12.3 # THE NOAA MODIS/NEAR-REAL-TIME SYSTEM

Paul Haggerty, Science and Technology Corporation, Suitland, Maryland *
Kristina Sprietzer, Science and Technology Corporation, Suitland, Maryland
Gene Legg, National Oceanic and Atmospheric Administration, Suitland, Maryland
Ray Luczak, Computer Sciences Corporation, Suitland, Maryland

1. INTRODUCTION

There are more than 50 U.S. operational weather satellites orbiting the earth [Kelso 2000]. Although these satellites are maintained by a variety of government agencies, this valuable data resource is difficult to access. Dr. D. James Baker, Administrator of the National Oceanic and Atmospheric Administration (NOAA), and Mr. Daniel Goldin, National Aeronautical and Space Administration (NASA) Administrator, challenged their respective communities to exchange satellite data to improve weather forecasting for high-impact weather emergencies. A new source of high-resolution weather data is now available via the MODIS (Moderate Resolution Imaging Spectroradiometer) instrument aboard the Terra satellite.

Using this data in near real time (NRT) increases the accuracy for weather forecasting and assessing natural disasters. The NOAA MODIS/NRT system will attempt to accomplish this by providing products to the National Weather Service within three hours of data acquisition. Producing data in NRT is not a simple task, considering time dependencies on the receipt of various data sources and the hardware and software constraints when dealing with complex and large volume high-resolution data.

2. THE BAKER (NOAA) AND GOLDIN (NASA) CHALLENGE

At the January 2000 American Meteorological Society Meeting, Dr. D. James Baker, Administrator of the National Oceanic and Atmospheric Administration (NOAA), and Mr. Daniel Goldin, Administrator of the National Aeronautical and Space Administration (NASA), challenged their respective communities to work together to improve weather forecasts for high-impact weather emergencies (such as hurricanes and precipitation) and develop program strategies to advance the U.S. Weather Research Program (USWRP). This collaboration seeks to maximize the utilization of satellite data. To have an impact on improving weather forecasts, it is ideal to have satellite data as soon as it is available, a process referred to as NRT data. Satellite data from QuikSCAT and GOES have already demonstrated that one of the benefits of NRT processing is more reliable predicted data.

Paul D. Haggerty, NOAA/NESDIS/IPD Federal Center FB-4, RM 0303, 4401 Suitland Road, Suitland, MD 20746; e-mail: phaggerty@nesdis.noaa.gov

3. THE MODIS INSTRUMENT ON TERRA

The MODIS sensor is sensitive in 36 different wavelengths of the spectrum, ranging from the visible to the thermal infrared, with spatial resolutions of 250 meters, 500 meters, and 1 KM, depending on the particular band. MODIS has a greater spectral and spatial resolution than previous instruments and is the most modern technological instrument on the Terra satellite. The Terra satellite is a part of the EOS (Earth Observing System) Program, which is designed to catalog many of the complex factors that combine to produce the earth's climate [King et al. 2000]. Another MODIS instrument will be launched on the Aqua Satellite, currently scheduled for May 2001.

4. NRT MODIS IMAGING OF MONTANA FOREST FIRES SEPTEMBER 2000

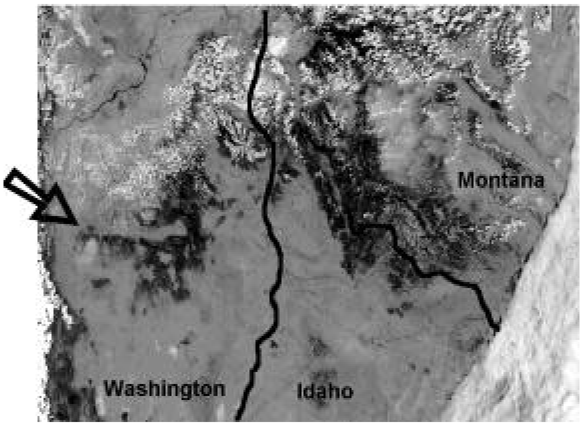
Although the MODIS instrument on Terra was designed to be used by scientists to understand the Earth's climate system in detail, it has also already been shown to be applicable to the average US citizen by being instrumental in helping to save lives and property during a national disaster.

