2.6 EVOLUTION OF NANOOK, THE NOAA/NESDIS NEAR-REAL-TIME MODIS PROCESSING 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

NANOOK is a combined scheduler and processing system for MODIS (Moderate Resolution Imaging Spectroradiometer) data. NANOOK is a hybrid system containing both a task scheduler and processing control system developed under the CSC Central Satellite Data Processing Center (CSDPC) contract, and a suite of NASA-built product generation software.

NANOOK accepts Rate Buffered Data from EDOS (Earth Observing System Data Operations System) and converts it to the Level 0 data format required by MODAPS (MODIS Adaptive Processing System) science software packages. The scheduler, as written by NOAA, is then in charge of managing all incoming and ancillary data, passing the data through the various modules of the science code (and additional NOAA produced production code), and proper distribution of all output products.

In the year since the NANOOK went on-line there have been a great number of changes to the system, both in capabilities, and in requirements. There are also a number of problems that still need to be overcome to effectively utilize the system.

2. THE OLD NANOOK

When NANOOK originally went on line in December of 2000, it was tasked with producing a limited set of products over the continental United States, plus Alaska and Hawaii. It was developed on an SGI ORIGIN 2000 computer attached to a Clarion RAID unit with 550GB of storage, 275GB of which were allocated to NANOOK, the remainder being allocated to AIRS (Atmospheric Infrared Sounder), a related NOAA project which is also engaged in proof-of-concept and operational demonstration.

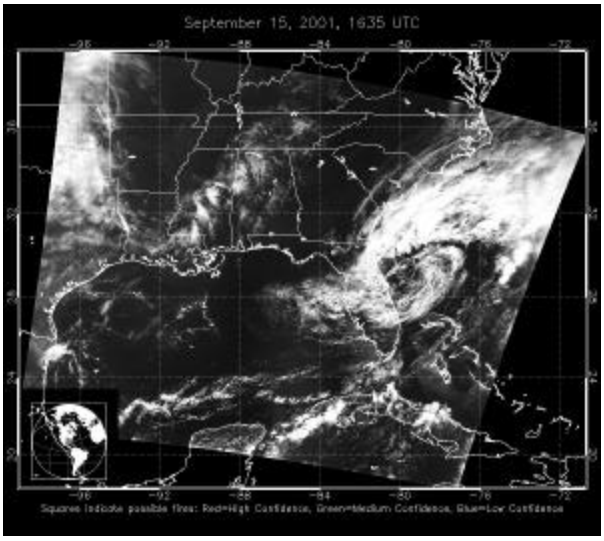
The RAID held two copies of all original data, intermediate data, ancillary data, final products, and processing code. Storage requirements were then doubled by the implementation of two separate but parallel development and production processing systems. The developmental system allowed for new software to be tested without disrupting the flow of data.

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

Twelve CPUs (from a total of 32) running at 250MHz were made available to process data for the designated area of coverage. The remainder of the CPUs were allocated to AIRS.

In order to allow for parallel processing, MODIS data was broken up into granules that consist of five minutes of contiguous data. Image 1 below shows the coverage of a typical granule. The limited daily coverage requirement resulted in approximately 18-20 of these granules for the continental US, 6-10 granules for Hawaii, and 6-10 granules for Alaska. Fortunately, since TERRA is a polar orbiting spacecraft, the incoming data load was limited to approximately 3-4 granules per orbit, which allowed for the small number of CPUs to keep up with the data flow.

IMAGE 1. RGB composite
Sept. 15th, 2001 : 1635 UTC
Southeastern United States



On a typical day, this data load averaged approximately thirty-five granules. Since each granule generated two gigabytes of products in each of the development and production systems, the storage system very quickly filled to capacity. This necessitated purging the data and the products immediately after being processed. Unfortunately, this left little to no opportunity to study the data in depth or to go back and reprocess specific areas for more detailed information.

Since this was a pilot project for processing near-real-time data using data ported from its original design environment, the primary concern was installing the software and acquiring ancillary data. The initial investment provided an environment in which data could be processed, but at a limited trial capacity. Although storage capacity concerns were envisioned and expected at the start of this project, there were sufficient resources allocated to show that the project could be successfully tested. And thanks to assistance from NASA, the system has now been expanded not only to fulfill it's original designs, but to assist another NOAA project, and to provide risk reduction for the upcoming NPOESS (National Polar-orbiting Operational Environmental Satellite System) project, and it's precursor NPP (NPOESS Preparatory Project) which will test three of the four major new instruments.

3. THE NEW NANOOK

3.1 Physical Growth of the System

The original physical components of NANOOK consisted of three full sized cabinets, one of which held the RAID, and the other two holding thirty-two CPUs and associated electronics.

Despite using SGIs newly modularized system to reduce space, the expansion to the new system has tripled the floor space and power requirements. The new NANOOK uses an ORIGIN 3800, which has two full sized cabinets for the sixty-four 400MHz CPUs dedicated to MODIS, and a third cabinet for the ORIGIN 3400 with thirty-two CPUs now dedicated to AIRS. Both projects now share a fourth cabinet housing a five-terabyte RAID storage system.

To reduce the volume of bandwidth for data distribution, a new ORIGIN 3200 system was added. This fileserver has five terabytes of RAID storage which will be used for daily data storage and distribution. This brings the total number of cabinets for the entire system to six which, in addition to the original hardware, triples the floor space and power requirements for the new configuration.

