A Web-Based Subsetting Service for Regional Scale MODIS Land Products

Suresh K. Santhana Vannan, Robert B. Cook, Susan K. Holladay, Lisa Mai Olsen, Upendra Dadi, and Bruce E. Wilson

***Abstract*—The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor has provided valuable information on various aspects of the Earth System since March 2000. The spectral, spatial, and temporal characteristics of MODIS products have made them an important data source for analyzing key science questions relating to Earth System processes at regional, continental, and global scales. The size of the MODIS product and native HDF-EOS format are not optimal for use in field investigations at individual sites (100 × 100 km or smaller). In order to make MODIS data readily accessible for field investigations, the NASA-funded Distributed Active Archive Center (DAAC) for Biogeochemical Dynamics at Oak Ridge National Laboratory (ORNL) has developed an online system that provides MODIS land products in an easy-to-use format and in file sizes more appropriate to field research. This system provides MODIS land products data in a nonproprietary comma delimited ASCII format and in GIS compatible formats (GeoTIFF and ASCII grid). Web-based visualization tools are also available as part of this system and these tools provide a quick snapshot of the data. Quality control tools and a multitude of data delivery options are available to meet the demands of various user communities. This paper describes the important features and design goals for the system, particularly in the context of data archive and distribution for regional scale analysis. The paper also discusses the ways in which data from this system can be used for validation, data intercomparison, and modeling efforts.**

***Index Terms*—Data management, data processing, MODIS, remote sensing, visualization.**

I. Introduction

THE ORNL DAAC is one of the nine data centers housed under the NASA Earth Observing System Data and Information System (EOSDIS) program [1]. Each of the nine NASA EOSDIS data centers specialize in archiving and distributing data related to a specific earth science discipline. The ORNL DAAC archives and distributes terrestrial biogeochemical dynamics data. This includes field campaign, land validation, regional and global data, and model products that are archived and distributed through the ORNL DAAC's Web portal [http://daac.ornl.gov]. The ORNL DAAC's data sets are primarily from ground-based field investigations and are augmented by data collected through remote-sensing techniques. Among the remote sensing data available at present, Moderate Resolution Imaging Spectroradiometer (MODIS) sensor data is particularly useful for studying biogeochemical dynamics and earth system science. MODIS is a sensor on board two remote sensing satellites: Terra and Aqua. Terra MODIS and Aqua MODIS view the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands. The data are acquired at 1-km, 500-m, and 250-m spatial resolutions [2]. These spectral, spatial and temporal characteristics make MODIS data valuable for studying regional and continental scale dynamics.

MODIS Level 3 Land products in their native format are available in approximately 1200 × 1200 km tiles. MODIS Level 3 Land products consist of derived geophysical variables gridded and/or composited into some earth-based coordinate system [11]. For regional scale analysis, however, user communities require remote sensing data of land characteristics at extents of about 100 × 100 km or even smaller. In some cases, as in flux tower measurements, satellite data are needed for regions less than 25 km^2 [3]. To facilitate analysis at these scales, the ORNL DAAC has developed a value-added system that delivers subsets and visualizations of MODIS land products data for regions at these scales. The MODIS subsetting and visualization tool complements the existing sources of MODIS land products by providing a fast, easy-to-use, and efficient system of delivering MODIS land products. The MODIS Adaptive Processing System (MODAPS) is the production center of the MODIS land product subsets and the Land Processes DAAC (LPDAAC) is the primary distributor of MODIS land products data. Most MODIS data available through these data centers are written in Hierarchical Data Format—Earth Observing System (HDF-EOS) format [4], [5]. While many tools and services are available to access MODIS data and to work with the HDF-EOS data files, users requiring land characteristics information for smaller regions would still have to download many billion bytes of data to retrieve information for their regional scale. For instance, to study a 9-year trend in Normalized Difference Vegetation Index (NDVI) over a field site, a user would have to download nearly 50 gigabytes of 200 individual MODIS HDF-EOS tiles each covering a 10 × 10 degree region (approximately 1200 × 1200 km). The magnitude and volume of data files needed for regional scale analysis constrains users with limited computing resources, especially for users needing subsets for multiple sites and also if the site-subsets require mosaicing from many individual MODIS

Manuscript received February 20, 2009; revised June 12, 2009 and October 02, 2009; accepted October 29, 2009. First published December 22, 2009; current version published January 20, 2010.

S. K. Santhana Vannan, R. B. Cook, S. K. Holladay, L. M. Olsen, and B. E. Wilson are with the Environmental Science Division, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN 37830 USA (e-mail: santhanavans@ornl.gov; cookrb@ornl.gov; wilsonbe@ornl.gov; holladaysk@ornl.gov; olsenlm@ornl.gov).

U. Dadi is with the Center for Spatial Information Science and Systems, George Mason University, Greenbelt, MD 20770 USA (e-mail: udadi@gmu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2009.2036585

1939-1404/$26.00 © 2009 IEEE

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

TABLE I
MODIS PRODUCTS SUBSETTED

MODIS Product	Composite Period	Resolution
Surface Reflectance	8 day	500m
Surface Temperature	8 day	1 km
Land Cover	Annual	500m
Normalized Differential Vegetation Index (NDVI)/ Enhanced Vegetation Index (EVI)	16 day	250m
Lead Area Index(LAI) / fraction of Photosythetically Active Radiation (Fpar)	16 day	1 km
Net Photosynthesis	16 day	1 km
Annual NPP	Annual	1 km
Albedo (Model and Calculated)	Produced every 8 days with 16 days of acquisition	500m
Reflectance – BRDF Adjusted	Produced every 8 days with 16 days of acquisition	500m
Gap-Filled, smoothed Lead Area Index(LAI) / fraction of Photosythetically Active Radiation (Fpar)	16 day	1 km

HDF-EOS tiles. Also, an understanding of the HDF-EOS file format is necessary and software tools and packages have to be downloaded and installed to process the data files for analysis. These constraints, coupled with the demand from the user community to provide Terra MODIS and Aqua MODIS data for smaller regions in readily usable formats, led to the development of the MODIS subsetting and visualization tool.

II. ORNL DAAC MODIS SUBSETTING AND VISUALIZATION

The initial goal of the MODIS subsetting and visualization tool was to provide MODIS land product subsets for field validation sites including Earth Observing System (EOS) core validation sites and ground measurement sites such as FLUXNET. The EOS core validation sites provide users access to field, aircraft and satellite data [6]. FLUXNET is a global network of micrometeorological tower sites that use eddy covariance methods to measure the exchanges of carbon dioxide, water vapor, and energy between terrestrial ecosystem and atmosphere [7]. These subsets were predefined and offered for a limited group of field sites. After the success and popularity of these field site subsets, the ORNL DAAC extended its goal to provide on-demand user-defined MODIS land product subsets for any location on land around the globe.

The ORNL DAAC MODIS subsetting activity is limited to providing MODIS land product subsets. Although a wide gamut of MODIS products are available, a smaller list of MODIS land products was chosen for subsetting. These offerings are those identified as most important for field research. A detailed list of all the MODIS land products subsetted through this activity is given in Table I.

