The Multi-angle Imaging SpectroRadiometer Science Data System, Its Products, Tools, and Performance

Graham W. Bothwell, Earl G. Hansen, Robert E. Vargo, and Kyle C. Miller

***Abstract*—Ground processing of data from the Multi-angle Imaging SpectroRadiometer (MISR) instrument, part of NASA's Earth Observing System (EOS), exploits new and unique science algorithms not previously used operationally. A range of data products from Level 1 through Level 3 is being produced. Because of MISR's unprecedented design, extensive prototyping was required from a relatively early stage. The data throughput is large, necessitating an innovative software design approach that maximizes performance. The systematic science processing software was developed at the Jet Propulsion Laboratory (JPL), with data processing occurring at the NASA Langley Research Center using the EOS Core System (ECS), a collaborative arrangement that works well. With the availability of actual mission data following launch on the Terra spacecraft in December 1999, MISR's computational needs have become better known, and many improvements have been made to both the science software and the production system to achieve a successful overall data processing capability. This paper provides information about MISR data for the science user, and describes the nature and scope of implementation and operations activities.**

***Index Terms*—Algorithms, data processing, data products, ground system.**

I. INTRODUCTION

THE Multi-angle Imaging SpectroRadiometer (MISR) [1] is one of the instruments aboard the Terra spacecraft, which is part of NASA's Earth Observing System (EOS), and was launched into earth orbit in December 1999. It provides multiple-angle, continuous imagery in reflected sunlight using nine separate pushbroom cameras observing the earth at nine discrete angles up to 70.5° relative to the local vertical, in four spectral bands.

The measurements of this instrument are designed to improve our understanding of the earth's ecology, environment, and climate. To facilitate this, a range of standard data products is available to the science community. These products range from raw instrument data to calibrated and geolocated radiances, geophysical retrievals of atmospheric and surface products, and global maps. This paper addresses those products and the system that produces them, both from the operational and the developmental viewpoints. It concentrates on the software and systems associated with the science algorithms for which the MISR Science team is responsible rather than the entire ground system of the Terra mission.

The paper begins with a description of the data products and then discusses the operational environment within which the products are generated. The remainder of the paper deals with implementation and design issues. It covers the software design and development process for the science algorithms, and the implementation of the software and of the processing system. It then looks at the most significant software tools that were developed for use by both developers and scientists.

The multiangle nature of the MISR instrument means that many of the algorithms and much of the production software have characteristics not encountered in earlier systems, and this characteristic is mentioned where it has major impacts on implementation and operation.

II. DATA PRODUCTS

This section provides an overview of the standard MISR data products, including an introduction to their content and the way they are constructed. More complete details of the MISR data products at the depth required for science utilization are given in the MISR Data Product Specifications document, which is accessible on the Internet [2]. The retrieval methodologies behind the data products are documented in a series of Algorithm Theoretical Basis documents, corresponding to the various products and supporting ancillary datasets. All of these are accessible on the Internet [3].

A. Product Designations

The MISR instrument delivers data in packets, which are recorded by the spacecraft and subsequently transmitted to the ground along with other data from the spacecraft. On the ground, the instrument packet stream is extracted from the larger mass of downlinked spacecraft and instrument data and reconstituted as the same packet stream originating from the instrument. This is Level 0, or raw data, and all of the MISR science and engineering data products are derived from it. Level 0 is the fundamental archive backup for MISR data; however, because of its complex, multiplexed format, it is not made available as a standard data product.

Beginning from Level 0, the generation of the standard science data products can be divided into five steps, as illustrated in Fig. 1. The numbering of these steps (1A, 1B1, 1B2, 2, and 3) conforms to the nominal product levels adopted for NASA's EOS missions [4]. Each step has at least one primary output product, but may have other secondary output products. It is

---

Manuscript received September 21, 2001; revised May 1, 2002. This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

The authors are with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109-8099 USA (e-mail: graham.bothwell@jpl.nasa.gov).

Publisher Item Identifier 10.1109/TGRS.2002.801152.

convenient to think of these five steps as occurring in sequence, with the predecessor producing at least one complete product, a portion of which is the primary input for the successor.

The five steps and their products follow.

1) *Level 1A instrument data reformatting and annotation*: The raw data from the instrument are reformatted into HDF-EOS computer files, and many quality checks carried out.
2) *Level 1B1 radiometric scaling and conditioning*: Radiance scaling converts each camera's digital number output to a measure of energy incident on the front optical surfaces of the MISR cameras. Conditioning modifies the radiances to remove instrument-dependent effects, such as focal-plane scattering.
3) *Level 1B2 geometric rectification and registration*: Images from the nine cameras are registered to one another and to the ground. This product, together with the Level 2 products, is mapped to the Space Oblique Mercator (SOM) projection [5], which is designed for continuous mapping of satellite imagery and minimizes spatial distortion. MISR's use of spatial resampling at Level 1 is unusual, but is necessary for coregistering the nine sets of images so that Level 2 retrievals can be done. This product has four constituent sets of parameters: a) radiances projected to the surface terrain, providing a common surface boundary condition for certain Level 2 geophysical retrievals; b) radiances projected to an ellipsoidal surface defined by the World Geodetic System 1984 (WGS84), where matching for Level 2 cloud stereo retrievals is done; c) a radiometric camera-by-camera cloud mask; and d) geometric parameters, including view zenith and azimuth angles, solar zenith and azimuth angles, and scatter and glitter angles.
4) *Level 2 science retrievals*: These products are geophysical measurements derived from the instrument data. a) The Level 2TC Top-of-Atmosphere/Cloud (TOA/Cloud) product contains measurements of top-of-atmosphere bidirectional reflectance factors, stereoscopically derived cloud heights and winds, top-of-atmosphere albedos, cloud fraction, and other parameters. b) The Level 2AS Aerosol/Surface product contains parameters such as tropospheric aerosol optical depth; aerosol composition and size information; surface directional reflectance factors; and other parameters.
5) *Level 3 global gridding*: These products are maps of parameters from the lower-level products, aggregated on monthly, seasonal, and other time scales, and using global grid cells of $0.5^\circ \times 0.5^\circ$ or $1^\circ \times 1^\circ$[6]. There are two parts to this product. Component products are simple statistical summaries of Level 2 geophysical and Level 1B2 radiance parameters. Joint products summarize interparameter relationships across the component parameters.

