

# Improving access to MODIS biophysical science products for NACP investigators

Robert E. Wolfe<sup>1,2</sup>, Feng Gao<sup>1,3</sup>, Jeffrey T. Morisette<sup>1</sup>,  
Gregory A. Ederer<sup>1,4</sup> and Jeffrey A. Pedelty<sup>1</sup>

<sup>1</sup>Terrestrial Information Systems Branch  
NASA Goddard Space Flight Center  
Greenbelt, MD, USA

<sup>2</sup>ERT, Laurel, MD, USA

<sup>3</sup>SAIC (GSC), Seabrook, MD, USA

<sup>4</sup>Corresponding author: robert.e.wolfe@nasa.gov

**Abstract**—MOSIS 4 NACP is a NASA-funded project supporting the North American Carbon Program (NACP). The purpose of this Advancing Collaborative Connections for Earth-Sun System Science (ACCESS) project is to provide researchers with Moderate Resolution Imaging Spectroradiometer (MODIS) biophysical data products that are custom tailored for use in NACP model studies. Standard MODIS biophysical products provide used to improve our understanding on the climate and ecosystem changes. However, direct uses of the MODIS biophysical parameters are constrained by retrieval quality and cloud contamination. Another challenge that NACP users face is acquiring MODIS data in formats and at spatial-temporal resolutions consistent with other data sets they use. We have been working closely with key NACP users to tailor the MODIS products to fit their needs. First, we provide new temporally smoothed and spatially continuous MODIS biophysical data sets. Second, we are distributing MODIS data at suitable spatial-temporal resolutions and in formats consistent with other data integration into model studies.

*Carbon research; moderate resolution; MODIS; NACP; NASA*

## I. INTRODUCTION

A NASA-funded Advancing Collaborative Connections for Earth-Sun System Science (ACCESS) project called MOSIS 4 NACP is supporting the North American Carbon Program (NACP). It is providing NACP researchers with Moderate Resolution Imaging Spectroradiometer (MODIS) biophysical data products that are custom tailored for use in NACP model studies.

The NACP is designed to provide the scientific underpinning to inform future policy decisions involving the carbon cycle, such as managing carbon sources and sinks by efficient and effective options to reduce emission or enhance carbon sinks. Information from earth observing satellites plays a major role in providing spatial and temporal information required to address the carbon accounting sought by the NACP. The NACP ecological and climate models require high quality continuous biophysical parameters as inputs or validation sources. The MODIS biophysical products [1] provide such data sources and have been used to improve our understanding on the climate and ecosystem changes. However, direct uses of the MODIS biophysical parameters are constrained by retrieval quality and cloud contamination. We have been working

closely with key NACP users to tailor the MODIS products to fit their needs.

New temporally smoothed and spatially continuous MODIS biophysical data sets from our processing provide ideal data sources for those applications. The temporal smoothing is done by weighting values from the quality assessment layers indicate poor quality and then by using filters and functions to derive an annual curve. Where temporal filling is not possible, a gap filling algorithm uses the temporal curve of high quality neighbor pixel of same land cover type from MODIS land cover products and high quality MODIS values of current pixel. A spectral similarity test is used in selecting the appropriate neighbor pixel to guaranty temporal similarity. This annual curve can then be used to derive phenology metrics that are useful for ecosystem and climate modeling. To ensure that temporal smoothing maintains the integrity of the original data sets, a quality metric has been developed that compares the high quality measurements from the original data to the smoothed data. The first four MODIS products that these techniques have been applied to are LAI, FPAR, VI and EVI.

We are also addressing a major challenge for researchers, namely, acquiring MODIS data at suitable spatial-temporal resolutions and in formats consistent with data integration into model studies. NACP users can select from a menu of product and image formats and services that can be performed on the products before they are delivered. The services include mosaicing of multiple MODIS tiles, reprojecting the products to a regional grid, and subletting the products spatially and by parameter.

In this paper we discuss the approach we use in our project to generate temporally and spatially filled MODIS products, as well as the approach we use to provide additional services to the products.

