synergy Advisor (MSDA) is designed to provide better access to the characteristics of datasets in Giovanni for a better understanding of potential sources of biases between datasets. This information is then used by Giovanni to inform users on the advisability of combining data sets. The MSDA is driven by an ontology of the sensors, datasets and processing algorithms. The ontology is used to populate a provenance for each dataset, allowing a provenance comparison of the two and highlighting where they differ.

Why two similar sensors,
MODIS on Terra and MERIS on
Envisat, reported very different
Aerosol Optical Thickness
(AOT) for March 2004?



• MODIS and MERIS sensors are very similar.
• MERIS aerosol product has the same spatial and temporal resolution as MODIS.
• However, MERIS misses high aerosol loading events while MODIS on Terra/Aqua see them because MERIS aerosols are reported only where ocean color retrievals are made, i.e., where the ocean surface is seen, i.e., where atmospheric optical thickness is small, i.e., there is an effective MERIS AOT threshold.

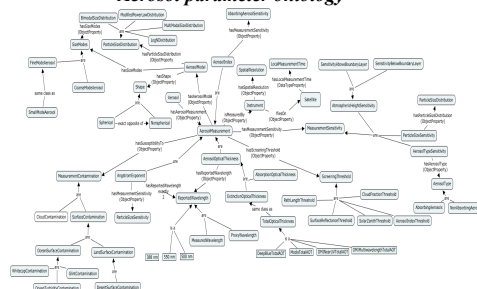
Use case: Fuse Level 3 Aerosol Optical Depth (AOD) data from multiple sensors for better monitoring of pollution transport and comparison with models

Approach: Use NASA Giovanni Data Fusion portal: select AOD from two or more sensors; compare them in Giovanni; view and assess their similarity by analyzing their provenance; if compatible, assess biases and fuse.

Methodology

- Semantic Web and ontologies to capture essential parameter details, quality and caveats
- Proof Markup Language (PML) and tools from the Inference Web project to capture inter-relations of the provenance
- Reasoners to automatically evaluate potential inter-comparisons as valid, speculative or invalid, with an explanation of the result
- Giovanni as testbed for implementation

Aerosol parameter ontology



Physical Phenomena

How close are two datasets to each other, that is, are they measuring the same thing? Need to evaluate the effect of the space-time alignment of two datasets. For example, slowly varying phenomena (chlorophyll) are less affected by a time offset than quickly varying phenomena (surface air temperature, fast changing cloud cover).

Data Measurements and Processing

Measurement methodology (spectral bands, total column or profile), the retrieval algorithm, instrument characteristics, and satellite observation pattern orbit, re-visit time, spatial and temporal resolution. Use of visible vs. UV bands to measure aerosols produces different sensitivity to different atmospheric heights and different perceptions of interfering clouds. Also, subtleties of the processing algorithm(s), such as the common bias toward very clear skies for aerosols computed jointly with ocean color, relative to aerosols computed for traditional atmospheric science uses.

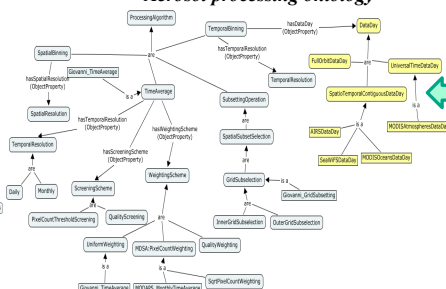
Space-Time Reference Frames

Gridded, orbital swath, geostationary, or *in situ*. Various types of projections, orbital patterns, etc. The temporal aspect must account for synoptic data, time-averaged gridded data and the space-time covariance of Sun-synchronous orbits, and even various definitions of a data day (different for AIRS, MODIS Atmospheric and Ocean Color datasets).

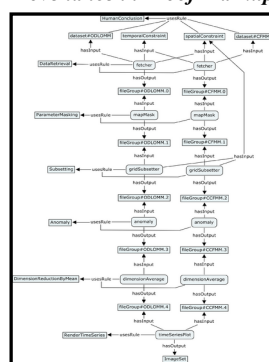
Data Quality Representations

Data quality ontology will account for the usual statistical and pixel based quality measures. Also, it should incorporate knowledge from the literature, such as documented spatial variations of quality for a given measurement, or differences over water or specific land cover types (such as aerosols over deserts). For example, the same (formal) quality levels for MODIS aerosol measurements over ocean and over land actually mean *different* quality, but this information is not captured anywhere but folk lore.

