An overview of MODIS Land data processing and product status

C.O. Justice a,*, J.R.G. Townshend a, E.F. Vermote a, E. Masuoka b, R.E. Wolfe c, N. Saleous c, D.P. Roy a, J.T. Morisette d

a Department of Geography and Institute for Advanced Computing Studies, University of Maryland, 2181 LeFrak Hall, College Park, MD 20742, USA
b Terrestrial Information Systems Branch, NASA’s Goddard Space Flight Center (GSFC), Greenbelt, MD 20771, USA
c Raytheon ITSS, 4400 Forbes Blvd, Lanham, MD 20706, USA
d Terrestrial Information Systems Branch and Biospheric Sciences Branch (joint appointment), NASA’s Goddard Space Flight Center, Greenbelt, MD 20771, USA

Received 1 March 2002; received in revised form 12 April 2002; accepted 24 April 2002

Abstract

Data from the first Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on the NASA Terra Platform are being used to provide a new generation of land data products in support of the National Aeronautics and Space Administration (NASA)’s Earth Science Enterprise, global change research and natural resource management. The MODIS products include global data sets heretofore unavailable, derived from new moderate resolution spectral bands with spatial resolutions of 250 m to 1 km. A partnership between Science Team members and the MODIS Science Data Support Team is producing data sets of unprecedented volume and number for the land research and applications. This overview paper provides a summary of the instrument performance and status, the data production system, the products, their status and availability for land studies.

1. Introduction

To continue a strong commitment to the US space program, the National Aeronautics and Space Administration (NASA) has undertaken a program of long-term observation, research, and analysis of the Earth’s land, oceans, atmosphere and their interactions, including measurements from the Earth Observing System (EOS) (Kaufman, Herring, Ranson, & Collatz, 1998; NASA, 1999). The EOS is funded by the NASA Earth Science Enterprise (ESE) Program and has three main components, (i) a coordinated series of Earth-observing satellites, (ii) an advanced data system designed to support the production, archival, and dissemination of satellite derived data products, (iii) teams of scientists who are developing the science algorithms to make the data products. The Moderate Resolution Imaging Spectroradiometer (MODIS) is a key instrument onboard the EOS Terra satellite, successfully launched in December 1999, and will be complemented by another MODIS on the EOS Aqua satellite to be launched in 2002.

The MODIS instrument data are converted on a systematic basis into derived atmospheric, terrestrial and oceanic products. This paper overviews the status of the MODIS instrument on the Terra platform, the production of land products derived from MODIS data, and availability of these products generated from algorithms developed by the land science team. More detailed descriptions of the individual MODIS land (hereafter termed “MODland”) products are found in the papers in this special MODIS Land issue of Remote Sensing of Environment.

Since the Terra launch in December 1999, the MODIS Science Team has made significant progress in characterizing the performance of the first MODIS instrument, providing the MODIS instrument data (Level 1B), generating and assessing the quality of higher order geophysical products (Levels 2, 3 and 4), initiating product validation and preparing for the launch of the second MODIS instrument on the EOS Aqua platform. The Team scientists responsible for the delivery of code and the creation of MODland products, has generated a new suite of moderate-resolution
land products in support of global change research and natural resource applications (Table 1). These products support directly NASA’s Earth Science Enterprise’s systematic measurement program (NASA, 1999). The overall goal of the Earth Science Enterprise is to determine how the Earth is changing and what are the consequences for life on Earth. The land products from MODIS lay the foundations for long-term land surface monitoring. This has been accomplished primarily through the generation of products that are prototypes for those from the next generation of operational sensing systems, supported by the US National Polar Orbiting Environmental Satellite System (NPOESS). Research on the Advanced Very High Resolution Radiometer (AVHRR) has provided much of the methodological underpinning and experience-base for MODLand product development. However, reliance on the AVHRR and its associated spectral and geometric constraints has limited the ability of the land research community to develop the range of products needed for global change research (Cihlar, 1997). The MODLand products will enable the global change research community to address a much broader range of questions associated with biogeochemical cycling, energy balance, land cover change and ecosystems (Table 3, WWW1). The MODLand products will also be used for regional and global applications.

2. The MODIS instrument

One of NASA’s roles is to provide a test bed for new technologies. With 36 spectral bands and 12-bit radiometric resolution, MODIS has the highest number of spectral bands of any global coverage moderate resolution imager. The design of MODIS was an inevitable compromise to satisfy the requirements of the three different disciplines: atmosphere, ocean and land, with spectral bands and spatial resolution selected to meet different observational needs and provide near-daily global coverage (Salomonson, Barnes, Maymon, Montgomery, & Ostrow, 1989). The land bands have a heritage from the Landsat Thematic Mapper, with capabilities added in the short-wave and long-wave infrared (Justice et al., 1998). The general capabilities of MODIS for land remote sensing are summarized in Table 2. The MODIS spectral bands are listed in Townshend and Justice (2002, this issue). Even though many of the bands used by the MODLand algorithms were designed specifically for the land community, there were compromises in the design of some bands used by both the land and oceans research communities; for example the saturation temperature of the 11-μm band was a compromise between the needs for monitoring sea surface temperature and active fires. The oceanic requirement was to maintain a high signal-to-noise ratio for sea surface temperature, whereas the land requirement was to avoid saturation over flaming fires. Although there is currently a trend towards smaller, single mission satellites (Committee on Earth Sciences, 2000a), operational imagers will remain multi-purpose for the foreseeable future, where similar compromises will inevitably have to be reached. Operational satellites are operated by NOAA and provide a continuous and reliable service of data provision to meet national needs, for exam-