The product list given in Table I covers essential parameters needed to characterize a field site. These products include data from March 2000 forward and are from the collection 5 reprocessing campaign of the MODIS land products. Raw MODIS satellite data are processed by MODAPS using various algorithms for use by researchers. These algorithms are usually updated and improved upon from time to time [8]. At the time this paper was written MODIS collection 5 data represents the latest reprocessed data produced with improved versions of level 1 science algorithms [9]. Special products such as Gap Filled and Smoothed Leaf Area Index are also offered through the subsetting and visualization tool [10]. Although the products listed in Table I are useful in characterizing field sites, the varied spatial, temporal resolution of the products in some cases prohibits the intercomparison of the products. Missing dates and missing values within the composite period might also present some issues for certain analyses, especially when integrating the subsets into models that require continuous data.

The MODIS subsetting and visualization system provides MODIS land product subsets for regions up to 201 $\times$ 201 km. The tool consists of two components that can provide MODIS land product subsets for:

- a collection of predefined field sites;
- any user-defined location on the globe.

These tools are available through Web pages hosted on the ORNL DAAC archive center Web portal. The main entry point for these tools is http://daac.ornl.gov/MODIS/modis.html. Detailed descriptions of each of these tools are provided in the sections to follow.

A. MODIS Land Product Subsets for Selected Field Sites

The MODIS land products subsetting tool for field sites provides land characteristics information generated from the MODIS sensor for selected sites distributed throughout the globe. Subsets are available for more than 1,000 sites (Fig. 1) that cover most major ground measurement networks, validation sites, and locations of interest to the earth science research community.

Sites for the MODIS Subset activity include the EOS core validation sites, FLUXNET sites, BigFoot project sites, LAINet sites, the Global Terrestrial Observing System (GTOS) Global Terrestrial Observing Network (GT-NET) Net Primary Productivity sites, Baseline Surface Radiation Network (BSRN) sites, and other field sites from around the world. The main requirements for inclusion in the MODIS subsets are that sites have accurate coordinates (usually GPS- based with 4 significant decimal places) and there is a commitment to share data and participate in analysis. Site location details and metadata about each of these sites were collected from individual researchers and are made available along with the subsetting tool. Site name, network information, elevation, and contact details are examples

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

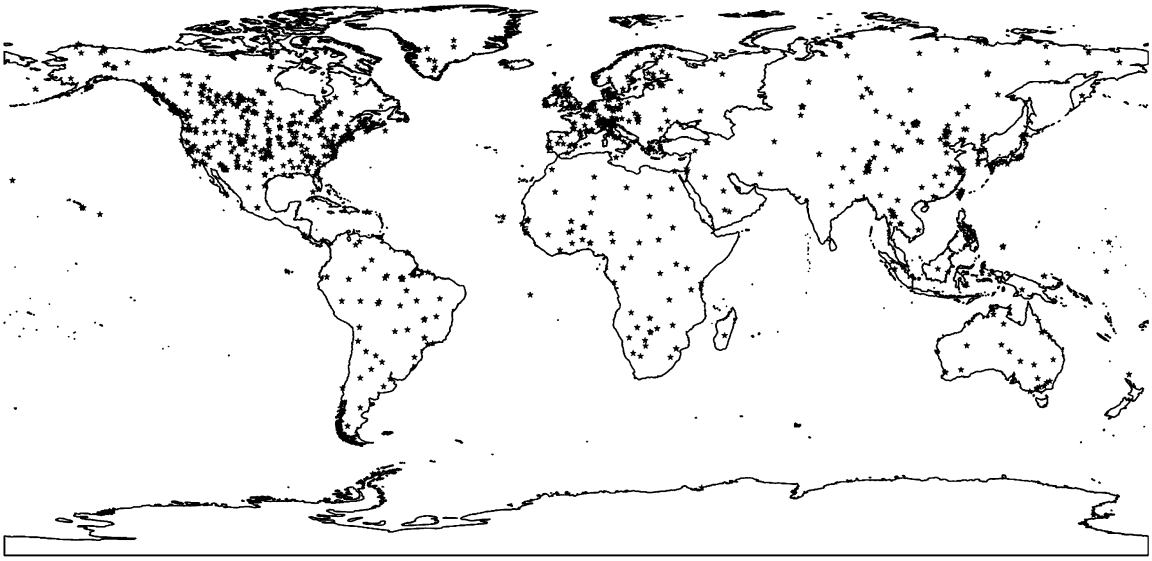


Fig. 1. Location of MODIS field site subsets.

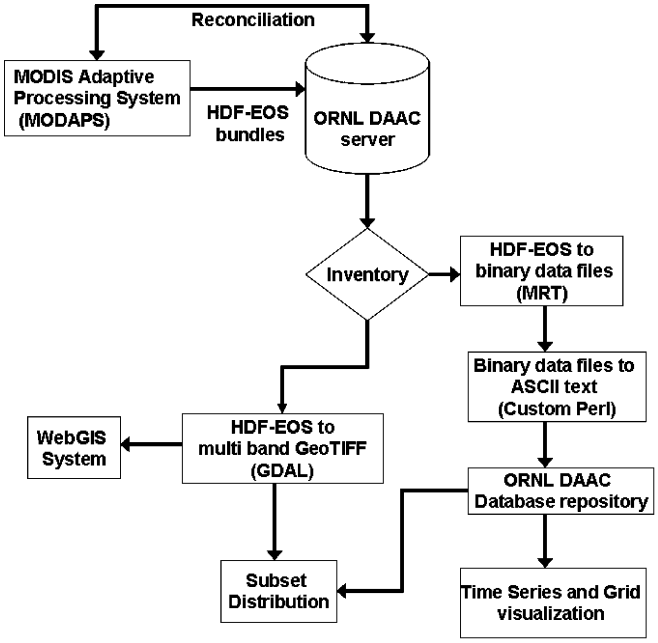


Fig. 2. Steps involved in generation of ASCII and GeoTIFF subsets.

of metadata available for each site. Where available, field data are also included.

To create MODIS land product subsets, the geographic latitude and longitude information for sites is first translated into MODIS level 3 gridded tile coordinates [11]. The gridded tile coordinates identify the MODIS tile containing the site. These translated coordinates are then provided to MODAPS for subset generation. As described in Fig. 2, colleagues at MODAPS prepare subsets of MODIS land products from level 3 MODIS tiles for locations provided by ORNL DAAC. These subsets are created in HDF-EOS data format and are pushed in bundles to ORNL DAAC's server. The bundles are zipped files that contain one or many HDF-EOS files subsetted around the field sites. For 10 years of data, we have received nearly 250,000 bundles containing nearly 6,700,500 individual HDF-EOS files. Since the subsets are created at the production center, the subsets are made available to the ORNL DAAC in near real-time. The HDF-EOS subsets being sent from MODAPS are 25 × 25 km. A subset size of 25 × 25 km was chosen to ensure coverage of the field site. To guarantee that no data were lost due to network and system failures, reconciliation mechanisms between MODAPS and the ORNL DAAC have been installed. The reconciliation mechanisms are a set of software codes that checks subset bundles pushed between MODAPS and ORNL DAAC servers and automatically recovers lost or corrupt bundles. After the subsets have been pushed to ORNL DAAC servers, the bundles are periodically processed into comma delimited ASCII files and multi band GeoTIFF files. The comma delimited ASCII format caters to users requiring satellite data in a tabular format for easy import into spreadsheet, statistical or graphing software packages. GeoTIFF is a spatial format targeted to the GIS community requiring geolocated data for geospatial analysis. The Web entry point for this tool is http://daac.ornl.gov/modisfixedsite.