3.2 The AQUA Satellite

Early in 2002, NASA is scheduled to launch AQUA (previously known as EOS-PM1), the second satellite to carry the MODIS instrument. NANOOK is already being reconfigured to be able to support this second sensor. The net result of this addition is a doubling of all data requirements, including processing, storage, and distribution.

3.3 The AIRS Project

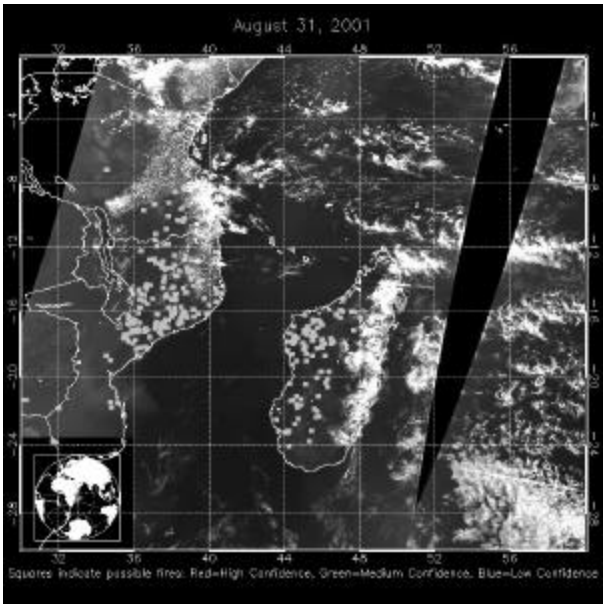
The AQUA satellite carries several different sensors, one of which is the new AIRS sensor. The NOAA AIRS project is also interested in processing near-real-time data, and will require the global cloud mask produced by NANOOK from the MODIS data. While this is not the full product suite of NANOOK, it will require global processing to the level 1B stage. This by itself will increase CPU needs from four CPUs per hour to 48 CPUs per hour, and increase data storage requirements from 35GB per day (just for Level1B) to 576GB per day.

3.4 Additional Products

The fire detection product, which was to be part of the Land Surface Reflectance product, has been re-written as a stand-alone product. On a global scale, this produces 15MB of data per granule, or an additional eight gigabytes of products. Due to interest from many parties, this product has been added to the NANOOK requirements. An example is shown below in Image 2. This image shows an early version of the product labeling possible locations of fires on the southeastern coast of Africa, and the island of Madagascar.

Other products created by NOAA include regional mosaics covering the Continental U.S, Alaska, and Hawaii. Images covering the current month, and the previous month are available at the NANOOK website: *http://www.osdpd.noaa.gov/MODIS*

Image 2: Fire Product Image.
31 August 2001.
Southeast African Coast, Island of Madagascar.



3.5 File Server for customer access

The original design for data distribution was for NANOOK to transmit all output products to NOAA's Central Environment Satellite Computer System (CEMSCS). However, this server is rapidly becoming saturated in bandwidth and storage due to the large variety of products from many other NOAA programs. Rather than overload the CEMSCS distribution server further, it was decided to procure a file server system that will allow all the products generated by NANOOK to be made available to the users without further burdening

CEMSCS. It was decided to have a file server as a separate machine from the processing system for two reasons. The first reason is that it will not compromise the security of the production servers by allowing public access. The second, that it will not overwhelm the network resources of the production server by shipping multiple copies of the products to multiple customers.

The Origin 3200, which serves as the fileserver, has an attached five terabyte RAID storage system which allows approximately one weeks worth of data to be kept.

4. Risk Reduction for NPOESS

In 2004, the NPOESS Preparatory Project Satellite will be launched as a test bed to validate the new sensors that will be launched as part of the NPOESS (National Polar-orbiting Operational Environmental Satellite System) constellation beginning in 2009. NPOESS combines NASA (National Aeronautics and Space Administration), NOAA (National Oceanic and Atmospheric Administration), and Department of Defense meteorological programs in to a single program for the entire nation. Among the instruments carried aboard NPP and NPOESS will be VIRS (Visible Infrared Imaging Radiometer Suite), and CrIS (Cross-track Infrared Sounder), the follow-on instruments to MODIS and AIRS respectively. These sensors will gather larger datasets than any before.

NANOOK has therefore been tasked as a risk reduction system in preparation for these new sensors. NANOOKs purpose in this regard is to identify the hardware, software, and network configurations that will allow NOAA to handle these extremely large datasets in near-real-time.

NOAA will use MODIS and AIRS data to simulate these high volumes and multi-spectral instruments. NANOOK will then provide real benchmarks for processor performance, data communications, and storage, thus allowing NOAA to generate findings and recommendations for NPP and NPOES planning. These findings will then allow NPP and NPOES to use real-world experiences when making decisions for ground system and instrument science designs.

5. SUMMARY

The NANOOK system has evolved from a basic proof of concept processing system to an operational products processing system for MODIS. Furthermore, not only has the processing increased, but also the physical system: more CPUs, increased storage, and a higher volume communications system. Because of the success of the proof of concept system, the requirements of the system have also grown and the system will be used in risk reduction for the NPP project which is scheduled of launch in 2004.

6. REFERENCES:

Images were created using Granule_Imager version 1.0, developed by Mark Gray (mag@climate.gsfc.nasa.gov) of the MODIS Atmospheres Group at the Goddard Space Flight Center. Modified for Fire Product overlay by Paul Haggerty (Paul.Haggerty@noaa.gov)