IN21C-1438

A Semantic Representation of Product Quality and Evidence for Satellite Data

Stephan Zednik[1] (zednis@rpi.edu), **Gregory Leptoukh**[2] (gregory.g.leptoukh@nasa.gov), **Peter Arthur Fox**[1] (pfox@cs.rpi.edu), **Christopher Lynnes**[2] (christopher.s.lynnes@nasa.gov), **Patrick West**[1] (westp@rpi.edu), **Suraiya P Ahmad**[2] (suraiya.ahmad@nasa.gov)

[1]Rensselaer Polytechnic Institute 110 8th St., Troy, NY, United States
[2]NASA Goddard Space Flight Center, Greenbelt, MD, United States

Abstract

There is growing interest within the broad research community to leverage satellite data for cross-disciplinary analysis and to make use of the data in ways unanticipated by the data provider. Poorly documented or publicized product quality information is a significant barrier to the successful or confident integration of satellite data for many users. Researchers seek clearly and consistently characterized product quality to facilitate assessment of product fitness-for-use. We argue that data product discovery mechanisms should be augmented with facilities to present product quality information; targeted to provide a condensed and clear view of product quality and to support comparison with quality of other like products.

We propose a method of provisioning product quality into aspects (e.g. completeness, consistency, accuracy, bias) and displaying computed and inferred facts as evidence to help characterize one or more aspects of the product quality. We describe the product quality ontology developed to facilitate this characterization of product quality. Finally, we illustrate the utility of this approach by showing how we have applied it to presenting product quality for the NASA MODIS Aerosol data product within a prototype implementation of the NASA Giovanni Data Access and Analysis Tool.

Inspiration – Quality Fact Label

What if quality information about scientific data products was as easy to view and use as a standard FDA nutrition label?

Quality Facts

Scope: Global, Daily Data

Product: MODIS Aerosol Optical Depth (τ)

Daily Spatial Completeness

Global Coverage: 50%	**Good Coverage**

Accuracy (vs Aeronet)

Slope of Linear Regression Fit* 0.91	**Low Underestimate Bias**
Expected Error (EE) (ocean)*	Δτ = **±0.03 ± 0.05**τ
Expected Error (EE) (land)**	Δτ = **±0.05 ± 0.20**τ
% Within EE (ocean)* 64%[2]	**Good Compliance**
% Within EE (land)** 67%[1]	**Good Compliance**

Measurement Characteristics

Platform: Terra

Instrument: MODIS

Collection: 5.1

Algorithm: Dark Target

Swath Width: 2330 km

Local Observing Time: 10:30

Wavelengths used for aerosol measurements (in nm):
ocean: 466, 553, 660, 860, 1240, 1640, 2120
land: 466, 553, 660, 2120

*MODIS vs Aeronet **QA = Very Good ***QA ≥ Marginal

References

[1]Hyer, E. J., Reid, J. S., and Zhang, J., (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

[2]Kleidman, R. G., A. Smirnov, R. C. Levy, S. Mattoo, and D. Tanré (2011), Evaluation and Wind Speed Dependence of MODIS Aerosol Retrievals Over Open Ocean, IEEE Transactions on Geoscience and Remote Sensing, 1-7.

Mock-up of Quality Facts label for MODIS AOD Terra

Finding Structure in the Complexities of Quality Information

Problem

- Perceived differently by data providers and data recipients
- 'Quality' is subjective and concerns fitness-for-use – which makes it hard to formalize
- Quality can be substantially non-uniform within a data entity

Solution

- Define aspects of quality about which you can make assertions
  - e.g. completeness, consistency, bias, compliance
- Quality assertions are subjective statements of fitness-for-use based on objective measures (evidence) of quality indicators
- present a scope against which evidence and assertions are held
- Aggregate quality assertions and related evidence based on relevant scope, present a report to data user

Quality Assertions

Regional & Seasonal Scope

Science Explanations

Quality Evidence

Title: MODIS Terra C5 AOD vs. Aeronet during Aug-Oct Biomass burning in Central Brazil,

***(General)* Statement**: ***Collection 5 MODIS AOD at 550 nm* during *Aug-Oct* over Central South America *highly* over-estimates for *large AOD* and in non-burning season *underestimates* for *small AOD*, as compared to Aeronet; *good* comparisons are found at *moderate AOD*.**

Region & season characteristics: Central region of Brazil is mix of forest, cerrado, and pasture and known to have low AOD most of the year except during biomass burning season

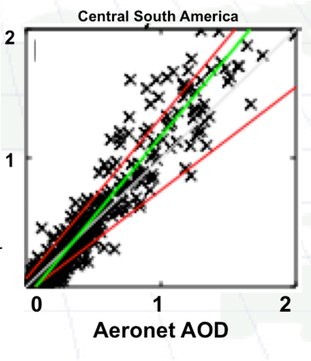
(***Dominating factors leading to Aerosol Estimate bias***):

1. Large positive bias in AOD estimate during biomass burning season may be due to wrong assignment of Aerosol absorbing characteristics. (Specific explanation) a constant Single Scattering Albedo ~ 0.91 is assigned for all seasons, while the true value is closer to ~0.92-0.93.

[Notes or *exceptions*: Biomass burning regions in Southern Africa do not show as large positive bias as in this case, it may be due to different optical characteristics or single scattering albedo of smoke particles, Aeronet observations of SSA confirm this]

2. Low AOD is common in non burning season. In Low AOD cases, biases are highly dependent on lower boundary conditions. In general a negative bias is found due to uncertainty in Surface Reflectance Characterization which dominates if signal from atmospheric aerosol is low.

