11.5 PROCESSING TERRA AND AQUA MODIS WITH NANOOK, THE NOAA/NESDIS NEAR-REAL-TIME SYSTEM

Paul Haggerty* and Kristina Sprietzer
Science and Technology Corporation, Suitland, Maryland

Gene Legg
National Oceanic and Atmospheric Administration, Suitland, Maryland

1. INTRODUCTION

NANOOK, the near-real-time MODIS processing system was conceived as a proof of concept program and implemented under the Computer Sciences Corporation (CSC) Central Satellite Data Processing (CSDP) contract. It was designed to provide NOAA with rapid access to MODIS (Moderate Resolution Imaging Spectroradiometer) environmental data from the Terra and Aqua satellites. Additionally NANOOK provides risk reduction for future high data rate satellite systems by identifying the risks and problems associated with high data volumes in near-real-time operations.

2. DIFFERENCES BETWEEN AQUA AND TERRA

The MODIS instruments on the Aqua and Terra spacecraft are near duplicates of each other. However, there is one major difference between the two platforms that has required a significant modification to the processing system. While Terra has attitude and ephemeris data imbedded in the data stream for use in navigation, Aqua lacks this ancillary information. Instead, NASA's Flight Dynamics Group computes predicted satellite position data and uploads it to the spacecraft once per day. This data is then returned to the ground stations as another data product called GBAD (Ground Based Attitude Determination) which is time coordinated with the other instruments. As an additional complexity, GBAD is a Rate Buffered Data stream product which must first be matched up with a separate set of orbital elements and processed into a separate ancillary product.

As with all ancillary products in a near-real-time system, this raises the problem of ensuring that the navigation data is available before attempting to process the MODIS data. Fortunately, NASA typically transmits the smaller products first, thus usually ensuring that the small GBAD products have been received and processed long before the much larger MODIS data stream can be ingested. In cases where processing still fails to deliver the required navigation data, the MODIS data can simply be restaged to be tried again after a short (1-2 minutes) pause for the navigation to catch up.

* Corresponding author address: Paul Haggerty, STC, NOAA/NESDIS/IPD Federal Center FB-4, 4401 Suitland Road, Suitland, MD 20746;
e-mail: Paul.Haggerty@noaa.gov

3. NANOOK IN A NUTSHELL

The NANOOK system is, at its core, about processing the most data in the least time. In this paper, we will discuss input data concerns, both in the structure of the Level 0 data and in how NASA's EDOS (EOS Data Operations System) gets the data to NANOOK. We will review the processing strategies for producing output and improvements gained from the use of new processing equipment. Finally we will mention some concerns related to the distribution of the data, based both on network and storage performance.

3.1 *Level 0 Data*

Version 1.0 of the program to converter NASA MODIS Rate Buffered Data to NOAA Level 0 Data was constructed with several assumptions in mind. These assumptions were based on preliminary talks with NASA engineers, long before the spacecraft was launched, and enhanced by experience gathered during the post-launch checkout phase. As with most assumptions, these proved to be less than accurate based on the data from the first few years of the project. Therefore, a new version of the converter needed to be designed to solve these problems. For a detailed description of the problems encountered, and the solutions developed, please refer to Ivan Tcherednitchenko and Paul Haggerty's Poster session (P1.15), a copy of which is located at: (http://www.osdpd.noaa.gov/MODIS/AMS/)

Initial assumptions indicated that there would be a problem syncing the ground station to the spacecraft. Thus, the ground station would receive playback from the satellite, reposition to the beginning of the solid state recorder, and then receive the first part of the data a second time to ensure that all data was received. In reality, however, the data stream received contains a variable number of rewinds, with occasional data gaps. This more complex than expected data stream caused the converter program to lose track of 20-25 granules of data per day. These granules are gathered from both the beginning and end of the Rate Buffered Files and from internal discontinuities. However, since NANOOK was originally conceived as a near-real-time processor for the continental United States (CONUS) only, this was not seen as an urgent problem. The odds of a missing granule being over the CONUS area were relatively small (6 out of 288) and the perceived improvements to the converter would place an unnecessary delay on the processing times. Furthermore, NASA's official MODIS processing system would have research quality data available in a one to three week time span.

