MODIS SCIENCE DATA SUPPORT TEAM PRESENTATION

June 28, 1991

AGENDA

1. Action Items
2. MODIS Airborne Simulator
3. CASE Tools
4. Product Granule Sizes
5. Scenarios for Level-1 Processing

ACTION ITEMS:

05/03/91 [Lloyd Carpenter and Team]: Prepare a Level-1 processing assumptions, questions and issues list, to be distributed to the Science Team Members and the MCST for comment. (The list, the executive summary, information on the EOS Platform Ancillary Data, and a cover letter are ready for signature and distribution.) STATUS: Open. Due date 06/07/91.

05/31/91 [Liam Gumley]: Establish a connection with the proper person at Ames Research Center for communication on MAS formats, an interface control document, agreements, etc. STATUS: Open. Due date 07/19/91

06/07/91 [Liam Gumley]: Speak to Alan Strahler, when he returns, regarding his MAS requirements. STATUS: Open. Due date 07/05/91

06/21/91 [Liam Gumley]: Obtain a copy of all available MAS Level-1B processing software and any existing documentation from the University of Wisconsin at Madison for porting to a system at GSFC. STATUS: Open. Due date 07/19/91

06/21/91 [Liam Gumley]: Generate a complete milestone schedule for conversion, installation and testing of all modules of the MAS Level-1B processing software at GSFC. Draw up an agreement between the SDST and Mike King of what will be done. STATUS: Open. Due date 07/19/91

05/31/91 [Al McKay and Phil Ardanuy]: Examine the effects of MODIS data product granule size on Level-1 processing, reprocessing, archival, distribution, etc. (A preliminary report is included in the handout.) STATUS: Open. Due Date 06/21/91

ACTION ITEMS FROM SDST MEETING 06/21/91 [Liam Gumley]

(1) Visit Wisconsin to obtain all software and documentation relevant to MAS Level-1B processing.

This will take place from July 10 to July 12. It is intended to obtain software, documentation, and MAMS test and verification data.

(2) Generate a complete milestone chart for development of MAS Level-1B processing software.

This will be done after consultation with Wisconsin.

(3) Establish a connection with Ames Research Center.

This will be done after consultation with Wisconsin.

(4) Talk to Alan Strahler (MODIS Land Group) about his requirements for MAS Level-1B processing.

This will be done when he returns to Boston in the first week of July.

QUESTIONS RAISED AT SDST MEETING ON 06/21/91

(1) Investigate MAS Level-1B data subsampling requirements.

I talked to Mike King about this and his feeling was that all MAS scan lines and pixels should be supplied to users. They can then decide on appropriate subsampling. I talked to Chris Moeller at Wisconsin and he obtains every 3rd MAMS scan line and every pixel by request from Ames. Ames will supply every scan line and pixel if requested. Chris Moeller also said that the size of the current MAMS Level-1A data record for one scan line is 10320 bytes. This implies that the data rate for MAS will be
10320 bytes/scan x 6.25 scans/second
= 64500 bytes/second
= 0.516 megabits/second
= 232.2 megabytes/hour.
A 2400 foot magnetic tape with 6250 bytes per inch will hold up to around 180 megabytes. An Exabyte tape holds over 2 gigabytes. It will therefore be useful to see if Ames can supply the Level-1A data on Exabyte tape.

(2) Confirm MAS Level-1B image review ("Browse") strategy.

I talked to Mike King and he confirmed that the imaging system should be a simple facility where users can look at part or all of the image data from a flight. The data could be subsampled by scan and by pixel, and several or all spectral bands might be required. The data is not required to be calibrated or navigated. Existing imaging software will be used. Whether the imaging system uses pre-made data sets or a Level-1B archive directly is yet to be determined.

The ability to make hard copies of image data is desired by Mike King and may be desired by other users. Facilities for hard copy generation will need to be investigated.

It may be desirable to review the use of the word "Browse" in relation to this proposed MAS image viewing system, because of the connotations associated with browse systems. It may be better to term the system "Quicklook", as long as users understand that it will be quick once the Level-1A data gets to GSFC, not while the sensor is in the field. This needs to be decided before a formal agreement is drawn up specifying the work to be done in developing the MAS Level-1B processing system.

