

# MODIS SCIENCE ALGORITHMS AND DATA SYSTEMS LESSONS LEARNED

*Robert E. Wolfe<sup>1,\*</sup>, Bill L. Ridgway<sup>2</sup>, Fred S. Patt<sup>3</sup> and Edward J. Masuoka<sup>1</sup>*

1. NASA Goddard Space Flight Center, Code 614.5, Greenbelt Rd., Greenbelt, MD, USA

\* Corresponding author: phone: 301-614-5508, e-mail: robert.e.wolfe@nasa.gov

2. NASA GSFC Code 613.2 & SSAI, Lanham, MD, USA

3. NASA GSFC Code 613.2 & SAIC (GSC), Seabrook, MD, USA

## ABSTRACT

For almost 10 years, standard global products from NASA's Earth Observing System's (EOS) two Moderate Resolution Imaging Spectroradiometer (MODIS) sensors are being used world-wide for earth science research and applications. This paper discusses the lessons learned in developing the science algorithms and the data systems needed to produce these high quality data products for the earth sciences community. Strong science team leadership and communication, an evolvable and scalable data system, and central coordination of QA and validation activities enabled the data system to grow by two orders of magnitude from the initial at-launch system to the current system able to reprocess data from both the Terra and Aqua missions in less than a year. Many of the lessons learned from MODIS are already being applied to follow-on missions.

**Index Terms**— EOS, MODIS, science data system, algorithm

## 1. INTRODUCTION

In the late 1990s when the initial version of the MODIS Data Processing System (MODAPS) was being developed, a research data system that produced such a large volume of research products was unprecedented. Over 40 calibration and geophysical science algorithms with complex interdependencies had to be integrated and tested before the first on-orbit MODIS data were available. Once the MODIS/Terra instrument data became available in early 2000 the data system helped NASA's science team to continuously improve the science algorithms. When MODIS/Aqua flew in mid-2002, most of the algorithms were stable and new algorithms were brought on-line to exploit the synergy of multiple data acquisition times provided by the two instruments. Since that time, multiple reprocessing has continuously improved the algorithms giving the community stable high quality validated earth science products. Many of these algorithms were also used by the direct broadcast and applications communities to

further exploit the MODIS data for the near real-time operational users.

Several factors were key to the success of the MODIS mission. The most important were strong science team leadership and good communication. These helped coordinate the large NASA science team and keep the mission focused on the science goals and objectives. Second was an evolvable and scalable data system that was developed through close interaction with key science team members. The data system evolved over time and grew by two orders of magnitude in terms of processing and storage to allow for the forward processing, science testing and reprocessing rates needed to meet of the science team's and the community's expectations. The third key to success was the central coordination of the quality assurance (QA) and validation activities. Even though these activities were primarily performed by the science team members, it is because of the complex linkages between the science algorithms that a coordination activity was needed to efficiently bring together information from multiple science team members and to help understand the algorithm interdependencies. Finally, active applications and outreach activities facilitated a rich and varied set of MODIS products that is widely used by the global community for near real-time and regional research and applications. The starting point for many of these applications was the standard science algorithms that were released early in the mission. Feedback from this community was also invaluable in the continual improvement of the standard algorithms.

In addition to the above overall lessons learned, each of the three MODIS disciplines, Oceans, Atmospheres and Land, has specific lessons learned which are described below.

## 2. OCEAN LESSONS LEARNED

The lessons learned regarding the science algorithms and data systems for the MODIS Ocean products are mainly focused on Ocean Color, as the Sea Surface Temperature (SST) algorithms have been significantly more stable and reliable during the mission. The MODIS Ocean data processing has been supported since 2004 by the Ocean

Biology Processing Group (OBPG) at NASA Goddard Space Flight Center (GSFC). The Ocean Color geophysical parameters include normalized water-leaving radiances (nLw) in visible wavelengths, chlorophyll-a concentration, fluorescence line height, and other in-water measurements such as the diffuse attenuation coefficient and calcite concentration.

