160-TP-005-001

Reducing Inter-DAAC Data Transfers Through Subsetting

Technical Paper

Technical Paper—Not intended for formal review or government approval.

October 10, 1995

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Andrew Endal /s/	10/23/95
Andrew Endal, Deputy Project Scientist EOSDIS Core System Project	Date

SUBMITTED BY

Mark Settle /s/ for	10/23/95
Joy Colucci, Science Office Manager EOSDIS Core System Project	Date

Hughes Applied Information Systems, Inc.
Landover, Maryland

This page intentionally left blank.

Abstract

We analyze the data transfers between EOSDIS DAACs required to support processing, reprocessing and archiving, based on the known data dependencies and currently assigned DAAC responsibilities for processing and archiving of the EOS Standard Products. Transferring of data between DAACs will be a major cost driver.

We analyze opportunities to reduce this data traffic by subsetting and/or data masking to eliminate fill values. We conclude that subsetting can reduce the inter-DAAC data traffic by at least 57%. This analysis does not address the increased processing capacity and re-allocation of working storage required to support subsetting and/or data masking.

Keywords: AHWGP, DAACs, data compression, networks, Standard Products, subsetting

Contents

1. Introduction

1.1 Scope 1-1
1.2 Organization 1-1
1.3 Review and Approval 1-2

2. Related Documents

2.1 Parent Documents 2-1
2.2 Applicable Documents 2-1
2.3 Information Documents 2-1

3. Baseline DAAC-to-DAAC Traffic

3.1 Distributed EOSDIS Data Processing and Archiving 3-1
3.2 DAAC-to-DAAC Traffic Assumptions 3-2
3.2.1 Processing Assumptions 3-3
3.2.2 Reprocessing Assumptions 3-5
3.3 Estimated DAAC-to-DAAC Transfer Rates 3-6

4. Analysis

4.1 GSFC-to-LaRC Data Transfers 4-1
4.2 GSFC-to-EDC Data Transfers 4-5
4.3 EDC-to-GSFC Data Transfers 4-7
4.4 GSFC-to-NSIDC Data Transfers 4-8
4.5 Net DAAC-to-DAAC Traffic Estimates 4-8

5. Implementation Options

5.1 Spectral Band Subsetting 5-1

5.2 Swath Subsetting 5-1

5.3 Compression for Fill Values 5-2

5.4 Geographical Masking and Compression 5-2

5.5 Remote File Access 5-2

5.6 Summary 5-3

Abbreviations and Acronyms

Figures

3-1 DAAC Processing and Archiving Responsibilities 3-1

3-2 TRMM Data Transfers 3-4

3-3 Phased ECS Capacity for SSI&T, Processing plus Reprocessing 3-5

4-1 Subsetting by MODIS Band Selection 4-3

4-2 Subsetting by MODIS Band Selection and Swath Width 4-4

4-3 Optimized Subsetting Strategy 4-5

4-4 Summed Circuit Requirements 4-10

5-1 Cumulative Effects of Subsetting/Masking (per Table 5-1) 5-4

Tables

3-1 Inter-DAAC Product Data Dependencies 3-2

3-2 Inter-DAAC Ancillary Data Dependencies 3-3

3-3 Transfer Rates for Processing 3-6

3-4 Transfer Rates for Processing and Reprocessing 3-7

3-5 Circuit Requirements for Processing 3-8

3-6 Circuit Requirements for Processing/Reprocessing 3-9

4-1 Largest GSFC-to-LaRC Data Transfers 4-2
4-2 CERES/MISR Requirements for MODIS Bands 4-3
4-3 MOD02_L1B_G Subsetting Options 4-5
4-4 Largest GSFC-to-EDC Data Transfers 4-6
4-5 Largest EDC-to-GSFC Data Transfers 4-7
4-6 Largest GSFC-to-NSIDC Data Transfers 4-8
4-7 Circuit Requirements for Processing with Subsetting 4-9
4-8 Circuit Requirements for Processing/Reprocessing with Subsetting 4-9
5-1 Volume Reduction Techniques 5-3

This page intentionally left blank.

1. Introduction

1.1 Scope

The purpose of this paper is to document the issue and to develop potential solutions with respect to DAAC-to-DAAC data transfers required to support processing and reprocessing of EOS Standard Products. This analysis does not include the DAAC-to-DAAC traffic required to support user queries; these typically represent less than 1% of the traffic required to support processing and reprocessing of EOS Standard Products.[1] This analysis also does not consider the increased processing and working storage capacities required for subsetting and/or masking[2] fill values in the data prior to transfer between DAACs

The analysis presented herein is based on the ECS Technical Baseline of August 1995, including Version 2.2 of the Ad Hoc Working Group on Production (AHWGP) data processing scenarios. These scenarios specifically represent the data processing required to produce Standard Products from the CERES and LIS instruments on the TRMM platform, the ASTER, CERES, MISR, MODIS and MOPITT instruments on the AM-1 platform, the Goddard Data Assimilation System (DAS), the SeaWinds Sensor (SWS) on ADEOS II, and the Dual Frequency Altimeter (DFA) on the Radar ALT platform. The analysis does not include processing/archiving of data from the EOS PM-1 instruments.

1.2 Organization

This document is organized into the following major sections:

Section 1	Introduction - Presents the scope, organization and approval process for this document.
Section 2	Related Documents - Provides a bibliography of parent, applicable and information documents related to the subject of this document.
Section 3	Baseline DAAC-to-DAAC Traffic - Describes the technical basis for the DAAC-to-DAAC data transfer rates required to support production and archiving of the EOS Standard Products, under the assumption that no subsetting, masking, or other strategies are adopted to reduce the data transfer rates.

---

1 Communications Requirements for the ECS Project, February 1995.

2 Throughout this paper, we will refer to the process of removing fill or unnecessary values from data arrays as "masking". In some instances, this function could be performed by standard data compression algorithms, as well as by custom bit masks. However, we do not consider the more general categories of data compression, which also have potential for reducing the data volumes to be transferred.

Section 4	Analysis - Provides estimates for potential reductions in the DAAC-to-DAAC data traffic, primarily by extraction of subsets of the data prior to transfer.
Section 5	Implementation Options - Describes the options available for achieving reductions in the DAAC-to-DAAC data transfers.

1.3 Review and Approval

This is an ECS Technical Paper, approved for release by the ECS Science Office. All comments on this paper should be provided to Andy Endal (e-mail: andy@eos.hitc.com).

2. Related Documents

2.1 Parent Documents

The parent documents are the documents from which this document's scope and content are derived.

423-41-02	Goddard Space Flight Center, Functional and Performance Requirements Specifications for the Earth Observing System Data and Information System (EOSDIS) Core System.
423-10-01-05	Goddard Space Flight Center, EOSDIS Core System Statement of Work.
505-41-19	Interface Requirements Document Between the EOSDIS Core System (ECS) and the National Oceanic and Atmospheric Administration Affiliated Data Center.
505-41-14	Interface Requirements Document Between the EOSDIS and the TRMM Ground System.
210-TP-001-004	Technical Baseline for the ECS Project.

2.2 Applicable Documents

The following documents are referenced herein and are directly applicable to this document. In the event of conflict between any of these documents and the present document, this document shall take precedence.

220-CD-001-003	Communications Requirements for the ECS Project.
543-TP-001-003	A Cost Comparison of Transferring Inter-DAAC Data via Media versus the ESN WAN.

2.3 Information Documents

These documents provide additional information or influence elements of this document.

not numbered	1995 MTPE EOS Reference Handbook, (NASA/Goddard Space Flight Center, 1995).

3. Baseline DAAC-to-DAAC Traffic

3.1 Distributed EOSDIS Data Processing and Archiving

The EOSDIS is a geographically distributed information system to process, store, and provide access to Standard Product data sets derived from EOS instruments. Data processing is performed at Distributed Active Archive Centers (DAACs), selected for their scientific expertise and interests in the EOS instruments and science disciplines. Archiving and distribution of these data products are performed at DAACs which have strong associations with the anticipated data users. Figure 3-1 shows the DAAC processing and archiving activities included in the current analysis. This includes the TRMM data and all EOS Standard Products up to (but not including) the EOS PM-1 launch, with the exceptions of SAR products at the ASF DAAC and SAGE III products at the LaRC DAAC.[3] The EDC DAAC is also responsible for archiving and distributing data from Landsat 7. This will not lead to any DAAC-to-DAAC traffic in support of processing.