During the weeks around Labor Day 2000, widespread wildfires burned through Montana and Idaho. Flights by reconnaissance planes were hazardous due to the heavy smoke from the fires, causing the Fire Lab of the National Forest Service to rely heavily on satellite imagery. As the MODIS sensor is sensitive to both visible and infrared radiations, it has the capacity to detect hot flames on the surface even through smoke and clouds. In addition, the MODIS instrument resolution (0.25 km) is greater than traditional AVHRR (Advanced Very High Resolution Radiometer) instrument data (1km). This resolution helped the firefighting command center monitor smoke dispersion, which is a critical issue for the health of those living in the area. The production of NRT MODIS data proved to be extremely useful to the Fire Response Team in reevaluating wildfire management on a daily basis (see Image 1.)

5. DESIGN OF THE NOAA MODIS/NRT

The NOAA MODIS/NRT system was designed under the Computer Sciences Corporation (CSC) Central Satellite Data Processing (CSDP) contract for the NOAA/NESDIS Information Processing Division (IPD). The original design of processing software was developed by the NASA Goddard MODAPS (MODIS Adaptive Processing System). MODAPS also processes MODIS data, but not in real time (typical turnaround times are about a week). Due to time constraints in processing, NOAA's NRT system produces fewer products than MODAPs produces, and they may differ slightly due to using predicted rather than actual ancillary data sets. The operation of NOAA MODIS/NRT involves the collaboration with EDOS (EOS Data and Operations

Image 1: Location of Fires (Montana/Idaho/ Washington) 30 August 2000, 18:50 UTC



System) for satellite data and NASA Goddard DAAC (Distributed Active Archive Center) for input ancillary data products.

A number of obstacles affect production of NRT data from MODIS. These obstacles result primarily from NASA's initial concept, which was to develop a system with a one-week turnaround time in mind. This concept is slanted towards bulk processing, with short-term storage of data to magnetic tape being used until all data is available, so that the system may run in the most efficient manner. This system is unrealistic for NRT processing, given NOAA's turnaround requirement of three hours for core products.

The obstacles faced in the NOAA MODIS/NRT system are the time requirements in executing the PGEs (Program Generation Executives) that create the products, the latency times in receiving raw and ancillary data, and the hardware constraints of local storage space and CPU availability.

However, for every obstacle there are solutions and compromises that overcome the problems. Changing the scope of the system was the first and foremost necessary change. With the currently installed hardware, it is not possible to be all things to all people, and therefore we began by reducing both the number of products to be produced and the area of interest.

The NOAA MODIS/NRT system is considered for "first look" use only. We have intentionally sacrificed a small amount of quality in exchange for speed. Visually, the reflectance images are the same; however, as our processes differ from NASA's stringent quality control, some small deviation is expected. In-house comparisons of NOAA- and NASA-produced Level1B datasets show a difference in radiance values of less than 0.5 percent on average.

5.1 Execution Times for PGEs (Program Generation Executives).

As NASA has a large system with more CPUs than are required to keep up with incoming data, it is not necessary for them to fully process incoming data in a linear mode. Multiple products and multiple datasets can be run simultaneously; older datasets can be reprocessed in case of failure. The NOAA MODIS/NRT system has more limited processing capacity and therefore must process all available data in a more rapid fashion. As the time to run a particular PGE is not adjustable, we have simplified our system by limiting the number of PGEs we process, and therefore the number of products produced. The NOAA MODIS/NRT system is oriented around producing a Level1B product and a cloud mask within a three-hour turnaround time, and Cloud Properties, Atmospheric Properties, Snow Cover, Sea Ice Extent, Ocean Color, and Sea Surface Temperature within 12 hours of observation. (See Table 1 for sample run times.)

