Development of an Operational System for the Retrieval of Aerosol and Land Surface Properties from the Terra Multi-angle Imaging SpectroRadiometer

Kathleen A. Crean
Jet Propulsion Laboratory, California Institute of Technology
Kathleen.A.Crean@jpl.nasa.gov

Abstract

An operational system to retrieve atmospheric aerosol and land surface properties using data from the Multi-angle Imaging SpectroRadiometer (MISR) instrument, currently flying onboard NASA's Terra spacecraft, has been deployed. The system is in full operation, with new data products generated daily and distributed to science users worldwide. This paper describes the evolution of the system, from initial requirements definition and prototyping through design, implementation, testing, operational deployment, checkout and maintenance activities. The current status of the system and future plans for enhancement are described. Major challenges encountered during implementation are detailed.

1. Introduction

The Multi-angle Imaging SpectroRadiometer (MISR) [1] is one of five instruments currently flying aboard NASA's Terra spacecraft [2], which was launched in December 1999 as part of NASA's Earth Observing System (EOS). The instrument provides multi-angle views of the Earth at nine discrete viewing angles up to 70.5°, in four spectral bands. A ground data processing system [3] installed at the NASA Langley Atmospheric Sciences Data Center (ASDC) [4] produces and distributes MISR data products to the user community. The data products, which provide information for studying the Earth's ecology, climate and environment, include estimates of aerosol and cloud properties, surface and top-of-atmosphere reflectance characteristics, and land cover properties. This paper focuses on the portion of the ground data processing system which produces the aerosol [5] and surface [6] science data products.

The evolution of the operational aerosol and land retrieval software is described in section 1, followed by a detailed characterization of the software in section 2. Section 3 recounts the technical challenges faced during development and deployment and their resolutions. The current status of the system is detailed in section 4, and future plans are discussed in section 5.

2. System evolution

Development of the software which performs the operational MISR aerosol and surface retrievals has spanned more than a decade. This section describes the evolution of the system over time.

Initial work began in the early 1990s. Early efforts focused on software requirements definition and algorithm prototype testing. Activities occurring during this time included science requirements walkthroughs with the MISR algorithm scientists, system requirements and design meetings with the MISR system engineers, and a series of science algorithm prototype tests. The major activity during the mid-1990s was the creation of an aerosol and surface product generation executable (PGE) which was based upon existing science algorithm prototypes. (The term PGE refers here to a set of compiled executables or scripts which produce a MISR data product. See [3] for details.) In contrast to the prototype codes, the PGE was required to run under the operational environment at the NASA Langley ASDC [3]. In this regard, the PGE software development effort required not only the implementation of the aerosol and surface science algorithms, but also the use of error handling, message logging, unit tests, inter-process communications, standard metadata and other features required by the software's operational environment. Furthermore, files used by the PGE were required to conform to the hierarchical data format (HDF) standard [7], or in some cases a special extension to HDF called HDF-EOS [8], for compatibility with the operational environment. The late 1990s saw the design, implementation, testing, delivery and installation of increasingly functional versions of the operational software to support the mission. One activity of significance that occurred during this time was a "tiger team" effort which corrected a significant performance issue in the software (described in section 4.3). MISR launched into orbit on December 18, 1999, and began collecting imagery of the Earth upon opening of the instrument cover on February 24, 2000 [9]. Post-launch activities included detailed product checkout and continued software upgrades to fix software bugs, implement algorithm improvements, and add new science functionality. A significant activity which occurred soon after launch was the formation of another "tiger team" effort which addressed spatial coverage issues in the aerosol and surface data products. The tiger team activity resulted in significant improvements to the spatial coverage of these data products.

A limited number of aerosol and surface sample products were released to the public in October 2000. Aerosol and surface standard products became routinely available beginning in February 2001, at a Beta (minimally validated; may still contain significant errors) maturity level. The aerosol and land products were promoted to a Provisional (partially validated) maturity designation in June 2002, except for a few areas, such as the land cover variables LAI (leaf area index) and FPAR (fraction of photosynthetically active radiation), which were promoted to Beta and then Provisional status in November 2002. Current work is underway to support the transition of products to a Validated (well-defined uncertainties) maturity level.

