MULTI-SENSOR VEGETATION INDEX AND LAND SURFACE PHENOLOGY EARTH SCIENCE DATA RECORDS IN SUPPORT OF GLOBAL CHANGE STUDIES: DATA QUALITY CHALLENGES AND DATA EXPLORER SYSTEM

by

Armando Barreto-Munoz

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL AND BIOSYSTEMS ENGINEERING

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2013

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Armando Barreto-Munoz, titled MULTI-SENSOR VEGETATION INDEX AND LAND SURFACE PHENOLOGY EARTH SCIENCE DATA RECORDS IN SUPPORT OF GLOBAL CHANGE STUDIES: DATA QUALITY CHALLENGES AND DATA EXPLORER SYSTEM and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ Date: 04/25/2013
Muluneh Yitayew

_____ Date: 04/25/2013
Kamel Didan

_____ Date: 04/25/2013
Donald Slack

_____ Date: 04/25/2013
Richard Hawkins

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 07/24/2013
Dissertation Director: Muluneh Yitayew

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Armando Barreto-Munoz

ACKNOWLEDGEMENTS

The author of this document would like to thank the following persons:

1. Dr. Yitayew and Dr. Didan for all their help, direction and time they spent with me. In addition a very special thanks to Dr. Didan for providing the funding of the research through the NASA grant #NNX08AT05A.

2. Dr. Slack, and Dr. Hawkins as committee members.

3. Javier Rivera and Abd Salam El Villaly for their help and support at the lab.

4. The IT support group at the Electrical and Computing Engineering.

5. The ABE Department professors and staff for their general support.

DEDICATION

I thank my parents Javier Barreto and Eva Munoz, my brother Reynaldo and sisters Evelia, Felisa and Teresa for their unselfish and unconditional support even in the midst of adversities.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

LIST OF FIGURES

LIST OF FIGURES - *Continued*

LIST OF TABLES

ABSTRACT

Synoptic global remote sensing provides a multitude of land surface state variables. The continuous collection, for more than 30 years, of global observations has contributed to the creation of a unique and long term satellite imagery archive from different sensors. These records have become an invaluable source of data for many environmental and global change related studies. The problem, however, is that they are not readily available for use in research and application environment and require multiple preprocessing. Here, we looked at the daily global data records from the Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS), two of the most widely available and used datasets, with the objective of assessing their quality and suitability to support studies dealing with global trends and changes at the land surface. Findings show that clouds are the major data quality inhibitors, and that the MODIS cloud masking algorithm performs better than the AVHRR. Results show that areas of high ecological importance, like the Amazon, are most prone to lack of data due to cloud cover and aerosols leading to extended periods of time with no useful data, sometimes months. While the standard approach to these challenges has been compositing of daily images to generate a representative map over a preset time periods, our results indicate that preset compositing is not the optimal solution and a hybrid location dependent method that preserves the high frequency of these observations over the areas where clouds are not as prevalent works better.

Using this data quality information the Vegetation Index and Phenology (VIP) Laboratory at The University of Arizona produced over 30 years of seamless sensor independent record of vegetation indices and land surface phenology metrics. These data records consist of 0.05-degree resolution global images for daily, 7-days, 15-days and monthly temporal frequency. These sort of remote sensing based products are normally made available through the internet by large data centers, like the Land Processes Distributed Active Archive Center (LP DAAC), however, in this project an online tool, the VIP Data Explorer, was developed to support the visualization, exploration, and distribution of these Earth Science Data Records (ESDRs) keeping it closer to the data generation center which provides a more active data support and distribution model. This web application has made it possible for users to explore and evaluate the products suite before download and use.

CHAPTER 1: INTRODUCTION

I. Explanation of the Problem and its Context

A very long data record suitable for studies dealing with long term environmental research and analysis is only possible by merging data from multiple sensors. The Advanced Very High Resolution Radiometer (AVHRR) data is available since July 1981, while data from the Terra and Aqua Moderate Resolution Imaging Spectrometer (MODIS) are available since February, 2000 and June 2002 respectively (Justice et al, 2002). However, both records, as is the case for all space-borne based observations, suffer from clouds, aerosols, spatial coverage gaps and inconsistent atmosphere correction approaches. To reach the desired long term and high quality data records required by global change these factors must be accounted. These issues are location and season dependent making it even more difficult to construct a consistent time series.

Clouds cover about 60% of the earth's surface at any given time (Kawamoto et al., 2000). Certain locations in the tropics are covered by clouds for most of the times making it difficult to obtain cloud free data at reasonable intervals supportive of terrestrial vegetation dynamic research.

Other atmospheric phenomenon that affects the collection of satellite imagery is high concentrations of aerosols. Their variation in time and space makes them difficult to quantify and characterize (Hansen et al., 1997; Kaufman et al., 2002.). They interfere

with the direct path of light by scattering (Levy et al., 2005), absorbing, and shading the earth's surface, creating a combined attenuation of the radiation (Kaufman et al., 2001) and causing the recordings to be inaccurate.

Considering the impact clouds and aerosols have on data from AVHRR and MODIS, this work was undertaken to quantitatively characterize these global data records, and to assess their quality and to establish the spatial and temporal probability of collecting suitable and frequent data to support the kind of research required by global change studies.

The Vegetation Index and Phenology Laboratory (VIP Lab) at The University of Arizona in collaboration with researchers from the University of Hawaii, Boston University and the USGS-EROS LP-DAAC, was supported by NASA Making Earth System Data Records for Use in Research Environments MEaSURES) project to generate over 30 years multi-sensor global long term data record of vegetation index and land surface phenology. In this project, the goal was to aid in the design and development of algorithms and applications to process these long term datasets and the creation of a web application to support these data records.

II. Explanation of Dissertation Format

The main body of this dissertation is contained in two appendices (A and B). These are individual research manuscripts that are logically connected and integrated into the dissertation as a whole. The manuscripts are based on the AVHRR and MODIS datasets processed at the Vegetation Index and Phenology Laboratory at the University of Arizona.

A. Relationship of the Appended Manuscripts

In the first manuscript, the 30-years daily global surface reflectance data from the AVHRR and MODIS sensors were analyzed to identify their per-pixel quality and guide the data filtering, data merging, and gap filling. The second manuscript, describes and online web application created with the purpose of supporting these data records by providing an online tool to explore, visualize, interact and distribute this 30-years long term data record.

B. Contribution of the Author

While the overall project was a multi-institution mission, the bulk of the work was undertaken at the VIP Lab at the University of Arizona, where this research work was carried. With guidance from the dissertation committee and project colleagues, the author devised a series of processes for data acquisition and algorithms coding for data processing and analysis. Additionally, the author implemented and coded the client/server

application part of this project. Generating a long term multi-sensor data record to support global change research is the hall mark of this research. The author's contributions are specifically in the design of novel techniques and approaches for the data processing and analysis and the overall system used to process and generate these data records. Both manuscripts describe novel ways of dealing with complex and long terms data sets and unique approaches for data manipulation. A large segment of the remote sensing user community will benefit from these efforts.

CHAPTER 2: PRESENT STUDY

I. Summary

The data, methods, results and conclusions of this research are presented in the papers appended to this dissertation. A summary of the most important results follow.

II. Data and Methods

Daily global surface reflectance data covering the period 1981-to present were acquired at the climate modeling grid resolution (CMG, 0.05 degrees ~5.6km). Daily data from the long term record of AVHRR and MODIS (Terra and Aqua) were acquired from the Vegetation Index and Phenology Lab (http://vip.arizona.edu). The original Long Term Data Record AVHR09 version 3 product was used in this research (Vermote & Devadiga, 2010) downloaded from the Long Term Data Record (LTDR) site (http://ltdr.nascom.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi).

Data files consisted of 10 layers of data structured into grids of 3600 rows by 7200 columns in the Latitude/Longitude projection system. These data sets include one quality assurance (QA) layer describing the surface reflectance state, a depiction of the atmosphere quality and other information about the observation. This QA pixel information is stored as a 16-bit unsigned integer. Bits are listed from the most significant bit (MSB) 15, to the least significant bit (LSB) 0.

Both MODIS Terra (MOD09CMG, 2000-2012) and Aqua (MYD09CMG, 2002-2012) were used in this research. The MODIS data was acquired from the LP-DAAC datapool (ftp://e4ftl01u.ecs.nasa.gov/). Similar to the AVHRR dataset, the MODIS data used in this work is the daily global collection 5, gridded on a simple 0.05 degree geographic projection. Every file consisted of 19 layers (2000-2001) and 21 layers (2002-2012) of data structured into grids of 3600 rows by 7200 columns.

MODIS products include two QA layers (Coarse Resolution Internal CM and Coarse Resolution State QA), except for years 2000 and 2001 where only the Coarse Resolution Internal CM is provided. This layer does not contain aerosol information. Both MODIS QA layers are 16-bit unsigned integer with bits listed from the MSB (bit 15) to LSB (bit 0) and fill value of 0.

In the first manuscript, we stratified the results per land cover in order to characterize the impact of data quality more specifically for the different land biomes. A single land cover file was used to establish land cover boundaries throughout the full period covering AVHRR and MODIS. The 2001 MODIS land cover data was selected as the mid temporal point between the two records. Although landscape vegetation changes from year to year, we only used a single 2001 map to keep constant the land cover boundaries throughout the 30 years period.

From a land surface point of view clouds are the single worst state for an observation (Huete et al., 2002). Hence, the emphasis was placed on clouds in this analysis. Furthermore, no aerosol information exists for the AVHRR product record and aerosol analysis was limited to the MODIS record. Although we devised a method to extrapolate the MODIS aerosol information to the AVHRR record, it was only done to assist with the data reliability analysis.

AVHRR clouds were extracted from the QA State Layer. Each daily QA pixel was decoded into individual bits to identify if the pixel contained valid information and to evaluate the cloud state. Similarly, the MODIS QA pixel analysis used one or the two QA layers available. Each QA value was decoded into bit fields and analyzed. Days of the year when only the 'Coarse Resolution Internal CM layer was available, cloud state was determined from that layer. If the Coarse Resolution State QA' layer was available QA pixel values were extracted from it instead.

Clouds and aerosols were extracted from the daily data records and analyzed at three different time steps: annually, seasonally and monthly. The annual analysis helps with identifying the general cloud patterns and percent cover; however, the spatial analysis used the daily observations.

Seasonal analysis considered the quarters January-February-March (JFM), April-May-June (AMJ), July-August-September (JAS) and October, November, December (OND), which correspond and relate well with the climate patterns.

To capture the spatial variability of data quality we considered three spatial levels; First, a per pixel analysis, second, by latitude band, and thirdly by land cover. Latitudinal bands were defined by zones of 5 degrees each, North and South of the equator from -45 to 75 degrees only since no significant data exist outside this range. Land cover type boundaries were kept fixed during the 30 years record and used only the MODIS Land cover from 2001 as the midpoint of the two sensors.

From the pixel statistical analysis of clouds and aerosol data reliability map was created. This map indicates the probability for any location on Earth to have reliable synoptic remote sensing based observation, or lack of cloud and aerosol cover in other words.

In the second manuscript, the AVHR09 version 3 and MODIS C5 datasets were processed into daily and multi-day vegetation indices and phenology metrics. Datasets were filtered to remove poor quality observations and continuity translation equations were applied to AVHRR to merge the data records. Finally, using the Inverse Distance Weighting Interpolation scheme spatial gaps were filled. The resulting data is a seamless, sensor independent and cloud free global record.

To support these data records, a web tool was developed. A MySQL database was created to inventory all the datasets created at the VIP Lab. This database is used to power the online application query engine. A php (Hypertext Preprocessor) script running on the Linux server was used to populate the database based on the files and folders of the file system holding the 30-years data records.

Images of NDVI, EVI2, Quality Rank, and Mir-Nir-Red false color images were created from the data record layers at full resolution and tiled for fast zooming in and out. Images were also resampled to 10% size to be used for video animations transmitted through low internet connections.

The creation of this VIP DataExplorer application software required the use of several technologies. The application is divided conceptually in two main parts, the client side and the server side. At the client side the application was developed using HyperText Markup Language (HTML5), JavaScript, Cascading Style Sheets (CSS), Extensive Markup Language (XML), Keyhole Markup Language (KML) and the Google Earth JavaScript Application Programming Interface (API). The server side is powered by a set of Linux CentOS servers and the MYSQL engine, with support for HyperText Preprocessing (PHP) and C programming languages. To access and manipulate the data stored as Hierarchical Data Format (HDF) files, the HDF-EOS library was used.

At the front end of the client side, an HTML5 web application allows the selection of datasets, executions of commands and display of results. Through JavaScript all user requests are sent to the server and processed, results are then returned for display or listing.

III. Results

The two manuscripts are appended to this dissertation (Appendices A and B), and the findings are summarized here.

In the first manuscript, the Terra MODIS sensor reports higher percent cloud cover for 2000 and 2001 relative to other years. The same was highlighted in the per latitude analysis and by land cover analysis. This is obviously an issues with the algorithms used to assign the data QA information during the earlier years. Starting the year 2002 the MODIS products include a new layer 'Coarse Resolution State QA' which is a more accurate assessment of the pixels QA. Since Aqua data is only available starting 2002 it did not show the same problem.

The global cloudiness distribution shows a consistent spatial pattern. Sparsely vegetated and closed shrubland areas have the lowest cloud cover in the order of 30-40%, while evergreen and deciduous forests cloudiness ranged from 60 to 75%. In general JMF and OND are the seasons with the highest cloud cover percentage, except for areas located between 5 and 15 degrees north. Individual season trends differ from the annual pattern making data availability and quality vary over time within a year.