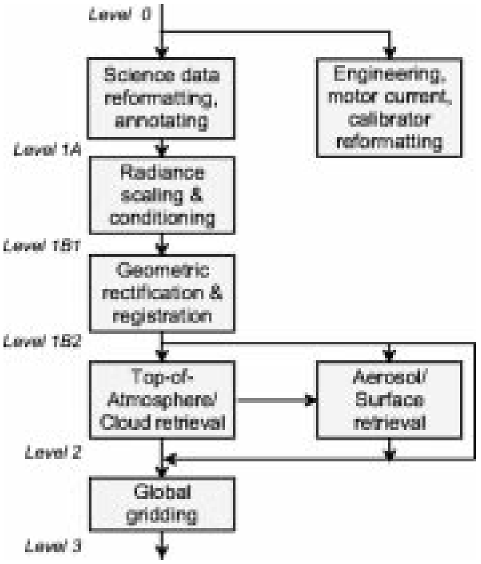

Fig. 1. MISR standard data product hierarchy.

B. Product Structure

In contrast with the Level 0 raw data, the MISR standard products conform with EOS mission requirements in using the HDF-EOS format. This is an extension to the native Hierarchical Data Format (HDF), originating from the National Center for Supercomputing Applications (NCSA) [7]. The extension allows for data structures pertinent to EOS satellite data, such as swaths and grids [8]. A special adaptation of HDF-EOS for MISR was necessary, and is discussed below in this section.

All of the MISR Level 1 and 2 science data is processed and archived as granules, each consisting of a single continuous swath. In this context, a swath consists of the entire two-dimensional data acquired on the illuminated part of the earth during one orbit. (MISR does not acquire science data on the night side of the earth.) The Level 1A and 1B1 products are based on the instrument data samples, and therefore employ the HDF-EOS Swath format. Level 1B2 and all Level 2 products are projected to the SOM grid, and use the HDF-EOS Grid format. Furthermore, all of the SOM products are constructed as a series of blocks, as shown in Fig. 2. These blocks provide a convenient segmentation for processing as well for subsetting by the user. Blocks must be wider than the swath width of any one of the nine cameras because the swaths do not exactly overlay. This arises because of earth rotation, and the degree of overlap varies with latitude. Within each block, the area either side of the swath is populated with fill values that can be readily distinguished from data values.

The spacecraft trajectory is constantly monitored, and orbital adjustments are made typically on a monthly basis so that the nominal orbit position is maintained to within ±20 km. As a result, spacecraft orbits and paths, and the MISR blocks, can be numbered and identified consistently with great precision. Thus, for each of the 233 paths in Terra's sun-synchronous orbit, MISR has defined 180 blocks in fixed geographic locations covering the full extent of the swaths. At any one time, 140 blocks are illuminated by the sun, and the position of these within the 180 moves up and down according to the season.

Because of the blocked nature of MISR's SOM-projected products, the HDF-EOS Grid data type required custom modifications by the EOS data system contractor to suit the MISR needs. HDF-EOS Grid is the implementation of HDF-EOS orig-

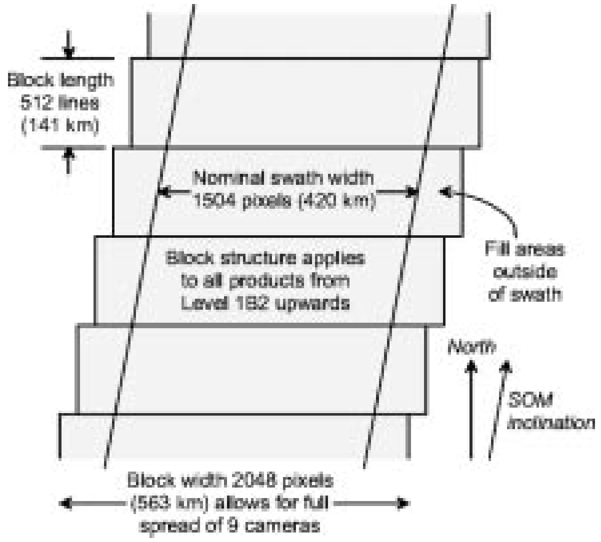

Fig. 2. Block structure of MISR SOM products.

inally intended for storing Level 3 and higher products; that is, products that have been "gridded" to a single earth-based map projection. The storage of map projection parameters is part of the format, and routines to access the data in Grid format by geolocation are supplied in the Grid API (Application Interface). On the other hand, MISR is required to store daylight-side orbit (swath) products at Level 1B2 and Level 2 in a georeferenced space-based map projection. In particular, as described above, MISR breaks up the L1B2 and L2 swath into equal-sized blocks. Changes were made to the Grid implementation, to handle these blocks as an additional dimension to a Grid dataset. This is referred to as the "stacked-block" Grid implementation.

In brief, the solution to meet MISR's needs is to "stack" all of the blocks of a swath into a single dataset, where the "third" dimension for the dataset becomes the block number. Groups of parameters of a product can be stored in these stacked-block Grid data structures, but each parameter in the dataset must have the same X and Y dimensions (i.e., same resolution). Within a Grid dataset, parameters can also be grouped into what HDF-EOS calls a "field," but each parameter in the field must be of the same data type (e.g., two-byte integer). The problem of each block having a different projection origin is handled by storing only the projection origin for Block 1, and saving in a separate dataset the integer pixel offsets from the upper left-hand corner of Block 1.

Several new data access tools were created by the MISR team to enable access to MISR data parameters stored in the stacked-block Grid format. These tools included the multiangle, multichannel viewer called *misr_view*, and the detailed file reader called *hdfscan*, which are described in Section VII.

C. Product Status

Standard practice for the Terra mission is that data products are classified by the maturity levels Beta, Provisional, and Validated. The definitions of these levels can be found on the EOS Terra data products Web page [9], and can be summarized as follows.

a) *Beta*: Minimally validated and may still contain significant errors.
b) *Provisional*: Partially validated.
c) *Validated*: Well-defined uncertainties.

The MISR products will progress through these maturity levels. As of April 2002, the Level 1 products are Validated while the Level 2 products are in Beta form, and the Level 3 Component product is being implemented.