OTHER ISSUES

MAS Level-1B data output format

The final format for the MAS Level-1B data is yet to be determined, however issues are emerging which will guide the final design. These issues are briefly summarized below.

(1) Original Level-1A data access. It is likely that some users of Level-1B data will want to look at original Level-1A data such as blackbody thermistor counts, aircraft pitch, yaw, roll etc. Therefore it is advisable to include all of this original data in the Level-1B dataset.

(2) Structure of output data record. The Level-1B data record is the item of data that makes up one "block" on an output tape. This would be the smallest item of data that a user would read at one time from a tape. An obvious choice for the size of this data record would be the complete data for one MAS scan line. This would include

- original engineering data for all channels,
- calibrated radiances/brightness temperatures for all channels,
- original aircraft INS data,
- slopes and intercepts, or other calibration parameters,
- geolocation data.

(3) Portability and simplicity. The users of Level-1B data will have many different computer systems with varying amounts of processing power, graphics capability, and disk space. These may range from IBM or DEC mainframes, to Unix workstations, to IBM-PC or Macintosh systems. Therefore the Level-1B data should be both portable, and in a simple, easily decoded format. This requires the use of integer data types, with proper care taken concerning byte-ordering. It also requires that the Level-1B data be in a format that can be handled both on small and large systems.

(4) Geolocation data. At present it is planned to include the latitude and longitude, sensor zenith and azimuth angles, and

solar zenith and azimuth angles for the MAS pixels. These would each be represented by a 32 bit integer. If these parameters were included for every pixel on a scan line the amount of data would be

6 parameters/pixel x 4 bytes/parameter x 716 pixels
= 17184 bytes.

By comparison, the amount of image data for one scan line (12 channels, 16 bit integers) would be

2 bytes/pixel x 12 channels x 716 pixels
= 17184 bytes.

Thus the geolocation data would almost double the size of the Level-1B dataset compared to the Level-1A dataset. This would also cause the output Level-1B record to be greater than 32768 bytes, and this may be a physical limitation with some magnetic tape systems (e.g. DEC VAX). Therefore it appears that a strategy for providing interpolated geolocation data is desirable. A simple cubic spline algorithm is recommended for this task.

CASE Presentation

27 June 1991

Notes on the current state-of-the-art in CASE tools and implementations as presented at the Hewlett-Packard (HP) CASE Technology symposium.

The first Presentation was given on HP's experience in object oriented design of HP's internal projects. The summary of the presentation results indicates that a first project needs 6-9 months to learn object oriented philosophy. The second project can be performed in about the same time as the first but with much fewer errors and will result in a much easier to maintain product. The remaining projects are faster to implement in addition to the maintenance benefits. Object orientation works best on larger projects. Because a philosophical change must take place to realize the benefits of object oriented design, not all projects will be successful.

The second presentation discussed data repositories as used for data management and interfacing. Data is used here to mean source code, structured diagrams, data dictionaries, etc. The Portable Common Tool Environment (PCTE) under development in the ISO and CASE supplier arenas was discussed as a means to achieve data repository transparency.

Software metrics was the third topic. This topic describes tools that can be used to visually illustrate the 'cleanliness' of code, test validation paths, reverse engineering, and design complexity indices.

'Hands on demonstrations' from several vendors was performed in the afternoon. The vendors included IDE 'Software Through Pictures' CASE tools, Cadre's 'Teamwork' CASE tools, McCabe & Associates performance metrics tools, CCC configuration management, and others.

CASE Recommendations

For a front end development tool, either 'Software Through Pictures' or 'Team Work' are the industry leaders in the number of platforms and languages supported. Teamwork has a version for PC's while 'Software Through Pictures' does not. Both of these CASE tools are full implementations and are encapsulated under HP's 'Soft Bench' facility. 'Soft Bench' gives a multi-platform umbrella product for CASE front ends, configuration management, and on through code development tools. This is supported on HP, Apollo, Sun, IBM, DEC Ultrix machines currently, with a CDC license for SGI (to be confirmed). This looks like the obvious choice for code development. It supports C and FORTRAN on all machines and C++ and ADA on HP machines currently.