1) MODIS Land Product Subsets in ASCII Format: To convert the MODIS HDF-EOS subsets into comma delimited ASCII format, the MODIS Reprojection Tool (MRT) created by LPDAAC is used in conjunction with customized Perl scripts [12]. The MRT tool converts the HDF-EOS format MODIS land product subsets into raw binary format. Customized Perl codes are then used to convert the binary data into comma delimited ASCII representations. Each product and site combination is contained in one file, with each row containing an array of all pixels in the 7 × 7 km area for a given Scientific Data Set (SDS) and date. Thus, data for each pixel are represented as a single column in the data file. As data are acquired, it is added to each file. The comma delimited MODIS ASCII file is a plain text file that could be read by any text editor and is independent of operating system and software constraints. With the increasing number of unique data formats, it is essential to provide data that could be read by systems many years into the future [13].

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

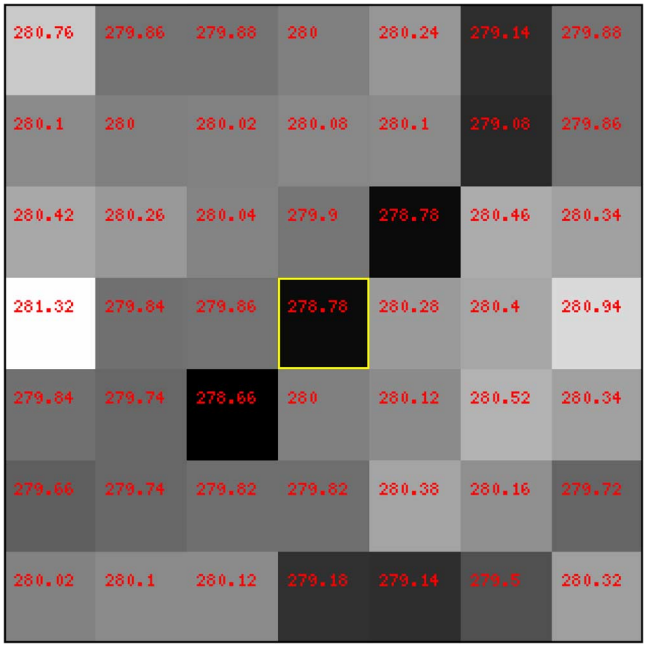


Fig. 3. Daytime MODIS land Surface Temperature for Chestnut Ridge, TN, field site. Data for composite period from November 9, 2007 to November 16, 2007 is shown. Scale factor of 0.2 was applied to the ASCII data values to obtain surface temperature in Kelvin.

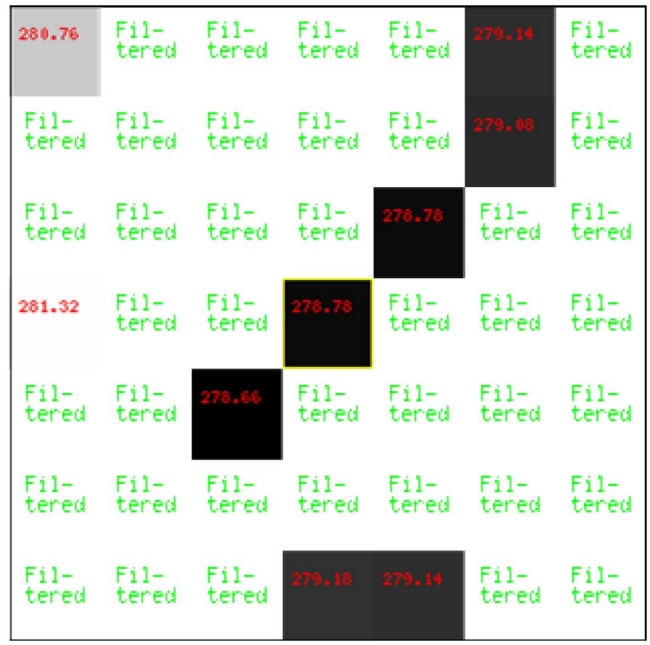


Fig. 4. Daytime MODIS land Surface Temperature for Chestnut Ridge, TN, field site. Data for composite period from November 9, 2007 to November 16, 2007 is shown. Scale factor of 0.2 was applied to the ASCII data values to obtain surface temperature in Kelvin. All pixels with LST error of more than 1 Kelvin have been filtered out.

For this purpose, and considering the relative ease of using the ASCII format, the comma delimited ASCII format was used to deliver the data. The ASCII subset data for field sites, inventory of subsets processed, site metadata, and product metadata are stored in a relational database for archiving and distribution. The ASCII data are delivered through Web pages connected to the relational database. The database tables are indexed and backed up periodically to ensure reliable and fast data delivery and visualization.

During the creation of the ASCII subsets from the MODIS HDF-EOS subsets, only pixels covering 7×7 km area centered on the field sites are considered. For most field sites, a 7×7 km area represents the area of interest to field investigators. A graphical representation of 7×7 km subsets is given in Fig. 3.

The ASCII data file is an aggregate of MODIS land product parameters for all composite dates for a particular field site. Users would have to download one ASCII data file per MODIS product per field site. These comma delimited ASCII files are available through Web pages connected to a relational database and through ORNL DAAC's FTP server.

Along with providing the subsets, the tool also offers Web-based visualizations for quick analysis of the data. The visualizations are generated using CGI scripts and Perl GD graphics library [14]. The visualization pages are generated on the fly using subset data stored in the relational database. Two types of visualization are provided to aid users in their analysis—composite grid visualization and time series visualization.

An example of composite grid visualization is shown in Fig. 3. This represents a MODIS land parameter, namely the daytime land surface temperature for a specific composite period, November 9, 2007 to November 16, 2007 in this example. The time series visualization, as the name implies, shows the trend of a land parameter over time along with some statistical summarization. All the data are filtered by quality control criteria as recommended by the product scientists.

In addition to the default quality control setting, advanced users also have the option of selecting their own quality control criteria for data filtering. Each MODIS data product has quality criteria defined by the product scientist. Product-specific quality controls are available for users to choose their data filtering criteria. For example, users can set the quality criteria for the NDVI/EVI product to extract the highest quality vegetation index pixels. For the NDVI/EVI, product users can also filter out water pixels, possible cloud shadows, snow/ice pixels, mixed clouds, high aerosol quantity pixels, etc. Fig. 4 shows an example of a grid visualization with a user-defined quality filter. In this example the grid visualization is for the same data as shown in Fig. 3; however, a stringent quality criteria has been set to filter out all Land surface temperature pixels with a LST error of more than 1 Kelvin. The LST error flag present within the QC_Day SDS of the $\mathrm{M*11A2}$ Land Surface Temperature product has been used to perform the filter. The tool provides a QC table that illustrates the SDS used for the filtering, the quality criteria used, the bits used for filtering, and a description of the quality criteria. The statistics table available as part of the download options also shows the number pixels in the subset area that passed the quality criteria.

Allowing users to define their quality criteria helps tailor the data and the visualization to meet users' needs. They also help define the usefulness of the data for a particular use case. Also, the ability to extract the highest quality pixels is an important feature of the QA filtering tool. The default and advanced user-defined quality control filtering is available within the grid and

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

time series visualization. The predefined/advanced quality filtered ASCII subset data can be downloaded from within the visualization tools.

Links to product and quality control documentation hosted at LPDAAC and product information such as units, scale factor, and quality control flags are also provided within the visualization pages. Many access routes to these data and tools are available to meet the needs of different user communities. For users preferring direct access to data files, an FTP site holding the ASCII data is provided. The ORNL DAAC also hosts Web pages to provide data files to users interactively. These Web pages allow users to select field sites on the basis of site network, country, land cover class, or continent. Users can also search for sites using the site name. Links to maps of the sites are also available. These maps link directly to Google maps or Google Earth and provide additional context to the data files.