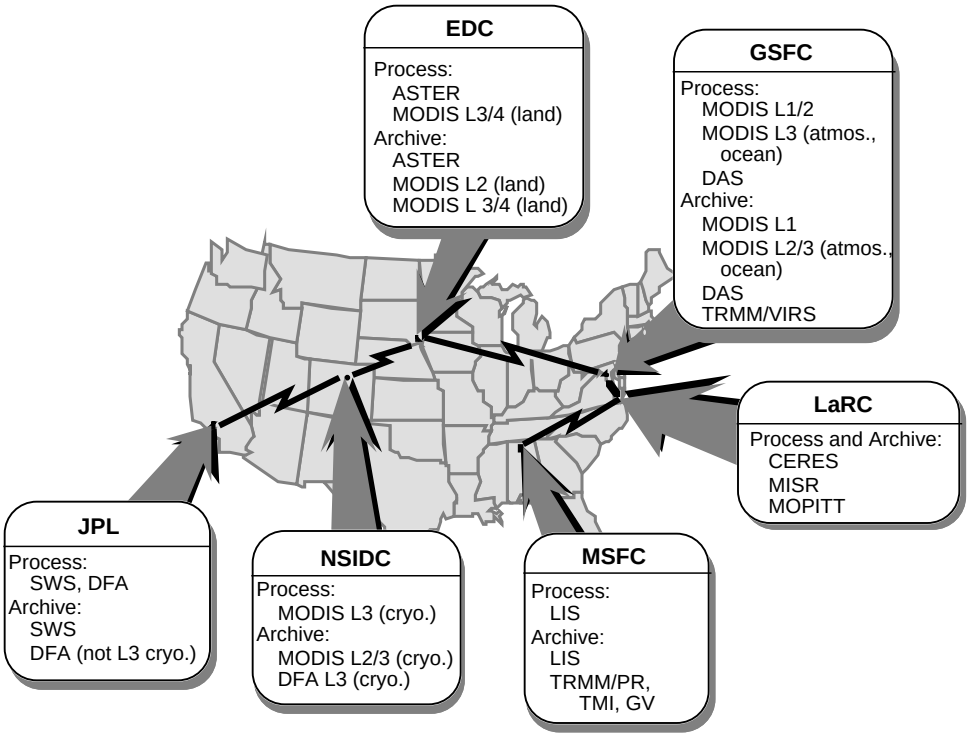

Figure 3-1. DAAC Processing and Archiving Responsibilities

[3] SAR processing is not expected to generate any DAAC-to-DAAC traffic. The SAGE III instrument team has indicated that they will not require data from other instruments for their processing. They will require NMC/NOAA data, which are already included in the transfers required to support other instruments.

In most cases, the processing DAAC is also the archiving and distribution DAAC. When the processing DAAC differs from the archiving DAAC, this requires DAAC-to-DAAC transfers of the data products. Data transfers more commonly occur because the data products processed and archived at a particular DAAC are required as primary or ancillary inputs to the data processing at one or more other DAACs. Processing dependencies among the instrument product sets are shown in Table 3-1. This may not be a complete set of dependencies. Several instrument teams (MOPITT, for example) have expressed interest in using the DAS products for their processing, but this is not yet reflected in their AHWGP processing scenarios. Each of the instrument processing scenarios also contain internal dependencies. These are not shown in Table 3-1, except in the case of MODIS, whose processing is distributed among 3 DAACs (GSFC, EDC and NSIDC).

Table 3-1. Inter-DAAC Product Data Dependencies

Data from:	Required by					
	ASTER	CERES	DAS	MISR	MODIS	MOPITT
CERES			X			
DAS				X	X	
MISR	X				X	
MODIS	X	X		X	X	X
SeaWinds			X			
TMS		X				
VIRS		X				

In addition to the TRMM and EOS Standard Products, specific DAACs are responsible for developing and/or maintaining external ancillary data which are derived from non-EOS instrument data. Where these data are required for processing at other DAACs, this also leads to DAAC-to-DAAC data transfers. Table 3-2 lists the external ancillary data sets which contribute to the DAAC-to-DAAC data traffic. Table 3-2 does not include those ancillary data dependencies which are satisfied within a single DAAC. The ancillary data traffic is low in volume as compared to the Standard Product traffic generated by the instrument dependencies shown in Table 3-1.

3.2 DAAC-to-DAAC Traffic Assumptions

DAAC-to-DAAC data traffic is estimated by analysis of Version 2.2 of the AHWGP information as expressed in the ECS Technical Baseline of August 1995. The AHWGP information provides quarterly processing scenarios describing, for each instrument, the planned Standard Product processing with required input data and output products.

The baseline DAAC-to-DAAC traffic estimates assume that no subsetting or compression of the data occur prior to or during transfers (*i.e.*, entire granules are transferred). Thus, no processing and/or working storage capacity is provided in the current ECS baseline in support of subsetting or compression of the data to be transferred across DAACs.

Table 3-2. Inter-DAAC Ancillary Data Dependencies

Data Set	Description	Volume	Archive	Needed by
ANC_EDC_DEM	Digital Elevation Map	200 MB	EDC	CERES MODIS MOPITT
ANC_EDC_LANDCOVER	Surface Land Cover and Vegetation Type	250 MB	EDC	CERES DAS MODIS
ANC_NMC_SURF	NMC Analysis - Surface Parameters	12 MB/day	GSFC	CERES DFA MODIS MOPITT SWS
ANC_NMC_PROF	NMC Analysis - Temperature, Moisture and Ozone Profiles	12 MB/day	GSFC	ASTER CERES MOPITT
ANC_NMC_4DA	NMC Assimilation Model Parameters	12 MB/day	GSFC	MODIS
ANC_GSFC_O3TOMS	TOMS Ozone Profiles	0.5 MB/day	GSFC	ASTER CERES
ANC_EPA_ECOSYSDB	EPA Ecosystems Map of 59 Classes at 10 Minute Resolution	10 MB	GSFC	ASTER
ANC_NESDIS_SNOW/ICE	Daily Level 3 Product Created at NOAA using FNOC Algorithm	10 MB/day	GSFC	ASTER CERES MISR SWS
ANC_NESDIS_GPI	Gridded Monthly Satellite Derived Estimates of Rainfall	1 GB/ month	MSFC	TSDIS
ANC_GPCC_RainProd	Gridded Monthly Rainfall from Gauge Data and Other Sources	1 MB/ month	MSFC	TSDIS
MSFC_SSM/I_BT	SSM/I Brightness Temperature	160 MB/day	MSFC	TSDIS

3.2.1 Processing Assumptions

The analysis assumes that all DAAC-to-DAAC transfers required to support processing are coordinated so that multiple instrument teams processing their data at a single DAAC require only one transfer of the same data.

Data transfers to support processing also include the ancillary data listed in Table 3-2. Some of these data sets are relatively static. For these data, we have assumed a weekly update to synchronize across DAACs. Due to the small volume of these data sets, total DAAC-to-DAAC data traffic is not sensitive to this assumption.

TRMM data operations leading to DAAC-to-DAAC transfers are indicated by the heavy lines in Figure 3-2. Data from the TRMM instruments are received by the Science Data Processing Facility (SDPF) at GSFC. The SDPF forwards the CERES instrument data to LaRC, and forwards the LIS instrument data to MSFC. This transfer occurs via the NASCOM Operational Network (NOLAN), which is not considered part of the DAAC-to-DAAC network. The SDPF forwards the Precipitation Radar (PR), TRMM Microwave Imager (TMI) and Visible Infrared Scanner (VIRS) data to the TRMM Science Data and Information System (TSDIS) facility, at GSFC, for processing. The TSDIS also ingests and processes TRMM Ground Validation (GV) data. The VIRS data products are maintained at the GSFC DAAC. None of the above flows contribute to DAAC-to-DAAC traffic. The PR, TMI, GV, and Combined Instruments products from the TSDIS are transferred to MSFC for archiving and distribution to the user community. VIRS and TMI products are required as ancillary data for CERES processing at LaRC, so the resulting transfers from GSFC and MSFC to LaRC are included in the DAAC-to-DAAC traffic. Finally, external ancillary data required to support TSDIS processing and CERES processing are transferred from MSFC.

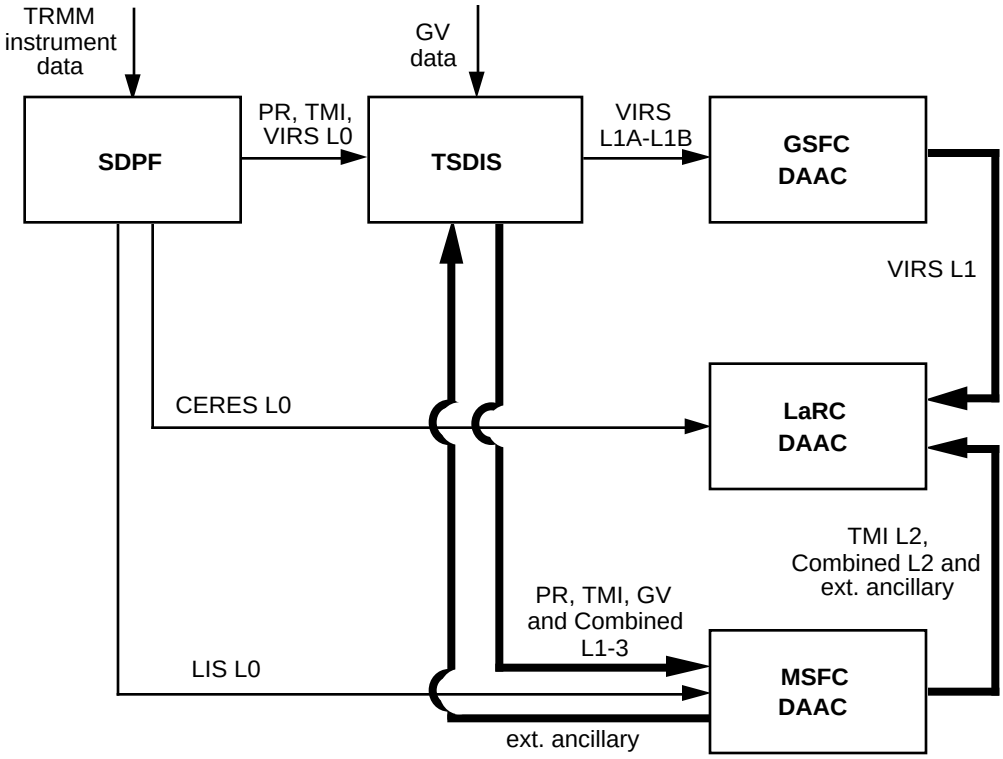

Figure 3-2. TRMM Data Transfers; heavy lines show inter-DAAC transfers.