D. Obtaining MISR Products

Users can obtain MISR products from the Atmospheric Sciences Data Center (ASDC) at NASA Langley Research Center, where they are generated. The ASDC website [10] contains a great deal of information of interest to the user. For newcomers, there is a Project Guide that describes top-level product structure, and a set of Frequently Asked Questions. Details of the product formats can be found in the previously mentioned Data Products Specifications document [2]. The ASDC Web site also has a description of product versioning; a set of Product Quality Statements; and references to software tools for reading the products. Tools are addressed in Section VII. The Product Quality Statements are of paramount importance to the users as they list all the known problems and limitations of the respective products. They also include discussion of the remaining pieces of product content yet to be implemented.

III. OPERATIONAL ENVIRONMENT

For the Terra mission, ground data processing, including the generation of MISR data products, is accomplished within the EOS Data and Information System (EOSDIS), which is the end-to-end data system for all of the current EOS missions [11].

Those parts of EOSDIS relating to MISR are summarized in Fig. 3. This extensive system embraces mission operations, ground data processing, archiving, and data distribution. Spacecraft data received by ground stations is passed to the EOS Data and Operations System (EDOS) at NASA Goddard Space Flight Center (GSFC), where spacecraft telemetry is divided into its component packet streams. High-rate science packets are forwarded to the respective Distributed Active Archive Centers (DAACs), or equivalent processing locations. In MISR's case, the DAAC is part of the NASA Langley ASDC, where data are processed into standard products, archived, and distributed.

Users requiring MISR data are assisted by the User Services facility at the ASDC DAAC, accessed through their Web site at http://eosweb.larc.nasa.gov. Actual product orders are currently placed through the EOS Data Gateway (EDG), which is a web-based facility centered at GSFC, with distributed functionality at the respective DAACs.

MISR science data processing operates within a system software environment known as the EOS Core System (ECS), which uses primarily Silicon Graphics, Inc., (SGI) Origin mini-supercomputers running the IRIX operating system. The functionality of the ECS includes systemwide EOS application features embracing data ingest, data archiving, data staging, job control, science processing, product distribution, and user services.

Performance requirements of the system include that it keep up with the rate at which the instrument produces data, and that

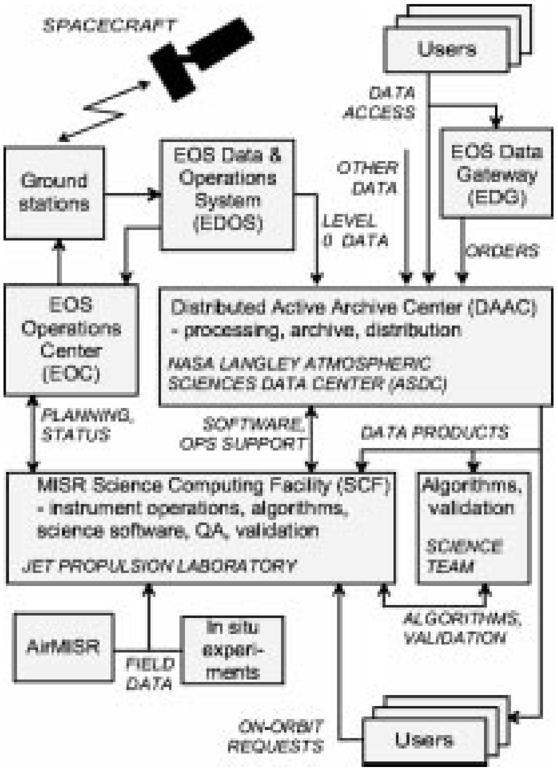

Fig. 3. MISR science data system components.

there be additional capacity for reprocessing of earlier data with revised algorithms. The first complete reprocessing of the MISR dataset is anticipated to begin in late 2002.

While routine processing activities are associated primarily with the DAAC and EOSDIS, there is also an important operational role played by the MISR Instrument Team. The Instrument Team, which is roughly analogous to the MISR Science Team plus MISR Project staff, is centered around the MISR Science Computing Facility (SCF) at the Jet Propulsion Laboratory (JPL). Science Team members are located not only at JPL, but also at other facilities, such as universities, both in the U.S. and Europe. The Instrument Team's data responsibilities include instrument calibration (primarily radiometric and geometric); data processing algorithms; data processing science software; data product validation; data processing operations support and quality assessment; and instrument-related science. The team is also responsible for an airborne MISR simulator called AirMISR [12], together with field instrumentation, which are used to support MISR vicarious calibration and validation campaigns. Data from AirMISR are processed by the MISR team into georegistered and rectified radiances, and are available to general users through the ASDC DAAC. All of these Instrument Team activities, together with observational and engineering support of MISR on-orbit activities, contribute directly to the operational success of the MISR experiment.

IV. SCIENCE SOFTWARE DEVELOPMENT PROCESS

At the present time, the MISR Instrument Team's largest single activity and largest deliverable is the software used at the ASDC DAAC for generating MISR standard products. All of the science production code was developed by MISR staff at the MISR SCF at JPL, based on algorithms devised by the MISR Science team. As with other Terra instruments, MISR development began in the early 1990s, initially with a very small development team, and reaching a peak of more than 25 software-related staff at launch.

The software development effort was divided broadly into a prototyping stage followed by a more formal software implementation phase, but with an ongoing iteration to accommodate the evolving algorithms, an evolution that continues. Prototyping of algorithms for the initial products continued through 1995. The initial functional versions of operational code appeared in 1996 for Level 1 and 1997 for Level 2 algorithms.

Prototyping was necessary in all major product areas because of the lack of heritage of the algorithms. Many of the processing concepts and algorithms were complex and some implementations were first-of-a-kind; for example, automated ground location and coregistration of multiangle images to the extent and accuracy required by MISR had never been done before.

Within JPL, the major prototyping was for the Level 1B2 and for the Level 2 Aerosol/Surface products. In these areas, the one team of implementers was responsible for both prototyping and implementation; that is, there were not separate groups dedicated to prototyping and implementation. The responsible scientists worked directly with this team in defining the algorithms and confirming from the prototypes that the proposed techniques would work.

The other major Level 2 product area, TOA/Cloud, was prototyped by the MISR Co-Investigator team at University of Arizona, incorporating codes supplied by other Co-Investigators at the University College London and the University of Illinois [13]. Additional algorithm development support came from co-investigators at Los Alamos National Laboratory (clear-sky albedo retrievals), Boston University (surface products), the Joint Research Center, Italy (bidirectional reflectance model used in surface retrievals), and the University of Miami (ocean products); in these cases, developers implemented new processing codes based on specifications and prototypes by the outside team members and their associates.