2) MODIS Land Product Subsets: GeoTIFF Format: Although the comma delimited ASCII file format provides the MODIS land product subsets in an easy-to-use format and in a smaller file size, it loses the spatial context of the data file. The ASCII files are plain text files without any internal geolocation information. The location information is stored externally in a relational database and is made available to users through the visualization pages. For users who require geolocation embedded in the data files in order to perform overlay analysis and other GIS operations, the ORNL DAAC also processes the HDF-EOS MODIS land product subsets into multi band GeoTIFF files. GeoTIFF is a powerful nonproprietary format for storing spatial image data in which the files internally store the coordinate information coupled with basic metadata [15]. The GeoTIFF files are generated in the native projection of the MODIS land product subsets. MODIS level 3 land products are generated in Integerized sinusoidal grid projection [2]. Integerized sinusoidal grid cells within a row have the same area but the spatial resolution may vary by up to 0.5% between grid rows [2]. Because the MODIS level 3 sinusoidal grid cells maintain the area, they look skewed when displayed on a geographic coordinate system. The processing flow diagram for generating the GeoTIFFs is shown in Fig. 2. The MODIS land product GeoTIFF subsets for field sites are generated from the HDF-EOS files sent from MODAPS using the Geospatial Data Abstraction Library (GDAL) software tools [16]. Perl codes are used to execute the GDAL software in batch mode and are also used for inventory tracking. The Perl codes perform inventory tracking by identifying the files that require conversion from HDF-EOS format into GeoTIFF and also store information regarding files that have been processed in a relational database. The Perl codes and GDAL generate one GeoTIFF file per product per date per field site. The GeoTIFF files can be overlaid on soil, elevation, land cover, or any other spatial layer using a GIS system. GIS systems can be used to perform overlay analysis, pixel extraction, spatial aggregation, etc.

Web-based visualizations are also provided for the GeoTIFF subsets using WebGIS, which is an internet-based mapping system. WebGIS is used to display the GeoTIFF subsets in the context of other data layers such as soil type, elevation, and land cover. The instance of the WebGIS system is in sinusoidal projection and it allows users to view rendered images of the GeoTIFF files without the need to download each of the data files. Fig. 5 shows a screenshot of the WebGIS system. Download options are also available to access the GeoTIFF files by field site. The GeoTIFFs displayed on the WebGIS system are color coded and the corresponding legends are also displayed. Open Source technologies—Map server [17] and GDAL, and commercial software—ESRI's ARCIMS [18] are used simultaneously for visualization of millions of GeoTIFF subsets. Specialized tools and services have been developed to help users perform some basic queries on the GeoTIFF files, all within the WebGIS environment. Tools to query individual GeoTIFF files and to filter pixels matching specific criteria (such as filtering all pixels with land surface temperature of $<$300 K) are available from within the WebGIS system. The ability to query and visualize the subsets without having to download huge amounts of data, allows users to check if the data products meet their scientific needs before downloading them.

B. MODIS Land Product Subsets for the Globe

Following the success of the MODIS land product subsets for selected field sites, a need for obtaining regional scale ($> 100 \times 100$ km) MODIS subsets for any location on the globe was identified. This demand led to the development of the MODIS global subsetting and visualization tool. Periodic interaction with the members of the research community helped to refine the tool to meet users' needs. Fig. 6 shows the processing flow diagram for the MODIS land product subsets for the globe.

The source data for the MODIS land product subsets for the globe are MODIS collection 5 data from LPDAAC. As explained above, the raw MODIS data are reprocessed time to time with updated and improved algorithms. Collection 5 represents the latest reprocessed MODIS data at the time this paper was written. Collections 4 and 3 are data from earlier reprocessing campaigns. The MODIS data files sent from LPDAAC are level 3 MODIS land products in HDF-EOS format and each file covers a 10×10 degree area (approximately 1200 $\times$ 1200 km area). The total volume of HDF-EOS data files is anticipated to be 35 terabytes to cover 10 years of 12 MODIS land products. These data files are organized in Redundant Arrays of Independent Disks (RAID) and the directory structure has been designed for optimal retrieval, quick subsetting, and data delivery. The directory structure used for storing the files is (Product)-(Year)-(Composite Date)-HDF-EOS File(s). A relational database is used to maintain inventory of the files and necessary metadata related to each of the files. The files are periodically backed up for archival and duplicate/corrupt files are routinely removed from the archive. Although the original MODIS land product HDF-EOS files are archived within the ORNL DAAC, they are not made available to the public directly. LPDAAC is the primary distributor of the HDF-EOS MODIS land product data files. The ORNL DAAC uses the HDF-EOS files as a source from which subsets of the MODIS land products are prepared in ASCII and GIS compatible formats at the regional scale.

MODIS land product subsets are offered through an on-demand processing system that accepts user orders through

