A Review of EOS Terra Quality Assessment

Bob Lutz, Code 423, NASA Goddard Space Flight Center, Greenbelt MD 20771
301-614-5037 / Fax 301-614-5270 / rlutz@ltpmail.gsfc.nasa.gov
David P. Roy, Department of Geography, University of Maryland College Park, MD 20742
301-614-5571 / Fax 301-614-5269 / droy@kratmos.gsfc.nasa.gov
Craig Leff, Jet Propulsion Laboratory**,** MS 169-315, 4800 Oak Grove Drive, Pasadena, CA 911109
818-354-8885 / Fax 818-393-1935 / cleff@haleakala.jpl.nasa.gov
Scott Lewicki, Jet Propulsion Laboratory, MS 169-315, 4800 Oak Grove Drive, Pasadena, CA 911109
818-354-3534 / Fax 818-393-4802 / scalle@jord.jpl.nasa.gov
Erika Geier, Mail Stop 423, NASA LaRC, Hampton, VA 23681-2199
757-864-1854 / Fax 757-864-8295/ e.b.geier@larc.nasa.gov
Dan Ziskin, NCAR, PO Box 3000, Boulder CO 80307
303-497-2920 / Fax 303-497-2920 / ziskin@ucar.edu
Katherine Kilpatrick, University of Miami, RSMAS, Miami, FLA 33149
305-361-4615 / Fax 305-361-4622 / kay@rrsl.rsmas.miami.edu
Allen Chu, SSAI, Code 913, NASA Goddard Space Flight Center, Greenbelt MD 20771
301-286-4080 / Fax 301-286-1759 / achu@climate.gsfc.nasa.gov

ABSTRACT

Terra is the flagship platform of NASA's Earth Observing System (EOS) carrying the ASTER, CERES, MISR, MODIS, and MOPITT instruments in a polar sun-synchronous orbit. These instruments sense approximately 200 gigabytes of data per day which are processed to produce a suite of standard products. QA involves the identification and labeling of those products which obviously and significantly do not conform to their expected accuracy/performance. This paper overviews the components of EOS QA and the different QA strategies developed by the Terra science teams.

INTRODUCTION

The NASA Earth Observing System (EOS) [1] consists of space flight, science and data production, archive and distribution components. Terra is the first comprehensive EOS satellite and has an associated science team (ST) for each Terra instrument. The STs are located at Science Computing Facilities (SCFs) and are responsible for developing the science algorithms and processing software to convert instrument data into derived geophysical products, at several processing Levels [2]. The EOS data production, archive and distribution components are facilitated by the EOS Data Information System (EOSDIS) and its infrastructure, the EOSDIS Core System (ECS) [3].

EOS products are stored as data granules consisting of one or more physical files defined in a common structure according to the ECS data model [4]. A granule is the smallest aggregation of data that is independently described and inventoried by the ECS. A major objective of any scientific data processing and distribution system is to identify suspect and bad data. The STs are responsible for ensuring the quality of their products through instrument characterization/calibration, validation, and quality assessment (sometimes termed quality assurance) [5] activities. EOSDIS supports these activities.

COMPONENTS OF EOS QA

All Terra products are stored using the ECS data model and include metadata that provide descriptive information that may be queried as part of the data order process. The results of QA undertaken by the production code, by the processing facility staff, and by the science team are documented in the core metadata (common for all EOS products) listed in Table 1. Additional core metadata include the percentage of pixels in the granule that are: missing, out of bounds, interpolated or cloudy. Product specific QA metadata are also stored, but are too numerous to list in this paper.

Table 1: EOS QA Metadata

QA Metadata Name	Valids
AutomaticQualityFlag	Passed, Failed, Suspect
AutomaticQualityFlagExplanation	Text (255 Characters)
OperationalQualityFlag	Passed, Failed, Suspect, Inferred Passed, Inferred Failed, Being Investigated, Not Being Investigated (default)
OperationalQualityFlagExplanation	Text (255 Characters)

ScienceQualityFlag	Passed, Failed, Suspect, Inferred Passed, Inferred Failed, Being Investigated, Not Being Investigated (default)
ScienceQualityFlagExplanation	Text (255 Characters)

Production Code QA - Production code QA, is performed by the software that generates the science data product. All metadata (except the Operational and Science Quality flags) and all sub-granule QA information are set at this stage. The Automatic Quality Flag is designed to alert the ST of granules that should be investigated in more detail. It may also be used by the processing system to mediate production (e.g., if the Automatic Quality Flag is set to Failed, do not use that granule as input to produce other granules).