3.2.2 Reprocessing Assumptions

The ECS DAAC Standard Product production facilities (including archive retrieval, local area networks, working storage, and processors) are sized to provide capacity for Science Software Integration & Test (SSI&T) and reprocessing, in addition to the initial Standard Product processing. While the processing requirements are based on estimates provided by EOS instrument teams through the AHWGP, the total capacities (including SSI&T and reprocessing) are specified as multipliers of the basic processing capacity, as illustrated in Figure 3-3. These multipliers are described relative to the launch epochs, which are assumed to be the third quarter of 1997 for TRMM and the third quarter of 1998 for AM-1.

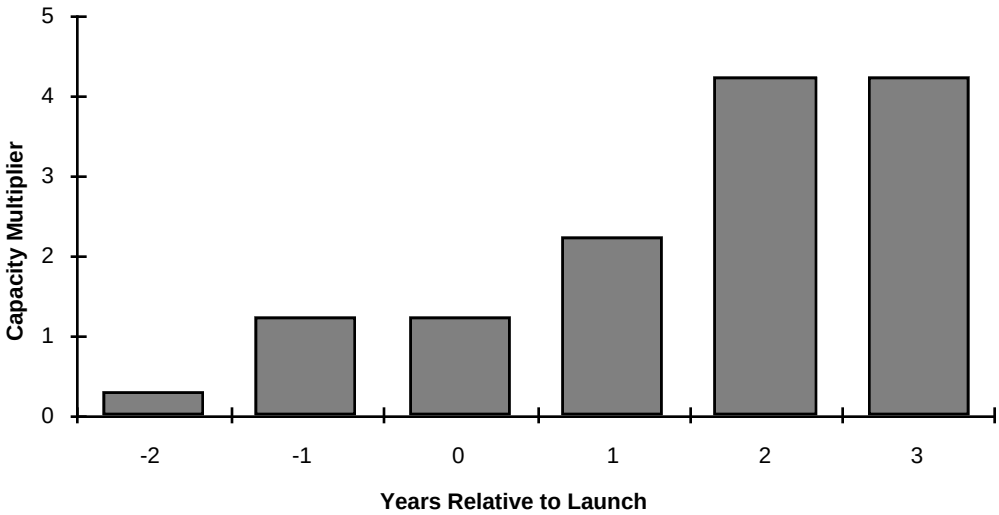

Figure 3-3. Phased ECS Capacity for SSI&T, Processing plus Reprocessing

We use the symbol **X** to indicate the capacity required for initial processing of the EOS Standard Products. Two years prior to launch of a given platform, the DAACs are sized at 0.3**X** in order to support early SSI&T. One year prior to launch, the capacity is increased to 1.2**X** to support pre-launch SSI&T and other operational testing. One year after launch, the capacity is increased to 2.2**X** to provide for reprocessing of the data with improved calibrations and algorithms. Two years after launch, the capacity is again increased, to 4.2**X**. These multipliers are specified in the ECS Technical Baseline. They are based on experience from previous scientific missions, primarily the Upper Atmospheric Research Satellite (UARS), where the capacity requirements for reprocessing grew as more and more data are accumulated and reprocessed, and as the algorithms became more complex.

Reprocessing requires DAAC-to-DAAC data transfers for the same reasons as the initial processing (see Section 3-1). The ECS Technical Baseline does not specify what part of the ECS capacity is allocated for reprocessing. In order to estimate the DAAC-to-DAAC transfers required to support reprocessing, we assume that reprocessing campaigns begin one year after launch and proceed at the same rate (1**X**) as the processing rate. At two years after launch, the reprocessing increases to twice (2**X**) the processing rate. These assumptions are consistent with (but not required by) the baselined total processing capacities indicated in Figure 3.3.

The DAAC-to-DAAC traffic analysis assumes that reprocessing is coordinated to minimize the required transfers. For example, at the LaRC DAAC, Level 2 processing of the CERES and MISR data both require the MODIS Level 1B data as inputs. MISR and MOPITT Level 2 processing require MODIS Level 2 atmospheric products as inputs. Thus, minimizing the transfer of MODIS data from GSFC to LaRC requires that CERES, MISR and MOPITT Level 2 reprocessing campaigns all be closely coordinated. Because the instrument teams are likely to be at different stages with respect to their Level 2 algorithm refinement, it is not clear that such optimization will always be possible. Nevertheless, we have made this assumption for the purpose of estimating DAAC-to-DAAC traffic. For Standard Products which are not produced until some time after launch, we assume that reprocessing starts one year after the initial processing and grows to twice the processing rate at two years after the start of processing.

DAAC-to-DAAC transfers to support TSDIS processing are specified in the ECS/TRMM Ground System Interface Requirements Document (IRD). TSDIS reprocessing occurs at twice the processing rate, starting in the same year as the TRMM launch. Each reprocessing requires that existing TSDIS archival products be transferred from MSFC to GSFC, and that a new set of TSDIS products be transferred from GSFC to MSFC. Ancillary data required to support TSDIS reprocessing must also be transferred from MSFC to GSFC.

3.3 Estimated DAAC-to-DAAC Transfer Rates

Tables 3-3 lists the yearly DAAC-to-DAAC data transfer rates required to support processing and archiving of data products.

Table 3-3. Transfer Rates for Processing

		GB/day			
From	**To**	**1997**	**1998**	**1999**	**2000**
EDC	**GSFC**		24.8	24.8	24.8
	LaRC	0.06	0.06	0.06	0.06
GSFC	**EDC**		118.8	118.8	118.8
	JPL			0.02	0.02
	LaRC	1.4	195.2	211.2	225.5
	MSFC	13.3	13.3	13.3	13.3
	NSIDC		14.7	14.7	14.7
JPL	**GSFC**			0.05	0.05
	NSIDC			<0.01	<0.01
LaRC	**EDC**		5.0	5.0	5.0
	GSFC			11.5	11.5
MSFC	**GSFC**	0.2	0.2	0.2	0.2
	LaRC	1.4	1.4	1.4	1.4
NSIDC	**EDC**		0.4	0.4	0.4

The rates in Table 3-3 differ from estimates provided at the time of the ECS Preliminary Design Review (PDR). The PDR estimate for the GSFC-to-EDC transfer rate included MODIS Level 1B data to support higher level MODIS processing at the EDC. The MODIS team has dropped this requirement, reducing the GSFC-to-EDC transfer rates. The PDR estimate for the GSFC-to-LaRC transfer rate assumed that the full MODIS Level 1B data were transferred to support MISR processing, and that the MODIS Level 1B cloud bands were separately transferred to support CERES processing (*i.e.*, the MODIS cloud bands were transferred twice). We now assume that a single set of MODIS Level 1B bands are transferred to LaRC, and any subsetting to support a particular instrument's requirements is performed at LaRC. This reduces the estimate for the GSFC-to-LaRC traffic. Other (smaller) changes have occurred as the instrument teams have updated their requirements and we have added more instruments to the AHWGP database.

Table 3-4 lists the yearly DAAC-to-DAAC data transfer rates required to support both processing and reprocessing. We should note that, while the processing requirements are based on scenarios supplied by the instrument teams, the reprocessing requirements are based on the simplifying assumptions specified in Section 3.2.2. In some cases, the instrument teams have indicated their reprocessing plans, which are not necessarily in accord with the multiplying factors which we have assumed for reprocessing. In reality, reprocessing will be governed by a combination of instrument team desires and project/DAAC allocations of processing and WAN capacities.