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

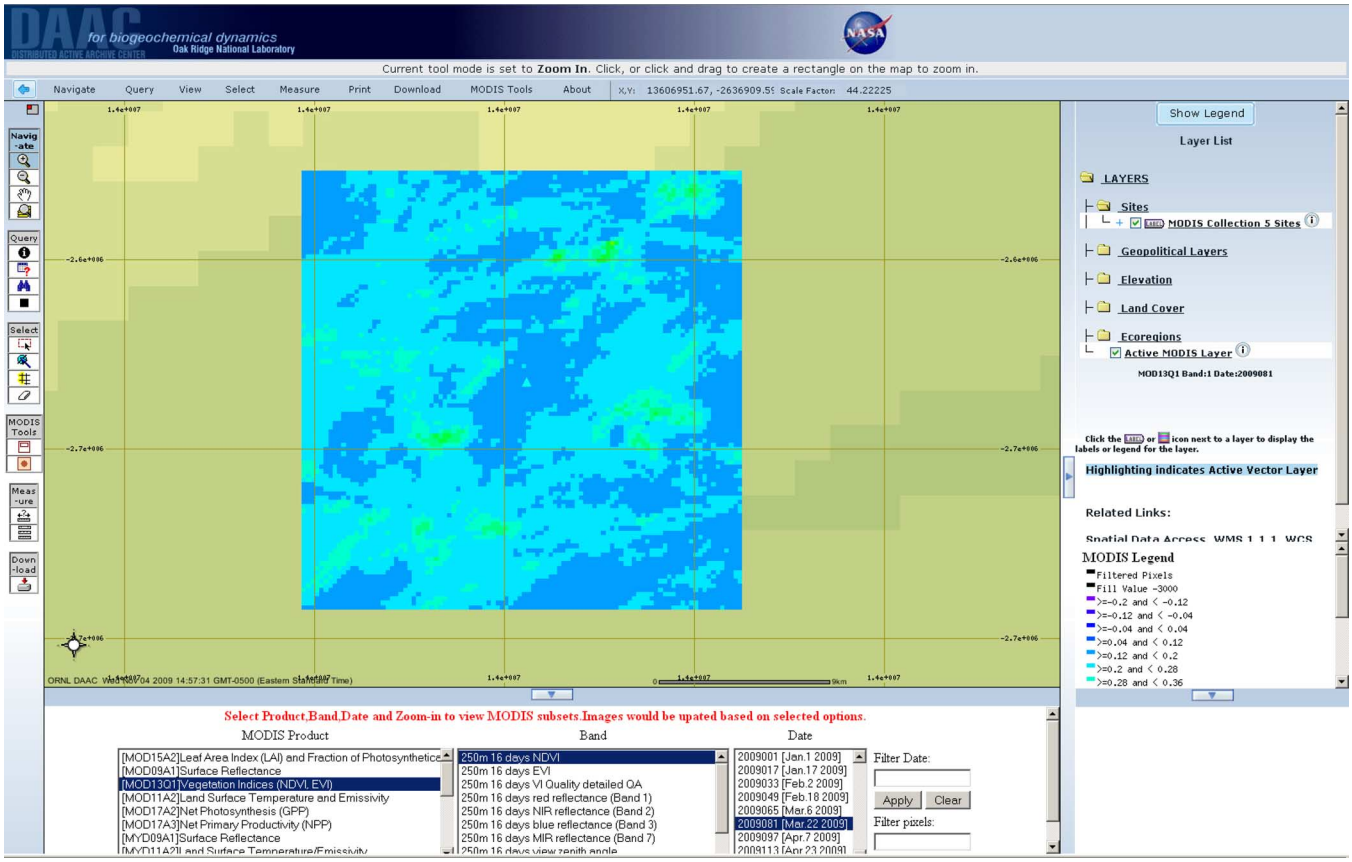


Fig. 5. MODIS NDVI GeoTIFF image for a site in Australia available through the WebGIS.

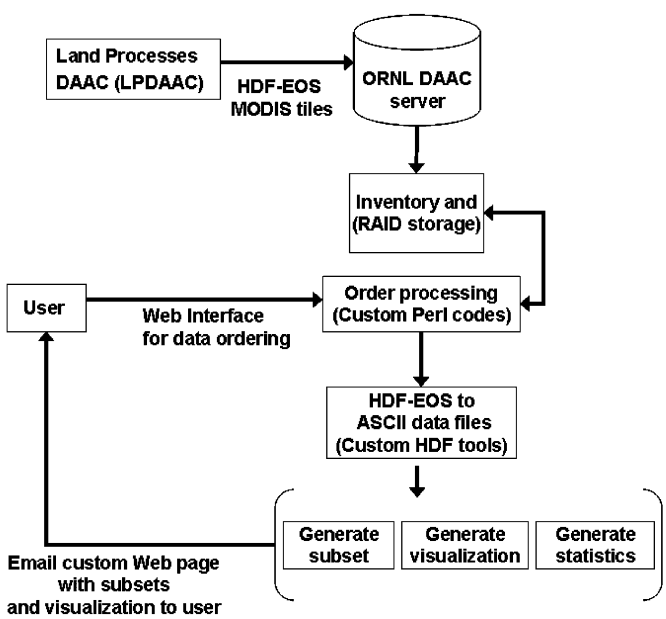


Fig. 6. Components of the on-demand MODIS land product subsetting and visualization tool for the globe.

a Web interface. The Web entry point for this tool is http://daac.ornl.gov/modisglobal. Users define the location of their desired subset, size, and time period of interest and provide an email address for data delivery through the Web interface. After an order for subset has been submitted by the user, the order is fulfilled by a series of Perl codes. The MODIS land product subsets for the globe creates mosaics of the MODIS land product where necessary. For example, locations falling at the edge of 10 × 10 degree tiles are mosaiced from adjacent data files. The order processing system first creates the ASCII subsets using a customized C program. The C program was created using HDF libraries available from National Center for Supercomputing Applications (NCSA) and has been optimized for speed by improving file reading performance and utilizing memory caches effectively. Custom Perl code, Perl GD graphics library, and Perl statistics package are then used to generate the MODIS land product ASCII subsets, GeoTIFF files, a statistical summary of the subset, MODIS land cover filtered data, and time series visualizations of the subset. Detailed documentations of all these files are also provided.

The system generally provides the subsets within 60 min. The time taken to deliver the subsets could be in the order of seconds up to several hours depending on the size of the subset and product spatial resolution. 500-m and 250-m resolution data take longer to process. Many optimizations have been made to offer the subsets as quickly as possible. Details regarding the optimizations and performance are provided in the evaluation section of the paper.

The data files and time series graphs are packaged into a customized Web page and delivered to the user through an email. The email sent to the user contains a URL link to the customized

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

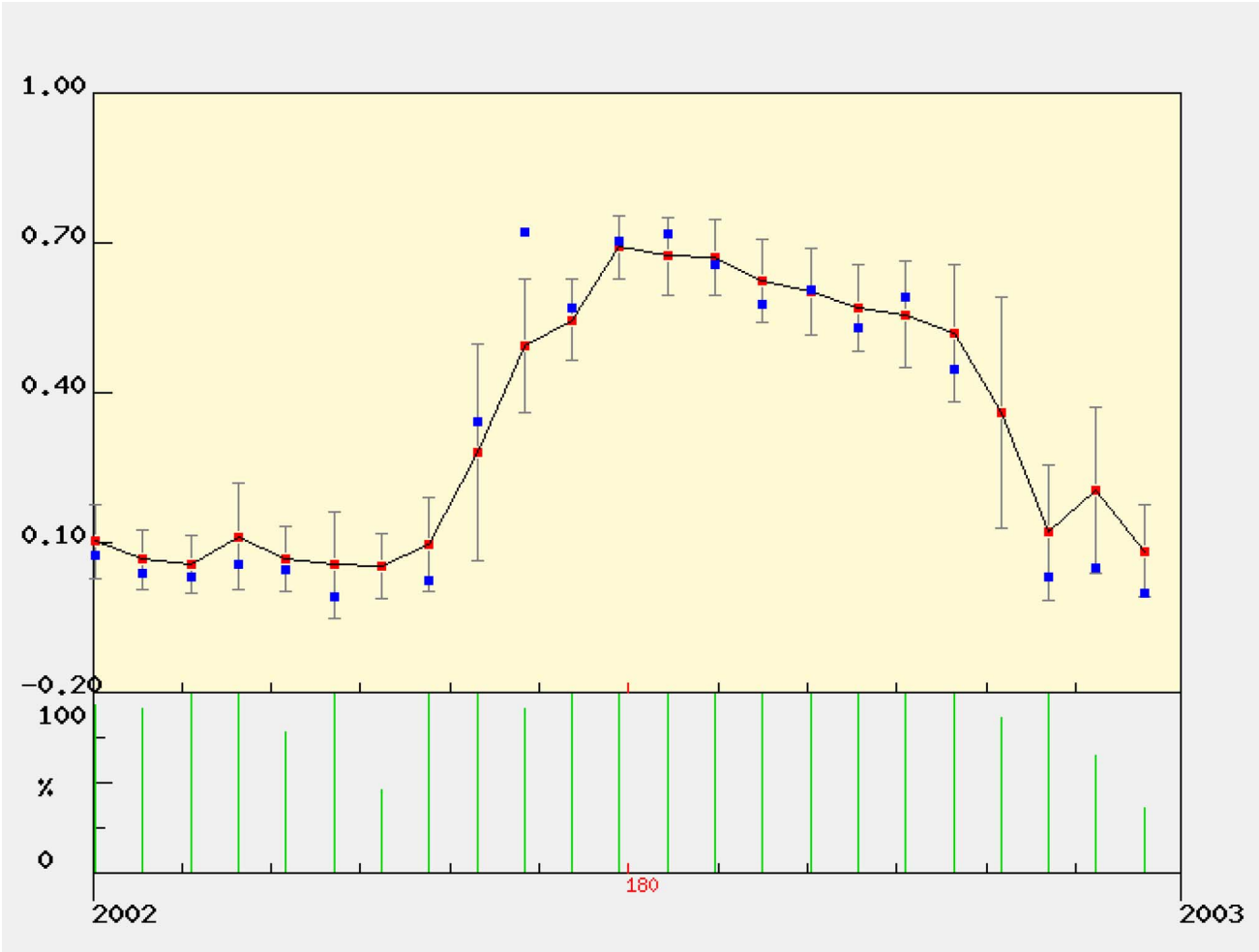


Fig. 7. Example time series graph generated by the on-demand subsetting service for Latitude, Longitude location 44.17, −110.52 and for a 6.25 × 6.25 km area. The graph shows the Normalized Differential Vegetation Index (NDVI) trend during year 2002. The vertical bars at the bottom of the graph denote percentage of pixel passing the default quality control criteria as defined by the science team. Standard deviation (grey lines), mean value of all pixels in the subset area (red) and pixel value of the center pixel (blue) are also shown.