The Ocean Color specific lessons learned derive from these principles: Ocean Color products generated from satellite remote sensing data are very sensitive to the quality of the sensor radiometry, much more so than for other disciplines; and therefore the overall success of the effort depends critically on collaboration between the Ocean and Calibration teams and on a robust Calibration and Validation program. In addition, the success of the Ocean Color products is also based on general approaches that apply to the other disciplines as well: the development of Climate Data Records (CDRs) from multiple instruments is best performed in discipline-specific measurement-based systems; the products themselves should be developed and validated in stages, with the initial products validated and mature before the downstream products; a well-defined, rigorous process is needed for adding new products in cooperation with external team members; the data system needs to be flexible to accommodate both regular algorithm updates and continual technology advancements; and the entire effort needs to be visible and linked to the external science community.

The quality of Ocean Color data products has always been extraordinarily sensitive to the accuracy and consistency of the radiometric measurements in the visible and near-infrared (NIR) wavebands that are used for this discipline. The reason is that the majority of the top-of-atmosphere (TOA) signal for ocean data is due to atmospheric scattering, with only 7% to 15% contributed by the ocean surface. Thus, the requirements on the TOA radiometry are about an order of magnitude more stringent than those on the ocean surface measurements. In addition, small artifacts (e.g., detector-to-detector or mirror side variations) are greatly magnified at the ocean surface.

For MODIS, a large part of the solution to this issue has been close collaboration between the Ocean Team and the MODIS Characterization Support Team (MCST), who have overall responsibility for the calibration of the instrument. These teams have regularly shared analyses of the radiometric calibration and characterization of the instrument and meet periodically to discuss significant results. In several cases, previous characterization results have been reviewed and the data re-analyzed in order to resolve issues with the data products. One particularly significant example is the polarization sensitivity. The initial Aqua MODIS Ocean Color products showed substantial variations geographically, temporally and spectrally compared to the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), which is known to be stable and has minimal

polarization sensitivity. Since the MODIS visible and NIR bands are sensitive to polarization, this was suspected as a possible cause, although a polarization correction was included in the Ocean Color processing. The Ocean Team obtained the prelaunch polarization test data from MCST and reanalyzed the measurements. This analysis revealed an error in the interpretation of the polarization angle for the test source. After this error was corrected, and the results included in the polarization correction, the Ocean Color products were much more consistent.

This is just one case in which a re-examination of a previous result from a fresh perspective produced a significant improvement in the algorithms and products. Other examples include the development of an improved Sun angle correction for the solar diffuser measurements, radiometric corrections for sub-frame striping in the high-resolution bands, and the analysis of the response vs. scan angle (RVS) correction using surface measurements for the blue (412 and 443 nm) bands, which have shown the largest overall degradation in radiometric response.

The Calibration/Validation (Cal/Val) program has always been an integral and substantial part of the OBPG activity at NASA/GSFC. The Ocean Color Cal/Val program spans all phases of the instrument life cycle, starting with a thorough understanding of the instrument design; through instrument pre-launch characterization and post-launch evaluation; algorithm development, implementation, testing and evaluation; and data product quality control and validation. A major element of the Cal/Val effort is the collection, analysis and archiving of the in situ measurements that are used for product validation, either by direct support for field programs or ongoing collaboration with other institutions performing this activity.

The general approaches listed above have also been applied to the MODIS Ocean Products with very good results. The change from mission-based to discipline-based processing has been critical to the success of the MODIS Ocean products. A major reason for this is that climate data quality and consistency can only be achieved with multiple instruments if the same algorithms are used. This type of consistency is best accomplished within a single organization that already possesses the scientific expertise and information technology infrastructure. Additional benefits are realized with a highly integrated project structure with co-located elements, facilitating continuous communication among all project elements. In fact, the OBPG is one of the earliest and most successful discipline-based satellite data processing organizations at NASA/GSFC. When this group assumed responsibility for the MODIS Ocean processing, one of the first actions taken was to implement the Ocean Color processing within the existing data processing software, thus removing a major source of discrepancies between the MODIS and heritage data products.

A second general principle is to develop and validate the geophysical parameters in stages that reflect the upstream dependencies for each parameter. In the case of Ocean Color, the most fundamental “feedstock” measurements are the nLw, and all of the other parameters are derived based on a combination of analytical and empirical algorithms. Thus, the success of the entire product set depends upon the nLw quality. The OBPG approach has been to focus initially on the algorithms (primarily atmospheric correction) that determine the nLw quality before attempting to implement and validate the downstream parameters. In the case of MODIS, this led to the discovery of the polarization characterization error described above.