The results showthat the Aqua MODIS sensor reports higher values of cloud distribution than Terra MODIS. AVHRR cloud seasonality in general shows higher percentage when compared to Terra and Aqua sensors. Regions with high probability of clouds show

between 10-15% more cloudiness in AVHRR. Still, cloud patterns distribution show that cloud percentage have higher values in the range of 80-90% during JFM and OND for evergreen forest than during AMJ and JAS seasons. Desert areas always have lower probabilities (<30%) of cloudiness.

We looked at the aerosol distribution by analyzing the percent high and average aerosol loads cover, since they are the ones that cause problem being hard to correct for effectively. For all seasons barren and desert areas showed less than 1% high aerosol. AMJ is the season with the lowest values of high aerosol for all regions, mostly below 10% and a few areas experience up to 18%. During the JAS season values remain below 10%, with the exception of a high peak values up to 70% in the region of Angola and the Democratic Republic of the Congo, in Africa; up to 35% in South America and up to 25 % at Hubei China, these high numbers correspond to the slash and burn season in the tropics. The OND seasons values are below 17%, except for a spike of 60% in the region contained between Delhi and Bihar in India in the shadow of the Himalayas. Terra and Aqua showed similar results.

The second manuscript describes the VIP Data Explorer online tool. It is a web application that runs in a web browser. It is independent from the operating system or computing processing power of the client computer. It requires an internet connection and it has been optimized to work with minimum internet speed and band width. The application web address is http://vip.arizona.edu/viplab_data_explorer/

The Data Explorer interface is simple and intuitive to users because of its resemblance to Google Earth and similarities to geographic information systems and remote sensing applications. The main areas of the application are the map/image viewer, data sources, layers management list and legend. The map/image viewer, which occupies the majority of the screen, is where the datasets are displayed. User can zoom in and out, pan the map and view the coordinate's location.

The Data Explorer provides for inspecting all the VIP datasets. Once the user selects a dataset, it can be visualized on top of the Google Earth Imagery viewer. By stacking different datasets, directly from the VIP Products or user custom data, Data Explorer gives the user the capability of comparing different datasets and observe any spatial trends or patterns. The system also provides the option to visualize multiple images over a year period. These images are displayed as a movie animation or a mosaic of dates for any of the layers available. Tools are provided to allow users to explore these datasets quantitatively by displaying the pixel value of a given dataset, plotting a transect or creating time series of vegetation indices.

The Data Explorer is an easy to use application that supports the quick and efficient access of large data records online. The purpose is to assist the users with the inspection, manipulation, and evaluation of these large data records prior to ordering and using them.

IV Conclusions

Clouds are the most severe inhibitor to synoptic remote sensing data acquisition. Although aerosols are problematic as well, they can for the most part be measured and corrected. The low performance of the cloud masking of AVHRR resulted in higher rates of cloud commissions when compared to MODIS. Because AVHRR Quality Assurance flags do not report aerosol information, aerosol analysis was only limited to MODIS.

The probability of obtaining cloud free data for a particular geographic location can be assessed and even predicted based on this work. Our results indicate that studying biomes like the rainforest and boreal forests with satellite imagery may be very challenging and would require accurate preprocessing to eliminate the pitfalls of using poor quality data. It is important to note that the areas with lowest vegetation cover (deserts, barren soil) are the least subject to cloud and aerosol issues.

Using a set of community maintained algorithms the Vegetation Index and Phenology Lab has produced three decades of global sensor independent and seamless long term data record of vegetation index and phenology metrics. These data records, currently at version 3 (V3), are provided at monthly, 15-day, 7-day and daily time intervals and at 0.05 degrees spatial resolution. In support of these records the Lab. also developed a highly interactive online tool to enable the user community to explore and manipulate these records, to search and order, and to perform basic data analysis operations of these data.

The system has built in on demand processing capabilities and tools, especially the time series and pixel inspection tool. The Data Explorer puts three decades of remote sensing data at the fingertips of the global long term land surface vegetation user community. This tool enhances the user experience and frees them from devoting large resources to the task of data inspection and preprocessing.

REFERENCES

Abel, D. J., Taylor, K., Ackland, R., & Hungerford, S. (1998). An exploration of GIS architectures for Internet environments. *Computers, Environment and Urban Systems*, *22*(1), 7-23.

Boelman, N. T., Stieglitz, M., Rueth, H. M., Sommerkorn, M., Griffin, K. L., Shaver, G. R., & Gamon, J. A. (2003). Response of NDVI, biomass, and ecosystem gas exchange to long-term warming and fertilization in wet sedge tundra. Oecologia, 135(3), 414-421.

Braswell, B. H., Schimel, D. S., Linder, E., & Moore, B. I. I. I. (1997). The response of global terrestrial ecosystems to interannual temperature variability. Science, 278(5339), 870-873.

Brown, M. E., Pinzón, J. E., Didan, K., Morisette, J. T., & Tucker, C. J. (2006). Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-Vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. Geoscience and Remote Sensing, IEEE Transactions on, 44(7), 1787-1793.

Didan, K., & Huete, A. (2006). MODIS vegetation index product series collection 5 change summary. MODIS VI C5 changes, 17.

Google Earth API Reference. (2013). https://developers.google.com/earth/documentation/reference/ (Accessed April 15, 2013).

HDF-EOS. http://hdfeos.org/ (Accessed April 15, 2013).

Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing, 7(11), 1417-1434.

Hu, S. (2008). Advancement of Web Standards and Techniques for Developing Hypermedia Maps on the Internet. In International Perspectives on Maps and the Internet (pp. 115-124). Springer Berlin Heidelberg.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote sensing of environment, 83(1), 195-213.

Huete, A. R., Miura, T., Kim, Y., Didan ,K., Privette J. (2006). Assessments of multisensor vegetation index dependencies with hyperspectral and tower flux data. Remote Sensing and Modeling Ecosystems for Sustainability III. Proc. of SPIE. Vol. 6298, 6298814.

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment, 112(10), 3833-3845.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., ... & Morisette, J. T. (2002). An overview of MODIS Land data processing and product status. Remote Sensing of Environment, 83(1), 3-15.

LP DAAC. https://lpdaac.usgs.gov/ (Accessed April 15, 2013).

LTDR. Land Long Term Data Record. http://ltdr.nascom.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi (Accessed April 15, 2013).

LTDR Data Product Description, version 3. 2010. http://ltdr.nascom.nasa.gov/ltdr/docs/LTDRDataFormatDescriptions_03_2010_30.pdf. (Accessed on April 15, 2013)

Miura, T., Huete, A., Yoshioka, H., & Kim, H. J. (2002). An application of airborne hyperspectral and EO-1 Hyperion data for inter-sensor calibration of vegetation indices for regional-scale monitoring. In Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International (Vol. 6, pp. 3118-3120). IEEE.

Murphy, R. E., Barnes, W. L., Lyapustin, A. I., Privette, J., Welsch, C., DeLuccia, F., ... & Kealy, P. S. (2001). Using VIIRS to provide data continuity with MODIS. In Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International (Vol. 3, pp. 1212-1214). IEEE.

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., ... & Pinheiro, A. (2007, July). Generating a long-term land data record from the AVHRR and MODIS instruments. In Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International (pp. 1021-1025). IEEE.

Reed, B. C., Schwartz, M. D., & Xiao, X. (2009). Remote Sensing Phenology: Status and the Way Forward. Phenology of ecosystem processes, 231.

Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., ... & Zhang, C. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. Remote Sensing of Environment, 114(1), 35-49.

Standart, G. D., Stulken, K. R., Zhang, X., & Zong, Z. L. (2011). Geospatial visualization of global satellite images with Vis-EROS. Environmental Modelling & Software, 26(7), 980-982.

Tsend-Ayush, J., Miura, T., Didan, K., Barreto-munoz, A. (2011). Compatibility Analysis of Global Vegetation Index from Terra MODIS and SPOT-4 VEGETATION for Continuity Studies. American Geophysical Union. Fall Meeting 2011. #IN21A-1412.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., ... & El Saleous, N. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing, 26(20), 4485-4498.

USGS EarthExplorer. http://earthexplorer.usgs.gov/ (Accessed April 15, 2013).

USGS Global Visualization Viewer. http://glovis.usgs.gov/ (Accessed April 15, 2013).

The HDF Group. http://www.hdfgroup.org/ (Accessed April 15, 2013).

Van Leeuwen, W. J., Huete, A. R., & Laing, T. W. (1999). MODIS vegetation index compositing approach: a prototype with AVHRR data. Remote Sensing of Environment, 69(3), 264-280.

Vermote, E., & Kaufman, Y. J. (1995). Absolute calibration of AVHRR visible and near-infrared channels using ocean and cloud views. International Journal of Remote Sensing, 16(13), 2317-2340.

Yaheng, C., Junmei, Z., & Mingxin, M. (2012, August). WebGIS-based farmland ecosystem spatial decision support system. In Agro-Geoinformatics (Agro-Geoinformatics), 2012 First International Conference on (pp. 1-9). IEEE.

Yoshioka, H., Miura, T., & Yamamoto, H. (2005, January). Relationships of spectral vegetation indices for continuity and compatibility of satellite data products. In Fourth International Asia-Pacific Environmental Remote Sensing Symposium 2004: Remote Sensing of the Atmosphere, Ocean, Environment, and Space (pp. 233-240). International Society for Optics and Photonics.

Yoshioka, H., Miura, T., & Obata, K. (2012). Derivation of relationships between spectral vegetation indices from multiple sensors based on vegetation isolines. Remote Sensing, 4(3), 583-597.

White, M. A., Brunsell, N., & Schwartz, M. D. (2003). Vegetation phenology in global change studies. In Phenology: An integrative environmental science (pp. 453-466). Springer Netherlands.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., ... & Huete, A. (2003). Monitoring vegetation phenology using MODIS. Remote sensing of environment, 84(3), 471-475.

Zhang, X., Friedl, M. A., & Schaaf, C. B. (2006). Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. Journal of Geophysical Research: Biogeosciences (2005–2012), 111(G4).

APENDIX A - SYNOPTIC GLOBAL REMOTE SENSING OF LAND SURFACE VEGETATION: DATA QUALITY CHALLENGES AND OPPORTUNITIES

Armando Barreto[1], Kamel Didan[1,2], Muluneh Yitayew[1,]

1. Agricultural and Biosystems Engineering Department. The University of Arizona, Tucson AZ.
2. Electrical and Computer Engineering Department & Institute of Environment. The University of Arizona. Tucson, AZ.

Paper was prepared to submit to journal of Remote Sensing of Environment, Elsevier B. V.

Abstract

Synoptic global remote sensing provides a multitude of land surface state variables and serves as a major foundation for global change research. Continuous collection of global satellite imagery throughout the years, has contributed to the creation of a unique and long-term satellite imagery archive from Landsat, AVHRR, MODIS, SPOT-VEGETATION, and other sensors. These records account now for more than 30 years of continuous observations, and have become an invaluable source of data for many environmental and global change related studies. The problems with these imagery archives however, are that they are not readily available for use in research and application environments and require multiple preprocessing phases. An analysis of data quality was performed on daily global AVHRR and MODIS observations, two of the most widely available and used data records, with the objective of assessing their quality and suitability to support studies dealing with global trends and changes at the land surface. Cloudiness, aerosol thickness, and other data quality traits' statistics were analyzed for both AVHRR and MODIS sensors over time and space. We have also developed a method for assigning snow/ice and aerosol quality to AVHRR and used this information to derive a global reliability index map about these data. Findings show that

clouds are the main data quality inhibitor, and that AVHRR snow/cloud mask is quite poor making using the data challenging if users do not invest in data pre-processing. Results show that areas of high ecological importance, like the Amazon and the Northern high latitude taiga forests, are most impacted by lack of data due to clouds and aerosols leading to extended periods of time with hardly any useful data, sometimes months. On the other hand, deserts and arid lands are the least prone to clouds and data issues. When analyzing mornings (AM) verses afternoon (PM) sensor, we observed that PM sensors are subject to more cloud cover on average.

While the standard approach to these challenges has been traditionally compositing of the daily images over a preset period of time to generate a representative cloud free image, our results indicate that compositing the data over a fixed global preset period is not the optimal solution and a hybrid location dependent method that preserves the high frequency of these daily observations over the areas where clouds are not as prevalent, can be more efficient.

Introduction

Continuous collection of global satellite imagery throughout the last four decades has contributed to the creation of a long record of satellite observations. Data from the Advanced Very High Resolution Radiometer (AVHRR) aboard the National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites became available since July 1981. Data from the Moderate Resolution Imaging Spectrometer (MODIS) onboard of Terra and Aqua spacecrafts are available since February 2000 and June 2002 respectively. Terra with a 10:30 AM local time equator crossing and Aqua with daytime equator crossing of 01:30 PM (Justice et al., 2002) were designed to provide complementary and identical data streams, and were originally part of a 5 Earth Observation System (EOS) spacecraft endeavor (Justice and Townshend, 1994; Running et al., 1999; NASA EOS, 1993).

By combining AVHRR and MODIS a potentially very long data record suitable for studies dealing with long term environmental change studies is achievable. However, both records have multiple problems related to clouds, aerosols, spatial coverage gaps, variable viewing geometry, inconsistent atmosphere corrections, multiple reprocessing, in addition to the differences in sensors' characteristics, making it difficult to obtain frequently the desired high quality data everywhere and every time to support the kind of research these instruments were intended to sustain. These factors must be considered in order to understand the limits of these records, the uncertainty associated with their use, and the need for preprocessing to support proper and accurate scientific results. To complicate matters further, these issues are location and season dependent making it even more difficult to construct a consistent time series.