However, shortly after NANOOK went into full production, NOAA received requests from both NASA and the Department of Defense (DoD) for weather data relating to other areas of the globe. With the expected addition of the Aqua spacecraft, for which the NANOOK team was tasked with providing global cloud masks to the AIRS (Atmospheric Infra-Red Sounder) project, the probability of missing important granules would become unacceptable. It was a fore-gone conclusion that NASA and DOD would be interested in data from Aqua

Therefore, a new effort was put forth to redesign the converter program to recapture these non-contiguous datasets. Version 2.0 of the converter incorporated many new checks, as well as a system for saving partial granules and stitching pieces together. As expected, this increased the time required to process data, however, it was not as large a penalty as was previously considered. Algorithms were added to detect the "rewinds" and skip processing of data already handled in previous portions of the data stream. Therefore, the new converter avoided significant chunks of disk I/O. The net result was an increase in latency of only 2-3 minutes on average for a 6 Gigabyte Rate Buffered Data file.

3.2 *EDOS*

NASA's EDOS group has worked closely with the NANOOK team to improve data latency from observation to delivery at the NOAA processing computers. In 2002, NASA tested a change to its distribution system that used two contacts per orbit via the TDRSS (Tracking and Data Relay Satellite System) network. In theory, this changes the delivery pattern from 6GB every 100 minutes, to 3GB every 50 minutes. Thus the oldest data on board the spacecraft gets delivered up to 50 minutes earlier then otherwise expected. As the data approaches the end of the orbit, the time improvement drops off accordingly. Currently, NASA is planning an experiment to stream the data. Under the nominal plan, the ground station links up with the spacecraft and receives a dump of the flight recorder. Once this dump is completed, the data is sent across the internal NASA networks. Under the new plan, once a certain amount of data is received at the ground station (10-15 minutes worth), it is immediately forwarded while the next installment of data continues to be received. This approach continues the improvements realized by the multiple TDRSS contacts and gets the oldest data to NANOOK faster.

3.3 *CPU performance and processing strategy*

In designing NANOOK, one of the founding principles for increasing production was the use of scheduled CPUs. The SGI computer doesn't have a mechanism for specifying which CPU should be used for a particular job. In fact it has built-in optimizers that spread multiple jobs over the available CPUs in order to maximize efficiency. However, it is still important to limit the total number of jobs submitted so as not to overload the limited resources.

To simulate individual addressable CPUs, a set of sixty directories were created, with a small text file set up as a database for control. While there are actually 64 CPUs on the system, four CPUs have been held in reserve for system use, operator use, and other miscellaneous operations that can't be easily quantified.

The twelve PGEs (Program Generation Executables), which comprise the science code that produce the various output products, are then sorted by priority and assigned to the "free" CPU's on a FIFO (first-in, first-out) principle. (See Table 1 for a description of the PGEs). Thus, once all 60 CPU slots have been filled, all future jobs are required to wait for one of the previous jobs to finish and empty a slot.

Table 1:
Program Generation Executables:

PGE	Product Description
1	Level 1A Data Counts and Geo-location
2	Level 1B Radiances
3	Level 2: Cloud Mask, Atmospheric Profiles, Water Vapor, Ozone
4	Level 2: Aerosol, Precipitation
6	Level 2: Cloud Products
7	Level 2: Snow Cover
8	Level 2: Sea Ice
9	Level 2: Ocean Color
10	Level 2: Sea Surface Temperature
30	Level 2: Fire Detection
17	Ancillary: NCEP GDAS data format conversion
19	Ancillary: NCEP OZN data format conversion

The PGEs represented in Table 1 have various priority levels assigned. PGE 17 and 19 have the top priority as they generate ancillary products, run only five times per day, and take minimal resources and time. PGEs 1, 2, and 3 are the next highest priority as these must run sequentially to reach the level 1B products that all the other PGEs require as input. At this stage, all other PGEs can run in parallel, as all required input datasets have already been created. However, the system of prioritizing continues so that the most desired products can be produced in the least amount of time, with other less urgent products following when processing power is available. PGE 30 runs next to generate the fire detection product and imagery. PGEs 4, 7, and 8 run after that since, as seen in Table 2, they take very little time and can be quickly cleared through the system. Finally PGEs 6, 9, and 10 are processed. These three PGEs take the longest time to run and are the least urgent.