The formal development cycle used for the operational versions of the software follows a waterfall pattern adapted for incremental functionality occurring over a sequence of builds. Code development occurs in phases, each one corresponding to a complete system-level loop in Fig. 4, and resulting in a specific software build. The top-level prelaunch builds required by EOS were called Beta, Version 1, and Version 2, and by having three major deliveries prior to launch the operability of the software for the start of the mission was significantly enhanced. Note that "Beta" in this context is not the same as its usage in the product maturity definitions discussed above in Section II-C.

Besides the production software, there are significant software subsystems that reside within the MISR SCF but are not delivered externally although they form a critical link in the development and operational chain for the standard products. They are used to prepare ancillary tables and databases that are referenced during standard product generation. Specifically, these include

a) for radiometric conversion, the coefficient files;

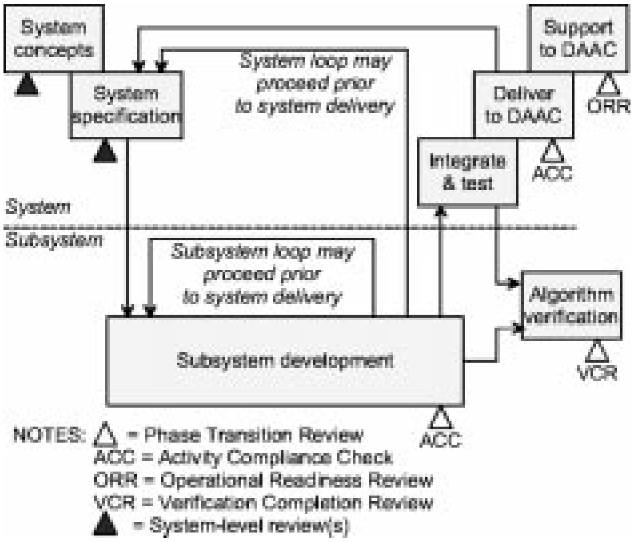

Fig. 4. Life cycle for software delivered to DAAC.

b) for georectification, the camera geometric model, projection parameters (for ground locating the MISR data), and reference orbit imagery (for image matching that provides accurate ground location);
c) for aerosol retrievals, the simulated aerosol radiance data, and aerosol climatology data;
d) for land surface retrievals, look-up tables used in the geophysical retrievals.

Other SCF software subsystems that support Science team operational activities include the In-Flight Radiometric Calibration and Characterization (IFRCC) subsystem [14]; the geometric calibration (Geocal) subsystem [15]; generation of threshold data for the cloud classifier algorithms; the validation subsystem; QA processing; and AirMISR data processing.

Because the SCF-based software subsystems have direct influence on the data products, they were all implemented using a rigorous development cycle that is a subset of the one used for the production software, and are subject to similar review, configuration management, and problem tracking practices.

V. SOFTWARE AND SYSTEM DESIGN

This section addresses the design of the MISR production software and MISR-specific drivers of the system in which it operates. It includes the key factors driving the software design and an overview of the software architecture, and gives greater detail in areas that were especially unique or challenging.

During the implementation years prior to launch, the MISR Project had the option of either using the ECS production system, or implementing its own production system in the form of an EOS Science Investigator-led Partner System (SIPS). MISR chose to use the ECS because of the significant investment already made in designing MISR processing within the ECS environment, and because that was regarded as the most cost effective route to take.

Before proceeding with a description of the software, here is a very brief introduction to science software architecture within the ECS context. The software executable entities are called Product Generation Executables (PGEs), each of which consists of a core service process and subsidiary child processes. A PGE has one or more compiled executables or scripts, and produces one or more standard or intermediate product files, or completes a processing step on a dataset. In most cases the PGE includes a script that links together the component executables to perform the desired processing. There is only one instantiation of a PGE for each granule that has been defined for the MISR instrument processing. A PGE cannot accept command-line arguments. No assumptions were made about the physical location of the files or the directory from which the PGE is being run. The files are accessed through logical file handles established in a PGE Control File (PCF), using the ECS toolkit access calls. Normal processing operations are performed automatically, and PGEs can be initiated manually. When all of the data needed for processing have been staged, and the computing resources needed for processing the PGE are available, the PGE is scheduled for execution.

A. Key Factors Driving the Design

In an extensive data processing system such as that used for MISR, the system design has many driving requirements. The ECS environment that is the heart of the DAAC production system was made for multiple instruments and missions. While such a system must accommodate all of the critical requirements dictated by individual instruments, it also must adopt standards and specifications to which the individual instrument software must conform. Here are the prime factors used in design of the MISR processing software, including factors arising from the MISR instrument characteristics and the requirements of the ECS.

1) EOSDIS performance degraded more quickly in response to increases in quantity of files staged than it did to the size of the files staged. For the same data volume, processing fewer large files proved more efficient than managing a large number of small files.
2) The amount of dynamic memory available for each processor would be limited to 512 MB, mostly due to cost constraints.
3) Level 1A, Level 1B1, and Level 1B2 science processing shared the following characteristics: a) very high I/O rates, b) MISR-only generated ancillary data, c) nearly identical locality of data access, and d) no data interdependence between cameras.
4) Level 1A CCD calibration, on-board calibrator, and motor data occur infrequently.
5) Level 2 TOA/Cloud and Level 2 Aerosol/Surface share the following characteristics: a) require data from all nine MISR cameras, and b) require ancillary data from non-MISR sources.
6) Level 2 TOA/Cloud has very large memory requirements.
7) HDF-EOS was the EOSDIS required data product format.

These drivers resulted in specific responses within the design of the production software, as follows.

1) The *fundamental* granule size adopted for MISR processing was the amount of data collected in one orbit, nominally one daylit swath for science data.

2) Every PGE was designed to require no more than 350 MB of dynamic memory.
3) A maximum size of approximately 1 GB was adopted for all files, whether input, ancillary, or output, in order to fit within the limitations of memory and the ECS toolkit.
4) The PGEs were designed to take advantage of opportunities for coarse-grained parallelism, to minimize disk accesses, enable scaling up or down, and to efficiently use system resources. A multiprocess PGE structure using dynamic memory-based interprocess communication services was implemented to reduce intra-PGE I/O. This resulted in intra-PGE chaining of Level 1A, and Level 1B processing. Level 1 processing was separated from Level 2 processing (not in the same chain). Level 1 processing used a serial approach within a PGE instance. Each camera is processed independently of every other camera. It takes nine instances of the Level 1 PGE to process all data from the nine MISR cameras.
5) Calibration PGEs were created to allow for separate scheduling based on their discrete acquisition schedule rather than the continuous acquisition of science data.
6) Each Level 2 PGE would require staging of data from each of the nine instances of Level 1 processing.
7) The above constraints caused the decomposition of Level 2 Top-of-Atmosphere/Cloud processing from one to three PGEs to remain under the 512 MB per processor allocation.
8) MISR requires that radiance data be colocated and georectified before geophysical parameter retrievals are performed. As described above, the HDF-EOS Grid data type, suitably modified, was selected for this purpose.