## II. OVERALL APPROACH

To help explain this approach, it is illustrative to go through an example of what is required for typical NACP investigators to utilize MODIS data. Consider the case where a research user needs a statistically clear signal of the year-to-year vegetation variability over North America. The user may want to utilize the 6 years of the standard 1km MODIS/Terra Leaf Area Index (LAI) product [2]. The user would like to include the MODIS

quality assessment (QA) information in the analysis to filter the data both temporarily and spatially. The end-product is a time series of the vegetation index signal across the US. The user wants the data in the standard map projection and grid to match their modeling system.

The old way of doing this involves setting up a processing environment to: 1) hold and process the data; 2) order the data and download it; 3) utilize tools, and building others, to generate the products; and 4) produce the time series of interest. Key obstacles the user encounters in getting this data are: 1) volume (both size and number of files) that impacts both download speed and local storage resources; 2) labor intensive data ordering and retrieval; 3) finding, installing and mastering or building the required tools; and 4) understanding detailed product QA information.

Just ordering the data is a major undertaking for most researchers. The volume for 6 years of 1 km MODIS/Terra LAI-FPAR data over North America is 50.8 GB stored in 11 316 individual files. MODIS 1km land data products are stored as 1200 by 1200 km tiles in a sinusoidal grid. The tiles covering North America are shown in the upper left of Fig. 1. Each data granule represents one time-step (e.g. an 8-day composite for the MODIS LAI/FPAR product). Ordering the data in manageable chunks (1100 files or 5 GB per day) would take the user 11 days to get the entire data set.

Once all of the data are local, the user must find or build the tools necessary to process the data. Tools do exist to reproject the data and perform some operations on the QA bits. The user has to spend resources downloading the tools, installing them, and learning to use them. Because most tools are not designed to work end-to-end, the user is likely to have to develop software to convert the output of the existing tools so that it matches the input format expected by another program. Further, it is likely that the ultimate application will require additional tools to perform some specialize functions. For example the user might want to filter the MODIS LAI data with poor quality and do some temporal interpolation to create a reasonable record. The result is very significant overhead for investigators to use MODIS products appropriately. Figure 1 provides a diagram for this example.

Our new way of doing this is much simpler. By having the entire time series of all of the MODIS atmosphere products globally and MODIS multi-day land products over North America available on-line and having a system architecture that allows multiple processing services, our system conducts the majority of processing steps in a way that is transparent to the user and delivers the use not just what they want, but also nothing the do not want.

The user first goes to our web site (<http://acclandweb.nascom.nasa.gov>) and interacts with it to describe the steps that are needed to produce the needed product. For typical users this step takes less than an hour. The user will be able to use the default values for each step or drill down and tailor the algorithm's user parameters so that it matches the user's requirements. Once this is done, the user submits the request and the system responds by sending an email notification of the request's acceptance. Based on the user's preference, the system sends email notifications to the

user relating the product progress and completion. The system is sized so that about 8-10 users can use the system simultaneously and so a typical job of this type should take about 2 to 3 days to complete. The user can then pick up the results from the system. Only the end product needs to be downloaded so for a mosaic of North America, the number of files is 41 times smaller. Data volume can also be reduced if the user requests that the system remove (subset) data from the files data that are not of interest. Note that we also use HDF internal compression the product we produce, so for MODIS Collection 4 land products there may be an additional factor of four reduction in volume of data to be downloaded.