Processing Facility QA - ASTER, MISR and MODIS (Level 1) products are generated using ECS functionality with CERES, MODIS (Level 2 and 3), and (initially) MOPPIT products generated using non-ECS functionality. All products are archived and distributed to the public from one of four EOS Distributed Active Archive Centers (DAACs): EROS Data Center (EDC), Goddard Space Flight Center (GSFC), Langley Research Center (LaRC), and the National Snow and Ice Data Center (NSIDC). The processing/archiving facilities are responsible for monitoring *non-science* quality aspects of EOS data production and archival. Checks are performed to ensure that the data are not corrupted in the production, transfer, archival, or retrieval processes. These may include checking that the file can be opened, and that the file size is correct; and examination of production history information. The results of these analyses are summarized in the Operational Quality Flag and the associated text explanation.

Science Team QA - The science teams are formally responsible for the *scientific quality* of their products. ST QA activities may be categorized into those of an investigative nature, principally analyzing suspect data, and those of a routine nature, involving regular screening of samples of data products. The results of ST QA are summarized in the Science Quality Flag and the associated text explanation. These metadata carry the most detailed quality information and should be examined by the general user.

EOSDIS INTERFACES

Earth Observing System Data Gateway (EDG) - The EDG allows the user community and ST to browse, search, and order products held at the EOS DAACs. The EDG supports searches on all core and product specific metadata. In times when granules fail specified QA criteria (e.g., when one of the core Quality Flags is set to Failed), the EDG generates a warning message stating that these EOS products may not be suitable for scientific research.

Subscription - The ST may also obtain data from the DAACs via subscription functionality which compares product metadata when products are first archived with pre-defined subscription criteria, using relational operators as qualifiers on the values of specified metadata. Upon successful matching of subscription criteria (e.g., Automatic Quality Flag = Failed), the ST is notified and the products are sent to the ST's facility.

TERRA QA APPROACHES

Each Terra science team has developed a different science QA approach. ASTER and MODIS Land strategies are presented first, as these teams advanced concepts that were adopted by the other Terra science teams.

ASTER - is an imaging radiometer, with 14 multi-spectral bands from visible through thermal, providing high spatial resolution images of the land surface, water, ice and clouds.

ASTER stores pixel level QA information within QA data planes [6]. The first QA data plane is mandatory and common in format and data length for all ASTER products. Quality information is contained in the first four bits and cloud information is in the last four bits. The second data plane is product specific, optional and may vary in length.

The ASTER production code performs extensive automated QA and utilizes an alert system. Summary statistics are calculated, evaluated to determine if an alert condition exists and are stored in product specific metadata. Each product has its own set of alerts. If there is no alert generated or if the alert is of a non-critical nature, the granule passes Automatic QA. If the alert is defined as being critical; the granule fails Automatic QA, the flag is set to Failed and the granule goes into the "Manual QA" queue for Investigative QA analysis. In this mode, an operator evaluates suspicious granules. A standard evaluation procedure is developed for each product, based on algorithm developer input and operator experience, where procedures are expected to evolve as algorithm performance is better understood.

For Routine QA, up to ten percent of the data product stream is sampled. This sampling is not completely random, as more complex products are sampled more intensively. The selected granules are manually evaluated until routines are developed to detect defects automatically. Operators look for "anomalous" characteristics within the product such as: ringing, gradients (shading), checker-boarding, speckles and other obvious defects.

MODIS - senses all of the earth's surface in 36 spectral bands (visible through long-wave) producing a suite of land, ocean and atmospheric products at various resolutions. The MODIS ST is divided into land, ocean and atmosphere disciplines.

MODIS Land QA is undertaken at eight Land SCFs and at the Land Data Operational Product Evaluation (LDOPE) facility that was established to provide a coordination mechanism for Land QA activities. Visualization, knowledge-based and statistical analysis procedures are applied by the LDOPE and SCF personnel to individual granules and to collections of granules at different spatial and temporal scales. Known product issues are posted on a web page to passively communicate issues within the ST and are flagged in the Science Quality metadata for public scrutiny via the EDG. Each product carries product specific QA information at the pixel and granule level. To enable consistent interpretation, all Land products carry two generic QA bits per-pixel, which are summarized over each granule as four QA metadata.

MODIS Land QA procedures are applied in a purposeful manner - where product metadata indicate low product quality, where algorithm understanding predicts expected problems, and after visual examination of low spatial resolution global browse products. A database populated with the metadata of every land product is used by LDOPE and SCF personnel to establish which products should be ordered for detailed analysis and to investigate the likely causes of suspected problems in these and related products. The database has a web interface and supports graphical display and querying of product metadata via relational and boolean operators. After the initial product shakeout QA procedures will be applied to product time series and to a regular sample of products to capture quality issues not evident by purposeful sampling. The MODIS Land QA approach is described in more detail in [7] and [8]. The LDOPE and SCF personnel are expected to examine no more than 10% of the daily averaged Land data production.