Clouds cover about 60% of the earth's surface at any given time (Kawamoto et al., 2000). For land studies, the presence of clouds limits the ground visibility and therefore makes the observation of limited or to no use. Certain locations in the tropics are covered by

clouds for most of the times making it impossible to obtain cloud free data at reasonable intervals supportive of terrestrial vegetation dynamics research (Justice et al., 2002). This inhibits effective and accurate research (Oreopoulus et al., 2011).

Another atmospheric contaminant that affects the collection of satellite imagery is the high concentration of suspended particles in the atmosphere, known as aerosols. Their variation in time and space makes them difficult to quantify and characterize (Hansen et al., 1997; Kaufman et al., 2002). They interfere with the direct path of light by scattering (Levy et al., 2005, Lyapustin et al., 2010 & 2012), absorbing, and shading the earth's surface, creating a combined effect of attenuation of the radiation being measured (Kaufman et al., 2001; Lyapustin et al., 2010) and causing the recorded radiation to be inaccurate requiring correction (Vermote et al., 2002). Although, observing the ground is still possible, particularly when atmosphere correction is applied to the observations (Vermote et al., 2002; Lyapustin et al., 2010), the issue remains however, that correction is very poor when aerosols are medium and thick. Further, AVHRR does not provide information about aerosols nor corrects for them; while MODIS has aerosol retrieval and correction algorithms (Kaufman et al., 2001; Vermote et al., 2002) that are fairly validated with ground-based sunphotometer data, particularly from the Aerosol Robotic Network (AERONET) (Holben et al., 1998).

In order to assess the impact of these issues on data collected by satellites, information about these issues' spatial and temporal distribution need to be well understood (Hansen and Lacis, 1990). This work aims at quantifying and characterizing these global AVHRR and MODIS data records, to assess their quality, and to reveal the limit of spatial and temporal probability of collecting suitable and frequent data to support the kind of research required for global change.

Data and Methodology

Daily global data covering the period 1981-to present were acquired at the climatic modeling grid resolution (CMG, 0.05 degrees). The CMG grid resolution was selected to address the sheer size of these records and permit a reasonable global analysis, that would have been prohibitive had we used the finer resolution data records. Additionally, the data quality to be used in this research was derived from the finer resolution native data record. Daily data from the long term record of AVHRR (spanning the N07,N09, N11 and N14 platforms) and MODIS (Terra and Aqua) were acquired from the Vegetation Index and Phenology laboratory at the University of Arizona. (http://vip.arizona.edu). The AVHRR Long Term Data Record (LTDR) AVHR09 version 3 product suite was used in this research (Vermote et al., 2010) and were downloaded from the LTDR site (http://ltdr.nascom.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi).

Although, there are sensor differences between AVHRR and MODIS related band passes and position our analysis looked only at the quality information (Roy et al., 2002) in a collective fashion, since the QA describe the atmosphere state and hence sensor independent (Roy et al., 2002). When using the Normalized Difference Vegetation Index NDVI data, derived information from both sensors was kept separately.

Because we are using AVHRR and MODIS records as a continuous data stream we only considered AVHRR data between 1981 and 1999 and MODIS starting 2000. Each of the input AVHR09 data files contained 10 layers of data structured into grids of 3600 rows by 7200 columns in the Latitude/Longitude projection system. These data sets contained the red (SREFL_CH1) and NIR (SREFL_CH2) channels, Zenith View Angle (VZEN), Sun Zenith Angle (SZEN), Relative Azimuth (RELAZ) and included one quality assurance (QA) layer describing the atmosphere state. Table 1 shows the QA bit flags and description. This QA information is stored in 16-bit signed integer. Bits are listed from the most significant bit (MSB) 15, to the least significant bit (LSB) 0.

Table 1. LTDR AVHRR09 surface reflectance data state QA description

Bit No	QA flag name	Bit value	QA State
15	Polar flag. Latitude over 60 degrees (land) or 50 degrees (ocean)	0	No
		1	Yes
14	BRDF-correction issues	0	No
		1	Yes
13	RH03 value is invalid	0	No
		1	Yes
12	Channel 5 value is invalid	0	No
		1	Yes
11	Channel 4 value is invalid	0	0
		1	Yes
10	Channel 3 value is invalid	0	No
		1	Yes
9	Channel 2 value is invalid	0	0
		1	Yes
8	Channel 1 value is invalid	0	No
		1	Yes
7	Channels 1-5 are valid	0	No
		1	Yes
6	Pixel is at night (high solar zenith)	0	No
		1	Yes
5	Pixel is over dense dark vegetation	0	No
		1	Yes
4	Pixel is over sunglint	0	No
		1	Yes
3	Pixel is over water	0	No
		1	Yes
2	Pixel contains cloud shadow	0	No
		1	Yes
1	Pixel is cloudy	0	No
		1	Yes
0	Unused	N/A	N/A

Both MODIS Terra (MOD09CMG, 2000-2012) and Aqua (MYD09CMG, 2002-2012) were used in this research to address the AM versus PM overpass differences in atmosphere state. The MODIS data was acquired from the Land Processes Data Active

Archive (LP-DAAC) datapool (ftp://e4ftl01u.ecs.nasa.gov/). Similar to the AVHRR data record, the MODIS data used in this work is the daily global collection 5, gridded on a simple 0.05^{o} geographic projection. Every file contained 19 layers (2000-2001) or 21 layers (2002 and later) structured into grids of 3600 rows by 7200 columns. For this study only the red (Coarse Resolution Surface Reflectance Band1), NIR (Band 2), Solar Zenith Angle, View Zenith Angle, Relative Azimuth Angle and QA information was used.

MODIS products include two QA layers: (1) Coarse Resolution Internal CM, and (2) Coarse Resolution State QA, except for years 2000 and 2001 where only the Coarse Resolution Internal CM was provided. As shown in tables 2 and 3, the Coarse Resolution State QA provides more pixel quality information. Therefore, for years 2000 and 2001, aerosol information was not available. Both MODIS QA layers are 16-bit unsigned integer with bits listed from the MSB (bit 15) to LSB (bit 0) and a fill value of 0.

Table 2.MODIS surface reflectance coarse resolution internal CM description

Bit No	QA flag name	Bit value	QA State
15	Unused	N/A	N/A
14	OAT has climatological values	0	No
		1	Yes
13	Criteria used for aerosol retrieval	0	Criterion 1
		1	Criterion 2
12	Pan flag	0	No salt pan
		1	Pan flag
10-11	Cirrus detected	00	None
		01	Small
		10	Average
		11	High
9	Pixel is adjacent to cloud	0	No
		1	Yes
8	Cloud Shadow	0	No
		1	Yes
7	Dust	0	No
		1	Yes
6	Sunglint	0	No
		1	Yes
5	Fire	0	No
		1	Yes
4	Snow	0	No
		1	Yes
3	Low clouds	0	No
		1	Yes
2	High clouds	0	No
		1	Yes
1	Clear	0	No
		1	Yes
0	Cloudy	0	No
		1	Yes

Table 3. MODIS surface reflectance coarse resolution state QA description.

Bit No	QA flag name	Bit value	QA state QA
15	Internal snow algorithm flag	0	No
		1	Yes
14	BRDF correction performed	0	No
		1	Yes
13	Pixel is adjacent to cloud	0	No
		1	Yes
12	MOD35 snow/ice flag	0	No
		1	Yes
11	Internal fire algorithm flag	0	No
		1	Yes
10	Internal cloud algorithm flag	0	No
		1	Yes
8-9	Cirrus detected	00	None
		01	Small
		10	Average
		11	High
6-7	Aerosol Quantity	00	Climatology
		01	Low
		10	Average
		11	High
3-5	Land/water flag	000	Shallow ocean
		001	Land
		010	Coastlines and lake shorelines
		011	Shallow inland water
		100	Ephemeral water
		101	Deep inland water
		110	Continental/moderate ocean
		111	Deep ocean
2	Cloud Shadow	0	No
		1	Yes
0-1	Cloud State	00	Clear
		01	Cloudy
		10	Mixed
		11	Not set, assumed clear

Due to data reprocessing issues, it is important to note that for years 2003 and 2007 there were days where only the Coarse Resolution Internal CM layer was available.

In parts of our data analysis we stratified the results per land cover in order to characterize the impact of data quality more specifically on the different land biomes. A single land cover file was used to establish land cover boundaries throughout the full period covering AVHRR and MODIS. The 2001 MODIS land cover map was selected as the mid temporal point between the two records. Although landscape vegetation changes from year to year, we only used a single 2001 map to keep constant the land cover boundaries throughout the 30 years period. Any land use change over the full study period, although significant, should not bias the results. The 2001 land cover map MOD12C1 is derived from the MODIS record (Friedl et al,, 2002) and has the same spatial resolution as the rest of our data records.

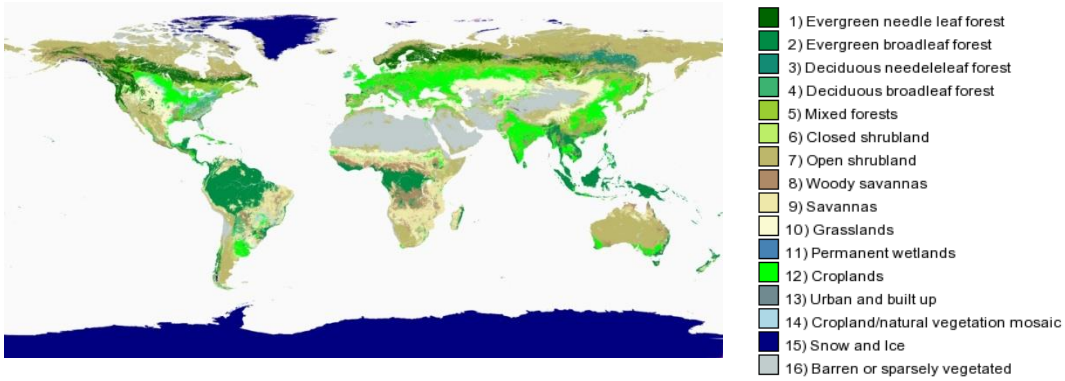


Fig 1. MOD12C1 global land cover types

From a land surface vegetation perspective, clouds are the worst contaminant (Holben et al, 1986; Huete et al., 2002; Miura et al, 1999) hence the emphasis placed on clouds in this analysis. Furthermore no aerosol information existed for the AVHRR product record hence aerosol analysis was limited to the MODIS record.

AVHRR clouds were extracted from the QA State Layer, by decoding each daily QA

pixel into individual bit fields. Bit 7 (see Table 1) was first used to identify if the pixel contained valid information, then bits 1 and 2 were evaluated for the cloud state.

In a similar fashion, the MODIS QA Pixel analysis used the two QA layers when available. Each QA value was decoded into bits and analyzed. Days of the year when only the 'Coarse Resolution Internal CM layer was available, the cloud state was determined from bits 0 and bit 8 when the QA pixel value was not a fill value. If the Coarse Resolution State QA' layer was available the QA values were extracted from it instead. Cloud state is stored in bits 0-1. If the bit value was 01 (cloudy) or 10 (mixed) the observation was considered cloudy. This status was further constrained by the internal cloud flag from bit 10 (Vermote et al., 1999) and cloud shadow from bit 2. The worst case scenario was used to mitigate omission of clouds (Didan and Huete 2005).

Only MODIS reports information about the aerosol quality in the 'Coarse Resolution State QA' layer. Bit fields 6-7 contains this information (Table 3). Only average and high aerosol categories are considered detrimental to data quality, since correcting for them is very challenging and leaves a large error (Vermote et al., 1999). Aerosol climatology correspond to instances when an aerosol quantity could not be estimated from MODIS aerosol bands (Kaufman et al., 1997, Vermote et al., 1999) and correction was based on ancillary or standard profiles (Vermote et al., 1999). This corresponds to observations over clouds and bright targets usually, which are captured during the cloud screening step. In other words if an observation is cloudy other quality information is no longer used to further categorize the observation. In case of low aerosol loads the atmosphere correction algorithm is capable of addressing it (Vermote et al., 2002; Kaufman et al., 1997; Lyapustin et al, 2010).

This quality information were extracted from the daily data observations and analyzed per year, season, and month. Seasonal analysis was considered to characterize the patterns association with the prevalent climate. A detailed monthly analysis is carried to

capture the dynamic behavior of quality in a finer temporal step. The annual analysis, although temporally coarse, can help capture any signature in the QA pattern particularly inter-annual trends.

To capture the spatial variability of data quality we considered three spatial aggregation levels; (1) Per pixel analysis, (2) Analysis by latitude band, (3) Analysis by land cover. Latitudinal bands were defined by zones of 5 degrees, North and South of the equator from -45 to 75 degrees only (no significant land and/or data exist outside this range). Land cover type boundaries were kept fixed during the 30 years record and used a single MODIS Land cover map from 2001 (Frield et al., 2002).

From the pixel QA statistical analysis a reliability map was created. The reliability map is a long term average map that identifies the probability for any location to have a reliable synoptic remote sensing observation, in other words this map corresponds to how likely an area is to be free of clouds and aerosol contaminants.