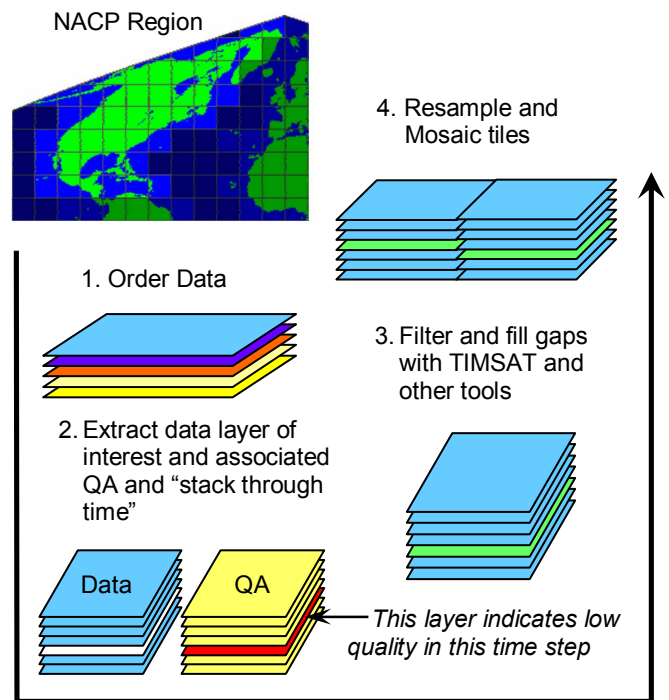


Figure 1. Steps needed to process MODIS data to obtain a LAI time series over North America.

### III. GAP FILLING APPROACH

Fig. 2 shows a six-year time series of the MODIS LAI time-series for a 1km pixel north of Ottawa in Quebec Canada (MODIS tile h12v04, row 395, column 1031 – N46.704, W74.959). Due to variations in viewing geometry, sky and cloud properties, and surface conditions, the original LAI time series (black crosses) clearly contains some noise. By using filters or functions to derive an annual curve we are able to remove this noise from the time series and fill any temporal gaps (solid line). This annual curve can then be used to derive phenology metrics that are useful for ecosystem and climate modeling.

Over the last year we have developed and tested a number of approaches to using the TIMESAT tool [3] filter and fit this data. The best approach involves two iterations based on the asymmetric Gaussian (AG) function [4]. First, based on the per-pixel information in the quality assessment layers (Fig. 2),

we identify the high-quality values (red circles) which are initially assigned higher weights, with lower weights assigned to low quality values. Using these initial weights, the first AG fit is performed (dashed line). Revised weights (blue squares) are then calculated for the high quality values based on each values difference from the initial fit. A second iteration of AG fit is then performed (solid line) using the revised weights. This two iteration approach shows better fitting result to high LAI values as well as high quality data.

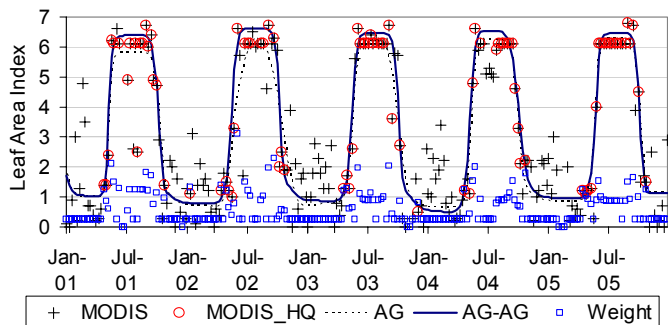


Figure 2. TIMSAT Fitting Example.

In more northern and mountainous areas, snow cover and/or low solar zenith angles during the winter causes a large gap in the temporal sequence of values that prevents us from performing the temporal fitting. In these cases we use the minimum of the high-quality values from before and after the gap to estimate the value in the gap.

In addition, for a small fraction of the pixels, temporal fitting is not successful because there are insufficient high quality values or because the high quality values are distributed unevenly. For these pixels a spatial gap filling approach has been developed. The spatial gap filling approach first looks for a nearby pixel with the same land cover class and a successful temporal fit. If there is not a nearby pixel that meets these criteria, then a temporal fit for that land cover class for the whole tile is used. This substitute fit is then scaled to more closely match the high quality values.