B. PGE Design

Based on the factors described above, there are nine PGEs used for Level 1 and Level 2 processing, as follows.

PGE1: Charge-Coupled Device (CCD) Science Data.
PGE2: Engineering Data.
PGE3: Motor Current Data (for the deployable calibration panels and the instrument cover).
PGE4: CCD Calibration Data.
PGE5: On-Board Calibrator (OBC) Data (for on-board photodiode measurements).
PGE6: Local Mode (high resolution) Geometric Processing.
PGE7: Geometric Parameters.
PGE8: Top-of-Atmosphere (TOA)/Cloud.
PGE9: Aerosol/Surface.

Level 3 processing uses an array of PGEs, known as PGE11-PGE16.

This breakdown of the PGEs, and the way they are chained together operationally are illustrated in Fig. 5. Not shown here are the various ancillary data flows used in the processing. The MISR Level 1 PGEs do not use any ancillary data other than that supplied by MISR, with the exception of navigation and attitude data, and the Detailed Activity Schedule (DAS), a file provided by the EOS Operations Center that contains the terminator crossing and corresponding camera-on times for the MISR instrument. The Level 2 science algorithms allow for the input of environmental data from meteorological sources and from MODIS, which would provide more accurate information than the climatology datasets assembled by the MISR Science Team.

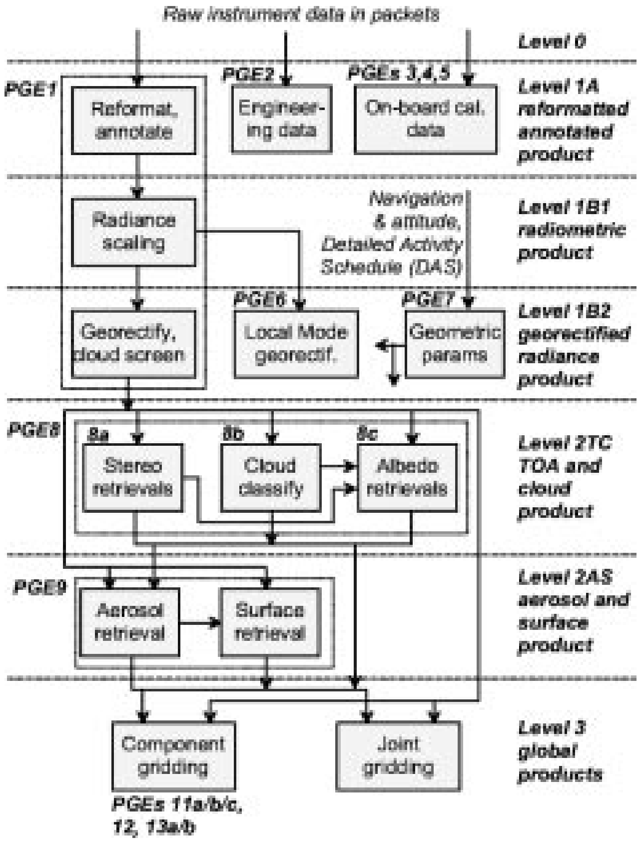

Fig. 5. PGE breakdown and chaining.

While the science data are processed in full swath granules, within the PGEs they are divided into segments to facilitate the transition between multiple concurrent processes. That is, each “data segment” is the minimum input data to a processing function, and must be understood in the context of MISR’s concurrent processing. To optimize production, the MISR processing design allows for the concurrent execution of PGE processes using the same input data. As the first process runs, a portion of the input data becomes available and may be handed off to the next process before the first process is entirely finished. Thus, the second process need not wait for availability of the entire granule dataset. Serial processing steps may thus occur between PGE processes in a pipeline fashion, i.e., process steps are chained. Segments are chosen to be at least as large as the minimum segment size for the along-track dimension, and equal to the swath width for the cross-track dimension. For Levels 1A and 1B1 processing, each segment is a 512-line portion of the swath. For Level 1B2 and Level 2, the segment consists of one or more blocks in the SOM projection. (Blocks are defined above in Section II and Fig. 2.)

The heavily computational nature of the production software was recognized early, and steps taken to mitigate it. Besides careful coding and the use of concurrent processing, algorithms were structured where feasible to use look-up tables prepared at the SCF rather than rigorous real-time analytical techniques. In

TABLE I
MISR LINES-OF-CODE COUNT

Category	*Operational code*		*Unit tests*	
	Executable	*Comment*	*Executable*	*Comment*
Product generation software	610K	434K	330K	206K
Specialized software at SCF	427K	291K	157K	73K

effect, this transferred heavy computation from the production system to the MISR SCF, where calculation of the look-up databases, while highly time consuming, does not need to be repeated for every granule of production. This technique was used in the Level 1B2 georectification and the Level 2 Aerosol/Surface and TOA/Cloud processing. Level 1B2 development required several workyears of effort to prepare a global Digital Elevation Model (DEM) based primarily on data from the National Imagery and Mapping Agency (NIMA); a Projection Parameter (PP) dataset used in translating navigation and pointing information to ground location; and Reference Orbit Imagery (ROI) that will allow improved ground location through image matching. The preparation of these datasets was computationally so intensive that care was needed to ensure the work needed to be done once only. One production run for Projection Parameters required approximately four months of continuous processing using 16 CPUs of a Silicon Graphics Origin-2000 computer.

The design of MISR processing required careful coordination with the development of the EOSDIS ECS production environment to account for unique characteristics of the MISR data stream. The most fundamental difference from other instruments is that MISR data are downlinked and front-end handled in the traditional fashion of time-aggregation, viz., processed in two-hour chunks, but must transition into an orbit-by-orbit handling because this is the way that the MISR ground processing is done. This unusual method of working is functioning very well.