The above analysis generates qualitative information for the most part, and in order to generate an equivalent quantitative map that shows the global distribution of error and uncertainty associate with synoptic data, we compiled a long term average NDVI from AVHRR and MODIS. Only good quality data were used (cloud free, low aerosol and low view angle). For each pixel the average long term value from the record was computed along with the standard deviation. Using this data, an envelope of maximum and minimum range was created for each pixel using 1.5σ (standard deviation). Any value can be compared to this dynamic range. We reason here that natural inter-annual change will always fall within this range and any deviation will either result from disturbance or noise which is the error in this context. Disturbance usually results from fire, infestation, or land cover land use change and is not widespread. Although the resulting map cannot separate between the two, the data is a first order estimate of noise related error in these synoptic data.

Because AVHRR data does not provide snow/ice and aerosol information, we designed a simple methodology to derive this information from the historic frequency of occurrence of this quality information in the MODIS record. The daily information about snow/ice and aerosol quality (low aerosol, average aerosol and high aerosol) was extracted for the entire MODIS period considered in this research (2000-2012), and averaged on a daily basis to compute the frequency (probability) of occurrence, numbers range from 0 to 100%.

These masks were then applied to the AVHRR data as a first guess to the likelihood of a pixel having snow/ice cover and the likelihood of aerosol contaminant presence. This is by no means an accurate representation, but it gives us a first order guess as to the likelihood of presence of these quality attributes. Snow and ice presence follows a quite fixed cyclic occurrence for the most part and during the bulk of the winter season. Similarly aerosols, especially around the equator and over the tropics are mostly related to the practice of slash and burn agriculture and occur episodically during a predetermined time of the year (ex: usually JAS over Brazil).

If this probability of occurrence mask indicates that a particular pixel was always covered with snow/ice during the MODIS record period of observation, then it was labeled as snow/ice in AVHRR and the snow/ice estimated flag was turned to YES to indicate it was estimated and not real. If the pixel was labeled as cloudy in the input, but the NDVI value was within the min-max boundaries it was re-ranked as snow/ice if the probability of snow/ice is above 80%, this was done to deal with the confusion between clouds and snow/ice by the AVHRR CLAVR algorithm.

Following the same procedure, aerosol quantities information was also estimated for AVHRR. If the pixel was cloudy, then aerosol quantity was set to climatology as was done with MODIS, otherwise the most occurring aerosol quality value was assigned to

the pixel QA. The category with the highest frequency of occurrence was applied to AVHRR. To reduce the error of this approach an NDVI constrainer was used. Using, once again, the long term NDVI envelope boundaries, and using the fact that aerosol tend to attenuate NDVI, then if the MODIS aerosol frequency was high but the AVHRR NDVI value is reasonably high we revert the aerosol quality back to average. Vice versa if the NDVI value was rather low compared to the long term and the MODIS aerosol frequency was low aerosol then the AVHRR was assigned an aerosol quality of high.

This approach is quite conservative, which is reasonable when dealing with synoptic remote sensing data as discarding a good observation has no serious implication on post processing and research when compared with retaining poor quality observations.

With this process we have created an AVHRR snow/ice and aerosol quality data set. This information was used to construct the data reliability map proposed earlier (Fig 2). Additionally we used the view zenith angle (VZA) and sun angle (SA) to further constrain the observation quality.

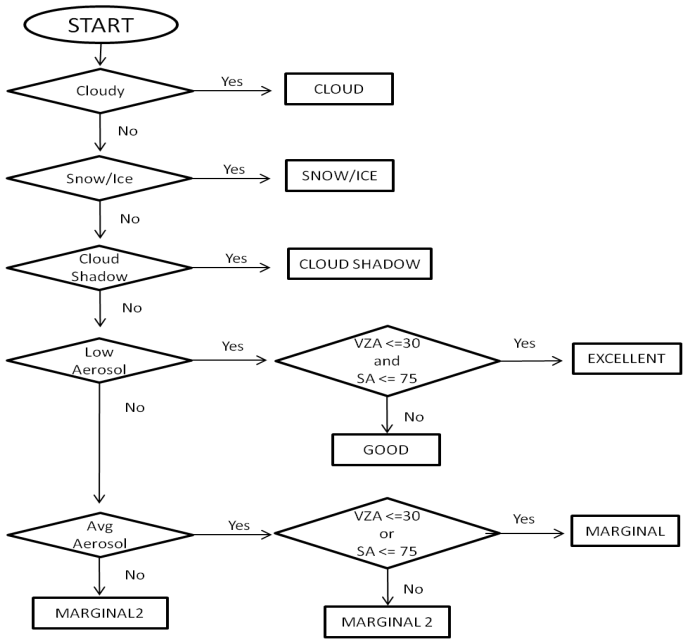


Figure 2. Daily images ranking methodology.

Every pixel from the AVHRR/MODIS dataset was ranked using the flowchart in Fig. 2. If data was not available, the pixel was labeled as NO_DATA, otherwise it was categorized into one of the ranks from Figure 2. The highest quality data is when aerosol is low, view angles low, classes with decreasing quality are then created to reflect how likely the sensor will view the land surface target. Using this method, reliability maps were generated to indicate the likelihood of good observation occurrence. These maps were generated for AVHRR and MODIS annually and seasonally.

Results

The cloud and aerosol distribution analysis is summarized by the figures and images. The latitudinal and land cover based results are summarized in charts, and pixel based results are illustrated by maps. These maps help indicate the availability of reliable data for a particular location at any given time, and vice versa can also inform the users about the locations and times of year were data analysis will need to consider the uncertainty resulting from data poor quality.

Missing Data

Ideally one would like to have data for each day; however it is not always the case. When missing data days is only for short periods of time or when the missing days are spread throughout the year the time series can still be reconstructed with a high degree of accuracy. However when long periods, long enough to miss phenological changes or trends, of missing data exist it becomes impractical to effectively generate the missing observations with enough accuracy to support change studies without introducing bias. Data is missing because either the satellite sensor had maneuver issues or due to processing issues. Table 4 lists the missing data days. Special consideration was applied to years with long periods of missing data which are 1981, 1994 for AVHRR and the early months of 2000 following MODIS launch. The year 1981 is missing the first 174 days of the year. The year 1994 does not provide information for days after day of the year 256. The Terra data record started on day of the year (DOY) 54.

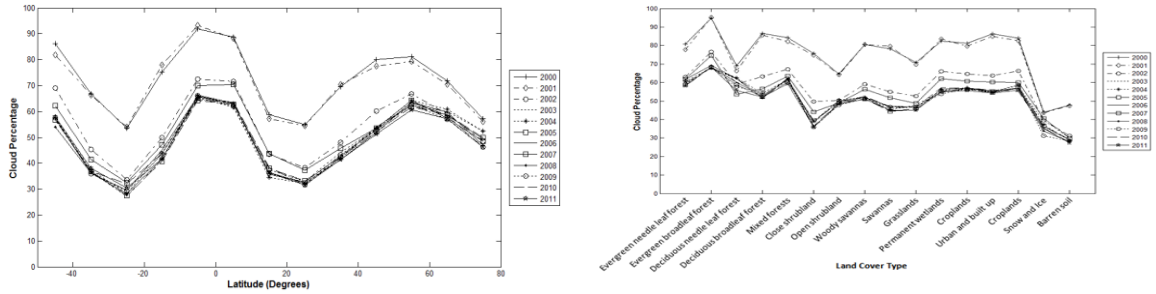
Table 4. List of AVHRR and MODIS missing data days.

Year	Data not recorded by sensor (Day of the year)
1981	1-174, 177, 178, 182-201
1982	22, 88, 104-107, 114, 119-121, 187, 202, 237, 268, 269
1984	14, 15, 51, 53, 62, 82, 101, 107, 205, 341, 342
1985	3, 18, 19, 39-42, 70, 310
1986	38, 73, 74
1988	4, 72, 73, 81, 90, 135, 136, 170, 197-199, 206-208, 235, 262, 281, 313-366
1989	80, 81, 96
1990	1, 3, 59, 201-205, 210-213, 307, 321
1991	1-4, 10-14, 41-43
1994	11, 257-365
2000	1-54, 117, 118, 219-230, 342-366
2001	167-182
2002	79-86, 105
2003	351-357
2007	316, 317
2008	356, 357

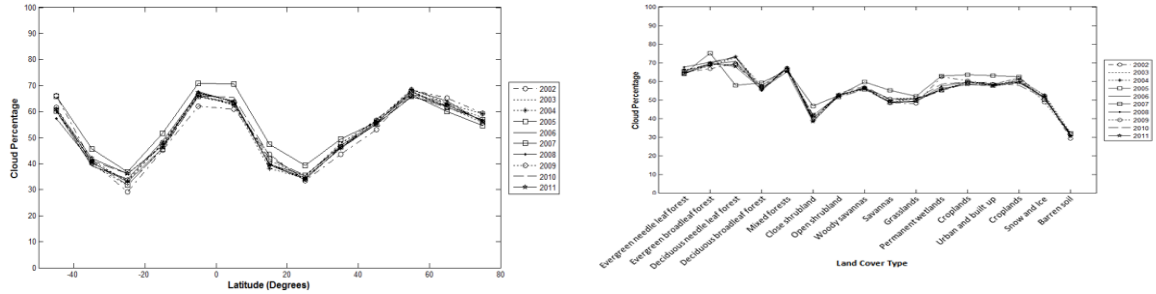
It should be noted here that the statistical analysis was adjusted for the missing days and subsequently the results were not biased.

Annual Cloud Distribution

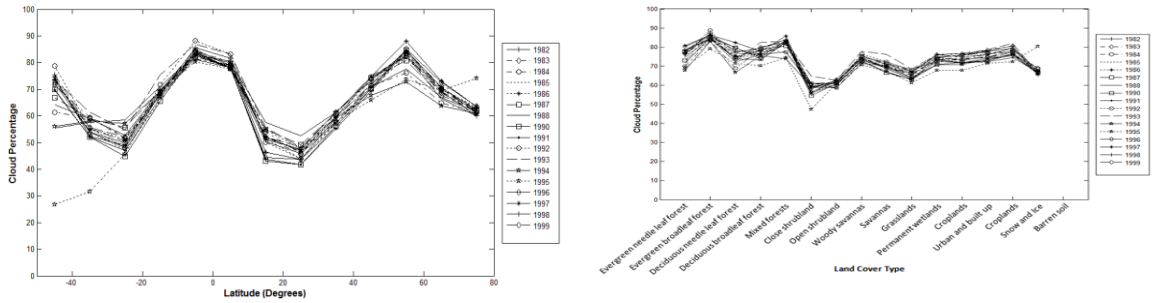
Mapping annual cloud distribution is useful and provides a general view of the data quality and year to year change in the cloud frequency. Figure 3 shows Terra years 2000 and 2001 with the highest cloudiness relative to other years. The same results were observed in the latitudinal and land cover based analyses. This is the result of using a QA layer with minimal information on pixel status. Starting year 2002 the MODIS products included a new layer 'Coarse Resolution State QA' which reports a more accurate and complete quality assessment of the pixels. Since Aqua sensor started reporting data in 2002 it used this new QA approach and was not subject to this error.



a) Terra



b) Aqua



c) AVHRR

Figure 3. Annual cloud distribution by latitude and land cover type for a) Terra, b) Aqua and c) AVHRR sensors.

Clouds distribution by latitude shows a clear and strong pattern. It reaches maximums between 60-70% at latitudes from -10 to 10 degreees, the equator band. A second peak of 50-60% at latitudes between 50 -70 degrees, corresponding to the Taiga forests. The lowest clouds percentage are observed in the area between -30 to -15 and 20 to 30 degrees which corresponds to the deserts and super arid regions of the world.

Distribution by land cover for all years shows also a consistent pattern. Barren and

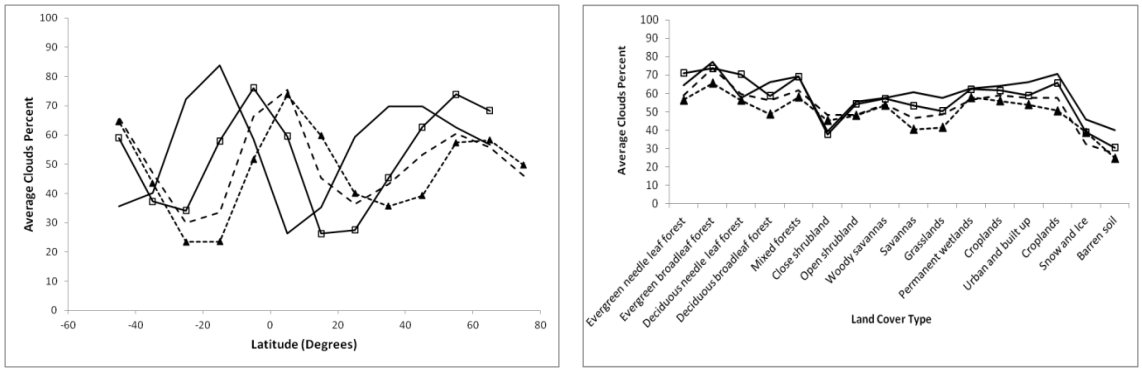
sparsely vegetated areas are on average coverd with cloud about 30% of the time. Closed shrubland areas have also a low cloud cover in the order of 40%, while evergreen forest and deciduous forest cloudiness ranges from 60 to 75%. This is not a surprise for the most part since it corresponds with well understood climate pattern, but the impact on these two key biomes is the issue here considering the impact of the derived data.