#### IV. INITIAL RESULTS

Results from the gap filling and temporal fitting approach are illustrated in Fig. 3 for a single MODIS tile (h11v04) centered on Wisconsin, USA. The original LAI product shown in image (a) is for the first 8-day period in 2004 and so has a low winter-like LAI over much of the image. Image (b) contains the gap-filled and temporally fitted (smoothed) values. Low-quality LAI values that are unseasonably high, e.g., the high values in the center of the image (a), are reduced in image (b). Image (d) is a “composed” image that contains the original high-quality values and replaces the low-quality values with the smoothed values. A separate new QA layer shown in image (c) identifies the smoothed values. The final product contains all three images (layers), along with the original and new QA layers, so that researchers can choose the approach best suited to their needs.

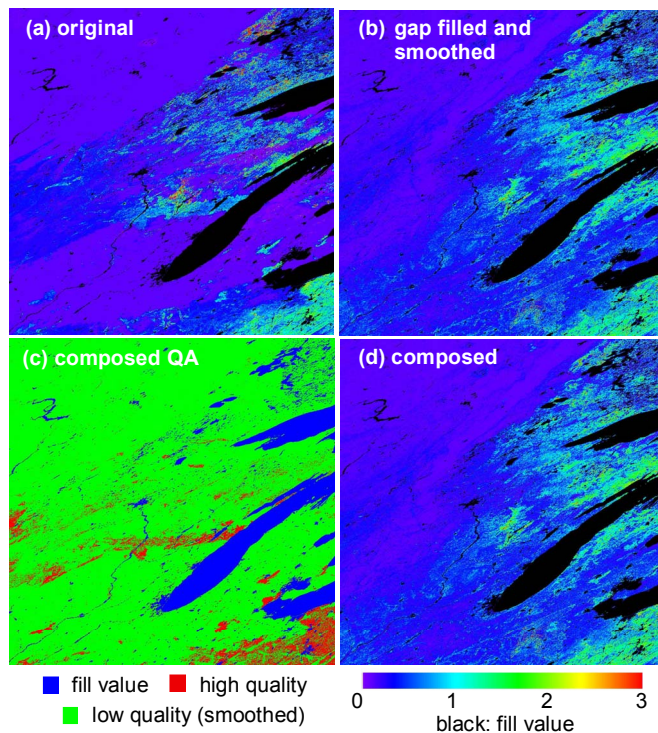


Figure 3. LAI Gap-filling and temporal fitting results (Collection 4, 2004001 – 1/1/2004).

We have applied our gap-filling and smoothing approach to MODIS data over all of North America. The first 8-day period of four seasons of gap-filled data are shown in Fig. 4 (note the change in scale from Fig. 3). White in the images represents unprocessed pixels such as water, urban or barren. When compared to the original LAI data (not shown), the smoothed and gap-filled values show smoother but similar spatial patterns. Also, the northern portion of the January image, which was mostly fill values in the original, has realistic values in the gap-filled and smoothed image.

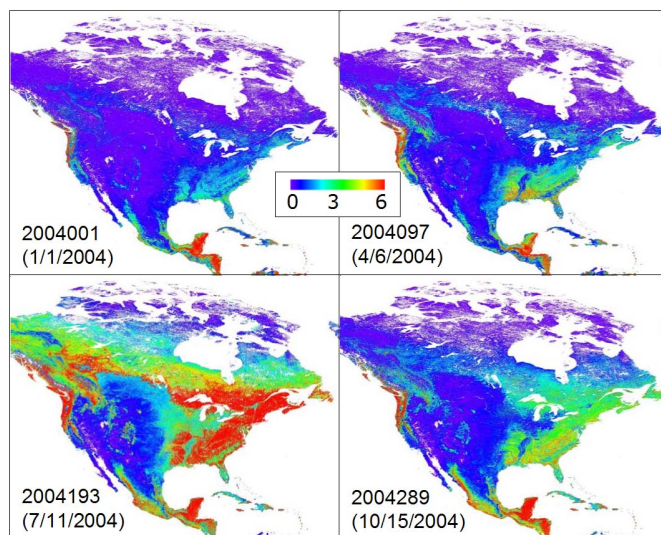


Figure 4. Four seasons of gap-filled and smoothed LAI over North America.