Web page. An example of a time series graph generated by the on-demand subsetting service and delivered through a customized Web page is shown in Fig. 7. The time series plot provides the percentage of pixels that passed the quality criteria, the mean value of all the good quality pixels in the subset area, standard deviation of the pixels and the value of the center pixel. The mean value of the pixels might vary significantly from the center pixel value due to land cover differences between the center pixel and the surrounding area. To extract pixels matching the land cover of the center pixel an additional time series graph is also provided. This time series graph plots values of pixels matching the land cover of the center pixel.

Although the MODIS land product subsetting tool for the globe supersedes the MODIS land product subsetting tool for fixed sites, there are two differences. The MODIS global tool allows users to create subsets for any location and any available time period. When the tool processes a user's request for subsets, only data available at the time of request are delivered; new data are not appended to existing orders as with the MODIS land product subsets for fixed sites. Also, the MODIS land product subsets for fixed sites is designed for field sites, and therefore collects metadata about each of the sites and also links to ground data where available.

Since the tool was made public, more than 11,000 subset orders have been fulfilled in a span of two years, generating nearly 1.8 terabytes of subset data. These subsets were generated from nearly 35 terabytes of compressed MODIS HDF-EOS tiles. The average size of the subset delivered is approximately 0.167 gigabytes. The small average data volume delivered is expected and is desired. The subsetting tool converts the original HDF-EOS files into ASCII files. During the process of file format conversion and subsetting, the size of the file is reduced by at least a factor of 7. Furthermore, if the subset files are greater than 4 gigabytes, they are compressed into zip files to reduce the network transfer time. For researchers to individually obtain the subsetted data equivalent to the ones delivered through the subsetting tool, they would have had to download nearly 500 terabytes of HDF-EOS files. These statistics were generated using logs of completed data orders and they highlight the significance of the on-demand MODIS subsetting and visualization tool.

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

III. Application Areas and Related Work

Comparisons of MODIS products to ground-based measurements are critical for evaluating and validating the accuracy of MODIS data. The ORNL DAAC MODIS subsets provide the data in a form that can be readily used in validation. Heinsch *et al.* have used MODIS land product subsets from 15 field sites to evaluate MODIS terrestrial productivity data using eddy flux network observation. Four years (2000 to 2003) of MODIS data were compared to eddy CO2 flux-bases estimates across diverse land cover and climate regimes. Eight-day mean, total seasonal, and total annual values of GPP were calculated for each of the 15 MODIS subsets and compared with the site estimates of GPP [19]. Zhao *et al.* validated improvements made to the MODIS terrestrial gross and Net Primary Production (NPP) data product by using the MODIS subsets. As noted by Zhao, the subsets are provided to researchers for validation efforts and these ongoing validation activities are vital for evaluations of model performance and refinements [20]. In other studies, tower-based LAI and NDVI measurements have been compared with MODIS data using the subsets [21].

MODIS land product subsets can also be used to drive decision support, ecosystem, and biogeochemical models. Models that require various land parameters such as NDVI, LAI, fPAR, etc., as input can utilize the subsets generated by the tool. MODIS subsets have also been used in field and satellite intercomparison studies. Cheng *et al.* used field data, AVIRIS, and MODIS subset data in a multiscale ecosystem intercomparison study [22]. *In-situ*, AVIRIS airborne and MODIS satellite data were compared for a small spatial area to identify performance of the MODIS NDVI, EVI, and fPAR data. Cheng *et al.* were able to extract a single pixel from the MODIS subsets for this study. MODIS subsets from ORNL DAAC have also been used in validation of algorithms developed for other satellite observations. MODIS fraction of absorbed photosynthetically active radiation (fPAR) subsets were used in validating fPAR generating algorithms from Meteosat Second Generation (MSG) satellite [23]. Huete *et al.* examined the seasonal vegetation patterns of the tropical forest sites with the subsets obtained from the tool [24]. Chasmer *et al.* used the MODIS subsets in their research to develop a lidar-based hierarchical approach for assessing the MODIS fPAR product [25]. Xia *et al.* used the MODIS subsets in conjunction with Ameriflux flux tower measurements to derive an estimate of the net ecosystem carbon exchange for the conterminous United States [26].

Besides such science applications, MODIS subsets have been widely used in class-room education and assignments. Examples of the class-room exercises and lectures using MODIS subsetting tool will be made available on the ORNL DAAC MODIS Web page. The subsets derived from the MODIS subsetting tool can also be spatially and temporally aggregated and the resulting data can be compared with other collocated datasets such as model simulation, remote sensing, and field data. Such inter comparisons aid in validation and also in improving models and algorithms that are used to process remote sensing products.

The MODIS subsetting tool is unique in its capability in the sense that it provides a tailored tool for serving MODIS subsets for regional scale analysis. Although the tool is unique in its overall functionality, there are many technically similar tools available. There are innumerable tools available for subsetting and visualization, most of which are based on standards such as Open Geospatial Consortium (OGC) and OPenDAP [27]. These standards are built to serve many different data formats and can be customized for a specific application. Giovanni, developed by the NASA's Goddard Earth Sciences Data and Information Services Center, utilizes the OpenDAP and OGC standards to both consume and serve data [28]. Physical Oceanography DAAC has developed a tool called POET that provides subsets and visualization of several ocean data products [29]. Geobrain, developed at the George Mason University, offers MODIS Level 1B products in HDF-EOS format through OGC standards. OGC standards-based subsetting is used in geobrain to alleviate the difficulty of accessing huge volumes of MODIS data using the HDF file format [30].

IV. Tool Evaluation

The following summary provides an evaluation of the tool's performance with respect to speed, efficiency and ease of use.

Evaluation of Speed of Data Delivery: We looked at 11,233 orders received September 2007 to present (June 11, 2009). The mean time of data delivery for these 11,233 orders was 25.8 min. The time for data delivery ranges from a few seconds to a maximum of ~25 h. The time taken for data delivery is a function of the size of the subsets, time range requested, orders in queue, and product resolution. Large temporal and spatial subset orders take longer to generate. Also, higher resolution MODIS products such as the 250 m NDVI/EVI product and the 500 m surface reflectance product take longer to generate. For the last two years, we have received an average of 16 orders each day. Delivering subsets in ~25 min in an easy-to-use format with various visualization options allow researchers to focus on analyzing the data and shifts the burden of data assembly to the data provider. To compare the performance of the tool, we downloaded 9 years of MODIS NDVI/EVI (MOD13Q1) tiles from LPDAAC. The total volume of the tiles downloaded was ~50 GB and took about 9 h to download on a high speed 100 MB/s network with a download rate of 1500 KB/s. Users wanting to extract the data for their region would also have to process the data. They would have to build special processing steps to handle subset creation using specialized HDF software. Comparatively, the MODIS subset tool offers just the subsets for their region of interest, minimizing download and processing time to a great extent. Additionally, since the subsets are offered in ASCII format, the volume of the data would be much lower; for example, a 200 × 200 NDVI/EVI 9 year subset would be less than 2 GB.