<table>
<thead>
<tr>
<th>Table 1</th>
<th>General characteristics of Terra MODIS for land remote sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbit</td>
<td>705 km, sun-synchronous, near-polar nominal descending equatorial crossing at 10:30 local time (specific overpass times can be found at <a href="http://www.earthobservatory.nasa.gov/MissionControl/overpass.html">http://www.earthobservatory.nasa.gov/MissionControl/overpass.html</a>)</td>
</tr>
<tr>
<td>Swath</td>
<td>2330 km ± 55° cross-track</td>
</tr>
<tr>
<td>Spectral bands</td>
<td>36 bands, between 0.405 and 14.385 μm with onboard calibration subsystems</td>
</tr>
<tr>
<td>Spectral calibration</td>
<td>band 1–4, 2% for reflectance, band 5–7, under investigation (some scene-dependent electronic crosstalk)</td>
</tr>
<tr>
<td>Data rate</td>
<td>11 Mbps (peak daytime)</td>
</tr>
<tr>
<td>Radiometric resolution</td>
<td>12 bits</td>
</tr>
<tr>
<td>Spatial resolutions at nadir</td>
<td>250 m (bands 1–2), 500 m (bands 3–7), 1000 m (bands 8–36)</td>
</tr>
<tr>
<td>Duty cycle</td>
<td>100%</td>
</tr>
<tr>
<td>Repeat coverage</td>
<td>daily, north of ~ 30° latitude, every 2 day for &lt; ~ 30° latitude</td>
</tr>
<tr>
<td>Gridded land products geolocation accuracy</td>
<td>within 150 m (1 sigma) at nadir</td>
</tr>
<tr>
<td>Band-to band registration for band 1–7</td>
<td>within 50 m in the along scan direction, within 100 m in the along track direction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>MODLand products, their release dates and their status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESDT</td>
<td>Product</td>
</tr>
<tr>
<td>Radiation balance product suite</td>
<td>MOD09 Surface reflectance</td>
</tr>
<tr>
<td></td>
<td>MOD11 Surface temperature and emissivity</td>
</tr>
<tr>
<td></td>
<td>MOD43 BRDF/Albedo</td>
</tr>
<tr>
<td></td>
<td>MOD10 Snow cover</td>
</tr>
<tr>
<td></td>
<td>MOD29 Sea ice extent</td>
</tr>
<tr>
<td>Vegetation product suite</td>
<td>MOD13 Vegetation indices</td>
</tr>
<tr>
<td></td>
<td>MOD15 LAI/FPAR</td>
</tr>
<tr>
<td></td>
<td>MOD17 NPP/PSN</td>
</tr>
<tr>
<td>Land cover product suite</td>
<td>MOD12 Land cover and change</td>
</tr>
<tr>
<td></td>
<td>MOD14 Thermal anomalies and fire</td>
</tr>
<tr>
<td></td>
<td>MOD44 Vegetation cover conversion/continuous fields</td>
</tr>
</tbody>
</table>
ple near real-time data for weather forecasting. NASA develops experimental satellite systems, including those that collect systematic measurements, such as MODIS. Experimental capabilities developed by NASA, which prove useful for enhancing the nations operational capabilities, can then be transitioned to the operational domain. The MODIS design team gave substantial emphasis to instrument calibration and characterization recognizing these activities as critical for generation of accurate long-term time-series products needed for global change studies. The electro-optical and geometric characteristics of MODIS are described in Table 3 (WWW2; Wolfe et al., 2002, this issue). Extensive pre-launch instrument characterization including radiometric, spectral, spatial and polarization sensitivities were performed. On-board calibration measurements monitor change in the pre-launch characteristics and establish post-launch absolute calibration in reflectance units in the reflective bands and in radiance units in the thermal bands (Guenther, Xiong, Salomonson, Barnes, & Young, 2002, this issue).

The absolute calibration in the MODIS Reflective Solar Bands (RSB) is achieved by measuring the signal over an onboard Solar Diffuser, and applying corrections (at-instrument temperature) accounting for spectral leakage, cross-talk and scan mirror reflectance. This procedure has been applied weekly for the first 2 years of on-orbit MODIS data. Uncertainty budgets estimate the absolute accuracy of the RSB calibration to be better than 2%. The MODIS Spectro-Radiometric Calibration Assembly (SRCA) and Solar Diffuser Stability Monitor (SDSM) have provided new on-board RSB calibration mechanisms (Guenther et al., 2002, this volume). The SRCA checks for stability of the band spectral responses up to ± 2 nm for wavelengths shorter than 1 μm. To date, the MODIS band spectral responses have remained stable. In addition, analysis of SRCA data help characterize electronic and optical cross-talk and provide an independent measure of the system degradation in the RSB. The SDSM is an independent ratioing radiometer that enables degradation of the on-board solar diffuser to be characterized. Monthly observations of the Moon are also used to monitor overall RSB degradation. Data are collected at different angles of incidence to the mirror, enabling monitoring of the degradation of the directional response of the scan mirror. Lessons learned from the Terra instrument will be put into practice for the Aqua version, where more emphasis will be given to calibration measurements parameterized for scan mirror angle.