Due to the unreliable and different QA assignment approach in earlier Terra MODIS data Fig. 3.a shows the years 2000 and 2001 are having more cloudy data in both the latidude and land cover analysis. These two years stand out from the rest and show a 15-20% more cloudiness. However, starting 2002 boht Aqua and Terra MODIS use a similar QA algorithm. On average, we observe that Aqua shows slighlty higher cloudiness that could only be xplained by the overpass time (PM). This makes the PM overpass cloudier than the AM overpass.

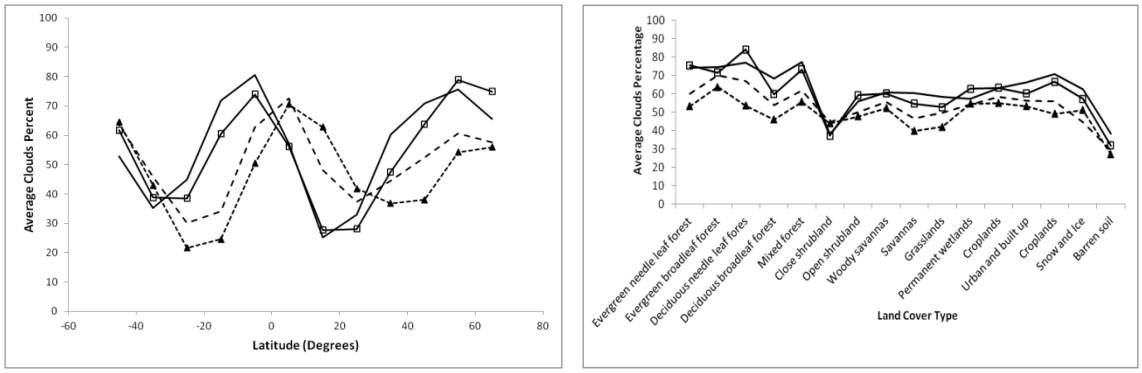
When comparing MODIS to AVHRR, the patterns are also similar both indicating the consistency of these atmosphere contaminant over space and time and the reasonnable algorithms used to derive them. However, AVHRR values are on average about 10% higher, mostly due to the poor performance of the AHRR cloud detection alogrithm CLAVR (Stowe et al., 1999).

Seasonal Cloud Distribution

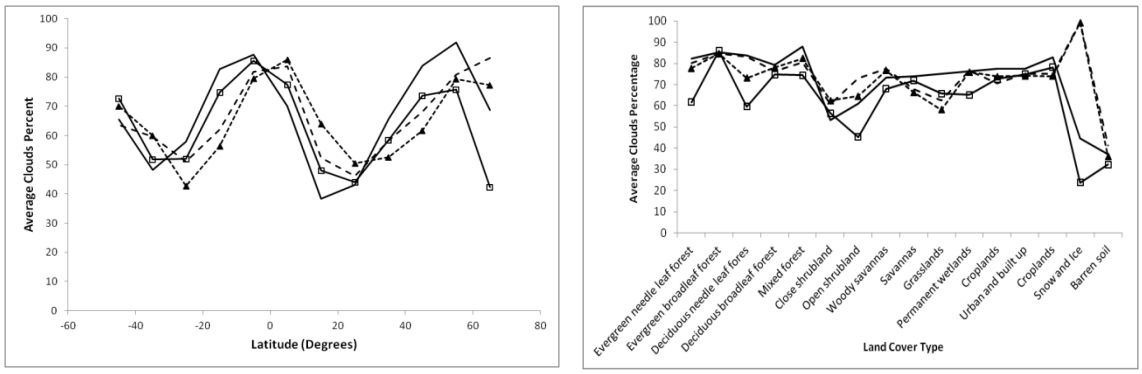
Analysis of the cloud distribution by season shows more useful patterns, since cloud distribution changes per season as a result of climate seasonal patterns. Figure 4 relates the seasonal cloud distribution by sensor at different latitudes and land cover types. In general JFM and OND are the seasons with the highest cloud cover except for areas located between 5 and 20 degrees Northern latitude.



a) Terra



b) Aqua



c) AVHRR

Figure 4. Seasonal cloud distribution by latitude and land cover type for a) Terra, b) Aqua and c) AVHRR sensors.

In areas located close to the equator AMJ and JAS seasonal values are higher. Individual

season trends differ from the annual pattern making data availability and quality vary over time within a year.

Among all the land cover types the barren and sparsely vegetated and closed shrublands have the lowest probability of cloudiness at just 30-40%. This is the case throughout all the seasons and for all sensors. Savannas and grasslands follow with 40% during the JAS season but up to 60% during the JMF season. Evergreen and deciduous needleleaf forests are the cloudiest areas with percentages between 55 and 85%.

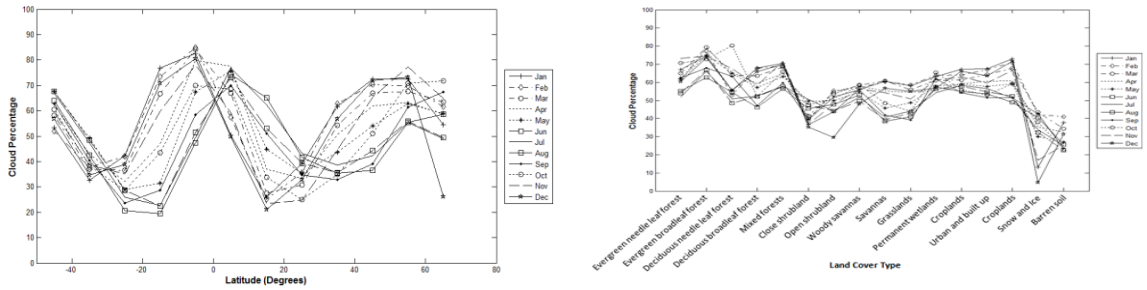
In general the seasonal trend by latitude can be indicated as follows: Below 35 deg south latitude, the values reported between seasons is very close, but AMJ and JAS are around 5% higher. From -35 to 5 deg, JFM and OND are cloudier by a margin of 30% over AMJ and JAS. From 5 deg to 25 deg latitude, AMJ and JAS are the cloudiest. From this region to the north the JFM and OND seasons report higher cloudiness. While the trend is similar for the three sensors, there are some differences. Mostly AVHRR reporting higher values than Terra and Aqua. When clustering by latitude, the results diverge from 55 to 75 deg North.

It is noticeable that the cloud percentages for the snow and ice land cover type, greatly differs from the Terra and Aqua trends. First, AVHRR does not report snow/ice QA and snow/ice pixels are usually counted as cloudy which explains the high values for AMJ and JAS.

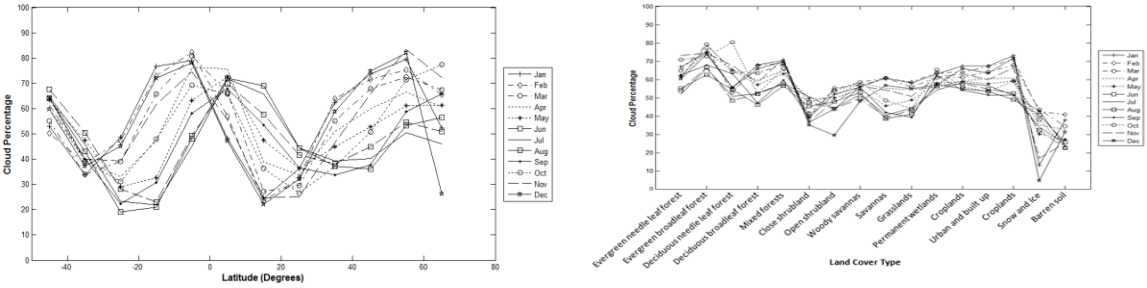
Monthly Cloud Distribution.

Monthly results from the different sensors when analyzed by latitude and land cover show, in general, similar results to seasonal distribution (Figure 5). Around 8 percentage increase can be observed for the months of November, December and January at high latitudes. Similarly when analyzing by land cover; which is due to the presence of snow

and ice. These differences are of little impact due to the fact that high latitudes are generaly covered by snow with little land surface and vegetation cover showing. Although, with global warming indicating changes in those areas, hence an improved cloud and snow/ice detection algorithms may be needed to eliminate this error.



a) Terra



b) Aqua

Figure 5. Monthly cloud distribution by latitude and land cover type for a) Terra, b) Aqua sensors.

The pattern followed by monthly cloud distribution (within the season) is similar to the trend from seasonal analysis. The gain with monthly analysis is that it reflects with more accuracy the cloud status.

Pixels by pixel analysis results are shown as maps to capture the spatial distribution. They are summarized by sensor and by seasons. Seasonal per pixel cloud maps reveal the spatial distribution of clouds.

AVHRR cloud seasonality (Fig 6), in general, shows higher percentage of cloud presence when compared to Terra (Fig 7) and Aqua (Fig 8). Regions with high probability of clouds are 10-15% higher. Still, cloud patterns distribution show that cloud percentage have higher values in the range of 80-90% during JFM and OND for Evergreen forest areas than during AMJ and JAS seasons. Desert areas always have lower probabilities (<30%).

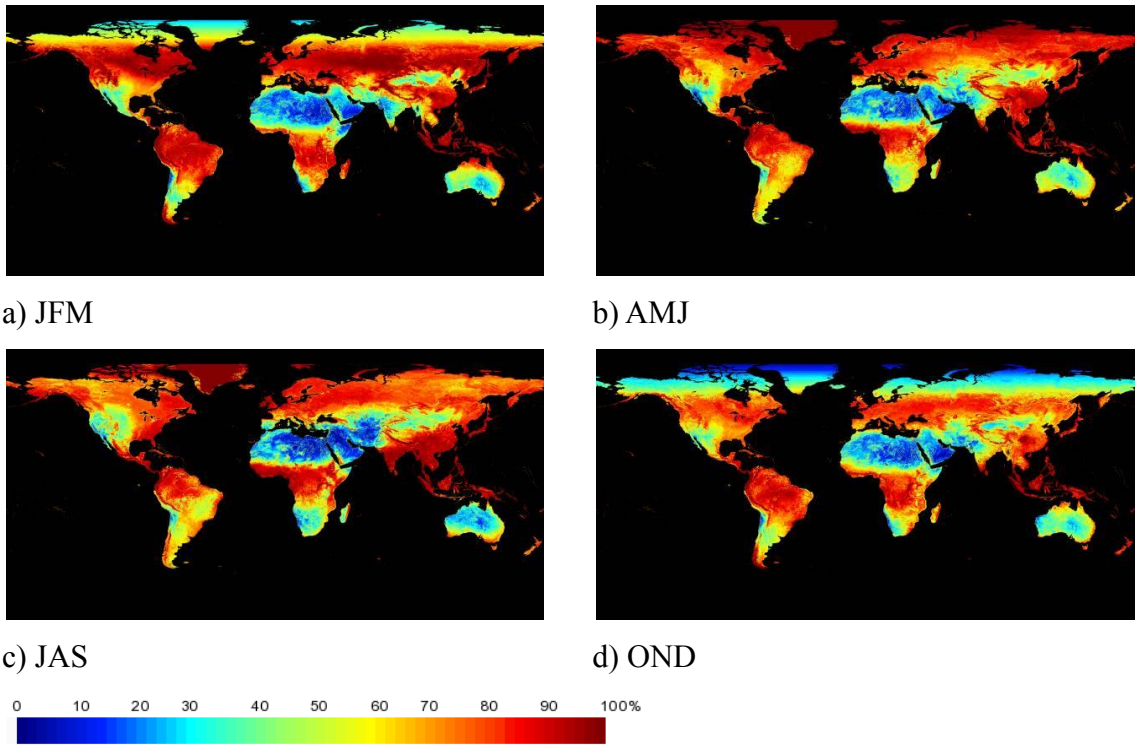


Figure 6. AVHRR seasonal cloud distribution.

Clouds spatial distribution is most clear and accurate in the MODIS sensors based results (Fig 7 and 8).