Comments Received from Dr. Ed. Chang:
MODIS Level-0 Data Completeness

If we can allow a few hours offset between the latest observations included in a batch delivery of Level-0 data and the clock time when the delivery is made, Ed believes that incomplete data will not be an important issue. In his opinion, the issue should not be made a driver in the granule size discussion.

Product Granule Sizes:
A Continuation of Earlier Discussions

Identity of MODIS Level-1 Data Users. The MODIS Level-1A products contain the basic digital count data generated by the instruments and serious use of the data is expected from only a few investigators. Level-1A data is of interest to those monitoring instrument behavior; Phil Slater's land calibration team at the University of Arizona and the MCST at Goddard Space Flight Center will occasionally access MODIS Level-1A data. Also, instrument calibration coefficients have historically been adjusted within the ocean science community to obtain agreement between data products and in situ observations, and the University of Miami is an expected user of the MODIS Level-1A product.

Level-1B data will be of wider interest. MODIS Level-1B and Level-2 data may be used for:

- Routine production of (other) Level-2 and Level-3 products
- Production of experimental and special Level-3 products
- Comparison of MODIS results and in-situ measurements
- Comparison of MODIS and "other instrument" results
- Instrument calibration monitoring
- Ancillary input to interdisciplinary models of Earth phenomena
- Ancillary input for "other instrument" products

- Field experiment support (quick look processing)

Facilities Available to Data Users. A review of the above list supports the idea that most users of Level-1A, Level-1B and Level-2 data products will be scientists and engineers with access to relatively sophisticated computer facilities. Exceptions may possibly occur for the comparison of MODIS results and in-situ measurements and for field experiment support, where users may be temporarily using portable data systems and may not have full access to their home compute facilities. However, relatively sophisticated facilities are expected at the home facilities. Since the MODIS data granule size has been related to the on-line disk storage available to an investigator (approximately 100 MB granule size for 5-10 GB of on-line storage), an estimate of disk storage capacity is needed to support the scaling of the MODIS granule.

Given the relative sophistication of the expected user community, and making allowance for price/performance increases in storage media during the time that will elapse before access to Level-1 and Level-2 data is required by future investigators, it is expected that the future on-line storage capability of a representative investigator may perhaps just equal the present on-line storage capability used to produce ocean products at the University of Miami (5-10 GB). This estimate is generally consistent with the few specific references to hard disk requirements included in the MODIS Science Team Member proposals (requests for 2 GB, 5 GB, and 10 GB were located).

Processing Granule Size. Assuming that a representative Level-1 and Level-2 investigator will have 5-10 GB of on-line storage available, and drawing corollaries on the University of Miami processing experience, the appropriate MODIS Level-1 and Level-2 processing granule would be about 100 MB.

The table below lists the MODIS-T (Level-1A) data granule volumes that correspond to several granule selection schemes. It is seen that for MODIS-T, the desirable data granule volume corresponds roughly to the "scene" used in the table.

MODIS-T GRANULE VOLUMES BY VARIOUS SELECTION SCHEMES

Granularity	Level-1A Granule Volume (MBytes)
1 Granule Per Day	17,890
1 Granule Per Orbit	1,229
1 Granule Per Scene	113 (Daytime)
1 Granule Per Scan	1.7 (Daytime)
1 Granule Per Packet	0.001 (Daytime)

As used here, a scene corresponds (approximately) to 1/20 of an orbit (defined by elapsed time) or about 5 minutes of data. At the speed of the platform, a scene corresponds to 2,000 km in the along track direction, and since the cross-track dimension of a MODIS-T scan (at normal incidence or tilted) is also roughly 2,000 km, a scene is roughly square. A scene contains 65 scan cubes, i.e. it corresponds to 65 scans of the MODIS-T instrument. All MODIS-T data obtained on the dark side of the Earth can be put in a single nighttime scene smaller than the daytime scene. A scene boundary could be defined to occur at the night-to-day transition when the instrument is switched from nighttime to daytime mode. Likewise, the day-to-night transition could be used as the boundary between the final daytime scene and the nighttime scene in each orbital pass.

Additional Questions to Be Addressed.

What might a 100 MB granule look like for MODIS-N? [Since the MODIS-N instrument has a much higher data rate than MODIS-T (approximate factor of three), the scene defined above for MODIS-T will not be appropriate for MODIS-N. It may be possible to define a square MODIS-N granule based on a quadrature of the MODIS-T scene into four roughly equivalent parts.]