The absolute calibration in the thermal bands is achieved by measuring the signal of an On-Board Calibration Black-body (OBC-BB), which is a high-quality reference with an emissivity close to 1.0. This procedure is applied for every scan-line. The instrument is also cycled through a thermal warm-up/cool down procedure every month to check for OBC-BB linearity. For all thermal infrared bands (except band 36, which has a noise level above specification), the required MODIS calibration accuracy (1%) at the typical level of radiance has been met. At launch, electronic and optical cross-talk were found to be quite problematic. The optical cross talk in the long-wave infrared (LWIR, principally 13–14.5 μm) bands involving bands 31–36 was corrected for all MODIS data after August 20, 2000. Several electronic configurations in the short-wave infrared (SWIR, bands 5–7) bands, the mid-wave infrared (MWIR, bands 20–25) bands, and the 1.38 μm band (band 26) were tried in the period from March through October 2000 to reduce the electronic cross-talk. By November 2000, a useable and stable instrument configuration (instrument operating on the “B-side” electronics) was achieved. That configuration remained until June 15, 2001 when the instrument malfunctioned. When the instrument resumed operation on July 2, 2001, another stable configuration (instrument operating on the “A-side”) was achieved and remains such to date. All data from November 1, 2000 to the present has been consistently calibrated and validated. As a result, the calibration stability for all of the land bands remains extremely good (better than 0.5%...
relative) and is derived using a rigorous approach developed and implemented by the MODIS Calibration Science Team (Table 3, WWW2).

The land group took on the responsibility for characterizing and improving MODIS geolocation (earth location) data (Wolfe et al., 2002, this issue). The MODIS geolocation algorithm operates as part of the Level 1 processing system for MODIS data. The MODIS geolocation data are stored in the Level 1 and some Level 2 products and explicitly describe the ground location (latitude, longitude and height) and the sensing geometry (sensor zenith and azimuth, solar zenith and azimuth and slant range) of each 1-km MODIS observation. MODLand science is the primary driver for geolocation accuracy and requires that the earth location knowledge be accurate to 50 m (1σ) at nadir. Since launch, a global network of ground control points has been used to operationally reduce MODIS geolocation errors to the accuracy required for terrestrial global change applications (Wolfe et al., 2002, this issue). MODIS band to band co-registration is monitored using the on-board SRCA and the co-registration has remained stable since launch. Research is also being undertaken through the EOS Validation Investigator program to better characterize the MODIS spatial response (Rojas, Schowengerdt, & Biggar, 2002, this issue).

3. The MODIS Land products

The land products for MODIS were selected in a peer review process in 1992, based on scientific priorities established in the late 1980s. This was at a time when Mission to Planet Earth was a major science theme of NASA (Committee on Earth Sciences, 1995; Running et al., 1994). The initial product suite and the algorithms have been externally reviewed twice and refined since initial selection, resulting in the loss of a polarization product and the addition of ‘vegetation continuous fields’ and ‘vegetation cover change’ products. These latter two products, selected through an open competition, were based on methodologies developed since initial product selection. Surprisingly though, the product needs have changed very little since initial selection and the suite of MODIS products remain central to addressing many of the high priority questions that make up NASA ESE Science Strategy (Table 3, WWW1). The land products are grouped in product suites in accordance with NASA’s Science Strategy; a vegetation product suite, a surface radiation product suite and a land cover product suite (Table 3).

The current NOAA operational products of vegetation index and snow cover are based on the current needs for weather forecasting and fall short of what is needed for climate and global change research (Committee on Earth Sciences, 2000b). MODIS products allow the land research community to move beyond its previous dependence on the Normalized Difference Vegetation Index, which has been the most widely used moderate resolution product for land vegetation studies (Justice & Townshend, 1994). A new MODIS product, the Enhanced Vegetation Index (EVI) offers improvements over the NDVI, reducing saturation at high vegetation cover and reducing soil background effects, but will take time to be fully evaluated in terms of its application and adopted by the community (Huete et al., 2002, this issue). Additional vegetation products of Leaf Area Index, FAPAR and Net Primary Production have been designed to address questions of ecosystem productivity, contributing to carbon cycle research (Myneni et al., 2002, this issue).

Work of the MODIS Science Team has resulted in a considerable new remote sensing research effort and several new developments in the use of the short-wave infrared, bidirectional reflectance, fire detection, land cover change and land surface temperature. The atmospheric correction of MODIS to generate surface reflectance represents a considerable advance on earlier systems and in particular provides systematic aerosol correction at 1 km (Vermote, El Saleous, & Justice, 2002, this issue). The middle infrared channels are also being exploited at moderate resolution (Petitcolin & Vermote, 2002, this issue). The Land Surface Temperature and the BRDF/Albedo products provide a new input for the climate modeling community (Schaaf et al., 2002, this issue; Wan, Zhang, Zhang, & Li, 2002, this issue). Areas for additional research remain and in particular improved methods are still needed for routinely compositing and combining products from MODIS and MISR to improve Leaf Area Index (LAI) and Albedo/Bidirectional Reflectance Distribution Function (BRDF) products (Jin et al., in press).