Aerosol processing ontology



Provenance in Proof Markup Language (PML)



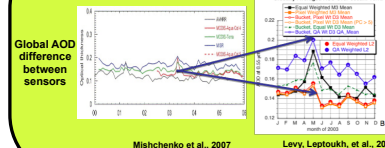
The diagram is a causality graph with PML-style rules associated with processing steps. Processes are represented as elliptical nodes; artifacts/information and rules are both represented by rectangle nodes. This diagram represents the dual processing of the two datasets as captured in the second example Giovanni Lineage XML. This diagram was used to verify the general provenance model and was constructed by analyzing the sample Giovanni Lineage XML documents and conversing with the Giovanni system engineers and scientists.

Sensitivity Studies

Once the main attributes of sensors, algorithms, quality, processing steps are identified and captured in ontology, it is important to understand which of these attributes or differences between products are significant.

Spatio-temporal Aggregation

Q: How sensitive are AOT time-series to different aggregations from Level 3 Daily to Monthly products and then Regional or Global means?



Single sensor
MODIS Terra
AOD: difference
between
different
aggregations

A: Very sensitive. For MODIS-Terra alone, AOD difference can be up to 40%

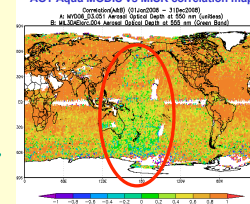
- Pixel count weighting (correctly) applied to gridded Level 3 data represents observational (Level 2) sampling, and has spatial and temporal (mostly clear sky) bias in the result.
- Applying Confidence weighting to Level 3 leads to a different Confidence-biased L2 sampling result.
- Grid number weighting (correctly) applied to L3 data during spatial or temporal aggregation represents L3 sampling and a lesser spatial and temporal (mostly clear sky) bias.

Dataday Definition

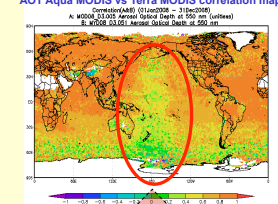
Level 3 daily products are generated by binning Level 2 data belonging to one day onto a certain spatial grid and according to a dataday definition.

1. Calendar UTC: all pixels between 00:00 - 24:00 UTC
MODIS Atmospheric products, OMI L2G
2. Calendar (local time): 24 hours centered at the Equatorial Crossing Time at 180 deg longitude
Intermediate case
3. Spatial: uses local date/time and ensures spatial continuity.
TOMS, AVHRR, AIRS, OMI, MODIS Ocean, SeaWiFS
4. MISR: full 14 or 15 orbits depending on a day within the 16-day cycle
MISR

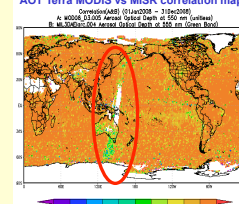
AOT Aqua MODIS vs MISR correlation map



AOT Aqua MODIS vs Terra MODIS correlation map

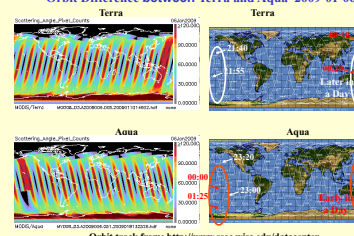


AOT Terra MODIS vs MISR correlation map



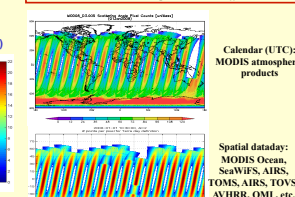
Correlation between Terra and Aqua MODIS AOT drops 0.8-0.9 → 0.4-0.1 around the dateline due to large Δt between Terra and Aqua measurements in these areas (up to 23 hours)

Orbit Difference between Terra and Aqua 2009-01-06



Area of the orbit overlap
(separated by more than 22 hours)

Even on Terra, MODIS and MISR have different dataday definition leading to a drop in correlation over Australia from globally very consistent 0.9 down to 0.1
Note: Globally MODIS-MISR correlation is better than of two MODIS on Terra and Aqua due to time difference



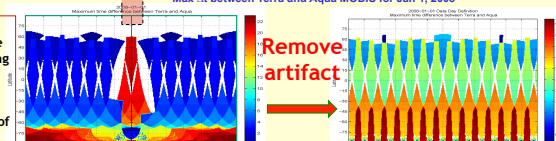
Calendar (UTC):
MODIS atmospheric
products

Spatial dataday:
MODIS Ocean,
SeaWiFS, AIRS,
TOMS, AIRS, TOVS,
AVHRR, OMI, etc.

Max Δt between Terra and Aqua MODIS for Jan 1, 2008

When comparing Level 3 AOT data from MODIS and MISR, the UTC Dataday definition, differences in the Equatorial Crossing Time and Orbital node (descending vs. ascending) lead to artifacts related to aerosols being measured at times separated by many hours.

Consistent Spatial Dataday definition removes some of these artifacts near the dateline



Remove
artifact

Acknowledgement: The project is funded by NASA ROSES 2009 ESTO AIST-08-0071