Closely linked to this is the process followed by the OBPG for implementing new geophysical parameters within the products. As new algorithms and parameters are proposed by external team members, the NASA Ocean Team Leader prioritizes their implementation according to the needs of the scientific community. For each new algorithm, a plan is then developed between the OBPG and the investigator for implementation, verification, evaluation and validation strategies. An OBPG programmer works closely with the investigator through the initial stages through the evaluation of the initial test products. Finally, the products are generated on an evaluation basis and made available to the entire community for review and comment before being accepted as standard products.

The Ocean Data Processing System (ODPS), implemented and operated by the OBPG, has been another major element in the success of the MODIS Ocean Products. This system allows the flexibility for new algorithms, parameters, or even entire processing streams to be added with minimal effort. At the same time, it provides substantial reprocessing capacity, which allows for extensive testing of all new and modified algorithms before they are accepted. A typical algorithm test run consists of processing four days of data per month over the entire mission. This allows the effects of calibration and algorithm changes or the characteristics of new parameters to be evaluated over the full temporal and geometric range of the data set before being accepted for routine processing. For MODIS/Aqua alone, a total of 35 mission-length test runs have been conducted over the past five years. The test descriptions and results are posted on the OBPG web site <http://oceancolor.gsfc.nasa.gov/> for public access. Also, the ODPS provides the long-term archive and data distribution for the MODIS Ocean products.

Finally, the OBPG has a long history of communication and collaboration with the Ocean Color science community. In addition to the involvement of individual investigators in the algorithm process as described above, this group has placed a high premium on soliciting community input, making regular presentations of ongoing work at meetings and conferences, publishing in the peer-reviewed literature,

and making all results and data products public as soon as they are available. In addition, the OBPG has provided a standalone data processing and analysis package, the SeaWiFS Data Analysis System (SeaDAS), to the general public free of charge, allowing anyone with a desktop workstation to analyze and process MODIS Ocean data.

### 3. ATMOSPHERE LESSONS LEARNED

MODIS products provide a daily snapshot of very dynamic atmospheric parameters. At the mission outset, the high spatial and temporal variability of cloud cover meant that direct validation of cloud properties was limited to field campaigns using satellite under-flights by the high altitude MODIS Airborne Simulator, in-situ aircraft cloud sampling and coincident surface measurements. The recent availability of A-Train data from CloudSat and CALIPSO provides the research community with a more direct validation of cloud top height, cloud particle phase and cloud optical properties. However, we do not yet make routine operational comparisons of data from these multiple A-Train sensors.

Aerosol optical thickness data are much less variable (except near point sources), but daily measurements are constrained to cloud-free regions. The large network of AERONET ground stations has been critical in validating MODIS retrievals. The MODIS project has facilitated this validation by creating two special aerosol parameter subsets. We provide HDF subsets of retrieved parameters near AERONET surface stations for detailed inter-comparisons. The team also creates MODIS “best value” text spreadsheets to facilitate simpler desktop analysis. Together, these tools are used to validate retrievals by region, season and aerosol model type among many factors.

Atmospheric product packaging illustrates one approach to the trade-offs that are inevitable between ease of use and completeness. At Level 2 (L2), retrieved parameters are bundled into five products with 1 to 10km spatial resolution, namely, cloud mask, atmospheric profiles, aerosol, water vapor and cloud parameters. There are no gridded or remapped products. All products are combined at Level 3 (L3) into one-degree statistical parameters for daily, eight-day and monthly periods. The primary disadvantage of this packaging choice is that product users must acquire large product bundles that include diagnostic parameters when only one science data set may be of interest. This overhead is particularly serious at where hundreds of statistical quantities are grouped within the daily and monthly products.

Parameter bundling in large products does of course offer significant advantages. The L2 products include a large number of diagnostic and QA parameters together with shared geolocation data. We make extensive use of these embedded QA flags and diagnostic parameters in doing product quality analysis. For clouds, this means

distinct analysis by particle phase (water, ice, undetermined) and identification of challenging multi-layer cloud cases. For aerosols, consistency in the solution index, path radiance and number of pixels used for optical parameter determination will determine confidence in the results.