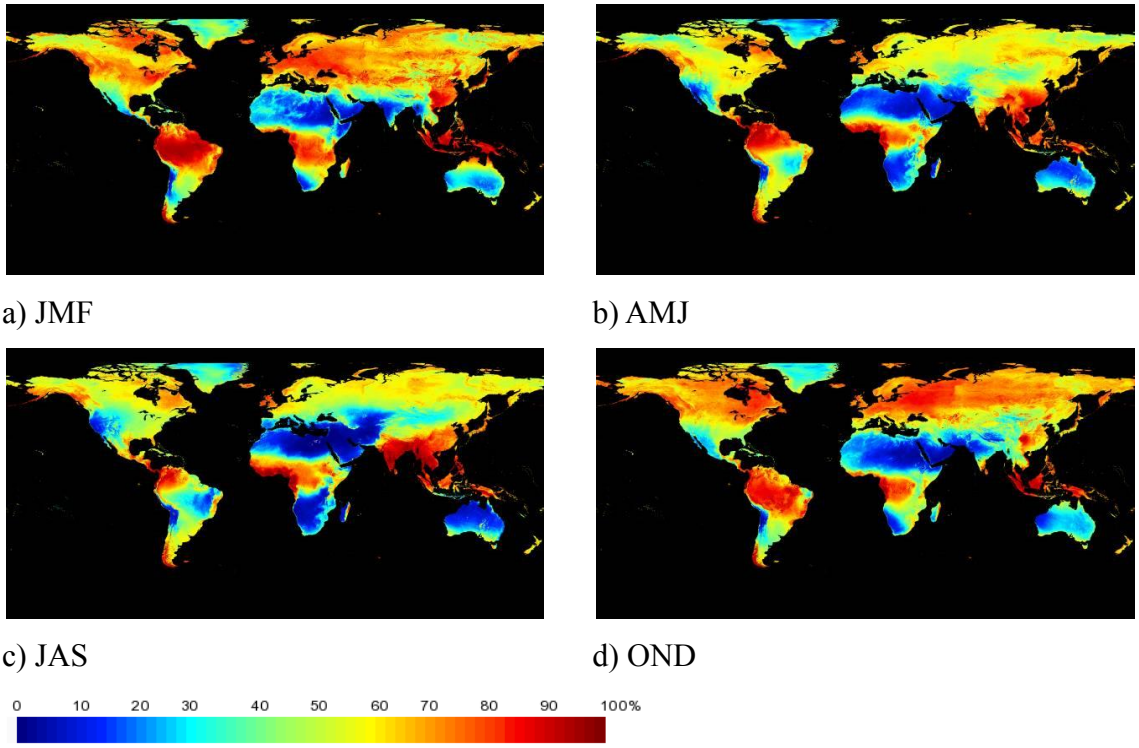


Figure 7. Terra MODIS seasonal cloud distribution

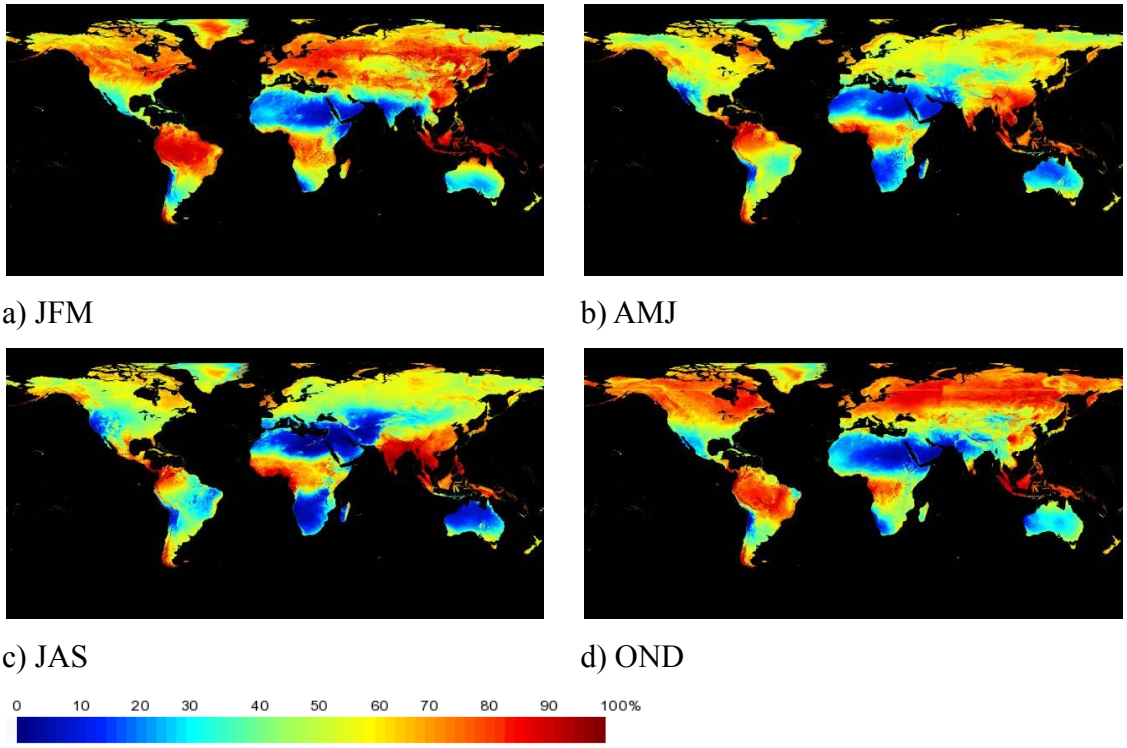


Figure 8. Aqua MODIS seasonal cloud distribution

Combined results of Terra and Aqua MODIS cloud distribution by seasons are shown in figure 9. These are consistent with the seasonal cloud analysis.

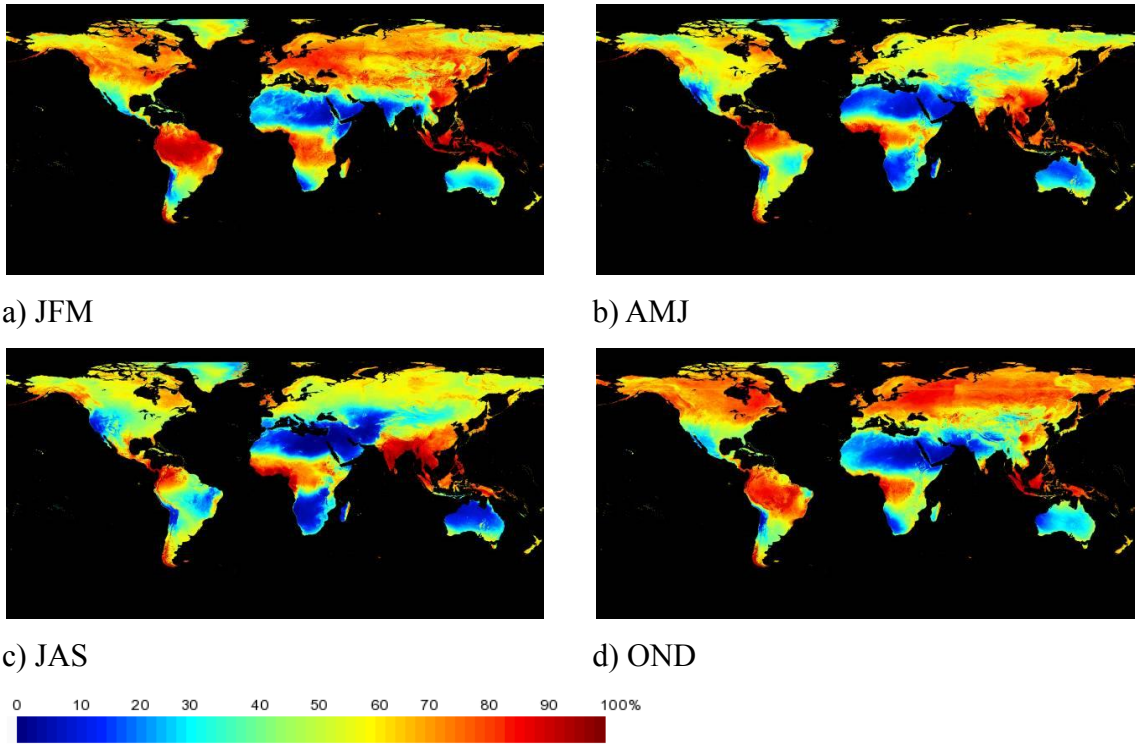
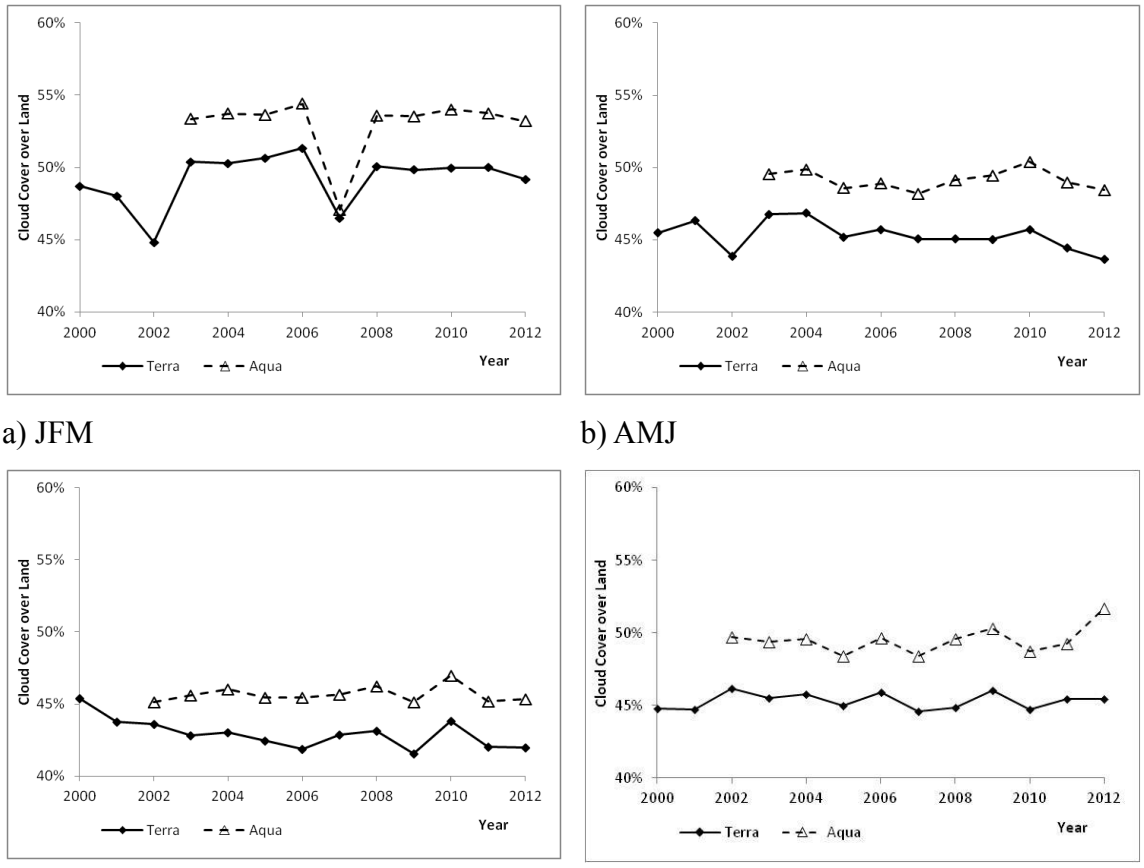


Figure 9. Combined Terra and Aqua MODIS seasonal cloud distribution.

The above results provide insights that data from AVHRR and MODIS cannot be used without quality analysis and preprocessing.

Comparison of Terra versus Aqua clouds distribution

While Terra and Aqua MODIS sensors report relative similar cloud distribution (Fig 7 and 8). The different overpass time (morning vs afternoon) makes the viewing subject to different cloud cover dynamic. An analysis of average cloud distribution (Fig 10, 12 and 13) shows that Aqua overpass is cloudier than Terra.



a) JFM b) AMJ

c) JAS d) OND

Figure 10. Terra and Aqua MODIS average seasonal global cloud distribution.

At a global scale (Fig 10), Aqua reports about 5% higher cloudiness than Terra. During the JAS season this difference drops to about 3%.

In Figure 10a, the year 2007 shows a different pattern, where the Aqua cloudiness drops

to a similar level to Terra. To investigate the source of this change, global daily values for the years adjacent to 2007 were extracted (Figure 11).

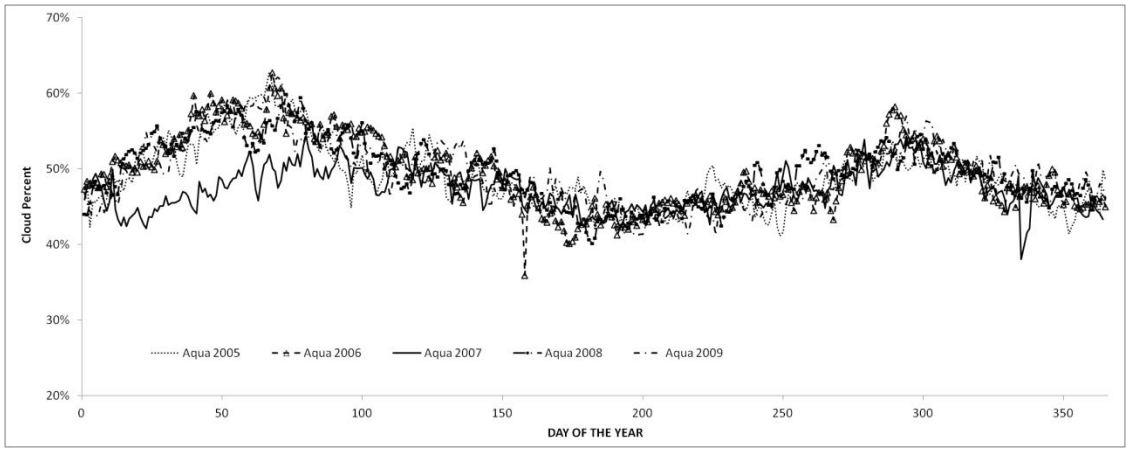
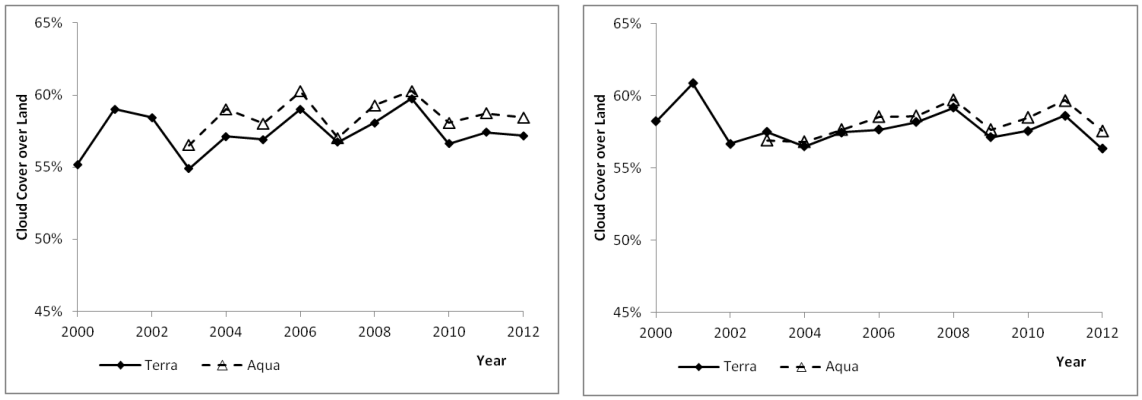