Table 3-4. Transfer Rates for Processing and Reprocessing

		GB/day			
From	**To**	**1997**	**1998**	**1999**	**2000**
EDC	**GSFC**		24.8	49.6	74.4
	LaRC	0.06	0.1	0.2	0.2
GSFC	**EDC**		118.8	237.6	356.4
	JPL			0.02	0.04
	LaRC	1.4	196.6	423.9	676.5
	MSFC	40.0	40.0	40.0	40.0
	NSIDC		14.7	29.3	44.0
JPL	**GSFC**			0.05	0.1
	NSIDC			<0.01	<0.01
LaRC	**EDC**		5.0	9.9	14.9
	GSFC			11.5	23.0
MSFC	**GSFC**	27.3	27.3	27.3	27.3
	LaRC	1.4	2.8	4.1	4.1
NSIDC	**EDC**		0.4	0.7	1.1

To convert from GB/day to units more commonly used for circuit analysis (Mbps), the values in Tables 3-3 and 3-4 should be multiplied by:

1000 (MB/GB) * 8 (bits/Byte) / 86400 (seconds/day) * 2.5 (effective overhead) = 0.23,

where the effective overhead includes circuit utilization, protocol overhead, and scheduling contingency.[4] Table 3-5 provides the circuit capacity requirements to support the processing data transfers indicated in Table 3-3. Circuit requirements which exceed T1 capacity (1.544 Mbps) are indicated in *italics*, while requirements which exceed T3 capacity (44.736 Mbps) are indicated in ***bold-face italics***.

The DAAC-to-DAAC transfer rates indicated in Table 3-5 (and subsequent tables) are based on logical flows required to support EOS Standard Product processing and archiving. The actual bandwidth required to support these DAAC-to-DAAC flows may be different, depending on the connectivity and existing network infrastructure utilized to implement the logical DAAC-to-DAAC connectivity.

Table 3-5. Circuit Requirements for Processing

		Mbps			
From	**To**	**1997**	**1998**	**1999**	**2000**
EDC	**GSFC**		*5.7*	*5.7*	*5.7*
	LaRC	0.01	0.01	0.01	0.01
GSFC	**EDC**		*27.5*	*27.5*	*27.5*
	JPL			<0.01	<0.01
	LaRC	0.3	***45.2***	***48.9***	***52.2***
	MSFC	*3.1*	*3.1*	*3.1*	*3.1*
	NSIDC		*3.4*	*3.4*	*3.4*
JPL	**GSFC**			0.01	0.01
	NSIDC			<0.01	<0.01
LaRC	**EDC**		1.1	1.1	1.1
	GSFC			*2.7*	*2.7*
MSFC	**GSFC**	0.05	0.05	0.05	0.05
	LaRC	0.3	0.3	0.3	0.3
NSIDC	**EDC**		0.08	0.08	0.08

Table 3-6 provides the circuit capacity requirements to support the processing and reprocessing data transfers indicated in Table 3-4. At current projections for Wide-Area Networks, these data transfer rates represent significant cost-drivers on the EOSDIS program. Transferring the data by media rather than networks is unlikely to significantly reduce these costs.[5]

The rates in Tables 3-3 through 3-6 do not include transfers from the SDPF (at GSFC) of CERES and LIS Level 0 data (to the LaRC DAAC and the MSFC DAAC, respectively). As mentioned in Section 3.2.1, these transfers are currently expected to use the NOLAN. If the CERES and LIS Level 0 data are to be transferred by the same DAAC-to-DAAC networks as the main EOS

[4] Communications Requirements for the ECS Project.
[5] A Cost Comparison of Transferring Inter-DAAC Data via Media versus the ESN WAN.

data, then 0.090 GB/day should be added to the GSFC-to-LaRC transfer rates in Tables 3-3 and 3-4, and 0.065 GB/day should be added to the GSFC-to-MSFC transfer rates in these tables. (No additional transfers are required for reprocessing, since the Level 1A data are archived at the LaRC and MSFC DAACs.) The additional circuit capacities to be added to Tables 3-5 and 3-6 are: 0.02 Mbps for the GSFC-to-LaRC transfers and 0.01 Mbps for the GSFC-to-MSFC transfers.

Table 3-6. Circuit Requirements for Processing/Reprocessing

		Mbps			
From	**To**	**1997**	**1998**	**1999**	**2000**
EDC	**GSFC**		*5.7*	*11.5*	*17.2*
	LaRC	0.01	0.02	0.05	0.05
GSFC	**EDC**		*27.5*	***55.0***	***82.5***
	JPL			<0.01	0.01
	LaRC	0.3	***45.5***	***98.1***	***156.6***
	MSFC	*9.3*	*9.3*	*9.3*	*9.3*
	NSIDC		*3.4*	*6.8*	*10.2*
JPL	**GSFC**			0.01	0.02
	NSIDC			<0.01	<0.01
LaRC	**EDC**		1.1	*2.3*	*3.4*
	GSFC			*2.7*	*5.3*
MSFC	**GSFC**	*6.3*	*6.3*	*6.3*	*6.3*
	LaRC	0.3	0.6	1.0	1.0
NSIDC	**EDC**		0.08	0.2	0.3

4. Analysis

In the following sections, we estimate the potential data traffic reductions achievable by transferring only those subsets of the data required at the target DAAC (for processing or archiving). The analysis is restricted to the four largest DAAC-to-DAAC transfers. Ordered by volume, these transfers are: GSFC-to-LaRC, GSFC-to-EDC, EDC-to-GSFC, and GSFC-to-NSIDC. The values cited in Sections 4.1 to 4.4 are for processing, only. In Section 4.5, we summarize the overall reductions in the DAAC-to-DAAC transfers, for processing and reprocessing.

4.1 GSFC-to-LaRC Data Transfers

Table 4-1 lists the largest GSFC-to-LaRC data transfers. The first column provides the data set name, as listed in the AHWGP data base; the second column describes the data set. The third column identifies the process(es) requiring the data set as input, using the AHWGP designation; a description of each process is given in the fourth column. Finally, the fifth column gives the average daily transfer rate. These data sets account for 99.3% of the GSFC-to-LaRC traffic.

Since the MODIS Level 1B radiances (MOD02_L1B_G) is largest data set to be transmitted, a detailed analysis of the subsetting possibilities is warranted. These data will be used by the CERES and MISR algorithms. However, they do not need all 36 of the MODIS spectral bands. CERES needs 11 bands and MISR needs 7 bands; 3 of these bands are common to both CERES and MISR. The simplest strategy is to transfer only the 15 bands needed by either CERES or MISR, as illustrated in Figure 4-1. Since the MODIS bands vary by a factor of 4 in spatial resolution (hence, a factor of 16 in data volume), calculation of the subsetted volumes must consider the specific bands to be transmitted. Table 4-2 lists the MODIS bands, their wavelengths, and spatial resolution, and indicates the bands required for CERES and/or MISR processing.

The reflected bands ($\lambda < 3\ \mu$) are collected during daytime only (50% duty cycle), whereas the thermal bands ($\lambda > 3\ \mu$) have a 100% duty cycle. If we use a single thermal channel (1000 m resolution, collected for 100% of each orbit) as a data unit, a 250-m reflected band is equivalent to 8 unit bands (16 times as many pixels, collected for 50% of each orbit), a 500-m reflected band is equivalent to 2 unit bands (4 times as many pixels, collected for 50% of each orbit), and a 1000-m reflected band is equivalent to a half unit band (same number of pixels, collected for 50% of each orbit). The full set of MODIS bands is equivalent to 48.5 unit bands. By the same method, the 15 bands required by CERES and/or MISR are equivalent to 23 unit bands. Thus, the volume to be transmitted using the subsetting strategy illustrated in Figure 4-1 is

$$(23/48.5) * 180.1 \text{ GB/day} = 85.4 \text{ GB/day}.$$

MISR has a narrow swath (370 km for the nadir camera), as compared to the 2,300 km swath of MODIS, so further reductions can be achieved by spatial subsetting. In fact, the MISR team has stated that they assume that MODIS data will be provided as a subset covering 200 km to either

side of nadir.[6] A MODIS swath has 1354 pixels to cover 2,300 km, with 1 km resolution at nadir.[7] If we (conservatively) assume that 400 MODIS pixels are required to cover the MISR swath, the potential data volume reduction factor is 400/1354. A possible strategy for reducing the GSFC-to-LaRC traffic would be to transmit the 11 bands needed by CERES, and transmit a separate, narrow-swath subset of the 7 bands required by MISR, as illustrated in Figure 4-2.