The dominance of requirements for global data sets by the Global Climate Modeling (GCM) Community in the early 1990s and their requirements for better parameterization of the land surface, led to a request at the MODIS products review in 1996, for products to be provided in a GCM grid at 1° resolution ( Sellers et al., 1994). The MODLand team has been waiting for products to stabilize, prior to the generation of coarse resolution climate modeling time-series data products. These GCM grid products are now starting to be generated (Hall, Riggs, Salomonson, DiGirolamo, & Bayr, 2002, this issue). The continued need for spatially degraded data sets by the modeling community was confirmed recently at the MODIS Data Processing Review (Table 2, WWW10). This recent request for a 10-km grid indicates a shift towards higher resolution modeling and relates to the development of the global change research agenda to address questions of regional impacts, prediction and adaptation, rather than solely climate change prediction requiring regional and nested modeling. The full resolution products from MODIS at 250, 500 and 1000 m are well suited to regional studies and when combined with Landsat data provide the basis for monitoring and modeling of land cover and land use change and providing carbon observations (Janetos & Justice, 2000; Skole, Justice, Janetos, & Townshend, 1997).

The gridded MODLand products (Levels 2G–Level 4) are defined in a global nonoverlapping tile grid in the equal
area Integerized Sinusoidal projection (ISIN) (Rossow & Garder, 1984; Wolfe, Roy, & Vermote, 1998). In the planning process and to harmonize production across disciplines, the land group agreed to adopt the ISIN projection, rather than the Goedes Homosoline Projection, which was adopted for the AVHRR Pathfinder projection (Eidenshink & Faudeen, 1994). This has caused some problems for the land user community as the ISIN projection is not supported by most image processing packages and is not easily converted to other projections. In response to the need for reprojection of the data, the EDC DAAC has developed a software tool for Unix and Windows, which converts the MODIS data to a number of different projections (Table 3, WWW3). This is worth noting that obtaining a consensus on projections is always difficult and that even after considerable debate as to the projection of choice, one of the most frequent user complaints about the AVHRR Pathfinder data set was the Goedes projection (Teillet et al., 2000).

All EOS data products are stored using an enhanced hierarchy data format (HDF-EOS) that allows implementation-transparent data access via an HDF toolkit. HDF-EOS files are self-describing, making it possible to fully understand their structure and content from the information stored in the file alone. However, a recent unpublished survey by User Services at the EDC DAAC identified the HDF-EOS data format as being one of the largest obstacles to data use (Dwyer, personal communication). This initial reaction to HDF-EOS is understandable because of its complexity compared with other earlier data formats. Similarly, the slow provision of HDF-EOS analysis software and support in commercial software has made it difficult for scientists to manage and analyze the MODIS data products.

The MODIS products are distributed, along with other EOS products from the DAACs, using a world wide web interface developed by ESDIS. The interface, the EOS Data Gateway (EDG) has also provided some challenges to the user community in obtaining the data that they need. The EDG is a generic web access interface that is still being developed and supports many different EOS products. There are currently several aspects of the EDG that need improvement. It is impossible to order very large volumes of data in a single request or alternatively to subset data to reduce volumes and it is difficult to identify and order data from multiple sensors for a given location and time period. Alternatives to the EDG are being developed, for example the GSFC DAAC has developed a simplified, ‘no frills’ system for ordering MODIS L1B data from the GSFC DAAC (Table 2, WWW3).

4. The data production system

The EOS Data Information System (EOSDIS) is designed to provide the computing and network facilities to support NASA’s EOS research activities, including processing, distributing, and archiving EOS data; exchanging research results among scientists; and commanding and controlling the spacecraft instrumentation. The EOSDIS Core System (ECS) provides the computing architecture to accomplish these goals. Data Active Archive Centers (DAACs) are responsible for running the ECS. At the beginning of the Terra program, the Goddard Space Flight Center’s (GSFC) DAAC was given responsibility for the archive, production and distribution of MODIS Level 1 products and distribution of the MODIS ocean and atmospheres products (Table 3, WWW3). The EROS Data Center (EDC) in South Dakota was given responsibility for the archive, production and distribution of MODLand products (Table 3, WWW4) and the National Snow and Ice Data Center (NSIDC) in Colorado (Table 3, WWW5) was given responsibility for the archive, production and distribution of the MODLand cryospheric (Snow and Sea Ice) products. Budget constraints and a desire from the science community for more flexibility led to consideration of alternative processing solutions to augment the ECS. The responsibility of generating Level 2 and higher products was given to the MODIS Science Team to develop a Science Investigator-led Processing System (SIPS). The resulting MODIS Adaptive Processing System (MODAPS) was developed under guidance from the science team. The MODAPS system produces the MODIS land, atmosphere and ocean products on a combination of SGI Origin supercomputers and commodity Intel Pentium processors. The MODAPS job scheduler and operators graphical user interfaces (GUIs) were written specifically to handle the 17,000 jobs and 2.5 TB of output products generated daily during MODIS production and provide an overview in spreadsheet format of the state of the production jobs in the system (Masuoka, Tilmes, Devine, Ye, & Tilmes, 2001). Maintaining a flexible data production system under the supervision of the MODIS Science Team has been beneficial in allowing rapid updates of the product generation codes ensuring improvements in the quality of the MODLand products (Table 3, WWW6).

- MODAPS generates land products in a hierarchy of processing levels: retrieved geophysical parameters at the same location as the MODIS instrument data (Level 2),
- earth-gridded geophysical parameters (Level 2G and Level 3),
- earth-gridded model outputs (Level 4).