One test of the integrity of the gap-filled smoothed product is to compare the original high-quality LAI values to the smoothed values. When we do this for all of North America, we see very good agreement (a slope close to 1, an intercept near zero, and high cross-correlation value). Comparing individual tiles for one year of data, we also see good agreement, especially for the tiles with more than 40% high quality retrievals (spatially and temporally). Based on this we believe these new smoothed products maintain the quality of the original LAI product.

## V. OTHER POST-PROCESSING OPTIONS

The MODIS 4 NACP product distribution system is based upon the MODIS Level 1 and Atmospheres Archive and Distribution System (LAADS) that relies on the capabilities of the MODIS Adaptive Processing System (MODAPS) [5]. In LAADS, users interact with a HTML web page that uses Java scripts for active components. The user first searches for products which meet specific criteria, i.e., geographic coverage and acquisition time. The user may then order one or more files or request some or all of them be transformed by post-processing. If the post-processing option is selected, an estimate of the length of time required to transform the files is given the user. The order is then submitted to the processing back end which schedules the job to be run on one or more processing nodes. Once the post-processing completes, the resulting files are staged to an ftp area and the user is notified. The LAADS web-based interface is designed to minimize the need for the user to be aware of the individual steps needed to generate the end product, but also allow sophisticated users to drill down and modify individual steps to obtain products that meet their more specific needs.

In addition to the gap-filling and smoothing post-processing options there are a number of other post-processing features. The features that have been implemented to date are: reprojection (including to a NACP regional map projection); file format conversion (to GeoTIFF); subset by data layer and spatially; mosaicing (from  $10^\circ \times 10^\circ$  tiles to regional coverage); and resampling (aggregating) to a coarser spatial resolution grid. Additional options that will be implemented over the next year are: additional format conversion (to ASCII and binary); modification of temporal composites (changing time steps), ancillary model data, and machine to machine requests (services).

Programs that perform each of these post-processing steps (including gap-filling) are integrated into the MODAPS system as Product Generation Executives (PGEs). Each PGE contains a Perl script that passes run-time parameters to the program,

executes the program and any supplemental programs, and applies a standard naming convention to the output product.

Because of the time needed to regenerate the gap-filled and smoothed products, standard versions of these products are available for users to search and order, download directly via ftp, and/or apply other post-processing options.

## VI. DISCUSSION AND FUTURE WORK

We have developed a set of tools to improve access by NACP investigators to MODIS biophysical science products. We are now providing new temporally smoothed and spatially continuous MODIS biophysical data sets. NACP users can also select from a menu of product and image formats and services that can be performed on the products before they are delivered. These tools are now available to researchers in the general Earth science community.

We are continuing to work to add additional post-processing options and other MODIS Land products to the system. These capabilities will be also be make available for the improved MODIS Collection 5 products as they become available over the next year.

## ACKNOWLEDGMENT

The authors thank the MODIS Science Team, especially, Ranga Myneni, for their support. Feedback from a number of NACP investigators has also been very valuable in refining the algorithm for gap-filling and smoothing the MODIS data.

## REFERENCES

- [1] C. O. Justice, et al., "The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research," *IEEE Trans Geosci Remote Sens*, vol. 36, no. 4, pp. 1228-1249, July, 1998.
- [2] R. B. Myneni, S. Hoffman, Y. Knyazikhin et al., "Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data," *Remote Sens. Environ.*, vol. 83, pp. 214-231, 2002.
- [3] P. Jonsson and L. Eklundh, "TIMESAT - a program for analyzing time-series of satellite sensor data," *Computers & Geosci*, vol. 30, pp. 833-845, 2004.
- [4] F. Gao, J. T. Morisette, R. E. Wolfe, G. Ederer, J. Pedelty, E. Masuoka, R. Myneni, B. Tan and J. Nightingale, "Towards Temporally and Spatially Continuous MODIS Land Products," *IEEE Tran Geosci Remote Sens Letters*, submitted.
- [5] E. Masuoka, C. Tilmes, N. Devine, G. Ye and M. Tilmes, "Evolution of the MODIS science data processing system", *Proceedings of the July 2001 IGARSS Conference*, 1454-1457.