Figure 1 shows the timeline of software activities.

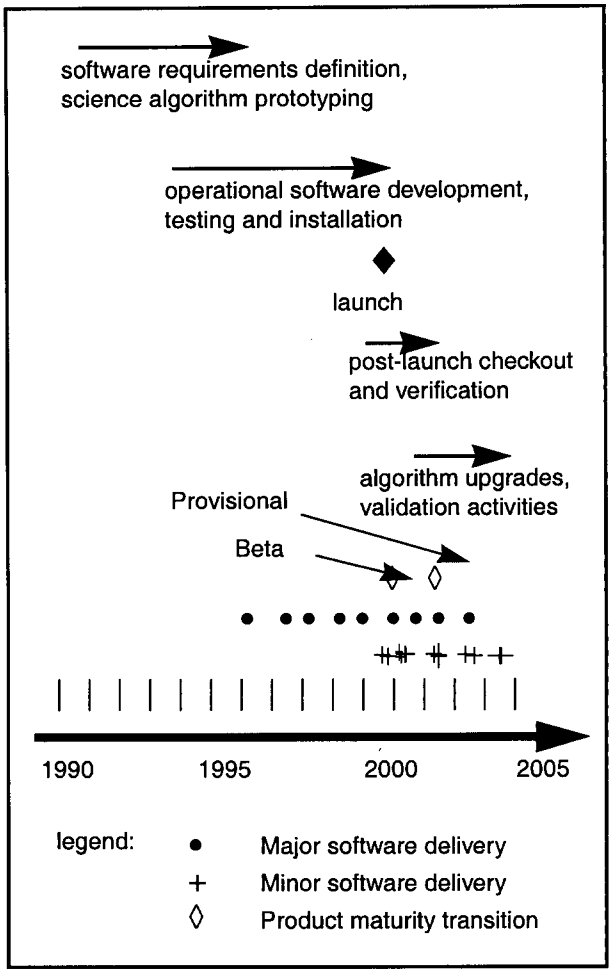

Figure 1. Activity timeline

As shown in Figure 1, major software deliveries have been performed once or twice per year. Several minor ("patch" or "increment") deliveries have been performed between major deliveries.

Staffing of the aerosol and surface software development effort grew steadily until launch, whereafter it declined steadily. Figure 2 shows a timeline of the staffing profile, with each data point representing the average number of full-time engineers (FTEs) used annually.

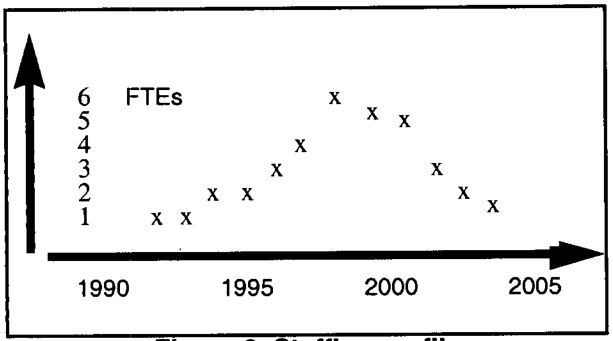

Figure 2. Staffing profile

3. System characterization

The software development staff spent considerable effort on the requirements and design phases. This section describes the final design which resulted from these activities.

The aerosol and surface PGE, known as "PGE9," consists of a core service process ("PCS_main") and two child processes ("AS_preprocess" and "AS_main"). The core service process handles PGE startup, shutdown, inter-process communications between child processes, and the interface with the production environment. The first child process (AS_preprocess) filters out cloud-contaminated pixels, pixels with poor radiometry, and other data of unacceptable quality. It also corrects the input data for ozone absorption and other factors. The second child process (AS_main) performs the aerosol and surface retrievals and writes the results to the output files.