The smallest unit of data processed at any one time is defined at Level 2 as a granule, and at Levels 2G, 3 and 4 as a tile. A granule corresponds to 5 min of MODIS sensing and covers approximately 2340 x 2030 km in the across and along track directions, respectively. MODIS senses 288 granules every 24 h. A tile corresponds to a 1200 x 1200 km geocoded area. Level 2 products sensed over a 24-h period data are binned into an intermediate tiled data format referred to as Level 2G. A tile corresponds to a 1200 x 1200 km geocoded area. The Level 2G format provides a convenient geocoded data structure for storing granules and
enables flexibility for subsequent temporal compositing and data projection (Wolfe et al., 1998). Currently, 45 land products are generated in MODAPS, but only 31 are shipped to the DAACs for archive and distribution. The remaining are intermediate products that are used in the processing of archived products and that are archived in MODAPS for a limited time as required by science Quality Assessment (QA) activities. In the processing stream, MODAPS currently creates over 10,000 files daily of standard and intermediate MODland products with a data volume exceeding 390 gigabytes.

4.1. Land data production

Each land product is generated by a Product Generation Executive (PGE), which consists of the science algorithm and the necessary scripts to ensure appropriate staging of inputs and archival of outputs. In order to optimize the input staging process, PGEs that share the same inputs and have similar production rules and frequency are bundled into recipes that are run at the same time in the production system. Fig. 1 summarizes the MODland data flows. The raw MODIS instrument data are broadcast to the MODIS receiving station at White Sands, New Mexico and then sent to the EOS Data and Operations System (EDOS) before being sent to the GSFC DAAC. The MODland products are derived from the calibrated MODIS data (Level 1B), MODIS geolocation data, and ancillary data that include National Center for Environmental Prediction (NCEP) meteorological data and Data Assimilation Office (DAO) data. These inputs are delivered to MODAPS from the GSFC DAAC (denoted C0 and D0 in Fig. 2) as soon as...
they are acquired and processed. Under normal circumstances, over 80% of the MODIS instrument data are delivered to the GSFC DAAC within hours of acquisition; the remainder of the data may take days to archive into the DAAC and may require multiple iterations with the data provider. Ancillary data delivery lags behind real time by approximately 24 h for NCEP predictions and by up to 2 weeks for certain DAO data.

The Level 2 products are generated first in the MODAPS as they are required to feed the higher order production streams. Given the science algorithm requirements, three different recipes are used in MODAPS to create the land Level 2 products (Fig. 2). The first recipe (R1) is used to create the snow, sea ice and fire products. It runs independently on each MODIS granule and so is run 288 times per day. The second recipe (R3) is used to create the Level 2 surface reflectance product. It requires one full orbit of data as input (20 granules) and creates up to 20 output granules. This recipe runs up to 15 times per day and requires all granules from one orbit to be available before starting. The third recipe (R4) is used to create the daily land surface temperature products. This recipe, which creates both Level 2 and Level 3 daily products, divides the global in six latitudinal zones and runs one processing stream per zone.

Level 2 data created by R1 and R3 are binned into the intermediate daily Level 2G data products that in turn are composited to create many of the daily Level 3 products. These operations are performed in the R5 and R5P recipes for the equatorial Integerized Sinusoidal (ISIN) and the polar Lambert’s Azimuthal Equal Area (LAEA) projections, respectively (Snyder, 1987). Only the MODland snow and sea ice products are produced in the LAEA projection. R5 creates up to 326 tiles per day and R5P creates up to 552 tiles. The Level 2G and Level 3 daily products are further temporally accumulated or composited to create: 8-day land surface reflectance, land surface temperature, snow, fire, and LAI/FPAR products (recipe R10), 16-day BRDF/Albedo, VI and intermediate land cover conversion products (recipe R12), and 32-day land cover products (recipe R14). These composited products are used to create the NPP/PSN and quarterly land cover products.

4.2. Evolution of processing priorities and algorithm changes

Processing priorities have evolved since the ‘first light’ MODIS data were sensed on February 25, 2000. There have been six main stages of data production: instrument checkout, algorithm checkout, golden month, preparation for the first reprocessing, replanning and reprocessing steady state. The ‘golden month’ was aimed at generating 1 month of continuous daily data with no data gaps and resulting in complete monthly products. There is a correlation between these stages and the number and nature of the code changes that have been made, with a decreasing number of code changes as the products have matured.

Periodic updates of the PGEs and reprocessing of products have been performed to ensure the quality of the products. Code updates are performed by the science team to rectify issues that have been identified by their quality assessment activities (Roy et al., 2002, this issue). Reprocessing involves applying the latest available PGE versions to the MODIS instrument data and using the best available calibration and geolocation information. A collection number is used to differentiate between different reprocessing runs and is apparent in the MODland product filename. This information is also stored in the PGE Version metadata that is stored in every product to reflect PGE changes. The PGE Version consists of three numbers, the first to indicate a major milestone in MODIS production (e.g., the collection number), the second reflects major science code changes (e.g., algorithm changes), and the third reflects minor non-science updates to the code (e.g., bug fixes). For example, Fig. 3 shows the different PGE versions for the Level 2 Surface Reflectance products. In the period from January 2000 to October 2001, there were two major milestones (Versions 2 and 3), four major science changes and 15 other changes. There were a number of data production gaps due to problems with the MODIS instrument and EDOS data availability. A summary of the major events in MODIS production to-date are shown as a timeline in Fig. 4.