The need for MISR software to conform with the ECS operational environment includes a prescribed data model, or metadata definition. Dedicated staff were necessary to complete this. The HDF-EOS metadata was designed to capture descriptive information for a granule, and be a common format used for all EOS instrument data. This generalized data model was insufficient to meet specific MISR needs, especially those regarding storage of detailed quality assessment parameters. MISR adapted by developing software to create separate quality assessment data files containing detailed measures of granule quality, there being one QA file per each product file. With these changes, MISR could record metadata at swath level, block level, line level, and pixel level as appropriate.

To complete this brief synopsis of the MISR software design, the size of the implemented software is given in Table I. This count of lines-of-code was made in April 2002. The total of over 1.5 million lines of executable code is divided between operational code and unit tests, and between the product generation code, and the specialized subsystems at the MISR SCF. The figures in the table include only code written by the MISR implementers, and exclude third party software used in the system.

VI. SYSTEM IMPLEMENTATION

The experience of implementing the MISR science data system and its science software, and of using the system to support an operational mission, spans a period of over ten years. This section presents the chronology of the implementation and operation, and examines the production issues and other challenges that arose.

A. Implementation and Operations Timeline

Implementation of the MISR science software by the MISR team began officially in 1991, with the inauguration of the MISR Project at JPL. System-level planning and algorithm prototyping characterized the period through approximately 1995, after which the formal software development was predominant.

The Beta version of prelaunch production software (not to be confused with the release of "Beta" data products discussed in Section II) to the DAAC in early 1996 was intended by EOS to exercise the system environment. It did this only to a limited extent because the MISR software at the time was largely stand-alone, i.e., it did not use the ECS system toolkit extensively, and did not use the HDF-EOS file format. However, it was a major milestone because it demonstrated that the MISR processing concepts were viable. Subsequent deliveries to the DAAC prior to launch included the Version 1 in mid 1997, Version 2 in early 1998, and Version 2.1 a few months prior to launch. Each of these included gradually greater levels of functionality and operability.

At the time of launch, MISR science processing capability integrated into the ASDC DAAC consisted of only the Level 1 functional string. For the first few months of operations, all MISR Level 0 data was transmitted immediately to the SCF so that data quality could be examined with quick turnaround, with special SCF versions of the Level 1 PGEs set up to run automatically upon receipt of data. On a typical day, between 30 and 50 GB of data arrived at JPL via a dedicated ATM network link. This processing began about three months before the MISR cover was opened on February 24, 2000.

During the weeks prior to opening of the cover, it was necessary for MISR's photogrammetry staff to prepare special navigation files because Terra had not yet reached its final orbit, and the standard day-night cycling of MISR had not started. During this time, with the cover closed, a global map of proton radiation was produced from this processing [16].

After the MISR cover was opened, normal processing operations began immediately, and the first public release of data files consisted of Level 1 products, on 28 June 2000. (A small number of selected images was released earlier at a press conference in mid-April 2000.) The availability of real instrument data heralded a major development push to complete the initial Level 2 products. For example, during the latter half of 2000, much work on the Aerosol/Surface algorithm was necessary to improve spatial coverage, and the cloud stereo algorithm required major refinements such as improving the accuracy of wind retrievals. Both of these products were demonstrated to the MISR Science Team at their meeting in December 2000, which permitted their public release in March 2001. Improvement has been and is ongoing. The Provisional status of Level 2 products is planned to

TABLE II
EVOLUTION OF MISR PROCESSING ESTIMATES

Processing Level	*Processing Capacity (MFLOPS)*		*Product Volume for 1x Throughput (GB/day)*	
	1995	*2000*	*1995*	*2000*
1	1364	2522	207	157
2TC	784	2896	6	2
2AS	1039	4656	8	8
3	12	300	0.05	2
Total	3199	10374	221	169

coincide with the start of the first complete reprocessing of the entire MISR dataset from the start of mission onward.

B. Challenges and Performance Issues

One of the larger lessons emerging from MISR's experience is that the development of the MISR/EOS kind of science software is unavoidably difficult, encompassing far more than a straightforward "bullet-proofing" of science prototype code and the use of a standard system environment. Instead, it is a major design and implementation challenge. Some of the issues include: error handling; unit testing; working through the data; using the system environment dependencies; automatic mass production; continually changing requirements; broad dependencies; pressure to write code before requirements are complete; lack of prototypes for some of the algorithms.

These challenges and related operational and performance issues are best illustrated by looking at the specific major issues characterizing the MISR experience.

1) Algorithm Efficiency and Evolution: The initial processing hardware capacity at the DAAC was based on estimates of algorithm requirements made by the MISR team in 1995, when many of the more demanding algorithms were still at the prototype stage, or only partly implemented. Table II shows those estimates (for a one-time processing of the dataset) and compares them with figures derived in mid-2000, after the real mission requirements became clear. The 2000 estimates remain essentially valid, and are currently used as a constraint within which the implementation must fit.

Having no heritage, there were no operational codes upon which to base reliable system capacity estimates until early versions of production software became available in mid-1999. Early estimates of processing needs, such as the 1995 figures in Table II, were based mainly on hand calculations, and included an application environment overhead of 100% to allow for metadata generation, product structure constructions related to the Hierarchical Data Formal (HDF) and HDF-EOS formats used for the EOS products, input staging, job control, and any other production environment requirements. Experience with the system since Terra launch has shown that because of system downtime for maintenance, Level 0 data delays, inefficiencies in data staging, limitations on maximal CPU utilization, and general system instability, a more realistic overhead allowance would be 175% to 200% with respect to science software execution times.

In some instances, initial implementation of MISR algorithms performed in ways that were very different from what was predicted earlier. The most notable example of this was that the initial measurements of Aerosol/Surface processing speed using actual code in mid-1999 suggested a deviation by approximately two orders of magnitude in computation time with respect to the estimates that were used to specify the installed processing hardware. This being an obvious threat to the success of MISR, it was decided to completely rework the software. Over 50 work-months were expended in this effort, and the results were dramatic, reducing the 100-fold increase to within the overall envelope defined in Table II.

Although performance improvements to the MISR software will continue to be made, it is believed that all of the major efficiency areas have already been addressed. Thus, the future evolution of the software will be driven primarily by evolution of the algorithms.