Is it possible to define a granule based on the geography of the observations and not upon orbit segment or elapsed time? [Probable answer, No!]

Is the processing granule as defined above also suitable for use as an archive granule? Specifically, what would be the effect on retrieval efficiency of a granule as defined above?

Is the projected metadata volume for the above granule definition reasonable and workable? I.e. is the storage volume for metadata reasonable? Can the metadata reasonably be searched as required?

What benefits, if any, accrue from the use of the same granule definition for Level-1A and -1B products?

Will the Level-1B granule be the "default" Level-2 granule? What aspects of Level-2 processing are important to consider?

What subsetting of the above granules would be required before Level-1 and Level-2 products are distributed?

How does the granule definition given above affect data retrieval efficiency?

Can missing data items reasonably be filled in with the above definition of a granule?

How can traceability be maintained across the granule size boundaries the occur between the various processing levels?

Is the granule definition given above suited to the available physical media for use in the archives?

SCENARIOS FOR THE MODIS LEVEL-1A AND LEVEL-1B PROCESSING

27 June 1991

1. Normal Processing:

Normal MODIS Level-1A and -1B processing will be done as required input data elements are received and as the necessary EOSDIS compute resources become available for MODIS use. In normal operations, processing of a MODIS data granule will begin when a complete set of instrument data packets for the granule has been received and required platform ancillary data for the granule is available.

Level-0 MODIS data (instrument data packets) will be received and stored at the local DADS until Level-1 processing begins. When MODIS processing begins, the EOSDIS scheduler will prefetch the required items from the heirarchical data storage system and initiate MODIS processing as processing hardware becomes available.

The CDOS will usually deliver Level-0 data in blocks of one, or at most, two orbits of data. During CDOS processing, data received at the ground station has been error corrected, if possible, has been error flagged, if correction is not possible, and has been bit order reversed if time reversal of received data is required. CDOS processing has also accounted for all received packets, identified any missing data items, and eliminated any duplicate data packets that may have been created during successive playbacks of the on-board tape recorders.

The MODIS Level-1A product is primarily intended to provide a permanent record of MODIS instrument data; it can be reversed to recover Level-0 data, if required. The Level-1A product receives minimal processing. During Level-1A processing, spacecraft ancillary data may be appended to the instrument data to provide a single, integrated source of all the data required to complete Level-1 processing, successive instrument packets are concatenated to produce the Level-1A data granule, and metadata is generated to facilitate data storage and user retrieval of data. By present plans, data will not be unpacked (byte-aligned) during Level-1A processing.

The primary functions performed during Level-1B processing include the Earth-location of MODIS pixels, the radiometric conversion of sensor outputs to obtain physical radiances at the sensor (sensor calibration), and data-handling and accounting functions required to control data flows and monitor data quality. Level-1B processing will also include quality checks to assure proper MODIS instrument operation. Results of the Level-1B instrument checks will not be routinely provided to instrument controllers or the MCST. Instrument controllers and MCST personnel may access the results obtained from Level-1B instrument checks on a demand basis.

2. Quick-Look Processing:

In most cases, the need for quick-look processing can be anticipated. For example, in planned field experiments, an early look at the MODIS data may be a key element in the decision making process directing the ground portion of the experiment. The investigators will coordinate their quick-look processing requirements with the MODIS Team Leader who will provide final approval and coordinate the effort with CDOS and other elements of the EOSDIS.

A MODIS quick-look processing request will be sent to the EOSDIS specifying which data are to be processed. The EOSDIS will adjust priorities and expedite the flow of the data through the CDOS. The SCA will initiate the MODIS Level-1A quick-look processing soon after the data are available.

Quick-look processing may require time-ordering, redundancy elimination, and quality control measures not normally required for standard MODIS processing. However, quick-look processing will be executed using the same version of software as normal processing, and the Level-1A software will be designed to perform these functions.

Platform ancillary data may not be available in time for quick-look processing. In this case an alternate source, such as a predicted ephemeris, will be specified and made available by the EOSDIS for the required platform position and attitude data.