MODIS Atmospheres generates common pixel level QA parameters for all Level 2 products [9]. Their QA plane consists of cloud mask flags, data quality flags and retrieval path flags. The Automatic Quality Flag is set using science-team defined thresholds of the successful rate of retrievals within a granule. This parameter is also saved as an important QA product specific metadata. Granules that possess a setting of the Automatic Flag = Fail, trigger an Investigative QA scenario.

Ten per cent of the data is routinely examined, with inferred quality to the rest of data when possible. This inferred quality is based on time, geo-location, atmospheric condition, transport pattern and sources and sinks. Routine QA analyses are to periodically select granules in geographic regions where scientists are more familiar with and more confident (e.g., the SCARs experiments). In this mode, the team checks Level 1B geo-location and ancillary data; QA flags and other metadata. Routine QA screening also entails generating Level 3 daily global maps to identify anomalies through comparison with climatological data. Once anomalies are identified, the input Level 2 granules are ordered and analyzed for suspicious quality. In addition, MODIS Atmospheres will attempt to compare Level 2 aerosol product retrievals with ground based measurements and perform statistical analyses in almost a real time mode**.**

MODIS Oceans stores extensive QA information at the pixel level for all Ocean products [10]. The first QA plane is common in format and contains: a) eight bits for eight quality related flags and b) four bytes for cloud mask flags. Contained within the second QA data plane is pixel level QA information unique for each product.

MODIS Ocean Level 2 product QA analyses focus on the examination of the pixel level flags contained in the QA planes. In addition, MODIS Ocean team members subscribe to their respective Level 3 products for daily review of processing correctness and reasonable data values.

Adopting the approach of MODLAND, MODIS Oceans has established a centrally coordinated facility, the MODIS Ocean Data Assessment Team (MODAT).. MODAT staff performs routine data visualization tasks, supervises automated QA procedures and investigates problems detected during Routine QA. In addition, the Miami SCF requests transfer of Level 1A, Level 3, geo-location and ancillary data to their facility, for parallel processing and algorithm development. The Miami SCF also receives summary QA reports from the MODAT to aid in their science QA analyses. Post-run time science QA results are derived by: application of visualization and statistical analysis procedures to the generated products, examination of run time QA results stored in the generated data products and analysis of temporal, zonal, meridian, secular and regional trends of Level 3 generated products

CERES - has two broadband scanning radiometers aboard Terra. CERES products include: atmospheric fluxes, surface radiation budget components, and cloud parameters (cloud height, cloud optical property determination, etc.)

To aid in investigating problems and to perform science quality assessment, CERES makes extensive use of Quality Control (QC) files [11]. Every CERES subsystem produces one or more of these QC files, which are archived at the DAAC. The files consist of various formats: ASCII reports, binary files and plot files. QC files may contain counts of

particular event occurrences, important intermediate values, mean values, standard deviations, trend information, etc.

CERES sets the Automatic Quality Flag to Failed when the production software exit code is not equal to zero. When this condition occurs, the science team is notified by Email, setting up an Investigative QA scenario. These output granules are then ordered with their QC files and placed in a special Error directory for investigation by the science team.

For Routine QA analysis, the science team subscribes and pulls QC files for science quality assessment. Based on the analyses of these QC files, the CERES team may then obtain selected data granules for further QA analysis. Every data product, which can be ordered by the general public, also has a Quality Summary associated with it. This Summary describes the data product, lists the validation and QA processes which were performed, lists uncertainties, states noteworthy cautions, and specifies scheduled product revisions.

MOPITT - is an eight channel radiometer that measures radiances in three infrared wavelengths. MOPITT's products include: atmospheric profiles of CO concentrations at several levels and total column CO and CH4 measurements.

Investigative QA is the primary mode of analysis for at least the first 6 months after launch [12]. To support this endeavor, MOPITT creates temporary files of data product and extended diagnostics during product generation. These diagnostic files are used to collect ancillary data and algorithm anomalies (e.g., numerical library error conditions). The diagnostic files are dynamic in nature, responding to the types of problems uncovered. In addition, QA summary files are also generated by post-processing running software.

A comprehensive set of graphical display tools has been developed that allow problems to be tracked from the diagnostic and QA summary files, through lower level data products. Results from these Investigative QA analyses bound "what is normal" and define a MOPITT Procedures Document. This document is utilized for Routine QA. It defines: the content within the QA summaries to be reviewed, QA criteria for Pass / Fail and the subsequent actions to be taken by the scientists in the cases of data granules that fail science team QA. To aid in Routine QA, graphical displays of trends in the data are analyzed, as well as utilizing historical data to detect changes in the instrument and products. Significant negative results in Routine QA analyses initiate the Investigative mode of QA.