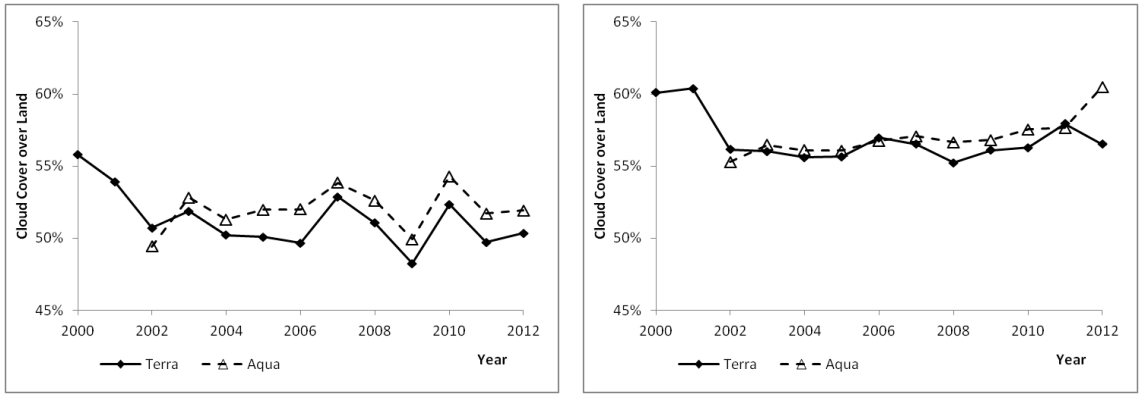
Figure 11. Aqua MODIS daily global average cloud percent cover

Daily values indicate that from DOY 1 to DOY 103 Aqua sensor for the year 2007 is reporting lower cloud cover percent in comparison with other years. When these values are averaged by season, the Jan-Feb-Mar season is the most affected confirming the source of this data QA issue.

As expected the highest cloud cover are found around the equator (Fig 12). A look at the 5S to 5N latitudinal band (Fig 13) shows that the difference observed by Aqua and Terra is minimum compared with the global analysis (Fig 12)



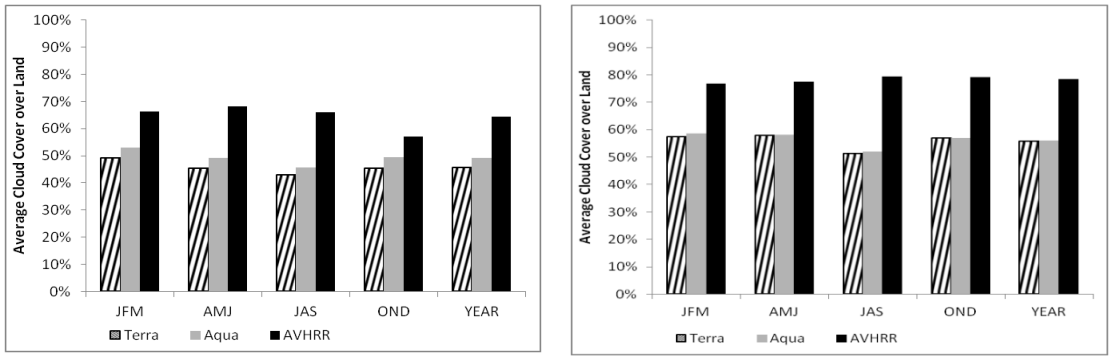
a) JFM b) AMJ



c) JAS d) OND

Figure 12. Terra and Aqua MODIS average seasonal cloud cover percent for the tropics (10 to -10 deg latitude).

On a global scale, Aqua reports 7.7% (JFM), 8.5% (AMJ) 6% (JAS) and 9% (OND) higher values than Terra (Fig 13a). On the other hand, at the tropics these values are only 2% (JFM), 0.5% (AMJ), 1.3% (JAS) and 0.2% (OND) (Fig 13b). The poor performance of the AVHRR cloud masking algorithm is noticeable here, where cloud frequency is consistently higher than that of Terra and Aqua MODIS. If any, this does not compromise research based on these data, since being over conservative does not introduce an error, however it leads to discarding a large amounts of data that could have been useful.



a) Global b) Tropics

Figure 13. Terra, Aqua and AVHRR seasonal and annual average cloud percent cover.

In general, Aqua MODIS reports up to 5% higher cloud cover than Terra. However to investigate if there were any spatial distribution patterns, seasonal spatial variability maps were generated (Fig 14). These maps capture the percent spatial difference and were obtained by subtracting Aqua from Terra and dividing by the average of the two.

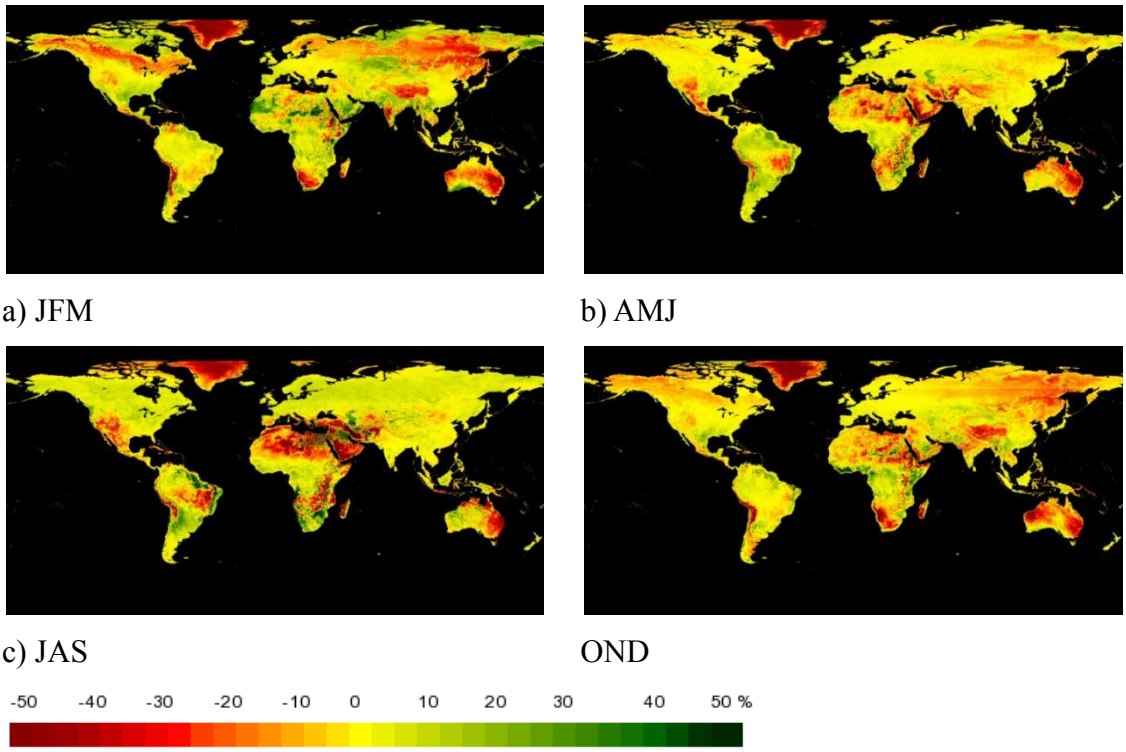


Figure. 14 Terra and Aqua MODIS long term average spatial seasonal cloud distribution.

According to Figure 14, Aqua reports higher cloud cover than Terra in arid areas like the Sahara desert (20-40% AMJ, JAS OND), US Southwest and Northern part of Mexico (25-35% AMJ, JAS, 5-20% OND, JFM), The desert of South America (10-45% AMJ, JAS; 3-10% JFM, OND), South Africa (25-45% JFM, OND; 5%AMJ), Australia (20-45% AMJ, JAS, OND) and areas with snow/ice cover like the Northern part of Canada (15-30%) and Russia (10-30%) during the JFM and OND seasons.

On the other hand, there are localized areas where Terra reports higher cloud values than

Aqua. Some of them are evergreen forest and croplands, Terra with values ranging from 3-15%. In the Sahara desert, during the JFM season, there is not a clear pattern. However most of the region Terra is reporting higher values between 17-30% and a region on the North West of Sudan with 63% higher than Aqua.

Aerosol Distribution

We looked at the aerosol distribution considering only high and average aerosol loads, being problematic for data quality. Fig 15 and 16 show Terra and Aqua high aerosol distribution respectively. For all seasons barren and desert areas are the least impacted at less than 1% by high aerosol. AMJ is the season with the lowest frequency of high aerosol for all regions, mostly below 10% and a few areas experience up to 18%. (Fig 15-b and 16-b). During the JAS season values remain below 10%, with the exception of a high peak values up to 70% in the region of Angola and the Democratic Republic of the Congo, in Africa, up to 35% in, South America and up to 25 % at Hubei region of China. The OND seasons values are below 17%, except for a spike of 60% in the region contained between Delhi and Bihar in India in the shadow of the Himalayas. Terra and Aqua show very similar results.

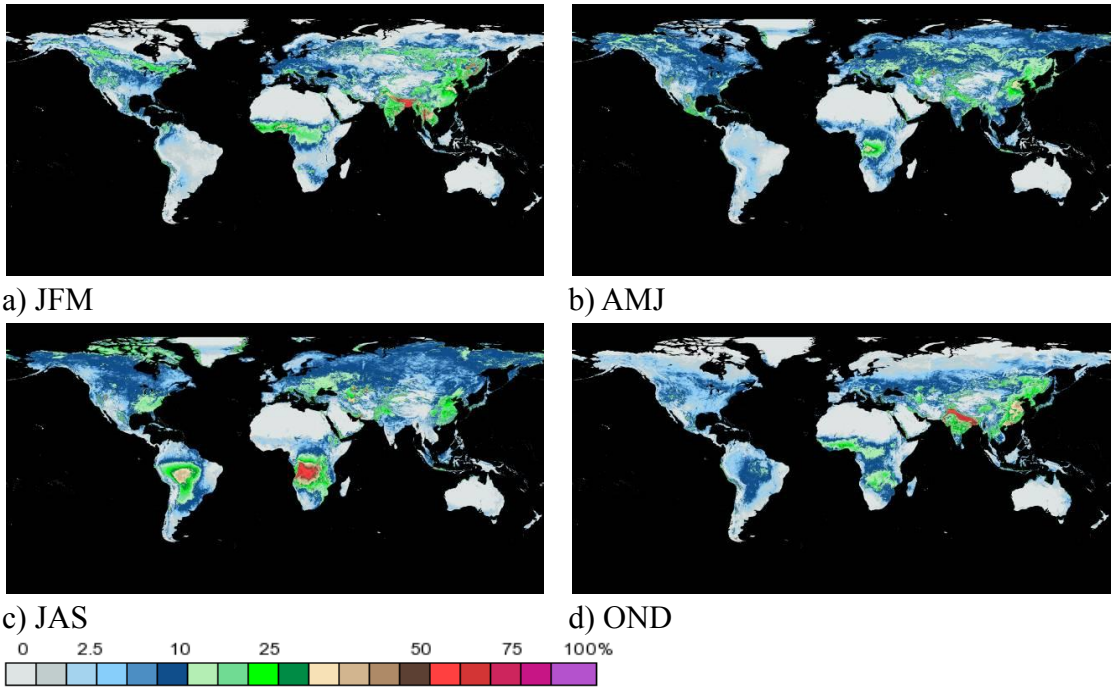


Figure 15. Terra MODIS seasonal high aerosol distribution.

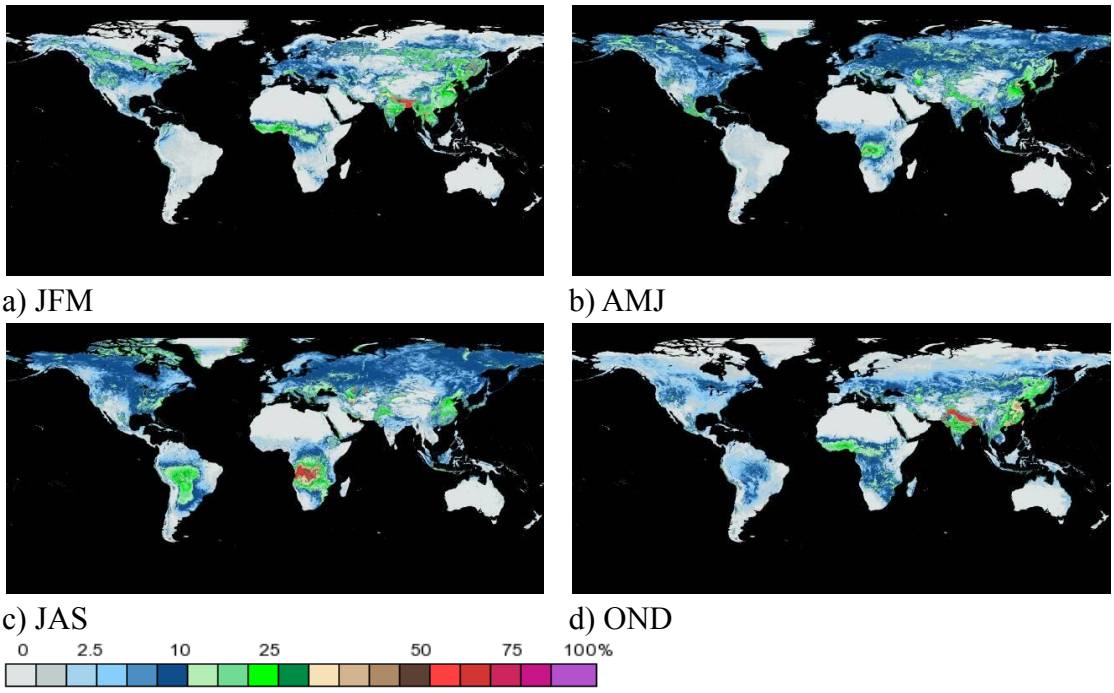


Figure 16. Aqua MODIS seasonal high aerosol distribution.