Table 4-1. Largest GSFC-to-LaRC Data Transfers

Data Set Name	Data Set Description	Process Name	Process Description	Transfer Rate (GB/day)
MOD02_L1B_G	MODIS Level 1B Calibrated Radiances	4bAV, 4bAF	Determine Cloud Properties (CERES Subsystem 4)	180.1
		MISP1B2	MISR Level 1B2 processing	
MOD03_L1A_G	MODIS Level 1A Geolocation Fields	4bAV, 4bAF	Determine Cloud Properties (CERES Subsystem 4)	13.5
		MISP1B2	MISR Level 1B2 processing	
		MISP2TC	MISR Level 2 Top≠ of-Atmosphere/ Cloud	
		MISP2AS	MISR Level 2 Aerosol/Surface	
MOD05_L2_G	MODIS Level 2 Near IR Precipitable Water	MISP2AS	MISR Level 2 Aerosol/Surface	11.1
MOD06_L2_G	MODIS Level 2 Cloud Product	MISP2TC	MISR Level 2 Top-of-Atmosphere/ Cloud	8.9
		MOPL2-E, MOPL2-H	MOPITT Level 2 Processing	
MOD30_L2_G	MODIS Level 2 Atmospheric Temperature and Moisture Profiles	MISP2AS	MISR Level 2 Aerosol/Surface	7.2
		MOPL2-E, MOPL2-H	MOPITT Level 2 Processing	
MOD35_L2_G	MODIS Level 2 TIR Precipitable Water	MISP2TC	MISR Level 2 Top-of-Atmosphere/ Cloud	3.2

6 D. Wenkert, e-mail message, July 23, 1995.
7 Off-nadir pixels are wider than the nadir pixels.

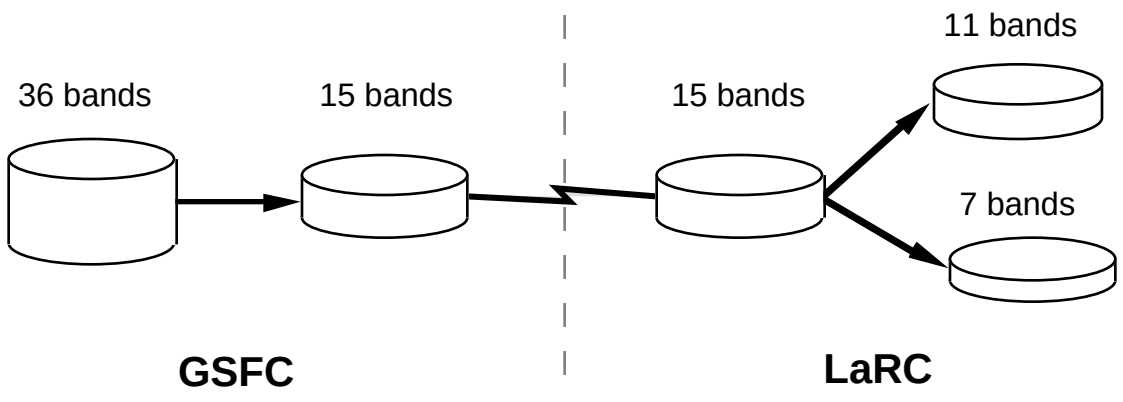

Figure 4-1. Subsetting by MODIS Band Selection

Table 4-2. CERES/MISR Requirements for MODIS Bands

Band	λ (μ	res (m)	CERES	MISR	Band	λ (μ	res (m)	CERES	MISR
1	0.645	250	√		19	0.940	1000		√
2	0.858	250			26	1.375	1000	√	√
3	0.469	500			20	3.75	1000	√	
4	0.555	500			21	3.96	1000		
5	1.240	500		√	22	3.96	1000		
6	1.640	500	√	√	23	4.05	1000		
7	2.130	500	√	√	24	4.47	1000		
8	0.412	1000			25	4.52	1000		
9	0.443	1000			27	6.72	1000		
10	0.488	1000			28	7.33	1000		
11	0.531	1000			29	8.55	1000	√	
12	0.551	1000			30	9.73	1000		
13	0.667	1000			31	11.03	1000	√	
14	678	1000			32	12.02	1000	√	
15	0.748	1000			33	13.34	1000	√	
16	0.869	1000			34	13.64	1000	√	
17	0.905	1000		√	35	13.94	1000	√	
18	0.936	1000		√	36	14.24	1000		

Using Table 4-2, the 11 bands required by CERES are equivalent to 19.5 unit bands, so the CERES subset volume is

(19.5/48.5) * 180.1 GB/day = 72.4 GB/day.

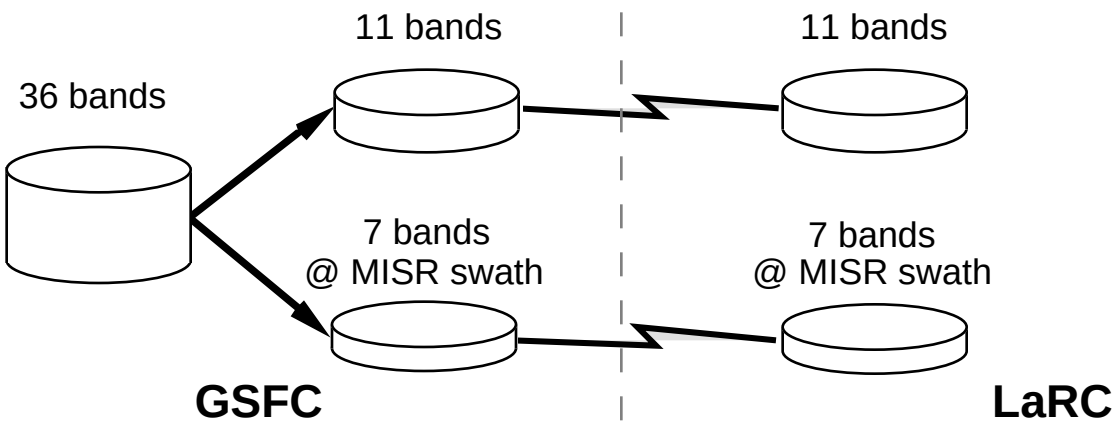

Figure 4-2. Subsetting by MODIS Band Selection and Swath Width

The 7 MISR bands are equivalent to 8 unit bands, so the MISR subsetted volume (400-km swath) is

(8/48.5) * (400/1354) * 180.1 GB/day = 8.8 GB/day.

Thus, the volume to be transmitted using the subsetting strategy illustrated in Figure 4-2 is

72.4 GB/day + 8.8 GB/day = 81.2 GB/day.

Additional savings can be achieved by re-using the 3 bands common to both CERES and MISR, as illustrated in Figure 4-3. The 4 MISR-unique bands are equivalent to 3.5 unit bands, and require

(3.5/48.5) * (400/1354) * 180.1 GB/day = 3.8 GB/day.

Adding in the 11 CERES bands (72.4 GB/day, see above) yields a total of 76.2 GB/day. Thus, the MOD02_L1B_G transfer can be reduced from 180 GB/day to somewhere between 85 and 76 GB/day, depending on the specific subsetting strategy.

The MOD02 subsetting strategies and the associated data rates are summarized in Table 4-3. The reduction factor is calculated relative to the full MOD02 volume of 180.1 GB/day. We show the CERES-only bands separately because CERES data processing requires the MODIS data immediately after launch, whereas the MISR team does not plan to incorporate MODIS data in their processing until eighteen months after launch. Thus, there is a period when MODIS data is only needed to support CERES processing.

Returning to Table 4-1, the MODIS Level 1A Geolocation Fields (MOD03_L1A_G) are required by both the CERES algorithms and the MISR algorithms. Thus, this data flow cannot be reduced by subsetting to a narrower swath.

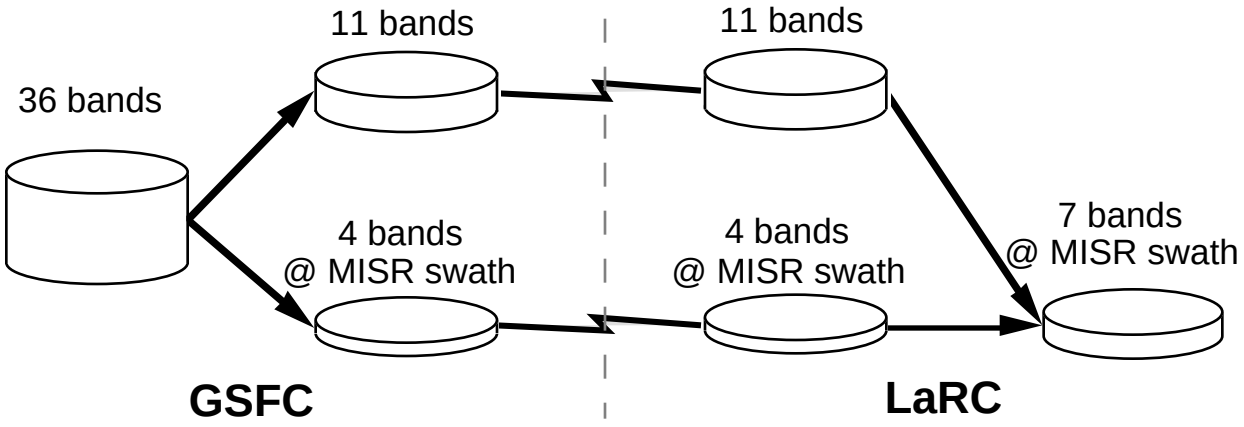

Figure 4-3. Optimized Subsetting Strategy

Table 4-3. MOD02_L1B_G Subsetting Options

Transmitted Data	Data Volume (GB/day)	Reduction Factor
15 bands required for CERES and/or MISR	85.4	0.474
11 bands for CERES + 7 bands for MISR @ narrow swath	81.2	0.451
11 bands required for CERES + 4 MISR-only bands @ narrow swath	76.2	0.423
11 bands required for CERES, only	72.4	0.402

The MOD05_L2_G and MOD35_L2_G data are only needed by the MISR algorithms, so this data transfer can be reduced by subsetting to match the MISR swath. The MOD06_L2_G and MOD30_L2_G products are used by both MISR and MOPITT. The MOPITT swath of 640 km is wider than the MISR swath, but both are small compared to the MODIS swath. Thus, significant savings can be achieved by spatially subsetting these MODIS Level 2 products to the MOPITT swath. The resulting transfer rate for all the MODIS Level 2 products is:

$$(400/1354) * (11.2 + 3.2) \text{ GB/day} + (640/1354) * (8.9 + 7.2) \text{ GB/day} = 12.5 \text{ GB/day}.$$

Thus, the GSFC-to-LaRC data transfers listed in Table 4-1 can be reduced from 224 GB/day to 101 GB/day (based on the subsetting scheme shown in Figure 4-3).