Table 1. PGE Storage Requirements and Processing Run Times.

PGE	Description	Storage (MB)	Processing Run Time
	Raw to L0*	170	10 minutes
1	L1A*	450	10 minutes
2	L1B*	1000	20 minutes
3	Cloud mask*	80	10 minutes
4	Atmosphere	25	5 minutes
6	Cloud Properties	70	25 minutes
7	Snow Cover	23	5 minutes
8	Sea Ice	20	2 minutes
9	Ocean Color	800	15 minutes
10	Sea Surface Temperature	180	5 minutes

(* indicates product with a 3-hour requirement)

5.2 Receipt of Ancillary Data Sets.

The MODIS processing system, as designed by NASA, requires a number of "first guess" solutions provided by ancillary data sets that are created using the latest satellite data from NASA and NOAA weather satellites. These "first guesses" are then refined by the processing software using the newly acquired MODIS data. Unfortunately, the latest complete data sets are no more recent than 24 hours old. Using these datasets as is means that under the standard processing environment, today's data cannot be processed until at least tomorrow.

While ancillary data created from the latest real data is preferred, and therefore used in the standard model, in order to produce products in near real time, we have turned to forecast ancillary sets. In practice, there is very little physical difference between the two. Instead of producing today's "first guess" data by analyzing yesterday's measurements; forecast files create the "first guess" data for tomorrow from yesterday's data. Data thus created is merely 48 hours old rather than 24 hours old.

For example, the Geolocation code requires today's orbital ephemeris data from the satellite in order to reference each pixel with its ground coordinates. However, assuming that nothing drastic happens to the satellite and that no orbital maneuvering needs to be performed, this data can be forecast weeks into the future with little to no deviation from the actual values.

5.3 Receipt of Raw MODIS Data.

MODIS data is first received at a NASA ground station via either direct download or via relay by TDRSS (Tracking and Data Relay Satellite System) satellite. Once at the ground station, each two-hour block of data, approximately 6 GB in size, is packaged by hardware formatters and forwarded to EDOS, which strips the MODIS data out of the entire satellite feed and feeds this as Rate-Buffered Data to the NOAA processing system co-located at the Goddard Space Flight Center in Greenbelt, Maryland. This process can take between 120 and 140 minutes, thus leaving something less than one hour for NOAA processing (See Figure 1).

Unfortunately, there is also no way to speed up delivery of the raw data to the NOAA MODIS/NRT processing system, as all users receive the data at the same time. However, as delivery via the TDRSS satellites is performed every 50 minutes, this mode can reduce delivery times by up to an hour and is therefore the preferred method. Unfortunately, this is not under our control.

NOAA's interest in the first-look MODIS data is limited to the United States only, rather than the entire planet. We are therefore able, via the satellite tracking program N3EMO [Berger 1990], to determine a priori which granules of data will contain the required data sets. This determination reduces the number of granules that need to be ingested into the system for processing from 288 in 12 orbits to approximately 35 granules within 6 orbits.

In the standard NASA model, each two-hour block of data is fed into the system and converted into 24 granules of five minutes each. Because the NOAA MODIS/NRT system has a much smaller area of interest that NASA, we disassemble the raw data during conversion into the base L0 data and ingest only the few granules that are of interest.

5.4 System Requirements.

In order to limit possible incompatibilities in porting the MODIS processing system, NOAA chose to implement the NOAA MODIS/NRT system on identical hardware running the same operating system version and with identical compilers and utilities. Changes to the code were only made in order to localize the system to the NOAA machine (IP address, Machine Name, Etc.), and in carefully selected areas in order to improve run time.