We routinely produce global daily L3 maps of all atmospheric parameters including statistical means, standard deviations, minimum and maximum values. In making these maps, parameters are further separated by key diagnostics such as cloud phase, scattering angle, aerosol retrieval band, QA weighting, etc. The availability of statistical parameters segregated by retrieval diagnostic flags makes possible this very detailed representation of atmospheric parameters.

#### 4. LAND LESSONS LEARNED

One of the biggest challenges that the data processing system for the MODIS Land discipline team faced was retrieving terrestrial geophysical values at 250m, 500m and 1km resolutions on a daily basis over the global land surface with the late 1990 computer technology. The twelve MODIS Land products include [1]: surface reflectance, vegetation indices, snow cover, land surface temperature, thermal anomalies (e.g. fires), bi-directional reflectance, albedo and leaf area index. When MODIS was launched on Terra in late 1999, the data processing system was primarily I/O constrained. Only 1TB of production disk and 10TB of fast intermediate tape storage were available for processing. To produce the multiday products, more than 1TB of data a day had to be written and retrieved from this tape system which stressed the capabilities of the overall system. To overcome this problem, two approaches were taken. The first was to reduce the volume by producing products at the 250m resolution only over the continental US and small areas outside of the US. The second was to acquire more disk storage so that the amount of I/O to tape was minimized. By the time MODIS on Aqua was launched in 2002, the data system had improved so much that the system was capable of producing the full set of Land products over the global Land surface for both instruments. Since that time, investments were made in commodity hardware so that the systems are now capable of producing the Level 1 and Land products at a rate of more than 10 times the real-time rate of each instrument.

During the initial startup phase, close interaction between the science team and the Science Data Support Team (SDST) allowed for timely decisions to be made on the production priorities and adjustments to be made to the production schedule so that early understanding of the product quality could be made while the data system was being improved.

A second Land lesson learned involved handling a large number of algorithm changes in the early Terra mission time-frame. Because MODIS was a research instrument, many of the Land product algorithms were new and being

produced globally for the first time. Complex inter-dependencies between the algorithms meant that each change rippled down the data stream. So to understand the single product's quality, it was necessary to not only understand the instrument characteristics but also the quality of the upstream products. For Land, a centralized QA group [2] expedited this process by investigating quality problems themselves and by coordinating the geographically dispersed science team members with a set of "Known Issues" web pages. These pages along with a QA database, a time-series tool and automatically generated global browse images allowed the science team to quickly localize issues and resolve them. These tools were latter very valuable in the science testing needed for each reprocessing. For instance, the time-series tool plots the product values for a location before and after a set of proposed algorithm changes. When MODIS/Aqua data was first acquired, these QA tools enabled quick inter-comparison between two sensors during the early characterization and testing of the MODIS/Aqua specific algorithm changes.

Finally, the validation of any new science product is a key factor in it being used by the research community. The MODIS Land science team recognized the importance of validation and set up a centralized activity to coordinate validation of the Land products [3]. The central validation group worked with the Land community selected a number of well instrumented core validation sites dispersed globally. The central validation group then acquired finer resolution (i.e. Landsat, Ikonos) imagery over the sites to enable the *in situ* data to be scaled to MODIS resolution.

#### 5. CONCLUSION

Many of these lessons learned from MODIS are already being applied to the follow-on operational Visible Infrared Imager Radiometer Suite (VIIRS) instruments aboard the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP) and future operational NPOESS missions. It is also expected that NASA's upcoming Earth Science Decadal Survey missions will build on MODIS experience.

#### 6. REFERENCES

- [1] C.O. Justice, J.R.G. Townshend, E.F. Vermote, E. Masuoka, R.E. Wolfe, N. Saleous, D.P. Roy and J.T. Morisette, "An overview of MODIS Land data processing and product status," *Remote Sens. of Env.*, 83 (1-2): 3-15, Nov. 2002.
- [2] D.P. Roy, J.S. Borak, S. Devadiga, R.E. Wolfe, M. Zheng and J. Descloitres, "The MODIS Land product quality assessment approach," *Remote Sens. of Env.*, 83 (1-2): 62-76, Nov. 2002.
- [3] J.T. Morisette, J.L. Privette, and C.O. Justice, "A framework for the validation of MODIS land products," *Remote Sensing of Environment*, 83(1-2), 77-96, 2002.