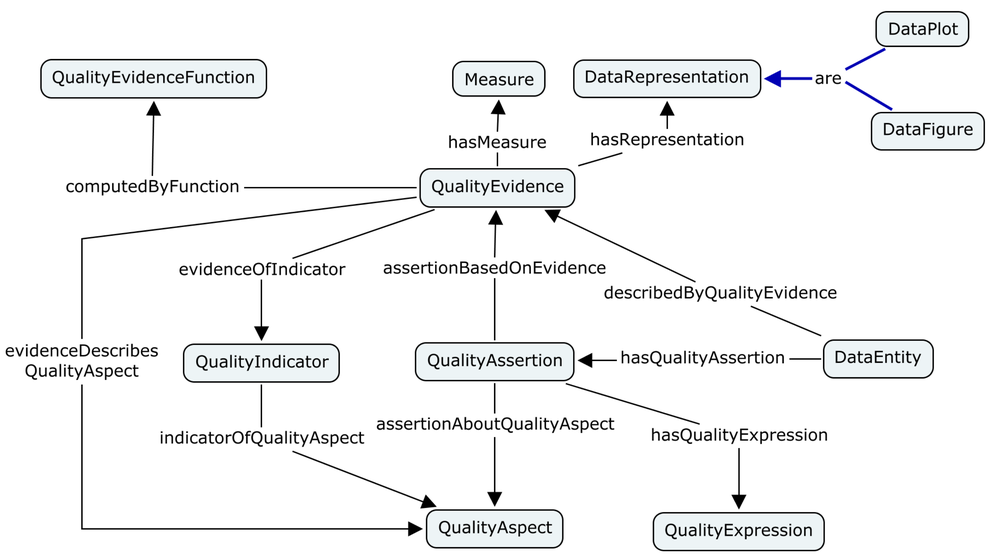
(***Example***) : *Scatter plot* of MODIS AOD and AOD at 550 nm vs. Aeronet from *ref.* (Hyer et al, 2011) *(Description Caption)* shows severe *over-estimation* of MODIS Col 5 AOD (dark target algorithm) at *large AOD* at 550 nm during *Aug-Oct* 2005-2008 over Brazil. *(Constraints)* Only best quality of MODIS data (*Quality =3*) used. Data with scattering angle *> 170 deg* excluded. *(Symbols)* **Red Lines** define regions of *Expected Error (EE)*, ***Green*** *is the fitted slope*
***Results:* Tolerance**= 62% within **EE**; RMSE=0.212 ; **r2**=0.81; **Slope**=1.00
For Low AOD (<0.2) **Slope**=0.3. **For high AOD** (> 1.4) **Slope**=1.54



Reference: Hyer, E. J., Reid, J. S., and Zhang, J., 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

Quality information for MODIS AOD bias over South America in Aug-Oct

MDSA Product Quality Model



Core concepts of the high-level quality ontology

Core Quality Model

- We defined a conceptual quality model based on the concept of making subjective quality assertions about data entities based on objective quality evidence
- Developed an OWL encoding for our conceptual model

Domain Extensions for Satellite Data

- We developed suite of domain-specific OWL ontologies that extend our core model for quality of aerosol products from satellite data
- Added concepts for science explanations, references, and visuals

MDSA Product Quality Interface

Quality Facts

Scope: Global, Daily Data
Product: MODIS Aerosol Optical Depth

Accuracy (vs Aeronet)	
slope of linear regression fit (MODIS vs Aeronet) = 0.93	low bias underestimate
67% within expected error (land)[1]	good compliance
64% within expected error (ocean)[2]	good compliance
69% within expected error (E. CONUS)[1]	good compliance
71% within expected error (W. CONUS)[1]	good compliance
40% within expected error (Central America)[1]	bad compliance
27% within expected error (South America)[1]	very bad compliance
51% within expected error (S. South America)[1]	marginal compliance
Measurement Characteristics	
Platform:	Terra
Instrument:	MODIS on Terra
Collection:	MODIS Terra Collection 5.1
Algorithm:	dark target land
Local Observing Time:	10:30:00

References

1. Hyer, E. J., Reid, J. S., Zhang, J., (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 (2011), info:doi/10.5194/amt-4-379-2011

2. Kleidman, R. G., Levy, R. C., Mattoo, S., Smirnov, A., Tanré, D., (2011), Evaluation and Wind Speed Dependence of MODIS Aerosol Retrievals Over Open Ocean, IEEE Transactions on Geoscience and Remote Sensing, 1-7,

Early version of MDSA Advisor Quality Facts interface

- We are developing a dynamic presentation of product quality for the MODIS Aerosol data product
  - Domain extensions to the data entity concept used to define the scope for which evidence and assertions are valid
    - e.g. spatial extent, temporal extent, aerosol loading, processing history
  - A report is compiled based on user interest – e.g. only show regional-specific information for regions the user is interested in
  - Science explanations, visuals, and background information are linked from the main table
  - Clear attribution for evidence
- In future, assertion criteria could be determined by the user
  - e.g. show assertions that make sense for *my intended use*
- Presentation design guided by and validated by aerosol scientists

Would you like to know more?

Visit the data product quality ontology primer at http://bit.ly/uVqjNn

Get the poster at http://bit.ly/uMyFgF
(or snap the QR code above)

Glossary:

RPI/TWC –Rensselaer Polytechnic Institute / Tetherless World Constellation
NASA/GSFC – National Aeronautics and Space Administration / Goddard Space Flight Center
MODIS – Moderate Resolution Imaging Spectroradiometer
AOD – Aerosol Optical Depth

Sponsors:

- NASA Earth Science Technology Office (ESTO)

Acknowledgments:

This work was supported in part by the NASA ESTO AIST-08-071 project "Multi-Sensor Data Synergy Advisor" (PI: Gregory Leptoukh, NASA GSFC)