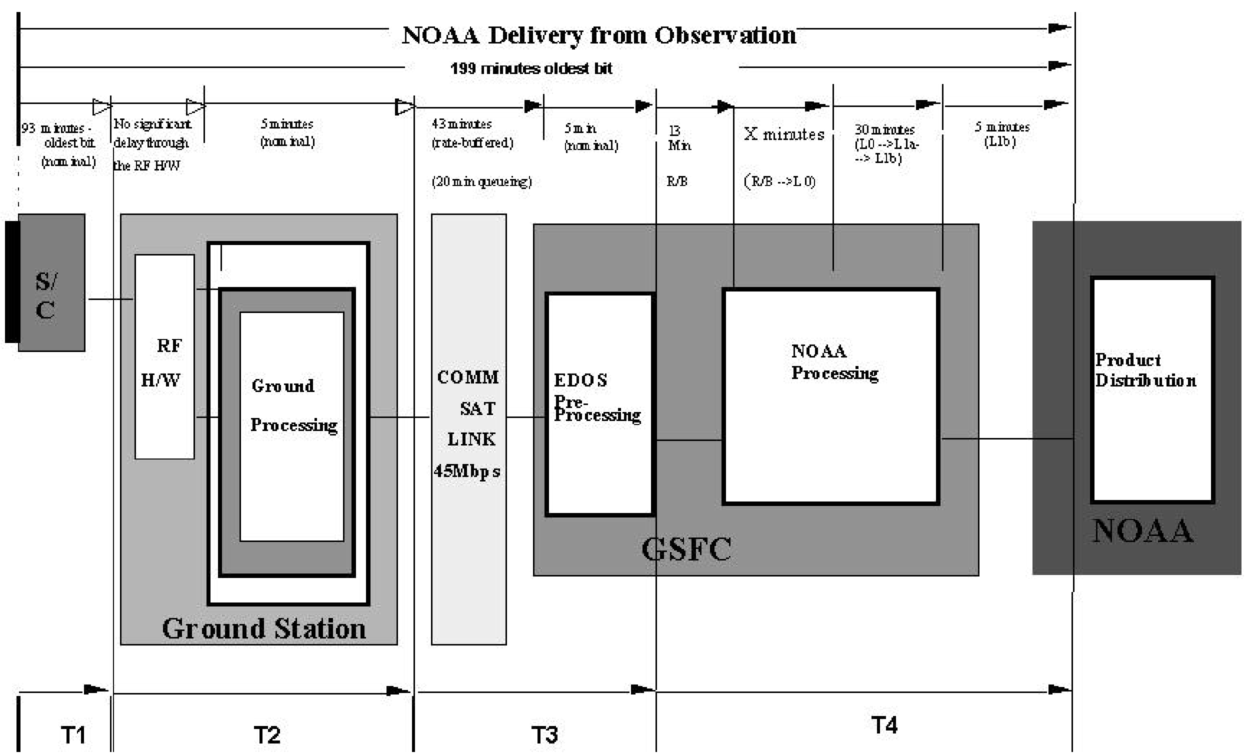


Figure 1: NRT Data Delivery Timeline (Ground Station Scenario)

However, NOAA's MODIS/NRT processing system hardware still differs from NASA in several respects. These modifications leave no opportunity for reprocessing in the case of failure. Any data that is not successfully processed in the first attempt must be dropped from the system and ignored.

5.4.1 Storage Capacity.

The NOAA MODIS/NRT computer system has no off-line tape storage system. It is necessary to keep all raw data, intermediate products, and final products located in local storage. NASA has listed 42 products and 51 variants, which accounts for a total daily volume of 590 GB [EOS Data Products Database Query]. This number is a rough approximation given that each PGE will produce files of differing sizes depending on time of day and geographical location (ex. there is less snow and ice at the equator than at the poles, and little need for visible light data during the night).