In the first 2 months after MODIS first-light data were sensed, the instrument data and Level 1B product underwent preliminary evaluation. In May 2000, the MODland Level 2 products were generated followed by the dependent downstream Level 3 products. Product quality assessment was of paramount importance in this period resulting in a large number of algorithm changes and code adjustments, as illustrated in Fig. 2. Many of these updates addressed meta-data and upstream input data format issues and code bug fixes. The initial MODland products released to the public in this period were designated as being “Beta” products. The Beta maturity designation is used by the MODIS Science Team to denote that products are available primarily to allow users to gain familiarity with data formats and parameters. Beta maturity products are unvalidated and may not be appropriate as the basis for quantitative scientific investigations. At the end of the initial checkout period, the production priority shifted towards processing complete data sets, allowing stable monthly composited products to be assessed meaningfully. As the data systems and flows started to stabilize, a “Golden Month” was produced with no missing data. This month of complete data was generated for the period following the second MODIS instrument configuration change on November 3, 2000. In this period, the frequency of code changes decreased and the science team concentrated on assessing the scientific quality of the products rather than resolving data format issues. The production managers at the GSFC DAAC and MODAPS concentrated on improving the communications and planning between their staff to increase the overall MODIS processing rate.
After approximately the first year of MODIS production (Collection 1), the science team had resolved the majority of issues in their algorithms and problems associated with dependencies between the different MODIS products resulting in the products being elevated to Provisional status. Provisional status means that the products are only partially validated and improvements are continuing. A first reprocessing with the objective of producing a complete year of products using consistent and improved algorithms was applied to the MODIS data sensed between October 31, 2000 and November 8, 2001. This first major reprocessing was named Collection 3. Prior to reprocessing, an algorithm configuration management process was implemented, which required both science discipline and team leader approval before algorithm changes could be made. The first MODIS reprocessing activity started officially mid-June 2001 and ended in January 2002. As shown in Fig. 3, the surface reflectance algorithm remained unchanged for the bulk of the reprocessing period. Some minor changes were necessary due to the instrument configuration change and coding errors, but these changes were of limited scope. Users are currently encouraged to order the Collection 3 MODland
products rather than the Collection 1 products. Collection 4, which will be produced by a second reprocessing, is scheduled for Fall 2002.

Testing new algorithms, performing quality assessment, and subsequently refining products is a time-consuming activity. The initial expectation by NASA management that science quality products would be available to the broader community immediately after the first data acquisition has proven unrealistic. The software that generated individual land products required between five and 30 changes during the first 2 years, thus hindering the broader community’s ability to utilize products and participate in validation activities. In retrospect, a phased release of higher order products would have been better, delaying release of anything other than sample products, until products had reached provisional status: inevitably given the novel characteristics of the sensor, many of the products substantial improvements are still emerging more than 2 years after launch. The overall goal for MODIS of delivering global time-series data of science quality measurements places an emphasis on data quality assessment, global product validation and data reprocessing.

4.3. Data production constraints

In the first 24 months, the EOS data system was hampered by various problems that constrained full data production. During the first year, the EDOS processing of data to Level 0 was affected by repeated data drops, resulting in gaps in global coverage and backlogs for L1/ L2 processing at the GSFC DAAC and MODAPS. These gaps can be seen in Collection 1 data. As a result of pre-launch budget reductions, the MODIS processing capacity at launch was limited to 1 × of the data volume baseline developed in 1996, while land product volumes had grown to twice this baseline as new parameters and peer-reviewed products were added. Processing at 1 × (i.e. producing 1 day’s worth of data in 1 day) means that when the system goes down, gaps in production will occur or production falls behind the current acquisition. This limited production capacity severely hindered the data system’s ability to cope with the inevitable system problems and delays associated with a large data system with multiple interfaces.

The 1996 volume baseline was an estimate established prior to the code being developed. Since that time, the land product data volumes have grown from an estimated 168 GB/day to 321 GB/day currently. This increase reflects a combination of new important intermediate products that are now being saved and the newly added products of Vegetation Cover Change and Vegetation Continuous Fields (Hansen et al., 2002, this issue; Zhan et al., 2002, this issue). The volumes of incoming data from Terra are large (194 GB/day). MODAPS is currently (February 2002) running at 2.25 ×, generating a total of over 2 TB/day, and is reliably shipping data to the DAACs for archiving and to the Science Team for data quality assessment. The MODAPS processing capacity will be enhanced during 2002 to 3 × for Terra for current processing and reprocessing and 1 × for Aqua MODIS processing. The 1 × processing capacity for Aqua MODIS for its first year will necessitate a slow phase-in of the Aqua products following launch. The data volumes for MODIS are unprecedented for the land community. In comparison, the 2001 average L0R and Level 1 production for Landsat 7 was 25.7 GB/day.

MODAPS is evolving as experience is gained with large volume processing but regardless of computer resources, some factors related to the production rules defined by the scientists or as implemented in the processing system, constrain the MODland data production rate. For the MODAPS in its current configuration, these factors include the following.

Availability of all inputs: Some of the Level 2 processing recipes require a set of granules to start. Any delay in receiving some granules will lead to delaying the production of other granules, which will subsequently delay the production of all daily products.

Use of ‘update’ files: Some science algorithms use intermediate ‘update’ files to collect data for use in the final products. These files can only be accessed by one processing stream at any given time, which limits the number of concurrent runs of these products. For example, the land surface temperature code can only be run at a rate of 2 × (2 times real-time) due to the use of update files.

Chronological sequencing: Products that use ‘update’ files require the production to be done in a chronological order. This requirement becomes a constraint that limits the land Team’s ability to reprocess periods out of order for certain products.