MISR - is an instrument that has nine cameras that view the Earth at nine discrete angles in four spectral bands. MISR provides top-of-the-atmosphere, cloud and surface angular reflectance functions; global maps of planetary and surface albedo; and aerosol and vegetation properties.

Since a single MISR granule contains an entire orbit's worth of data, the MISR team concluded that core and product specific QA metadata would be of little value to most users. Instead, the team stores QA statistics within the MISR products at several data levels: global swath, swath, block, line, pixel and grid cell [13].

For the immediate post-launch period, the Automatic Quality Flag is set to Suspect and will continue until the science team has studied MISR data in detail in an Investigative mode, validating instrument and algorithm performance by hand.

For every granule that MISR generates, a separate QA file is generated; consisting of core metadata, MISR-specific QA metadata, and sub-granule QA parameters. These QA granules are copied and stored in an external database within the DAAC. A search and query subsystem within the database provides MISR scientists with a user-friendly web-based GUI interface to search for QA granules. Preliminary Quality Assessments are performed on these files and as needed, additional data granules are ordered for further analysis. To cope with the large data volume, interactive visualization tools have been developed to inter-compare data values at many levels of aggregation, and to report graphical and statistical summaries of the results.

In addition, MISR staff places standing order subscriptions for Level 1 and Level 2 Browse products, to ensure that there are no gross errors in the output products.

CONCLUSIONS

This paper has presented the various approaches that the Terra science teams and their processing facilities have developed to ensure the quality of the Terra data products. Overall, QA methodologies tend to be more alike than different, which in part is due to the ECS data model that they employ and also reflects the exchange of information through the workings of the EOS QA Working Group. QA is an evolving element within EOS. Maturity in the understanding of the behavior of the instruments and revised science algorithms as a result of QA and validation campaigns should see a gradual change from Investigative QA to Routine QA. It is expected that during the first year post-launch, a greater emphasis will be placed on analyzing suspect data than routinely screening data products. The ST QA approaches will be described in more detail in separate publications.

REFERENCES

[1] G. Asrar and J. Dozier, *EOS Science Strategy for the Earth Observing System,* American Institute of Physics Press, Woodbury, N. Y., 1994.
[2] S. Wharton and M. Meyers, *1997 MTPE EOS Data Products Handbook*, National Aeronautics and Space Administration, Washington, D. C., 266 pp., 1997.
[3] G. Asrar and H. K. Ramapriyan, "Data and Information System for Mission to Planet Earth.", *Remote Sensing Reviews* **13**, pp. 1-25, 1995.
[4] C. Gross, *B.0 Implementation Earth Science Data Model, ECS Document #**420-TP-015-002**,* Hughes Information Systems Company, 1997.
[5] B. Lutz, "Development of EOS Quality Assessment (QA) methodology and EOSDIS' support for QA.", *Proceedings of the International Society for Optical Engineering (SPIE) Meeting, San Diego, Earth Observing Systems III*, pp. 531-540, 1998.
[6] C. Leff, *ASTER Higher-Level Data Product Quality Assessment Plan, Version 2.0, JPL D-**13841***, Jet Propulsion Laboratory, California, 1999.
[7] D. Roy, *The MODIS Land Quality Assurance Plan,. Version **2.0,*** University of Maryland, Maryland, 1999.
[8] D. Roy et al., "The MODIS land quality assessment approach", *IGARSS 2000*, Honolulu, Hawaii., 2000.
[9] A. Chu, K. Strabala, S. Platnick, S. Mattoo, R. Hucek, and B. Ridgway, *MODIS Atmosphere QA Plan. Version **2.0,*** Goddard Space Flight Center, Maryland, 1999.
[10] K. Kilpatrick, *MODIS Oceans Quality Assurance Plan, Version 2.**0***, University of Miami, Florida, 2000.
[11] P. K. Costulis et al., *CERES Data Management System Quality Assurance Plan, Release **3***, Hampton, Virginia, 1999.
[12] D. Ziskin and P. Bailey, *MOPITT Science Data Product Quality Assurance Plan, Version **2,*** National Center for Atmospheric Research, Colorado, 1999.
[13] D. Diner, et al., *MISR Processing Quality Assessment Plan (JPL D-**13496**) and MISR Science Quality Indicators (JPL D-**13965**)*. Jet Propulsion Laboratory, California, 1997.