The Level-1A process will notify the scheduler when quick-look processing is completed. The scheduler will then initiate the MODIS Level-1B quick-look processing, which will be executed using the same version of software as normal Level-1B processing.

3. Metadata Generation:

MODIS metadata consists of information describing the MODIS data which is obtained or derived from the data sets, and which provides an understanding of the content or utility of the data set. Metadata may be used to select and evaluate data for a particular scientific investigation.

Beginning at Level-0, each successive processing level will generate and append metadata as part of the data product. The metadata associated with the input product are updated to reflect further derived information. Previous metadata items are retained to allow backward tracking of information to the original source. This information can be used for debugging and quality assurance determination. For example, the CDOS Reed-Solomon error statistics can be maintained with the mapped Level-3 product as an indication of the quality of the original data that went into the product. Metadata derived in the beginning of the processing chain will provide information which is useful for the generation and assessment of products later in that chain.

Care must be taken in the interpretation of metadata in some cases. For example, error statistics for a granule of Level-1A data could be misleading when each Level-1A granule is subdivided into several granules at Level-1B. Some of the Level-1A statistical information may not be correct when applied to the subset of data which went into a particular Level-1B granule, etc. The reverse situation occurs, for example, when many Level-2 granules are used as input to a Level-3 process which generates an average value and a secular rate of change for some parameter. However, with a complete trail of metadata information, the user should be able to trace the heritage of his data and properly interpret his results.

4. Browse Data Generation:

Since the MODIS science team members have not indicated a requirement for Level-1A browse, no MODIS Level-1A browse products have been planned.

The MODIS Level-1B browse process will be designed for easy adaptation to future technology developments without affecting the standard Level-1B process. To this end, the MODIS Level-1B browse process will be separate from the normal Level-1B process. After each execution of the Level-1B process, the Level-1B browse process will be initiated to generate the standard Level-1B browse product.

Specifics of the Level-1B browse product will be defined by the MODIS science team. This product might include, for example, a time sequence of "scenes", each of which is generated by sub-sampling the pixels of a rectangular area on the earth's surface. For each pixel, the recorded signals would be included (at reduced resolution) for a specified subset of the available frequency bands. Users would "call up" these browse scenes at will.

If the EOSDIS supports an extended browse product, then users could, at any time, request that special Level-1B browse scenes be generated from archived Level-1B data. In this mode, for example, the user would specify times, earth locations, and frequency bands to be included in the browse scene. The success of this extended browse capability could reduce the need for archiving standard Level-1B browse products.

5. Processing Previously Missing Data:

In normal operations, processing of a MODIS Level-1A or -1B data granule includes all of the data which should be in the granule. However, there will be occasions when some of the data are missing at the time of normal processing. In fact, this will happen frequently when the time boundaries of blocks of Level-0 data scheduled for Level-1A processing do not correspond to Level-1A granule boundaries.

If the previously missing data correspond to one or more complete granules, they will be processed in the normal mode. Otherwise, special handling will be required.

One possible scenario for processing previously missing data would be to have all of the data processed to Level-1A as it is scheduled, and then have the incomplete granules combined by a utility process. A shortcoming of this scenario is that, in some cases, not all of the data in a granule would have been processed at the same time. In fact, data within a granule could have been processed using different versions of the software.

Another possible scenario would be to reprocess all of the data within a granule whenever previously missing data is received. If this causes excessive inefficiency, the reprocessing could be done for only those cases where different versions of the software or the ancillary data are involved.

6. Reprocessing:

A. New Version of Software:

It is unlikely that the original, at-launch, version of the software will be used throughout the mission without revision. When significant changes are needed, the Configuration Control Board (CCB) will authorize a new version of the software. Generally, this will require reprocessing of all of the previous data, starting at the level of the software revision, and continuing through all higher levels, to provide consistency. Each data granule will contain accounting information which will provide processing traceability through all levels.

B. New Version of Ancillary Data:

In the Level-1A processing, the platform ancillary data are appended to the Level-1A data without making any changes to the MODIS instrument data. If more accurate spacecraft position and attitude data become available later, and it is required that the more accurate data be used for earth location, etc., then the replacement can be made using utility software, without reprocessing the instrument data. This replacement will be reflected in the version number of the Level-1A product, and in the metadata.