4.2 GSFC-to-EDC Data Transfers

Table 4-4 lists the largest GSFC-to-EDC data transfers. The columns in Table 4-4 are the same as in Table 4-1. In some cases the transfer is required to archive and distribute data at the EDC; this is indicated in the fourth column. These archived data may also be needed for processing at

the EDC, but this is not considered the primary reason for transferring the data. The data sets listed in Table 4-4 account for over 99.9% of the GSFC-to-EDC traffic.

The MODIS land data products cover both land and ocean areas, with fill values used for pixels which fall on ocean areas. This specifically applies to the MOD09, MOD41, MOD13, MOD11, and MOD14 data included in Table 4-4. These fill values can be eliminated by either a land/ocean bit mask or by a suitable data compression algorithm. Since processing at the EDC is limited to land areas, only the land pixels need to be transferred, even if a data set includes real values over both land and ocean areas (*e.g.*, MOD03, MOD30 and MOD35). This will require a land/ocean bit mask. Approximately 70% of the Earth's surface area is covered by water. Masking out the pixels over water (prior to transferring the data sets from GSFC) can thus reduce the required transfer rate to 30% of the uncompressed rate.

Thus, the GSFC-to-EDC transfer rate can be reduced from 118.8 GB/day to 35.6 GB/day.

Table 4-4. Largest GSFC-to-EDC Data Transfers

Data Set Name	Data Set Description	Process Name	Process Description	Transfer Rate (GB/day)
MOD09_L2_G	MODIS Level 2 Surface Reflectance		transfer to EDC archives	41.2
MOD41_L2_H	MODIS Level 2 Land Surface Resistance Index		transfer to EDC archives	23.7
MOD13_L2_G	MODIS Level 2 Vegetation Indices		transfer to EDC archives	15.8
MOD03_L1A_G	MODIS Level 1A Geolocation Fields	MOD09:SUBS:L3:DY	MODIS BRDF Daily Subsetting of Level-2 Data and DB Compilation	13.5
		MOD11:L3:WK	MODIS Level 3 Weekly Compositing for Land Surface Temperature	
		MOD15:L4:10DY	MODIS Level 4 10-Day Production of Leaf Area Index & FPAR	
		MOD34:L3:10DY	MODIS Level 3 10≠ day Compositing of Vegetation Indices	
		MOD40:L3:DY	MODIS Level 3 Daily Compositing of Gridded Thermal Anomalies	

Table 4-4. Largest GSFC-to-EDC Data Transfers (cont.)

Data Set Name	Data Set Description	Process Name	Process Description	Transfer Rate (GB/day)
MOD11_L2_G	MODIS Level 2 Land Surface Temperature		transfer to EDC archives	12.6
MOD30_L2_G	MODIS Level 2 Atmospheric Temperature and Moisture Profiles	AST_PGE_04 AST_PGE_05	ASTER Atmospheric correction--VNIR, SWIR ASTER Atmospheric correction--TIR	7.2
MOD35_L2_G	MODIS Level 2 Precipitable Water	MOD09:SUBS:L3:DY MOD12:L3:3MN:G	Daily Subsetting of Level 2 Database MODIS Level 3 Land Cover	3.2
MOD14_L2_G	MODIS Level 2 Thermal Anomalies		transfer to EDC archives	1.6

4.3 EDC-to-GSFC Data Transfers

The EDC-to-GSFC data transfers are dominated by the MOD09 Level 3 product, whose characteristics are shown in Table 4-5. In this case, the MOD09 Level 3 data from the previous 16-day period is used as input for the process to compute the current MOD09 Level 2 data. This data set accounts for 97% of the total EDC-to-GSFC traffic.

Table 4-5. Largest EDC-to-GSFC Data Transfers

Data Set Name	Data Set Description	Process Name	Process Description	Transfer Rate (GB/day)
MOD09_L3_16DY_G (previous period)	MODIS Level 3 BRDF/Albedo	MOD09:13:L2:G	MODIS Level 2 Production of Land Surface Reflectances and Vegetation Indices	24.0

As in the case of the MODIS land products in the GSFC-to-EDC flows, the data volumes assume that all pixels (including those falling over water) are included in the transferred data. Here again, data masking can eliminate the fill values for the pixels falling over water. Thus, these products can be reduced to 30% of the uncompressed volume. On this basis, the total EDC-to-GSFC traffic can be reduced from 24.8 GB/day to 8.0 GB/day, where we have not assumed any reduction of the other data transferred from EDC to GSFC.

4.4 GSFC-to-NSIDC Data Transfers

Table 4-6 lists the largest GSFC-to-NSIDC data transfers. These data sets account for 99.9% of the GSFC-to-NSIDC traffic.

Table 4-6. Largest GSFC-to-NSIDC Data Transfers

Data Set Name	Data Set Description	Process Name	Process Description	Transfer Rate (GB/day)
MOD03_L1A_G	MODIS Level 1A Geolocation Fields	MOD10:L3:DY:G	MODIS Level 3 Daily Compositing of Gridded Snow Cover	13.5
		MOD29:L3:DY:G	MODIS Level 3 Daily Compositing of Gridded Sea Ice Max Extent	
MOD29_L2_G	MODIS Level 2 Sea Ice Max Extent		transfer to NSIDC archives	0.8
MOD10_L2_G	MODIS Level 2 Snow Cover		transfer to NSIDC archives	0.4

The data sets listed in Table 4-6 are used in the computation of gridded snow/ice MODIS products at the NSIDC, with the MODIS Level 2 sea-ice and snow products archived at NSIDC. The data volumes in Table 4-5 assume global coverage. However, only approximately 20% of the Earth's surface is covered by snow or ice at maximum extent. Therefore, 80% of the data traffic can eliminated by masking out those areas of the Earth's surface which are never covered by snow or ice. On this basis, the GSFC-to-NSIDC traffic can be reduced from 14.7 GB/day to 2.9 GB/day.

4.5 Net DAAC-to-DAAC Traffic Estimates

Sections 4.1 through 4.4 described options for reducing DAAC-to-DAAC transfers by transmitting only the data which are actually required for processing and/or archiving. In the present section, we summarize the potential DAAC-to-DAAC traffic reductions. Table 4-7 provides the circuit capacity requirements to support data transfers for initial processing and archiving, including the reductions indicated in the previous subsections and utilizing the network efficiency factors described in Section 3.3. For the GSFC-to-LaRC transfers, we have assumed the most efficient subsetting option. For the LaRC-to-EDC transfers, we have assumed a 70% volume reduction, using the same logic as applied to the GSFC-to-EDC and EDC-to-GSFC transfers (*i.e.*, only land pixels are transferred). Table 4-8 provides the circuit requirements to support both processing and reprocessing. Circuit requirements in Tables 4-7 and 4-8 which exceed T1 capacity (1.544 Mbps) are indicated in *italics*, while requirements which exceed T3 capacity (44.736 Mbps) are indicated in ***bold-face italics***.

Table 4-7. Circuit Requirements for Processing with Subsetting

		Mbps			
From	**To**	**1997**	**1998**	**1999**	**2000**
EDC	**GSFC**		*1.7*	*1.7*	*1.7*
	LaRC	0.01	0.01	0.01	0.01
GSFC	**EDC**		*8.3*	*8.3*	*8.3*
	JPL			<0.01	<0.01
	LaRC	0.3	*20.2*	*22.0*	*23.0*
	MSFC	*3.1*	*3.1*	*3.1*	*3.1*
	NSIDC		0.7	0.7	0.7
JPL	**GSFC**			0.01	0.01
	NSIDC			<0.01	<0.01
LaRC	**EDC**		0.3	0.3	0.3
	GSFC			*2.7*	*2.7*
MSFC	**GSFC**	0.05	0.05	0.05	0.05
	LaRC	0.3	0.3	0.3	0.3
NSIDC	**EDC**		0.08	0.08	0.08

Table 4-8. Circuit Requirements for Processing/Reprocessing with Subsetting

		Mbps			
From	**To**	**1997**	**1998**	**1999**	**2000**
EDC	**GSFC**		*1.7*	*3.5*	*5.2*
	LaRC	0.01	0.02	0.04	0.04
GSFC	**EDC**		*8.3*	*16.5*	*24.8*
	JPL			<0.01	0.01
	LaRC	0.3	*20.6*	*44.3*	***68.9***
	MSFC	*9.3*	*9.3*	*9.3*	*9.3*
	NSIDC		0.7	1.4	*2.0*
JPL	**GSFC**			0.01	0.02
	NSIDC			<0.01	<0.01
LaRC	**EDC**		0.3	0.7	1.0
	GSFC			*2.7*	*5.3*
MSFC	**GSFC**	*6.3*	*6.3*	*6.3*	*6.3*
	LaRC	0.3	0.6	1.0	1.0
NSIDC	**EDC**		0.08	0.2	0.3

Figure 4-4 shows the summed quarterly flows between all DAACs, with various combinations of processing and reprocessing, with and without subsetting. The primary objective of subsetting is to reduce the cost of the DAAC-to-DAAC data transfers. The cost per unit of data transferred depends on the distance of the transfer, the size of the link (*i.e.*, T1 vs. T3), and the year in which the transfer occurs. The summed transfer rates in Figure 4-4 ignore these factors, but still give some indication of the potential savings.