A single run of PGE9 processes one granule of MISR data , where a granule corresponds to the amount of data collected during one complete MISR orbit (approximately half an orbit of data, since MISR collects data on the daylit portion of the orbit only). During a run, PGE9 processes the granule segment by segment, where a segment consists of a "block" of data (an area corresponding to 140.8 km in the spacecraft downtrack direction, and 563.2 km in the spacecraft cross-track direction). There are typically 141 segments in a granule. As soon as AS_preprocess finishes processing a segment of data, the results are sent to AS_main and AS_preprocess then begins processing the next segment. AS_main completes processing of the received segment, and then waits for the next segment to arrive from AS_preprocess. The granule is thus processed segment-by-segment in pipeline fashion, until processing of all segments completes. This design was chosen for its ability to meet PGE memory and run-time constraints [3]. Figure 3 illustrates the high-level PGE9 design. Figure 4 shows a snapshot of the data flow through the system, where the first segment has just completed processing and is ready to discard, the second segment is being passed from the first to second process, and segment is beginning processing.

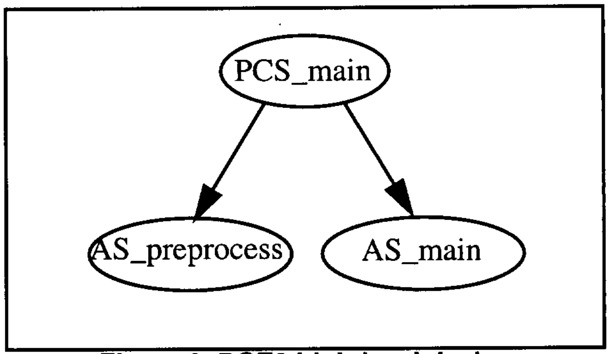

Figure 3. PGE9 high-level design

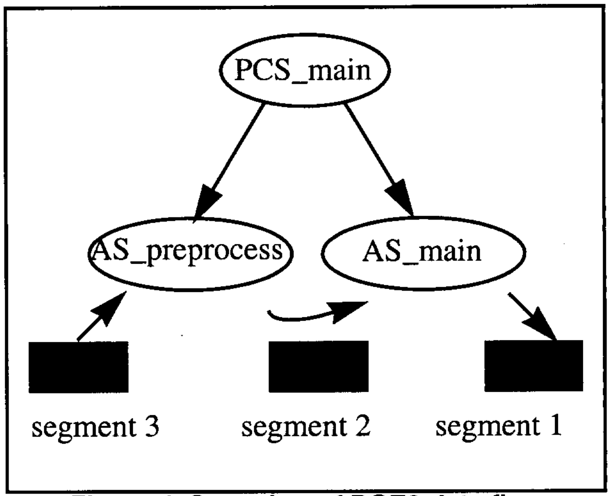

Figure 4. Snapshot of PGE9 data flow

PGE9 was written in the C and Fortran90 programming languages. The science algorithms were written in Fortran90, at request of the MISR algorithm scientists, while C was used for the remainder of the software.

In addition to PGE9, several other support codes are used at the MISR science computing facility (SCF) to generate static ancillary datasets [11] which are required inputs to PGE9. These include the Ancillary Climatology Product (ACP), the Simulated MISR Ancillary Radiative Transfer (SMART) dataset, and the Canopy Architecture Radiative Transfer (CART) dataset. These programs were also written in C and Fortran90. Each piece of software is characterized in table 1 below.

Table 1: Software characteristics

Program	Lines of code[a]	Run time	Number of input files
PGE9	248K [12]	1.8 hr/ orbit	52
ACP gen.	15K	negligible	7
SMART gen.	40K	26 hrs	5
CART gen.	3K	negligible	7

a. Includes comments, blank lines, and unit tests

4. Technical challenges

Many technical challenges were encountered and addressed over the course of development. This section describes several of the most significant hurdles which arose and were overcome. It can be viewed as a "lessons learned" for future projects, as well as a collection of examples of current space mission challenges for information technology.