In practice, all PGEs are capable of being run between orbital deliveries and latency is minimal for even the least urgent products.

The addition of new processing and storage hardware has also drastically improved performance,

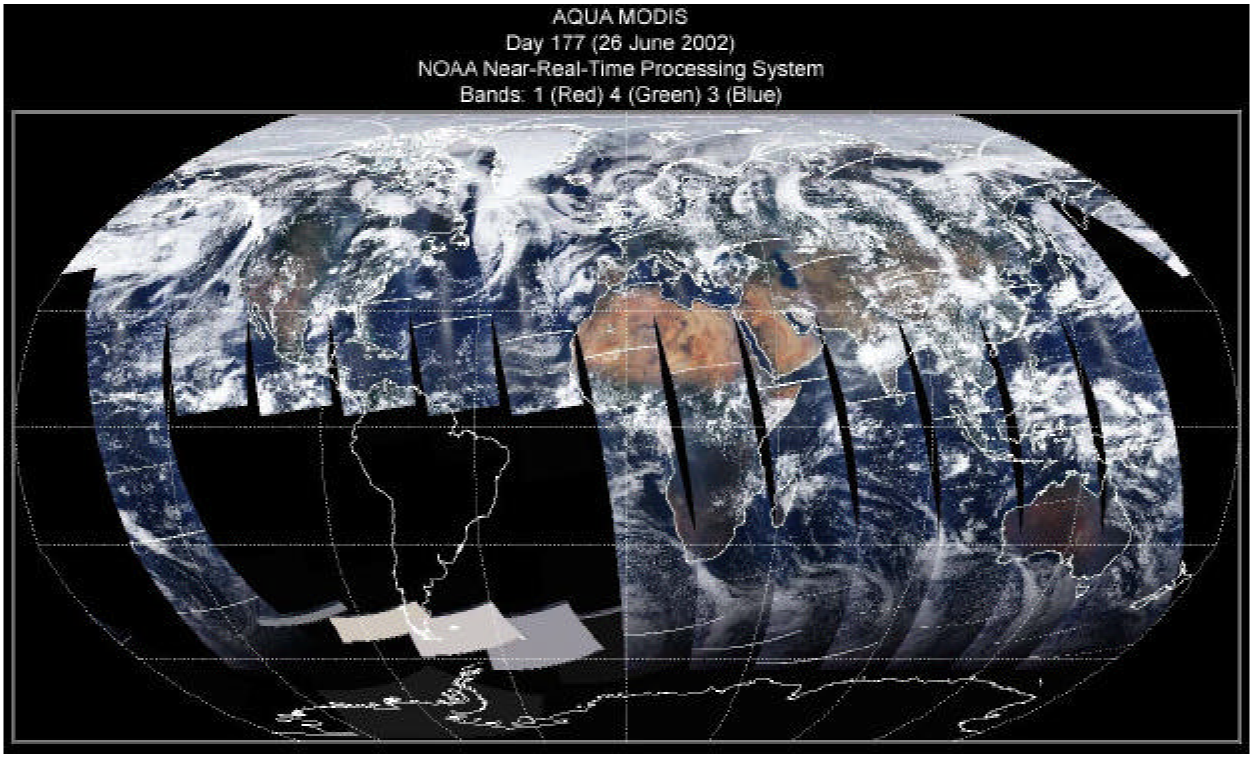

Figure 1: Global Mosaic. Aqua Data June 26, 2002

cutting 30% from our processing times (see Table 2), as well as allowed us to move from performing regional processing for a single spacecraft to performing global processing for two spacecraft. Figure 1 shows a global mosaic created from the first full days of data collected by NANOOK.

In addition to the simple improvement in CPU power alone (250MHz – 450MHz), a significant portion of the decrease in run times can be attributed to the acquisition of a new RAID storage system. The old storage pack had a single controller, which formed a chokepoint for all incoming and outgoing data streams. While the new system has eight controllers, allowing more data to be read and written simultaneously, thus distributing the workload over a greater set of data channels.

In Table 2, the row marked Serial Time represents the total average processing for a single granule (5 minutes) of data if each process were to be run separately. The row marked Parallel Time represents the total average time needed for running PGEs 1, 2, and 3 serially, followed by the others run in parallel. In essence, it's the total runtime of PGEs 1, 2, 3, and 6 (PGE 6 has the longest runtime of the remaining products.)