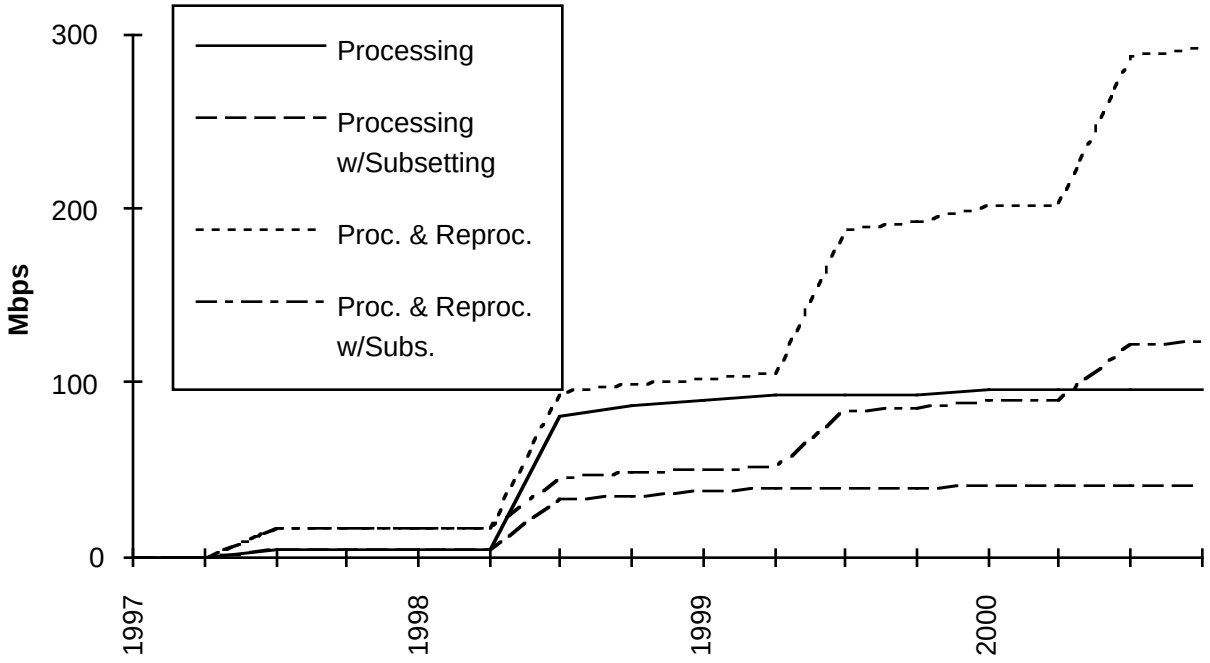

Figure 4-4. Summed Circuit Requirements

5. Implementation Options

Options available for reducing DAAC-to-DAAC data transfer requirements include:

1. subsetting radiance data at the sending DAAC to create smaller files containing only those spectral bands required for processing at the receiving DAAC;
2. subsetting swath data at the sending DAAC to create smaller files providing only the swath coverage required for processing at the receiving DAAC;
3. compressing files at the sending DAAC to remove fill pixels;
4. utilizing a geographical mask to insert fill values for pixels at locations which will not be processed, and then compressing these files at the sending DAAC to eliminate the fill pixels; and
5. using remote file access to extract the desired data under control of the receiving DAAC.

The DAAC-to-DAAC transfer requirements can also be reduced by re-allocating processing responsibilities among the DAACs to bring processing to the DAAC with the largest fraction of the required input data; and/or by re-allocating archiving responsibilities among the DAACs such that the processing DAAC is also the archive DAAC. Re-allocation of DAAC responsibilities involves programmatic issues, as well as technical/cost issues, which are beyond the scope of the present paper.

In Sections 5.1 through 5.5, we consider some of the implementation issues involved in the traffic reduction strategies enumerated above. In Section 5.6, we provide recommendations for steps to achieve the savings and/or further analysis to reduce the uncertainties associated with DAAC-to-DAAC data transfers.

5.1 Spectral Band Subsetting

As described in Section 4.1, subsetting the MODIS Level 1B data to include only those spectral bands required for CERES and MISR processing can reduce the GSFC-to-LaRC traffic to approximately 58% of the original volume. Subsetting by bands will be a standard Data Server process at ECS Release B, so this will not require any additional software development. This option will require some re-allocation of processing and data handling hardware among the DAACs. The major effect will be to reduce the inter-DAAC network costs.

5.2 Swath Subsetting

Further reduction in the GSFC-to-LaRC traffic can be achieved by geographically subsetting the MODIS swath data to match the relatively narrow swaths of the MISR and MOPITT instruments. This type of subsetting is not currently a standard Data Server function.[8]

[8] At Release B, the ECS Data Server will support subsetting of "point" and "grid" data within a polygonal region. For swath data, along-track subsetting will be supported, based on

Implementing a precise cross-swath subsetting for the MODIS data would require a detailed understanding of the MODIS data structures. In particular, this would require reading a MOD03 geolocation data granule and searching for the pixels within the specified geographic region, extracting the corresponding pixels out of the parameter granules, and constructing new granules for both the parameter data and the geolocation data. This would create a new geolocation granule for each subsetting operation, which would partially offset the data volume reduction.

If it is permissible to specify the swath subsetting as including a fixed range of pixels from each scan, the subsetting can be accomplished by using the existing Hierarchical Data Format (HDF) library. The pixel range would have to be specified liberally to ensure that the desired geographical range is included. This type of swath subsetting would involve relatively minor software development, and should be able to achieve most of the desired reductions in the DAAC-to-DAAC traffic. Acceptability of this option will need to be worked out with the MISR and MOPITT instrument teams.[9] Here again, this will require some re-allocation of processing and data handling capacities to accommodate the subsetting. The major effect will be a reduction in the inter-DAAC network costs.

5.3 Compression for Fill Values

As noted in Section 4.2, the MOD09, MOD11, MOD13, MOD14, and MOD41 data will include fill values for pixels falling over water areas. Eliminating these fill pixels can reduce the volumes of these data sets by 70%. This can be implemented through a data compression algorithm which eliminates repeated values.

5.4 Geographical Masking and Compression

For the remaining data products listed in Table 4-4, a land-sea mask would be required to achieve the volume reductions cited in Sections 4.2. For the MODIS Level 1 and Level 2 products, this will require comparing the pixel-by-pixel geolocation data (MOD03) with a land/sea demarcation file such as the Digital Chart of the World (DCW), and then nulling the corresponding values in the parameter granules. This development will require a detailed understanding of the MODIS data structures.

Additional reductions in the DAAC-to-DAAC traffic can be achieved by masking the snow/ice data exchanged between the GSFC and NSIDC DAACs, as described in Section 4-4. This will require development and/or identification of a suitable snow-ice mask, similar to the DCW land/sea data.

5.5 Remote File Access

A remote file system allows remote processes to access files on a disk as if the files were locally available. When partial files are needed, a remote file system may transfer less data, as

---

complete rows of the included scans. Subsetting the wide MODIS swaths to match the narrower MISR and MOPITT swaths will require a different subsetting algorithm.

9 The MISR team has indicated that this option is acceptable. We have not yet discussed this with the MOPITT team.

compared to transfer of entire files to local disks. This latter capability is of potential interest for reducing DAAC-to-DAAC traffic.

Remote file access transfers (caches) data in blocks. In the case of an application which needs access to contiguous data elements, block transfer reduces the number of read/write operations, thereby improving performance. When an application needs to subsample the file, this can actually increase the network load. For example, assume that the instrument team for a multi-spectral sensor (*e.g.*, ASTER or MODIS) elects to store all bands contiguously for each pixel (band-interleaved), and that a user/process wants to extract images at a single band. This would involve sampling 1 data element out of every n data elements, where n is the number of bands (14 for ASTER, and 36 for MODIS). Since each sampling would transfer a block of the neighboring data elements (bands), the result could be an increase in the network load, by comparison to a local subsetting operation which extracts precisely the desired band(s) and transfers the subsetted data.

The DAAC-to-DAAC transfers required to support Standard Product processing and archiving involve repeated transfers of identical subsets. For this type of application, it would appear that explicit subsetting at the sending DAAC will always be at least as efficient as remote file access. Subsetting does lead to additional files at the sending DAAC. This should be more than offset by the reduced residence time of the full data granules achieved through subsetting, as opposed to the case of remote file access where the full granules need to be maintained on disk throughout the data access period.