4.1 Lack of instrument and algorithm heritage

The MISR experiment is the first of its kind. Due to its uniqueness in using many view angles to derive geophysical information, there was little in-flight heritage on which to base the aerosol and land surface retrieval algorithms. Prior to launch, the MISR scientists had to rely on data from other sources such as the Advanced Solid-state Array Spectrometer (ASAS) aircraft instrument to test their new MISR algorithms [5].

Once the MISR experiment became operational and new science products became available, the MISR scientists were finally able to see the results of their new retrieval algorithms. It was then desireable for them to experiment on modifications to their algorithms to improve the retrievals. A workable way for them to try out their new ideas on the PGE9 code had to be established. Prior to launch, all PGE9 software development was performed by the software staff, under the direction of the scientists. However, this method eventually became too slow and ineffective, as the software staff declined post-launch (see Figure 2), and as the scientists desired direct access to the standard production code themselves. The solution lay in the creation of a "sandbox" area, in which the scientists could experiment with a copy of the actual production code. In the sandbox, the scientists could make algorithm modifications and run them on several test cases. This proved to be a more efficient and satisfying way of managing science algorithm experimentation after launch.

4.2 Data volume constraints

The MISR instrument collects roughly 36 GB of raw data per day. The data undergo several levels of processing [3], resulting in over 100 GB per day of higher-level data products which must be archived and made available to users. Given a nominal mission lifetime of six years, the amount of data which must be stored is significant. Therefore, considerable attention was devoted to minimizing data volume when designing data products. To achieve this end, many high-volume parameters were converted from floating point numbers to integers and stored in scaled form, as one- or two-byte integers. Another tactic employed was the packing of multiple flag values into a single field. Documentation [13] provides the user with instructions for unscaling and unpacking the values. Finally, data compression was also incorporated, using the GDsettilecomp routine from the HDF-EOS library [14].

4.3 Performance issues

The MISR instrument orbits the Earth every 99 minutes. The ground data processing system must keep up with this continual stream of new data. In addition, it must also support the reprocessing of earlier data to maintain currency with newly-updated algorithms.

The aerosol and surface operational software addressed performance issues in a number of ways. First, like all MISR PGEs, it utilized concurrent processes to allow a full swath of data to be processed in small chunks (i.e. MISR "blocks"), in assembly-line fashion. Second, many computationally-intensive radiative transfer calculations were pre-computed and stored in look-up tables, to avoid encumbering the standard production code with expensive on-the-fly computations. Examples of this include the SMART dataset, which stores information used in aerosol retrievals, and the CART dataset, which stores values used in the land surface retrievals. Finally, further tradeoffs between processing speed and numerical accuracy were made when it became apparent that the software was not performing at an acceptable rate, as described below.

The performance problem occurred within the SMART dataset access portion of the software. The SMART dataset contains more than 2 GB of precalculated quantities which are required by PGE9 during standard processing. Due to memory constraints, the SMART dataset cannot be read into memory all at once. Therefore, the PGE must access a chunk of SMART data as needed, and throw it away when finished, whereupon it then accesses another chunk of SMART data for the next processing segment. Initial science requirements dictated that PGE9 access many small chunks of SMART data for each MISR block processed. However, using the small chunk size resulted in PGE9 taking approximately two orders of magnitude longer than required, to complete processing of a single orbit. This being an unacceptable condition, the science requirements were relaxed with the consent of the MISR scientists, and the software was redesigned to access larger chunks of SMART data which would be used to process a MISR block. The chunks were large enough to allow for more acceptable input/output (I/O) performance, yet not too large as to exceed memory constraints and scientific accuracy requirements. To date, PGE9 is performing within requirements, taking on average less than two hours to process an orbit.