In addition to these factors, multiple changes to the PGEs have made integration and testing a time-consuming task and placed a burden on the production system. Due to funding constraints, testing has been undertaken on the production system and there is now a strong case for establishing a separate test system, which will not interfere with operational production.

The creation of temporal composites by selecting data from multiple images to reduce the impacts of clouds is an integral part of moderate resolution remote sensing but places a large logistical burden on the data system. To improve data throughput, compositing requires keeping multiple dates of daily data online. Approximately 380 GB of input data need to be kept online for the 8-day composited products. The volume of input data common to the 8- and 16-day products is about 174 GB/day and a total of 2.7 TB needs to be kept online to process 16-day products. This means that approximately 3 TB need to be kept online to process the 8- and 16-day products. Given the frequency of data gaps in the input data stream from EDOS and the GDAAC, MODAPS is often obliged to work on more that one 16-day period at a time, keeping 4.5 TB of daily data online. Having to keep such large volumes of data online makes it impossible to wait for extended periods of
time for missing inputs and leads to the processing of multi-day compositing products with missing data.

The 250-m land data volumes have also been a particular challenge to the data system. The land products archived at 250-m resolution for the land surface account for 77 GB/day or 25% of the current MODIS land data volume. To reduce the volume of land products archived to fit within the 1996 baseline, the land science team were had to forgo archiving the 250-m products at the DAAC and created a separate system for the generation, archival and distribution of a limited 250-m product volume until the data archive was augmented to allow ingest of the current product volume at more than double the 1996 baseline.

MODAPS generated Level 2 data were sent to a script-based production system to produce surface reflectance and vegetation index products. This system was scoped to generate 10% of the global 250-m coverage. A web-based distribution system was developed, demonstrating that low-cost systems can be developed rapidly (Table 3, WWW7). The Global Land Cover Facility at the University of Maryland extended the service, receiving the 250 m data and creating and distributing stitched 250 m Vegetation Index tiles of the continental US (Table 3, WWW8). There is a similar need for specialized, distributed services to provide other higher order products and the necessary tools for data manipulation, analysis and management. In October of 2000, when the land volume baseline was increased, MODAPS resumed partial production of 250-m products that are archived at the EDC DAAC. With reduction in the size of the 250-m products and approval to store larger data volumes at the DAAC, production of the 250-m data will reach full coverage in June 2002. In the planned Collection 4 reprocessing, 250-m land products will be processed with full global coverage from the start of the mission to the present and sent to the EDC DAAC for archive and distribution.

At times over the last 2 years, MODAPS production has fallen 2 months behind the acquisition date, although now that the production systems are more stable, production is running between 2 to 7 days behind Level 1B production at the GSFC DAAC, which in turn is operating approximately 1 to 3 days behind acquisition. In response to the need for timely data for forest fire monitoring, the MODland Rapid Response System has been developed (Table 3, WWW9). This system, which uses a MODIS data feed from EDOS to NOAA at GSFC, provides web-based MODIS corrected reflectance and fire imagery within 2–4 h of acquisition (Sohlborg, Descloitres, & Bobbe, 2001). The Rapid Response System provides data to the US Forest Service and also to several forest services around the World generating products customized to specific applications needs (Justice et al., 2002, this issue). This is part of a broader trend toward domain-oriented data services. The NASA data pathfinder projects showed the advantage of data systems targeted at a specific user domain, rendering large data volumes down to a manageable size aimed at specific user data needs (Maiden & Greco, 1994).

5. Land product quality assessment, validation and availability

The EOS product distribution policy is to make products available to the user community in a timely manner, with information on product errors, artifacts and accuracy. Members of the Science Team are responsible for the quality assessment (QA) and validation activities required to define these information. At the time of writing, all but two of the MODland products are available to the user community. The remaining two products (NPP and Vegetation Cover Change), which are about to be released, require a full year of internally consistent time-series data for their generation. Quality assessment information are available for all of the released MODland products (Table 3, WWW11). Product accuracy information will be available for the reprocessed Collection 4 products and will be available through the MODland Validation web site (Table 3, WWW13).

MODland QA is a near-operational activity that is playing an active role in documenting product quality and providing feedback to the algorithm developers to ensure the long-term quality of the MODland products. The system developed for quality assessment by the MODland team adopts a comprehensive approach and likely represents the most sophisticated level of operational quality characterization yet achieved for land products. Metadata attached to the product give an indication of product quality and known problems are advertised using the WWW (Table 3, WWW11) (Roy et al., 2002, this issue). This enables users to consult QA results when ordering and using individual product instances to ensure that they were generated without error or artifacts. Ancillary data fields, such as view or solar angles, land sea boundaries and snow/cloud presence are often included in the QA fields. Users are strongly encouraged to review the product quality summaries before using products and publishing results.