In general operations, the time is somewhere in between the two extremes as we lack sufficient CPUs to perform all operations in parallel. However, in practice, product generation times tend to average under one hour.

Table 2:
Program Generation Executable Runtimes (Minutes)

PGE	Old System (250 MHz)	New System (450 MHz)
1	8.70	5.16
2	20.00	10.58
3	13.22	2.73
4	4.05	2.65
6	35.02	26.49
7	1.38	3.52
8	1.71	1.42
9	25.15	15.16
10	18.48	13.80
30	0.26	0.19
Serial Time	114.90	81.70
Parallel Time	76.94	44.96

3.4 *Data Distribution*

Once the data has been processed, it is of little use unless it can be transferred to the user community. Data distribution can be every bit as challenging as ingesting and processing the data in the first place.

Ingesting MODIS data was facilitated by collocating the processing computers with NASA's EDOS distribution system. This allows us to receive the approximately 12GB of MODIS data every 100 minutes, at an average rate of 16 Megabits/sec. However, full product processing results in an output data size of nearly double the input data rate. Use of data compression mitigates this problem somewhat. Active tests show that on average the output data sizes can be reduced by 30% using the GZIP protocol. However, there is still a total daily data size of 768 GBs to be stored and transferred. Transfers of data not only require I/O channels, they also bog down CPUs somewhat to manage the traffic. This takes away from the limited, and expensive resources, needed for processing.

The more customers involved, the more I/O and CPU resources must be diverted. In order to limit the impact on our system, an SGI Origin 3200 dual CPU system with attached 5 TB RAID was acquired to handle data distribution. In this fashion, the only transfers the production server would be required to make would be a single push to the distribution server. However, as soon as the system was set up, modifications to the operational scheme needed to be made. Initially, a Gigabit Ethernet network was planned for distribution of the data; however delays in construction of this network forced us to utilize optional paths.

DoD customers had access to a private high-capacity network, but would not allow us to connect our public server to it due to security considerations. Therefore, the production server was still forced to handle data distribution for this set of customers.

Other customers are forced to use the normal Internet channels until the production network can be completed sometime in 2003. In one test over the internet, Level 1B data was transferred for a small area (12 granules total) over the course of a day. The files, averaging approximately 150MB each, required between 3 hours at night, to 11 hours during the day to be transferred. Until the production network is finished, customers will be force to forego raw data and accept the smaller processed products. (See Table 3 for product sizes.)

Table 3.
MODIS data product sizes

PGE	Description	File Size (MB)	Compressed File Size (MB)
1	L1A	450	NA
2	L1B	1000	500
3	Cloud Mask	61	25
4	Atmosphere	12	1.8
6	Cloud Properties	13.5	4.1
7	Snow Cover	32	6.8
8	Sea Ice	13.5	3.1
9	Ocean Color	278	14.8
10	Sea Surface Temperature	33	10
Total		1893	565.5

Setting up a network connection involves more than simply connecting two computers. Once the IP addresses were assigned and the cables connected various groups had to be contacted. For security reasons, these groups had implemented packet filters to prevent unauthorized use of the network. Each of these groups needed to make modifications in order to allow MODIS traffic to flow across the full length of the network.

Finally, the configuration of the destination computer had to be considered. It doesn't help to have a Gigabit Ethernet network if the destination computer can only write to its drives at 10 megabits/sec. Most computers have multiple network connections to maintain, thus splitting the bandwidth among clients. Additionally, the storage system of these servers are typically busy servicing the needs of many applications. In order to properly serve high load datasets, the slowest subsystem in the host computer, destination computer, and all network paths in between will ultimately dictate the speed at which the data can be transmitted. All of these must be optimized and maintained to prevent the data flow from being bogged down.

3.5 *Conclusions*

NANOOK is a continuously evolving system which, over the last few years, has expanded dramatically in both processing power, and client expectations. The system has been updated from serving a small geographic area to global coverage and from one spacecraft to two. Additional products have been added to the roster, and additional customers have come forward. NANOOK will continue to improve its capabilities over the next few years in order to accommodate new customers and newer technology. And finally, lessons learned from the implementation and evolution of the system will be used to help plan the next generation of Earth Observing sensors -- already in production for launch later this decade.