Evaluation of Efficiency and Ease of Use: Efficiency of the tool can be characterized in many different ways. We look at the following aspects to characterize the efficiency of the tool:

- frequency of failed orders and Web page issues;
- user comments;
- software bugs;
- usage statistics.

Over the last two years of operation the MODIS Global subsetting tool has had only one failed order. Various fail-safe mechanisms have been put in place to ensure the orders are fulfilled without any glitches. Web page problems such as broken

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

links and undesirable browser behavior have been reported and have been promptly fixed. We have received periodic positive feedback about the tool and the usage of the tool has been consistent. For the fixed sites tool, we have seen an average use of ~6,000 data downloads every month. The global tool usage is ~500 orders every month.

The few limitations of the tool include gaps in the subset data. The gaps in the data could be due to data not generated from MODIS due to sensor downtime or other processing issues. Some gaps are also caused by network failure when transferring the underlying data files from LPDAAC or from MODAPS (refer the processing flow diagrams shown earlier). Every few months, these gaps caused to due network issues are filled. In some cases, the gaps in the data cause significant issues—specifically, in the case of some models that require continuous data. The gap-filled and smoothed products hopefully help to address some of these issues. In the future, an automatic gap-filling technique and a network reconciliation technique may be considered to fill missing data and to prevent loss of data due to network failure.

V. Future Directions

There is scope for improvement in the current version of the MODIS subsetting and visualization tool. Specifically, many users have expressed interest in obtaining subsets of daily MODIS products, which would be particularly useful in studying short term phenomenon. Users have also expressed interest in obtaining subsets greater than 201 × 201 km. Another area for improvement is to include the option of providing co-located subsets with the exact same spatial dimensions. When requesting a subset through the MODIS Global tool, users cannot get the subsets for the exact same spatial area if the products they request have different spatial resolution. This limitation to the system has been identified and will be addressed in the future versions of the system. One of the main goals of the subsetting tool is to help with site characterization. Right now, we provide the MODIS subsets and relevant links to site data such as flux tower measurements. This can be improved by providing visualizations of the site data in conjunction with the MODIS subset time series. In addition to providing improved functionality to the tool, the tool can be expanded further by adding more data products to its subsetting suite. Gap-filled and smoothed data products, phenology, and long-term data records are being prepared by various research groups; these are candidate data products that can be added to the subsetting and visualization interface.

To help users understand the tool and its possible science application, we would be providing examples of application areas, FAQs, tutorials, help files, references to papers utilizing the MODIS subsetting tool, etc. We are also experimenting with providing subsets through Simple Object Access Protocol (SOAP)-based Web service (http://daac.ornl.gov/modiswebservice). The Web service allows users to obtain MODIS subsets programmatically. Based on user interest and usage, we will be expanding this Web service further. We will also be providing a bulk ordering interface to make it easier for users to order multiple subsets all at once.

Acknowledgment

The authors would like to thank S. Margle (formerly of the ORNL DAAC) for his significant contribution as lead developer through the early years of this activity. The authors would also like to thank all members of the ORNL DAAC User Working Group, steering committee, J. Dwyer, R. J. Olson, and MODIS land product scientists, especially J. Randerson, C. Schaaf, K. Didan, A. Huete, M. Friedl, and J. Morissette for their valuable suggestions; the LPDAAC and MODAPS team for their data support, in particular C. Jenkerson, C. Davidson, G. Fu, G. Ye, and N. Saleous; and all current and former members of the ORNL DAAC data archive for their team support.

References

[1] B. Kobler, "NASA Earth Observing System Data Information System (EOSDIS)," in *Proc. 11th Symp. Mass Storage Systems*, 1991, p. 18.

[2] E. Masuoka, A. Fleig, R. E. Wolfe, and F. Patt, "Key characteristics of MODIS data products," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, pp. 1313–1323, 1998.

[3] S. W. Running, D. D. Baldocchi, D. P. Turner, S. T. Gower, P. S. Bakwin, and K. A. Hibbard, "A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and eos satellite data—A comparison of estimates from a constrained quantum-use efficiency model and eddy correlation," *Remote Sens. Environ.*, vol. 70, pp. 108–127, 1999.

[4] E. Masuoka, R. Wolfe, S. Sinno, G. Ye, and M. Teague, "A disk-based system for producing and distributing science products from MODIS," in *Proc. IEEE Int. Symp. Geoscience and Remote Sensing*, 2007, pp. 3043–3046.

[5] C. Reber and H. Ramapriyan, "Earth science satellite remote sensing data from the EOS data and information system," in *Proc. Earth Science Satellite Remote Sensing*, 2006, pp. 11–17.

[6] J. L. Privette, J. T. Morisette, C. Justice, and D. Starr, "EOS global land validation network," in *Proc. IEEE Int. Symp. Geoscience and Remote Sensing*, 1999, vol. 5, pp. 2587–2589.

[7] D. Baldocchi, E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, R. Evans, J. Fuentes, A. Goldstein, G. Katul, B. Law, X. Lee, Y. Malhi, T. Meyers, W. Munger, W. Oechel, K. T. Paw, K. Pilegaard, H. P. Schmid, R. Valentini, S. Verma, T. Vesala, K. Wilson, and S. Wofsy, "FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities," *Bull. Amer. Meteorol. Soc.*, vol. 82, p. 2415, Nov. 2001.

[8] 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. Environ.*, vol. 83, pp. 3–15, 2002.

[9] X. Xiong, V. V. Salomonson, J. Kuyper, L. Tan, K. Chiang, J. Sun, and W. L. Barnes, "Status of the MODIS level 1B algorithms and calibration tables," in *Proc. IEEE Int. Symp. Geoscience and Remote Sensing*, 2005, pp. 2231–2234.

[10] F. Gao, J. T. Morisette, R. E. Wolfe, G. Ederer, J. Pedelty, E. Masuoka, R. Myneni, B. Tan, and J. Nightingale, "An algorithm to produce temporally and spatially continuous MODIS-LAI time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 60–64, Jan. 2008.

[11] R. E. Wolfe, D. P. Roy, and E. Vermote, "MODIS land data storage, gridding, and compositing methodology: Level 2 grid," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, pp. 1324–1338, 1998.

[12] J. Dwyer and G. Schmidt, "The MODIS reprojection tool," in *Proc. Earth Science Satellite Remote Sensing*, 2006, pp. 162–177.

[13] M. Hedstrom, "Digital preservation: A time bomb for digital libraries," *Lang. Res. Eval.*, vol. 31, pp. 189–202, 1997.

[14] S. Wallace, *Perl Graphics Programming*. Sebastopol, CA: O'Reilly, 2002, p. 25.

[15] N. Ritter and M. Ruth, "The GeoTiff data interchange standard for raster geographic images," *Int. J. Remote Sens.*, vol. 18, pp. 1637–1647, 1997.