The primary responsibility for validation was given to the Science Team, although independent validation studies were also funded by NASA (Morisette, Privette, Justice, & Starr, 2000). Validated products have well-defined uncertainties and are quality products suitable for longer term or systematic scientific studies and publication, although there still may be later improved versions. There is very little heritage for global land product validation. For example, there has been no explicit validation associated with the AVHRR NDVI products (Townshend, 1994). A group of international scientists, as part of Data and Information Systems for the International Geosphere Biosphere Program (IGBP-DIS), developed the first quantitative global validation for the Global Land Cover product (Estes et al., 1999). The international coordination needed for global Land Product Validation (LPV) has continued through the CEOS Calibration and Validation Working Group (Morissette, Privette, Guenther, Belward, & Justice, 2000) using the MODIS experience with a primary focus on the satellite products being developed in the context of Global Observation of
Forest Cover/Global Observation of Landcover Dynamics (GOFC/GOLD) program, e.g. LAI/NPP, Land Cover, Fire and Albedo. As the land products transition to the operational domain, it will be important to provide users with an assessment of product accuracy, indicating the likely performance for the range of conditions for which the product is provided and identifying conditions under which the product might not be as reliable. Distinct phases of validation activity can be identified: (i) product evaluation compared to other satellite products, (ii) comparison against independent measurements for a small number of selected locations and conditions, moving towards and (iii) collection of a representative global sample of independent measurements.

The MODland group developed a validation strategy based in part on the Global Hierarchical Observing Strategy (GHOST) concept developed by the Terrestrial Observation Panel for Climate (TOPC) (Cihlar, 1997). This concept has been further developed to provide the foundation for a Global Terrestrial Monitoring Network (Running et al., 1999). The approach adopted by the land group was to develop the EOS Validation Core Site initiative (Morisette, Privette, & Justice, 2002, this issue; Privette, Morisette, Justice, & Starr, 1999). This validation initiative, which has now become part of the international LPV program, is resulting in selection of a small number of sites from different biomes to provide a focus for multi-satellite data acquisition and in situ measurement. It is worth noting that developing and implementing a validation program is a large undertaking, taking both time and resources. Involvement of the user community in global product validation has advantages in that it takes advantage of regional expertise, it develops an informed user community and expands the validation experience base, thereby building a stronger user community. MODland validation efforts have developed a new paradigm for the earth science community, which will hopefully set a precedent for others making global terrestrial products (Table 3, WWW12).

6. Conclusion

It takes time for new land products derived from a new sensor system to be adopted by users. The user community was relatively slow to adopt products derived from the AVHRR. In retrospect, products from the AVHRR led arguably the most significant advances in land remote sensing during the late 1980s and 1990s and they gave users the ability to develop and use time-series moderate resolution data providing the basis for terrestrial global scale monitoring. However, widespread use of the products only occurred after the data sets had undergone several sets of reprocessing over a decade. For the AVHRR, there were several distinct stages of initial production, internal and external evaluation, identification of problems, improvements and subsequent reprocessing and research is still being undertaken to improve the AVHRR processing which started in 1984 (El Saleous et al., 2000; Townshend, 1994). In the initial stages of MODIS product generation, these stages have been highly compressed and are taking place in parallel. User Guides have been developed for each product, which, along with the Algorithm Technical Background Documents, help the user understand the products.

The MODland team members are currently developing pilot examples of MODIS science use to guide the community as to how the data can be applied. This is similar to the approach adopted for the AVHRR, when the NASA group funded to create the data sets was also supported to undertake 'pilot' science application studies (Justice, Townshend, Holben, & Tucker, 1985). In the first instance, there is a distinct advantage in having those who generate the data identify the potential and limitations of the data sets.

In summary, the above account presents the first 2 years of Terra MODIS from the land perspective. The instrument provides an important advance in moderate resolution remote sensing and the data for land studies are truly exceptional, providing new spatial and spectral data with exciting possibilities for land science. The team responsible for characterizing the instrument and generating the data products has met a number of challenges associated with the instrument performance and data production, some of which have been described above. The data throughput of MODAPS represents a major achievement in terms of land data processing from optical sensors. Concerns from the Science Team and the user community about the ease of ordering, large volume data distribution and tools for data analysis remain. Currently, responsibility for these aspects of the data system lie with the ESDIS Project and the DAACs. The MODIS Science Team and the associated Science Support Teams have worked hard to design and deliver a new suite of land products that have set a precedent for land remote sensing. This success speaks to NASA's approach of funding instrument science teams. With the impending launch of the second MODIS on the Aqua platform, there is now a body of hard-won experience that can be applied to provide the science quality multi-instrument time-series of data that were envisioned for EOS.

References

response: operational use of terra data for USFS wildfire management.


Teillet, P. M., Saleous, N. E., Hansen, M. C., Eidenshink, J. C., Justice, C. O.,

Townshend, J. R. G. (1994). Global data sets for land applications from the
Advanced Very High Resolution Radiometer: an introduction. Interna-

toring of terrestrial systems by moderate-resolution remote sensing.

correction of MODIS data in the visible to middle infrared: first results.

Wan, Z., Zhang, Y., Zhang, Q., & Li, Z. (2002). Validation of the land-
surface temperature products retrieved from terra moderate resolution
imaging spectroradiometer data. Remote Sensing of Environment, 83,
163–180 (this volume).

storage, gridding and compositing methodology: LEVEL 2 Grid. IEEE

Wolfe, R. E., Nishihama, M., Fleig, A. J., Kuyper, J. A., Roy, D. P., Storey,
support of MODIS Land Science. Remote Sensing of Environment, 83,
31–49 (this volume).

Zhan, X., Sohlberg, R., Townshend, J. R. G., DiMicieli, C., Carroll, M.,
Eastman, J. C., Hansen, M., & DeFries, R. S. (2002). Detection of land
cover changes using MODIS 250 m data. Remote Sensing of Environ-
ment, 83, 337–351 (this volume).