4.4 Team software development

The PGE9 software development effort spans nearly a decade, and has been performed by over two dozen members of the MISR team. Such an undertaking required extensive coordination and cooperation among team members, to maintain the integrity and consistency of the software. Notable practices which helped the PGE9 development progress smoothly included 1) the establishment of regular design team meetings by the MISR system engineering staff, in which all developers participated and agreed to common tools, templates, and coding conventions; 2) extensive use of software configuration management tools in a consistent manner across the project; 3) use of software peer reviews; and 4) close interaction with the customer (i.e. the MISR scientists) through weekly meetings. While conventional wisdom advises that "adding manpower to a late software project makes it later" [15], we found that we were able to add several developers from other MISR subsystems to fix the PGE9/SMART performance problem described in Section 4.3, and successfully meet a tight deadline. (Redesign efforts began in August 1998; implementation started in September, and the redesigned software was largely in place by the December deadline. At the height of the activity, 12 developers worked in parallel to complete the task.) The success of this effort can be attributed to the clean, well-documented redesign performed by a few team members and by the cohesiveness of the entire implementation team, which developed over time as a result of the factors mentioned above.

4.5 Ancillary dataset development

PGE9 requires the use of several ancillary datasets. The creation of these ancillary files, which include global meteorological climatologies, aerosol particle property descriptions, radiative transfer look-up tables, and science threshold configuration files, entailed significant development effort. One extreme example of this was the development of the SMART dataset. The SMART dataset contains precomputed radiative transfer calculations. The computationally-expensive calculations must be performed over a number of independent variables, each on a unique grid, making the runtime a significant issue. Furthermore, verification of the code and its results posed a real challenge, as many of the quantities were multidimensional in nature, and special visualization techniques had to be developed in order to perform the verification. Several man-months were spent on the development and refinement of the SMART generation system. Under the current system, generation of this 2.5 GB dataset takes approximately 26 hours (wall-clock time). Verification is performed using tools built with IBM's Open Data Explorer (OpenDX) [16]. The OpenDX tools proved to be invaluable in completing the SMART checkout [17].

5. Current status

The MISR experiment is currently halfway through its nominal mission lifetime of six years.

Current efforts in MISR's aerosol and surface operational software development focus on two key areas: 1) product improvements and 2) validation activities which will lead to a maturity status of Validated (well-defined uncertainties). Product improvements include further algorithm upgrades, and improvements to the datasets which are input to the software, most notably the datasets containing aerosol properties [11]. Validation activities include the systematic comparison of MISR data products to independent, external data sources, notably the aerosol products available from the Aerosol Robotic Network (AERONET) [10].

A system-wide reprocessing of all MISR standard data products is currently underway. This activity will ensure that data products from earlier years are available at the current maturity status.

The development of monthly and seasonal global summaries of the aerosol and surface parameters is underway. The summaries are due to become available this year.

6. Future plans

Discussions of future directions in work effort are in progress. However, the following activities are nominally planned: 1) further validation activities which compare MISR products with data from external sources in a systematic, comprehensive manner; and 2) further improvements to the aerosol particle models which are input to the aerosol retrievals in PGE9. Other activities which are desired, but will be pursued only as schedule and resource constraints allow, include 1) implementation of science algorithms which retrieve parameters over ocean surfaces in addition to land surfaces; 2) the ingest of ancillary meteorological data from real-time external sources (e.g. from the Data Assimilation Office) instead of the current practice of using climatological values; and 3) the production of data products at a higher spatial resolution, to aid in regional studies.

7. Conclusion

The MISR aerosol and surface standard production system was developed and deployed over the last decade by a dedicated team of individuals. Many challenges arose during the development and all have been successfully addressed. Provisional-quality data products are now available from the ASDC. Work is continuing to enhance the system and transition products to a Validated maturity status.

8. Acknowledgements

Many members of the MISR team at the Jet Propulsion Laboratory contributed to the development of the MISR aerosol and surface operational retrieval software over the past several years. Susan Paradise led the software development in the early years, and made significant contributions throughout the lifecycle. Earl Hansen provided system leadership and support throughout the task. Mike Bull and David Nelson provided significant contributions to design and implementation activities. Dave Diner and John Martonchik provided science guidance as the principal investigator and lead algorithm scientist respectively. Mike Smyth is currently implementing the aerosol and surface global data products. The following JPL individuals are also acknowledged for their contributions to development of the software: Mark Apolinski, Rick C de Baca, Brian Chafin, Peter Glover, Lucia Marino, Kyle Miller, Duncan McDonald, Ruth Monarrez, Brian Rheingans, Alec Shaner, and Jack Tuszynski. Additionally, the work of the JPL MISR project, system and operations staff, and the support of the operations and user services staff at the NASA Langley ASDC are recognized.