2) Simulated Data: A necessary part of premission development was the preparation of simulated datasets for testing both the algorithms and the software. Although the need for test data was recognized early, there were no prelaunch multiangle test datasets at comparable spatial resolution that could be used to provide input for algorithm testing or for performance measurement studies. Thus, the construction of simulated MISR data was a complex exercise. Approximately two workyears of effort was put into this, and the result was the equivalent of a single day of processing. This limited dataset restricted the range of operational stress testing that could be accomplished before launch although it is doubtful that it resulted in any serious implications for the postlaunch release of products.

Although prelaunch data from the MISR instrument was in the correct Level 0 packet format, the content for software testing purposes was minimal. Therefore, some level of realistic image content was needed to confirm the basic software and algorithm functions prior to receiving MISR data from orbit.

Software was developed to produce this image data using Landsat scenes as a base and creating MISR-like data in the appropriate form. This software read Landsat nadir-view radiance images and produced the nine angle MISR-like radiance images using rendering and projection techniques. MISR swath structures were then generated in two ways. Initially this was accomplished by mirroring the Landsat scene-based data in both cross-track and down-track foldings to achieve the desired spatial extent. To improve the overall content of the test data an extended mosaic of Landsat scenes was later used as input to the rendering step. This radiance form of test data was used as the input stream to the geometric processing steps.

Additional software was developed to generate Level 0 packet files from these radiance files to support the testing of the complete Level 1 processing chain.

3) Impact of Instrument Issues: Prior to the public release of Level 1 products in June 2000, many problems were resolved. Some of these arose from non-MISR sources, such as "bit flips" in the packet headers. The primary MISR-related issue to facilitate the public release was the delivery of software patches to mask out regions of data acquired while the MISR instrument was experiencing an out-of-sync condition [16]. Prior to this patch, the software often halted because of discontinuities in the data when out-of-sync packets were encountered.

The out-of-sync issue has been one of the main Level 1 processing issues, and continues to receive attention. Without corrections, the geolocation of the product could not be correctly determined when lines of data were no longer in their correct location subsequent to an out-of-sync event. Gradually, the Level 1 software has been made increasingly resilient with respect to out-of-sync packets. In May 2001, a major patch increased robustness to the point where Level 1 processing proceeded uninterrupted in more that 95% of cases. A concerted effort is planned in mid-2002 to provide a definitive solution to this problem and achieve a 98% or better success rate. This is necessary to attain a satisfactory global coverage with Level 2 products because Level 2 processing as currently structured requires input data from all nine cameras. (A 100% success rate may not be practicable because of the many variants of the problem.)

4) DAAC Issues: The MISR production system at the ASDC DAAC was based on a new design that needed many adjustments to achieve the necessary performance, and these adjustments are continuing. The early months of the mission revealed serious operational issues, many of which have been successfully addressed or are being addressed, e.g., the need for improved data staging between the archive and the processors; inadequate production planning capacity; rearrangement of archive files to optimize access times; susceptibility to operator errors; gaps in attitude and navigation data; internal system saturation; poor interprocessor communication; and an inefficient operating system.

By late 2001, which was two years after launch, an average throughput between 100 and 150% of daily instrument output was being realized, although not consistently. Since then, performance and reliability have been significantly enhanced by augmenting the computer processing hardware and by major revisions to the ECS software. At the time of writing, April 2002, the augmented system was in a test phase, and appears capable of supporting the reprocessing planned for later in the year, when there must be at least as much capacity available for that purpose as for the processing of new instrument data. At that point, the system will be based on SGI Origin 3800 machines with the equivalent of 96 CPUs running at 400 MHz.

An additional round of capacity augmentation is expected later in 2002, which should at least double the reprocessing capability, and this should for the first time realize the kind of system capability originally intended by EOS for full mission support.

VII. SOFTWARE TOOLS

The implementation of support tools was an important part of the MISR software development task. In general these tools filled the typical roles in support of software testing, data generation, and data analysis and visualization. Several specialized tools were built to reflect the MISR needs. Some were developed as extensions to commercial off-the-shelf (COTS) products, or used COTS products as a base. Others were written in their entirety.

The ASDC DAAC Web site [10] lists MISR data access tools available from various sources. The following paragraphs describe tools developed by the MISR Instrument Team that are available to the science user community through a no-cost software license agreement with JPL/California Institute of Technology (http://www-misr.jpl.nasa.gov/software).

A. GIS Extension

Toolkit extensions to the ERDAS Imagine GIS software package were implemented for viewing and analyzing MISR data. These were used by the MISR team as a means for specific manipulations of MISR data content and structure necessary to confirm the rectification of the images from the nine cameras and four bands per camera. They also provided the ability to verify the geolocation of the images. This step was aided by inserting measuring crosses into the data so that these tools could be used to examine the relative locations of these crosses among the camera views.

B. hdfscan

Two tools were built to specifically view MISR data products. The first is called *hdfscan*. This software consists of two components, a core program used to read and process MISR HDF files, and a graphical user interface (GUI) built as an interface and display wrapper around the core program. *hdfscan* can be initiated in two modes, command line driven using the core program only or using the GUI with X-windows. The core program is written in Fortran 90 and the GUI consists of routines written with Tcl/Tk, C and C++. This program permits examination of both the metadata and product parameters in MISR data products. Many data selection and display options are available which allow viewing parameters from any HDF structure. Data can be displayed as image, gray scale or color, tabular, or text where appropriate. Also both raw and interpreted representations of the data are available, e.g., conversion of numeric flag values to text descriptors.

C. misr_view

The other data analysis tool is called *misr_view*. This program was developed to provide advanced viewing and analysis capabilities applicable to those MISR products written in the HDF-EOS grid format, including the georectified Level 1 radiance data, the Level 2 retrieved geophysical parameter data, and AirMISR georectified radiance data. The software for this tool is based on IDL, a COTS product from Research Systems, Inc.

misr_view is a feature-rich analysis tool allowing comprehensive data viewing and investigation. It is controlled through a GUI that permits the selection of data products to be examined and analysis options. Some of these options include the ability to extract and stitch together a range of blocks for display and to load parameters having different spatial and bit-depth resolutions into the display planes of a viewing window. Several utilities exist providing data transforms to perform tasks such as data scaling and unpacking, histogram viewing, and stretch manipulation. It is possible to link multiple viewing windows via geolocation information and to use alternative map projections. Other capabilities include vector overlays, tools for constructing anaglyphs, automatic scrolling through a large range of blocks, and emulation of 24-bit color on eight-bit display hardware.