The NOAA MODIS/NRT storage system consists of a RAID disk system with 270 GB of storage. In addition to the data processing areas, this storage needs to store approximately 30 GB of Static Data, Ancillary Data, and processing logfiles. Having chosen only the United States (including Alaska and Hawaii) as our area of interest, reduces our incoming data from 288 granules over 12 orbits to 35 granules over 6 orbits per day. Furthermore, reducing the number of products to those given in Table 1., results in a total data volume of 90 GB per day on average. Again, the peak volume can be somewhat higher depending on circumstances. In order to accommodate this volume, each intermediate product is deleted as soon as it is no longer required, and each final product is transferred to the NOAA CEMCSC (Central Environment Satellite Computer System) and then deleted as soon as possible. The processing system has no available resources for archiving or for reprocessing products. Each product has one chance to be created. If processing fails for any reason, that granule is abandoned and deleted.

5.4.2 Processing Capacity.

A day's worth of data consists of 12 two-hour orbits, with each orbit consisting of 24 five-minute granules of data. In order to keep up with this influx of data, it is necessary to have enough CPUs to be able to fully process all required data granules in an hour's time, which in practice requires several CPUs for each data granule.

The NOAA MODIS/NRT processing system consists of 12 (250 MHZ) CPUs. As each orbit of MODIS data is comprised of 24 granules, it is necessary to highly subset the data to most efficiently exercise all CPUs. Given the localized nature of our area of interest, no more than four CPUs are needed for any given orbit, thus giving us three CPUs for each granule so that once basic level 1B processing has been completed, the remaining products can be produced in parallel. In contrast to the NASA model, which processes two hours worth of raw data in a linear model, the NOAA system processes each five-minute granule in parallel, thus reducing this first critical step by a factor of 24. Furthermore, with fewer orbits per day to cover, there is additional downtime between active orbits to complete processing of non-core products such as ocean color and sea surface temperature. As can be seen in Table 1, additional work must still be performed to bring all the processing times within requirements.

5.4.3 Bandwidth.

The MODIS processing system requires a large amount of data to operate and in turn generates an even larger amount of data in the form of products. All this data needs to be transmitted over internal and external networks, which must be large enough to handle the unusually high volumes. Even with a data volume reduction inherent in our streamlined system, each day 72 GB of raw data are transmitted to the processing system and over 90 GB of final data products need to be transferred to the CEMSCS storage system. Spreading transmission out over the day still requires a 15 MB/sec network line, which is far above standard internet capacities. In order to accommodate this, NOAA and NASA are jointly installing a new Gigabit Ethernet network linking the NOAA facility in Suitland, MD and the Goddard Space Flight Center in Greenbelt, MD. This network is designed to accommodate a wide variety of new satellite systems scheduled for the near future.

6. Conclusions.

Although there are many obstacles in creating a NRT system, the solution to these problems lies in optimizing processing based on essential priorities, using the latest technological advances and the cooperation of multiple government agencies to supply data.

The Baker (NOAA) and Goldin (NASA) challenge shows an interest in facilitating intercommunication of government agencies and a need for near-real-time data. The importance of this challenge, in aiding the creation of a NRT system, was shown to be invaluable during a natural disaster such as Montana wildfires. Weather recording instruments such as MODIS, that were originally designed to be used for long-term weather surveys, have been shown to be applicable in NRT surveys to help save lives and property during a national disaster.

7. References:

Berger, R. W., 1990: N3EMO Orbit Simulator. ftp://ftpamsat.org/amsat/software/source/emoat.zip_AMSAT The Radio Amateur Satellite Corporation.

EOS Data Products Database Query: http://spsosun.gsfc.nasa.gov/cgi-bin/eos-ksh/product.ksh/inst_name=MODIS EOS Science Processing Support Office.

Kelso, T. S., 2000: CelesTrak WWW two line element weather satellite data file. http://celestrak.com/NORAD/elements/weather.txt__CelesTrak.

King, M. D., and Herring, D. D., 2000: Monitoring Earth's Vital Signs. *Scientific American*, April 2000, 92-97.