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

9. References

[1] D. Diner, J. Beckert, T. Reilly, C. Bruegge, J. Conel, R. Kahn, J. Martonchik, T. Ackerman, R. Davies, S. Gerstl, H. Gordon, J.-P. Muller, R. Myneni, P. Sellers, B. Pinty, and M. Verstraete, "Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview," *IEEE Trans. Geosci. Remote Sensing*, July 1998, vol. 36, pp. 1072-1087.

[2] http://terra.nasa.gov

[3] G. Bothwell, E. Hansen, R. Vargo, and K. Miller, "The Multi-angle Imaging SpectroRadiometer Science Data System, Its Products, Tools, and Performance," *IEEE Trans. Geosci. Remote Sensing*, July 2002, vol. 40, pp. 1467-1476.

[4] http://eosweb.larc.nasa.gov

[5] D. J. Diner, W. A. Abdou, T. P. Ackerman, K. A. Crean, H. R. Gordon, R. A. Kahn, J. V. Martonchik, S. McMuldroch, S. Paradise, B. Pinty, M. M. Verstraete, M. Wang, and R. A. West, "Level 2 aerosol retrieval algorithm theoretical basis," Jet Propulsion Laboratory, Pasadena, CA, JPL Tech. Doc. D-11400, Rev. E., 2001.

[6] D. J. Diner, J. V. Martonchik, C. Borel, S. A. W. Gerstl, H. R. Gordon, Y. Knyazikhin, R. Myneni, B. Pinty, and M. M. Verstraete, "Level 2 surface retrieval algorithm theoretical basis," Jet Propulsion Laboratory, Pasadena, CA, JPL Tech. Doc. D-11401, Rev. D., 1999.

[7] http://hdf.ncsa.uiuc.edu/

[8] http://hdf.ncsa.uiuc.edu/hdfeos.html

[9] D. J. Diner, J. C. Beckert, G. W. Bothwell, and J. I. Rodriguez, "Performance of the MISR Instrument During Its First 20 Months in Earth Orbit," *IEEE Trans. Geosci. Remote Sensing*, July 2002, vol. 40, pp. 1449-1466.

[10] http://aeronet.gsfc.nasa.gov:8080/

[11] D. J. Diner, W. A. Abdou, H. R. Gordon, R. A. Kahn, Y. Knyazikhin, J. V. Martonchik, D. McDonald, S. McMuldroch, R. Myneni, and R. A. West, "Level 2 ancillary products and datasets algorithm theoretical basis," Jet Propulsion Laboratory, Pasadena, CA, JPL Tech. Doc. D-13402, Rev. B, 1999.

[12] M. Smyth, "V2.2/V2.2_i1 LOC," internal project email, 2002.

[13] S. A. Lewicki, K. A. Crean, S. E. Gluck, K. C. Miller, C. M. Moroney, and S. R. Paradise, "MISR Data Products Specifications Document," Jet Propulsion Laboratory, Pasadena, CA, JPL Tech. Doc. D-13963, Rev. E., 2002.

[14] L. Klein, "HDF-EOS Library Users Guide for the ECS Project, Volume 2: Function Reference Guide," Raytheon Systems Company, Upper Marlboro, MD, June 1999.

[15] F. P. Brooks, Jr., *The Mythical Man-month*, Addison-Wesley Longman, Inc., USA, 1995.

[16] http://www.opendx.org/

[17] D. Nelson, "SMART Generation Code - Changes from V2.1 to V2.2," MISR Internal Design File Memo 650-F, December 2001.