5.6 Summary

Section 4 considered options to reduce the DAAC-to-DAAC traffic by examining the traffic involved in each link and identifying appropriate data volume reduction techniques. Table 5-1 summarizes the subsetting/masking techniques, the associated links, and the implementation issues.

Table 5-1. Volume Reduction Techniques

Technique	Applies to	Implementation Issues
Subsetting by spectral **Band**	GSFC-to-LaRC	Hardware allocation
Subsetting by **Swath** width	GSFC-to-LaRC	Special algorithm (software); hardware allocation
Compressing fill values	EDC-to-GSFC GSFC-to-EDC GSFC-to-NSIDC	Algorithm selection; hardware allocation
Geographical **Mask**	GSFC-to-EDC LaRC-to-EDC	Software development; hardware allocation

Figure 5-1 shows the summed DAAC-to-DAAC circuit requirements assuming that each of these potential volume reduction strategies are, in turn, adopted (*i.e.*, reductions in data transfer

volumes in Figure 5-1 reflect the cumulative effects of applying the reduction techniques, in the order listed in Table 5-1).

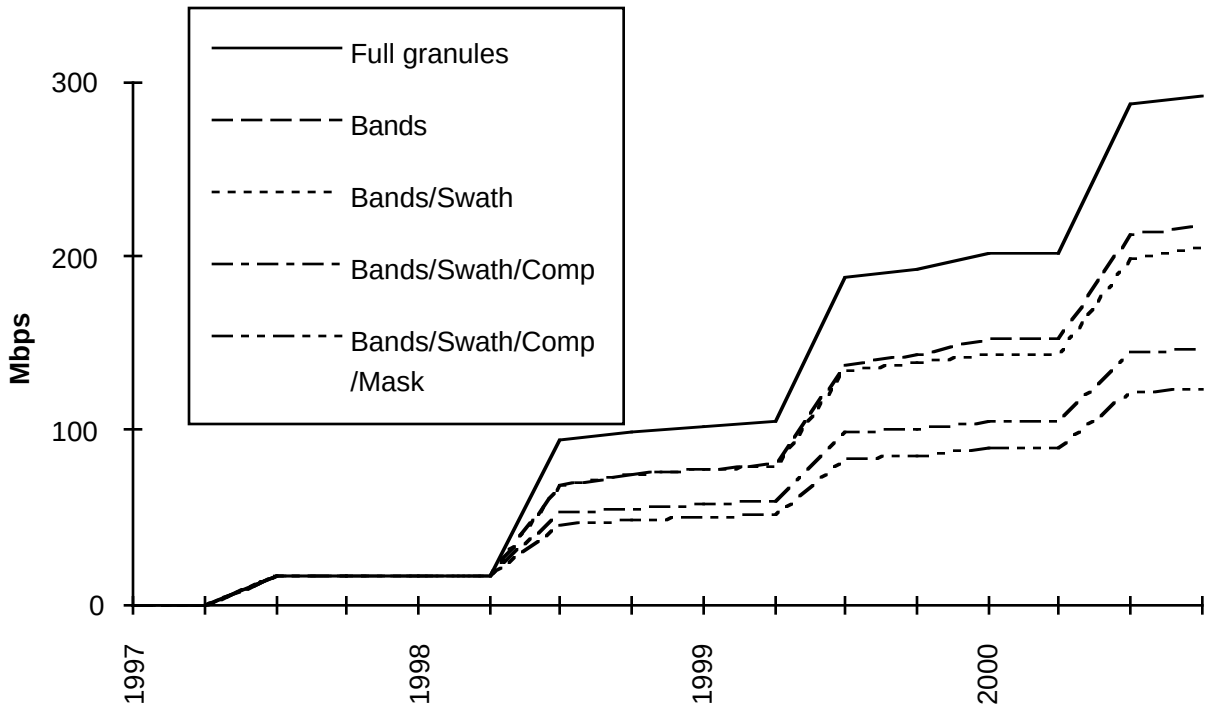

Figure 5-1. Cumulative Effects of Subsetting/Masking (per Table 5-1).

Figure 5-1 shows that the greatest leverage can be achieved by subsetting the MODIS Level 1B data by spectral bands (26% reduction in total network transfers for the year 2000) and by compressing the MODIS Level 2 and 3 land products to remove fill pixels over ocean areas (an additional 20% reduction in total network transfers for the year 2000). Applying all four methods can reduce network transfers by 57% for the year 2000.

There may be other opportunities for reducing network transfers. For instance, many of the Level 2 and 3 products contain multiple parameters. If only a subset of these parameters are required as inputs for processing at the receiving DAAC, then the DAAC-to-DAAC traffic can be further reduced by sending only the required parameters. However, this would not apply to the MODIS Level 1B data required to support processing at the LaRC DAAC, or to the MODIS Level 2 land products transferred for archiving at the EDC DAAC. Since these large products dominate the transfers, additional savings achievable through subsetting smaller products are likely to be marginal.

We have not attempted to estimate the potential savings from loss-less data compression, as normally applied to image data. This may offer significant additional volume reductions, depending on the nature of the data and the compression technique(s). Test data sets (in the same

format as the actual data, and with simulated values representing different types of scenes) will be required to identify the most cost-effective compression techniques and the resulting network savings.

Reprocessing assumptions (described in Section 3.2.2) are major drivers in the estimated DAAC-to-DAAC transfers. Reducing the uncertainties associated with reprocessing will require that the instrument teams develop specific reprocessing plans. However, much of the reprocessing will be driven by improved understanding of the scientific data, which cannot be precisely predicted.

Finally, the DAAC-to-DAAC traffic discussed in this paper only covers the processing and archiving of data from the EOS instruments on platforms scheduled for launch before the PM-1 platform. Instruments on PM-1 (to be launched in December 2000) are: AIRS/AMSU/MHS, CERES, MIMR, and MODIS. At the least, we can expect that the MODIS Level 1B radiances will be required as input to the CERES processing. If DAAC responsibilities remain the same as for the AM-1 versions of these instruments, we can expect a significant increase in the GSFC-to-LaRC traffic. Subsetting will be applicable to reducing these flows.

Abbreviations and Acronyms

ADEOS	Advanced Earth Observing System (Japan)
AHWGP	Ad Hoc Working Group on Production
AIRS	Atmospheric Infrared Sounder
ALT	altimeter
AM-1	EOS Morning Crossing (Ascending) Mission -1
AMSU	Advanced Microwave Sounding Unit
ANC	ancillary
ASF	Alaska SAR Facility (DAAC)
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BRDF	bi-directional reflective distribution function
BT	brightness temperature
CERES	Clouds and Earth's Radiant Energy System
DAAC	distributed active archive center
DAS	Data Assimilation System
DB	database
DCW	Digital Chart of the World
DEM	digital elevation model
DFA	dual frequency altimeter
ECS	EOSDIS Core System
EDC	EROS Data Center
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EPA	Environmental Protection Agency
ESN	EOSDIS Science Network (replaced by EBNet)
FNOC	Fleet Numerical Operations Center (U.S. Navy)
GB	gigabyte (10^9)
GPCC	Global Precipitation Climatology Centre, Frankfurt, Germany
GPI	Global Precipitation Index

GSFC	Goddard Space Flight Center
GV	ground validation data (TRMM)
HDF	hierarchical data format
IRD	interface requirements document
JPL	Jet Propulsion Laboratory
km	kilometer
λ	lambda
LaRC	Langley Research Center
LIS	Lightning Imaging Sensor
L0	Level 0 (data)
L1	Level 1 (data)
L2	Level 2 (data)
L3	Level 3 (data)
m	meter
μ	micron
MB	megabyte (10^6)
Mbps	mega bits per second
MHS	Microwave Humidity Sounder
MIMR	Multifrequency Imaging Microwave Radiometer
MISR	Multi-Angle Imaging SpectroRadiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MOPITT	Measurements of Pollution in the Troposphere
MSFC	Marshall Space Flight Center (DAAC)
MTPE	Mission to Planet Earth
NASA	National Aeronautics and Space Administration
Nascom	NASA Communications
NMC	National Meteorological Center (NOAA)
NOAA	National Oceanic and Atmospheric Administration
NOLAN	Nascom Operational Local Area Network
NSIDC	National Snow and Ice Data Center
PDR	Preliminary Design Review

PGE	Product Generation Executable
PM–1	EOS Afternoon Crossing (Ascending) Mission -1
PR	Precipitation Radar
res	resolution
SAGE III	Stratospheric Aerosols and Gas Experiment III
SAR	Synthetic Aperture Radar
SDPF	Sensor Data Processing Facility (GSFC)
SSI&T	Science Software Integration & Test
SSM/I	Special Sensor for Microwave/Imager
SWIR	Short Wavelength Infrared
SWS	SeaWinds Sensor
TIR	thermal infrared
TMI	TRMM Microwave Imager
TOMS	Total Ozone Mapping Spectrometer
TP	technical paper
TRMM	Tropical Rainfall Measuring Mission (joint US-Japan)
TSDIS	TRMM Science Data and Information System
UARS	Upper Atmosphere Research Satellite
VIRS	Visible Infrared Scanner (TRMM)
VNIR	visible and near infrared
WAN	wide area network