[16] G. Walter, F. Warmerdam, and P. Farris-Manning, "An open source tool for geospatial image exploitation," *IEEE Int.*, vol. 6, pp. 3522–3524, 2002.

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.

[17] R. R. Vatsavai, S. Shekhar, T. E. Burk, and S. Lime, "Umn-mapserver: A high-performance, interoperable, and open source web mapping and geo-spatial analysis system," in *Proc. GIScience*, 2006, pp. 400–417.
[18] C. S. Jensen, M. Schneider, B. Seeger, and V. J. Tsotras, Eds., "Advances in spatial and temporal databases," in *Proc. 7th Int. Symp. SSTD*, 2001, vol. 2121, Lecture Notes in Computer Science, p. 396, Springer.
[19] F. A. Heinsch, M. Zhao, S. W. Running, J. S. Kimball, R. R. Nemani, K. J. Davis, P. V. Bolstad, B. D. Cook, A. R. Desai, D. M. Ricciuto, B. E. Law, W. C. Oechel, H. Kwon, H. Luo, S. C. Wofsy, A. L. Dunn, J. W. Munger, D. D. Baldocchi, L. Xu, D. Y. Hollinger, A. D. Richardson, P. C. Stoy, M. B. S. Siqueira, R. K. Monson, S. P. Burns, and L. B. Flanagan, "Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1908–1925, Jul. 2006.
[20] M. Zhao, F. A. Heinsch, R. R. Nemani, and S. W. Running, "Improvements of the MODIS terrestrial gross and net primary production global data set," *Remote Sens. Environ.*, vol. 95, pp. 164–176, 2005.
[21] T. B. Wilson and T. P. Meyers, "Determining vegetation indices from solar and photosynthetically active radiation fluxes," *Agricult. Forest Meteorol.*, vol. 144, pp. 160–179, 2007.
[22] Y. Cheng, J. A. Gamon, D. A. Fuentes, Z. Mao, D. A. Sims, H.-l. Qiu, H. Claudio, A. Huete, and A. F. Rahman, "A multi-scale analysis of dynamic optical signals in a Southern California chaparral ecosystem: A comparison of field, AVIRIS and MODIS data," *Remote Sens. Environ.*, vol. 103, pp. 369–378, 2006.
[23] F. Camacho-de Coca, J. Garcia-Haro, J. Melia, and J. L. Roujean, "Prototyping algorithm for retrieving FAPAR using MSG data in the context of the LSA SAF project," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, 2007, pp. 1016–1020.
[24] A. R. Huete, N. Restrepo-Coupe, P. Ratana, K. Didan, S. R. Saleska, K. Ichii, S. Panuthai, and M. Gamo, "Multiple site tower flux and remote sensing comparisons of tropical forest dynamics in Monsoon Asia," *Agricult. Forest Meteorol.*, vol. 148, pp. 748–760, 2008.
[25] L. Chasmer, C. Hopkinson, P. Treitz, H. McCaughey, A. Barr, and A. Black, "A lidar-based hierarchical approach for assessing MODIS fPAR," *Remote Sens. Environ.*, vol. 112, pp. 4344–4357, 2008.
[26] J. Xiao, Q. Zhuang, D. D. Baldocchi, B. E. Law, A. D. Richardson, J. Chen, R. Oren, G. Starr, A. Noormets, S. Ma, S. B. Verma, S. Wharton, S. C. Wofsy, P. V. Bolstad, S. P. Burns, D. R. Cook, P. S. Curtis, B. G. Drake, M. Falk, M. L. Fischer, D. R. Foster, L. Gu, J. L. Hadley, D. Y. Hollinger, G. G. Katul, M. Litvak, T. Martin, R. Matamala, S. McNulty, T. P. Meyers, R. K. Monson, J. W. Munger, W. C. Oechel, U. K. T. Paw, H. P. Schmid, R. L. Scott, G. Sun, A. E. Suyker, and M. S. Torn, "Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data," *Agricult. Forest Meteorol.*, vol. 148, pp. 1827–1847, 2003.
[27] P. Cornillon *et al.*, "OPeNDAP: Accessing data in a distributed, heterogeneous environment," *Data Sci. J.*, vol. 2, pp. 164–174, 2003.
[28] S. W. Berrick, G. Leptoukh, J. D. Farley, and H. Rui, "Giovanni: A Web service workflow-based data visualization and analysis system," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 106–113, Jan. 2009.
[29] K. E. Case, A. W. Bingham, R. W. Berwin, E. M. Rigor, and R. G. Raskin, "Ocean surface topography data products and tools," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, 2004, vol. 1, p. 662.
[30] L. Di, "GeoBrain-A web services based geospatial knowledge building system," in *Proc. NASA Earth Science Technology Conf.*, Palo Alto, CA, Jun. 22–24, 2004, (8 pages. CD-ROM).

Suresh K. Santhana Vannan received the B.E. degree in geoinformatics from Anna University, Chennai, India, in 2001, and the M.S. degree in meteorology from the University of Maryland in 2003.

He was with the University of Maryland, Department of Geography, working primarily on WebGIS distribution of MODIS active fire locations. He joined the Oak Ridge National Laboratory, Oak Ridge, TN, in 2005 and has been working for NASA's Distributed Active Archive Center (DAAC) since then. His research interests are geospatial data processing and dissemination, GIS interoperability, and GIS-based knowledge discovery.

Robert B. Cook received the Ph.D. degree in geochemistry from Columbia University, New York.

He is currently a Distinguished Research Scientist at Oak Ridge National Laboratory, Oak Ridge, TN. His research interests include biogeochemistry, the global carbon cycle, and ecological informatics.

Susan K. Holladay received the B.S. degree in chemistry from the University of Tennessee in 1978.

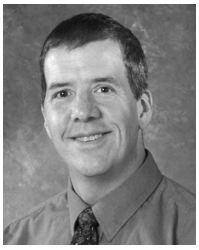
Beginning work at the Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, immediately after receiving the B.S. degree, she has worked in all areas of environmental sample analysis, starting with field work, chemical analysis, and data management over the last 31 years. Currently, she is the ORNL Bioenergy Science Center Laboratory Information Management Task Leader in the Biosciences Division at ORNL.

Lisa Mai Olsen received the B.A. degree in geography in 1993 from the University of Tennessee and the M.S. degree in geography in 1995 from the University of Wisconsin.

She was with the EROS Data Center as a Staff Scientist and as a GIS Analyst for the Knoxville-Knox County Metropolitan Planning Commission. She is currently a Research Associate in the Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN. She is interested in applications of remote sensing technology in geography.

Upendra Dadi received the B.S. degree in civil engineering from the Indian Institute of Technology, Kharagpur, and the M.S. degree in geodetic science from The Ohio State University. He is currently pursuing the Ph.D. degree in the Department of Georgraphy and Geoinformation Science, George Mason University (GMU), Fairfax, VA. His dissertation is about enabling semantic discovery of Earth Observation Data and Services.

He is a Graduate Research Assistant at the Center for Spatial Information Science and Systems, GMU. He was a Post-Masters Research Associate at Oak Ridge National Laboratory, Oak Ridge, TN. His research interests include various aspects of use of geoinformation technologies within cyberinfrastructure for enabling e-science.



Bruce E. Wilson received the Ph.D. degree in analytical chemistry from the University of Washington.

He is currently the Group Leader for the Environmental Data Science and Systems Group, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, He is currently the Group Leader for the Environmental Data Science and Systems Group.

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 13,2010 at 12:11:59 EST from IEEE Xplore. Restrictions apply.