VIII. CONCLUSION

Putting a complex data production system in place at the same time as research and development on the algorithms proceeds is a daunting task. Many factors must be managed for this to work, and in the MISR case the outcome was successful.

The first generation of MISR Level 1 data products became available to the science community in July 2000, and Level 2 products became available in March 2001. Validated Level 1 products are now available, and Provisional Level 2 products are expected later in 2002. This is a major accomplishment for such a new instrument concept launched little more than two years ago.

Although the MISR data products will be subject to ongoing revision and improvement during the Terra mission, the foundations of the design, implementation, and operation are well established, and highly successful. Many innovations and specialized approaches were necessary because of the newness of the MISR multiangle observing concept. These are groundbreaking achievements that will no doubt be the foundation of data processing for future on-orbit, multiangle instruments.

ACKNOWLEDGMENT

The authors gratefully acknowledge the dedication and inspiration of the MISR science data system team; the untiring efforts of the ASDC DAAC staff; the developers of EOSDIS; and all others contributing to the success of the MISR science data system. For further information, see the MISR Web site at http://www-misr.jpl.nasa.gov

REFERENCES

[1] D. J. 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*, vol. 36, pp. 1072–1087, July 1998.

[2] S. A. Lewicki, B. G. Chafin, K. A. Crean, S. E. Gluck, K. C. Miller, and S. R. Paradise. (1999) MISR Data Products Specifications. [Online]. Available: JPL Tech. Doc. D-13963, Rev. C. http://eosweb.larc.nasa.gov/PRODOCS/misr.

[3] MISR Science Team. Algorithm Theoretical Basis Documents. [Online]. Available: http://eospso.gsfc.nasa.gov/atbd/misrtables.html.

[4] EOS Project Science Office, "EOS Reference Handbook," NASA Goddard Space Flight Center, Code 900, 1999.

[5] J. P. Snyder, "Map Projections: A Working Manual," U.S. Government Printing Office, Washington, DC, USGS Professional Paper 1395, 1987.

[6] A. Braverman and L. DiGirolamo, "MISR global data products: A new approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 1626–1636, July 2002.

[7] National Center for Supercomputing Applications at the University of Illinois. HDF Homepage. [Online]. Available: http://hdf.ncsa.uiuc.edu/.

[8] NASA. (2002). HDF-EOS Tools and Information Center Homepage. [Online]. Available: http://hdfeos.gsfc.nasa.gov.

[9] NASA. (2002). EOS Terra Data Products Homepage. [Online]. Available: http://eosdatainfo.gsfc.nasa.gov/eosdata/terra/terra_dataprod.html.

[10] NASA. (2002). NASA Langley Research Center's Atmospheric Sciences Data Center Homepage. [Online]. Available: http://eosweb.larc.nasa.gov.

[11] NASA. (2002). Earth Observing Science Data and Information System (EOSDIS) Homepage. [Online]. Available: http://eosdismain.gsfc.nasa.gov/eosinfo/EOSDIS_Site/index.html.

[12] D. J. Diner, L. M. Barge, C. J. Bruegge, T. G. Chrien, J. E. Conel, M. L. Eastwood, J. D. Garcia, M. A. Hernandez, C. G. Kurzweil, W. C. Ledeboer, N. D. Pignatano, C. M. Sarture, and B. G. Smith, "The Airborne Multi-angle Imaging SpectroRadiometer (AirMISR): Instrument description and first results," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1339–1349, July 1998.

[13] C. Moroney, R. Davies, and J.-P. Muller, "Operational retrieval of cloud-top heights using MISR data," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 1532–1540, July 2002.

[14] C. J. Bruegge, N. L. Chrien, R. A. Ando, D. J. Diner, M. C. Helmlinger, W. A. Abdou, and K. Thome, "Early validation of Multi-angle Imaging SpectroRadiometer (MISR) radiometric scale," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 1477–1492, July 2002.

[15] V. M. Jovanovic, M. M. Smyth, J. Zong, R. Ando, and G. W. Bothwell, "MISR photogrammetric data reduction for geophysical retrievals," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1290–1301, July 1998.

[16] D. J. Diner, J. 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*, vol. 40, pp. 1449–1466, July 2002.

Graham W. Bothwell received the B.Eng. degree in electronics from the University of Queensland, Brisbane, Australia, in 1967.

He was with NASA's Deep Space Station, Canberra, Australia, from 1967 to 1968. From 1969 to 1985, he was with the Anglo-Australian Observatory, Sydney, Australia, where he was involved in computing and electronics development in support of construction and subsequent operation of the 4-m Anglo-Australian Telescope. He has been with the Jet Propulsion Laboratory (JPL), Pasadena, since 1985, initially as supervisor of a group developing image processing software for JPL's planetary program. Since 1988, he has worked with various projects associated with NASA's Earth Observing System, and led a group involved in instrument data processing, ground system development, mission operations, and airborne remote sensing. He is currently Manager of the MISR Project.

Earl G. Hansen received B.S. degrees in mathematics and physics from the University of Washington, Seattle, in 1989

He has been involved with data system development and operations since 1972 including roles as software engineer, systems programmer, and systems and operations manager with commercial companies and the Jet Propulsion Laboratory (JPL), Pasadena, CA. He has been at JPL since 1989, initially as Systems Engineer for the science data system of the AVIRIS project. Since 1992, he has worked as a senior systems engineer and task manager for the science data system of the MISR Project.

Robert E. Vargo received the B.A. degree in biology from the University of California, Los Angeles, in 1975, the B.S. degree in computer science from the California State University at Los Angeles in 1984, and the M.S. degree in computer science from the University of Southern California, Los Angeles, in 1987.

He has been with the Jet Propulsion Laboratory (JPL), Pasadena, CA, since 1984, initially developing application subsystems for a large military intelligence data system. He joined the MISR Project as a system engineer in 1994, later becoming the technical supervisor of a group developing the science data systems for MISR and the Tropospheric Emission Spectrometer. He currently is Deputy Manager of the Space Science Data Systems Section at JPL.

Kyle C. Miller received the B.S. degree in physics from Principia College, Elsah, IL, and the M.S. degree in computer science from the University of Southern California, Los Angeles.

He is a systems engineer on the MISR Project at the Jet Propulsion Laboratory (JPL), Pasadena, CA. He has been involved with MISR ground software development since 1995. Prior to that he worked on large-scale simulation software at JPL.