Average Aerosol distributions are shown for all seasons for Terra and Aqua (Fig 17 and

Fig 18). Values are mostly below 10% with some areas with up to 18%. Barren area and deserts show values below 1% (North Africa and Australia). In addition, it is noticeable that during the AMJ season South America has the lowest values below 2%, while during the JAS season it remains between 5 and 12%.

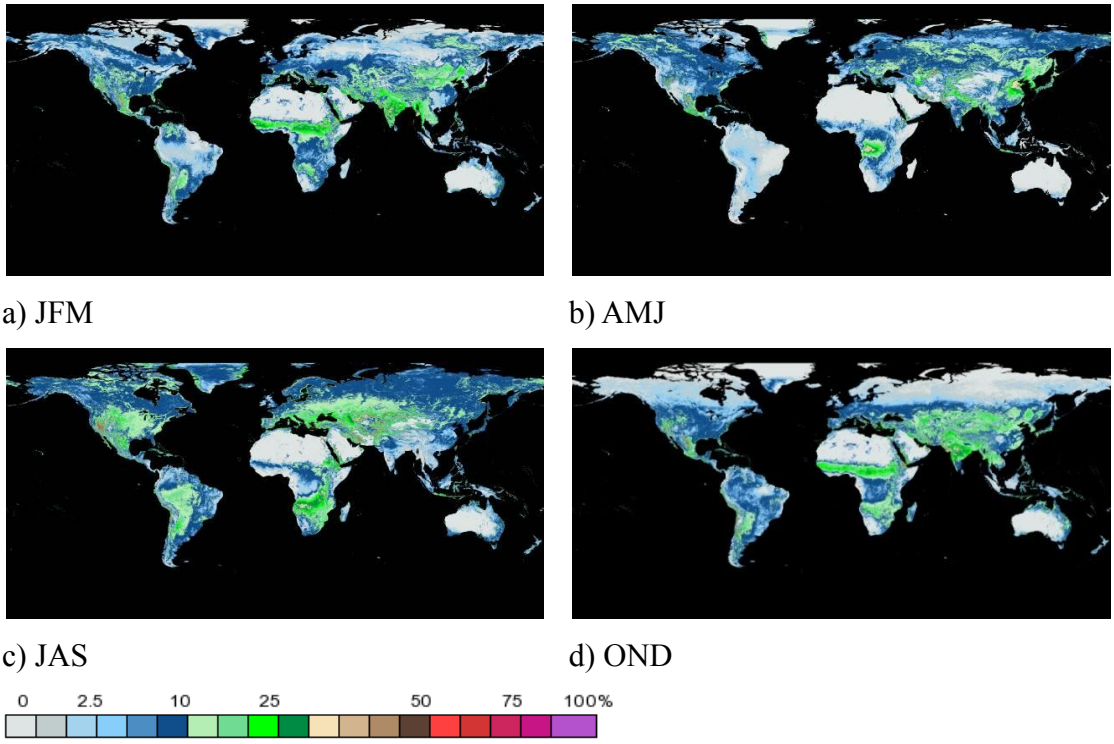
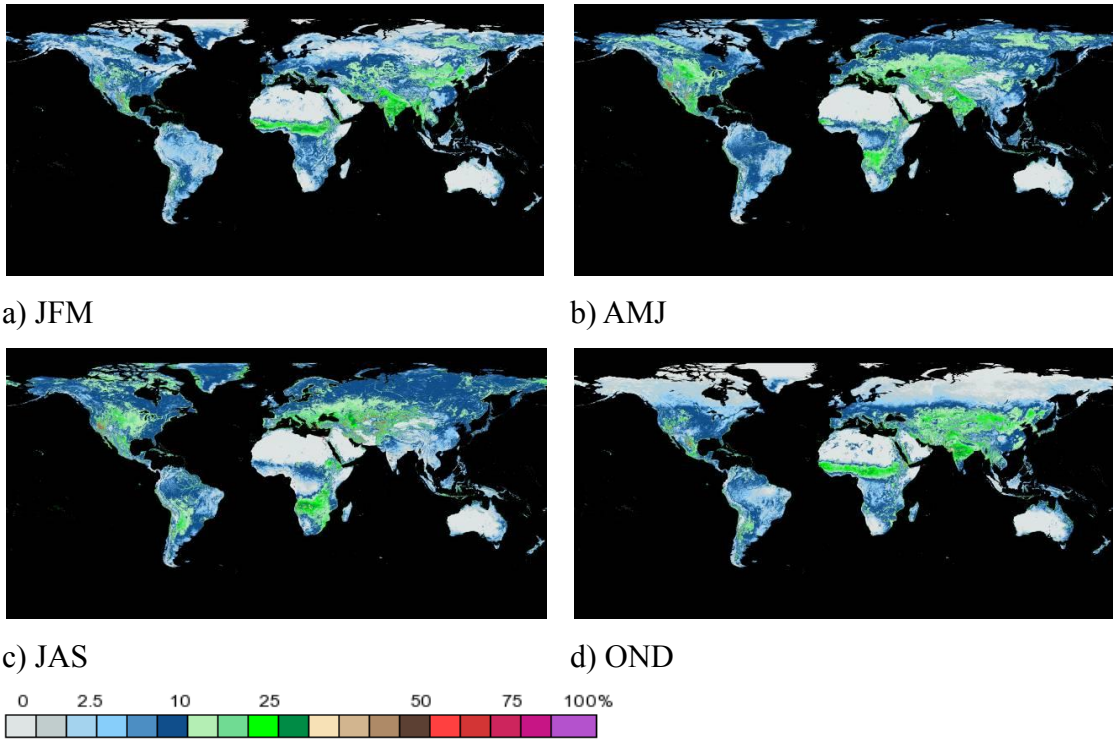


Figure 17. Terra MODIS seasonal average aerosol distribution.

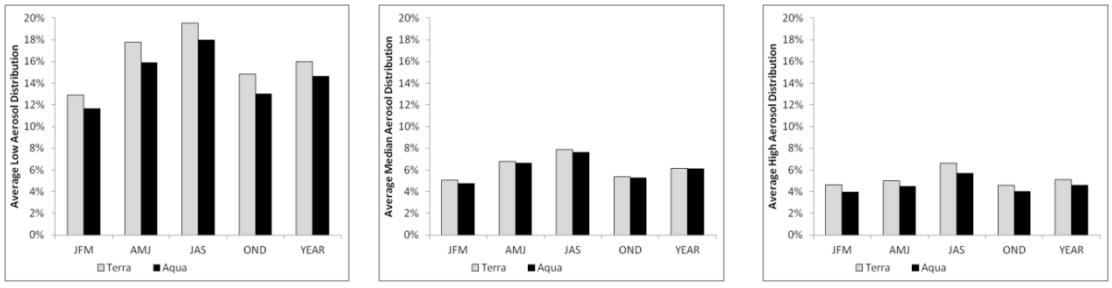


a) JFM b) AMJ

c) JAS d) OND

Figure 18 Aqua MODIS seasonal average aerosol distribution.

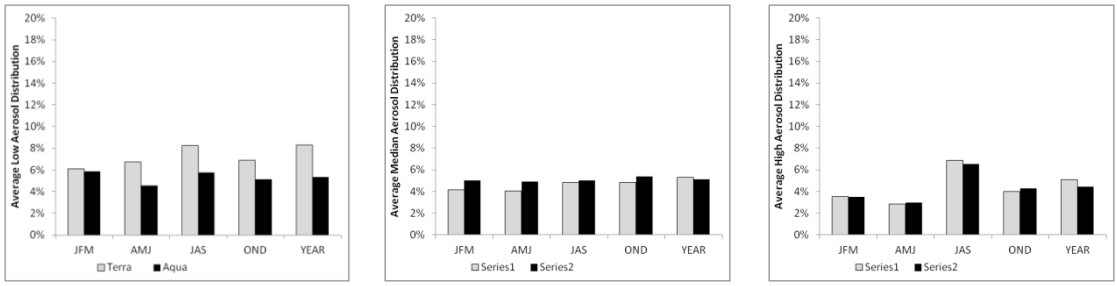
The per sensor global distribution of aerosols are shown in Figure 19 and Figure 20. The global and tropics aerosols frequency and annual distribution indicate Aqua with generally lower aerosol frequency than Terra, owing to cloudiness being higher in the case of Aqua.



a) Low Aerosol b) Median Aerosol c) High Aerosol

Figure 19. Terra and Aqua MODIS seasonal and annual aerosol distribution.

For low aerosol AMJ and JAS seasons are the seasons with highest frequency of low aerosol. In order of magnitude, average aerosols and high aerosols values are below 10 percent globally.



a) Low Aerosol b) Median Aerosol c) High Aerosol

Figure 20. Terra and Aqua MODIS seasonal and annual distribution of aerosol over the tropics.

Data Reliability

The previous results confirm the similar nature of Terra and Aqua MODIS data. Although differences existed related to land cover, location, and overpass time the three sensors report serious percentage of cloud and aerosol cover. To understand the impact of these contaminants on the data we looked at the data reliability, a measure of how likely a cloud and aerosol free observation is achieved. We only considered the Terra platform and AVHRR, separately (Fig 21, 22, 24 and 25) and combined (Fig 23 and 26); by seasons (Fig 21, 22 and 23) and annually (Fig 24, 25 and 26). These maps show a global per-pixel reliability map that captures the average seasonal and annual probability of acquiring data with minimum to no clouds and aerosol problems.

These seasonal reliability maps show that for vegetated areas of key importance the probability of observing good quality data is below 30%. The south west in North America, South Africa, Eastern China and Australia have between 25 to 50% chance of observing high quality data. While the Sahara desert shows a 60-85%. This correlates with the clouds distribution. Clouds are the most important factors impacting synoptic

remote sensing. Similarly to clouds seasonality, these index maps reflect the influence of seasonality.

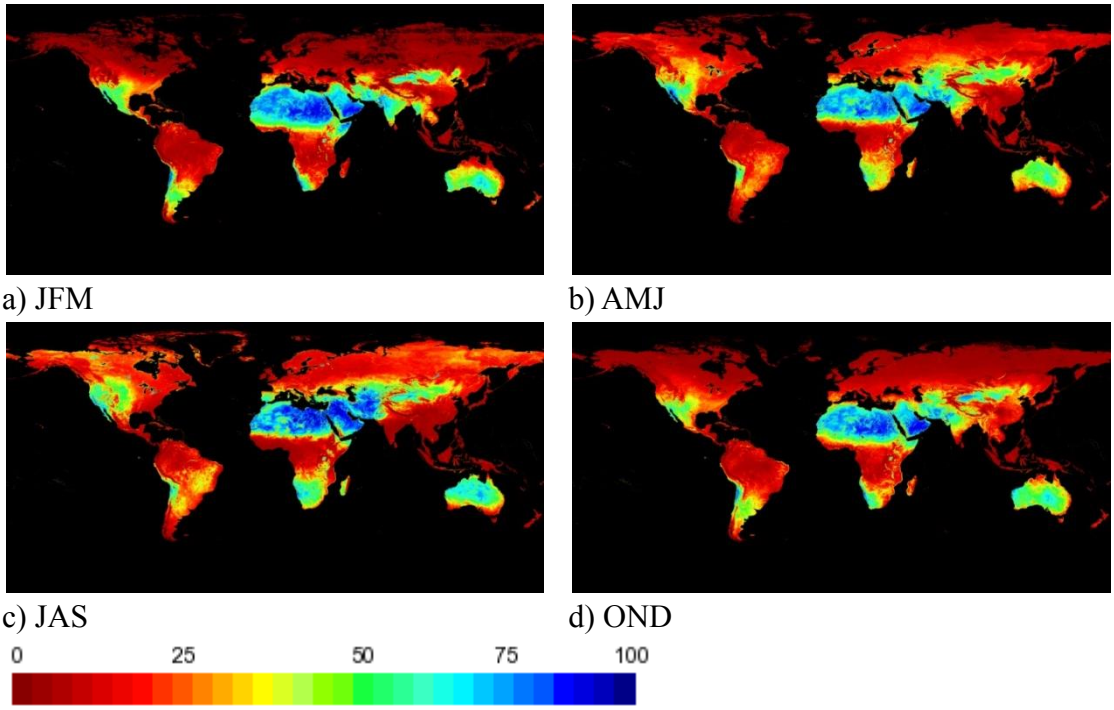


Fig 21. AVHRR seasonal distribution of data reliability.

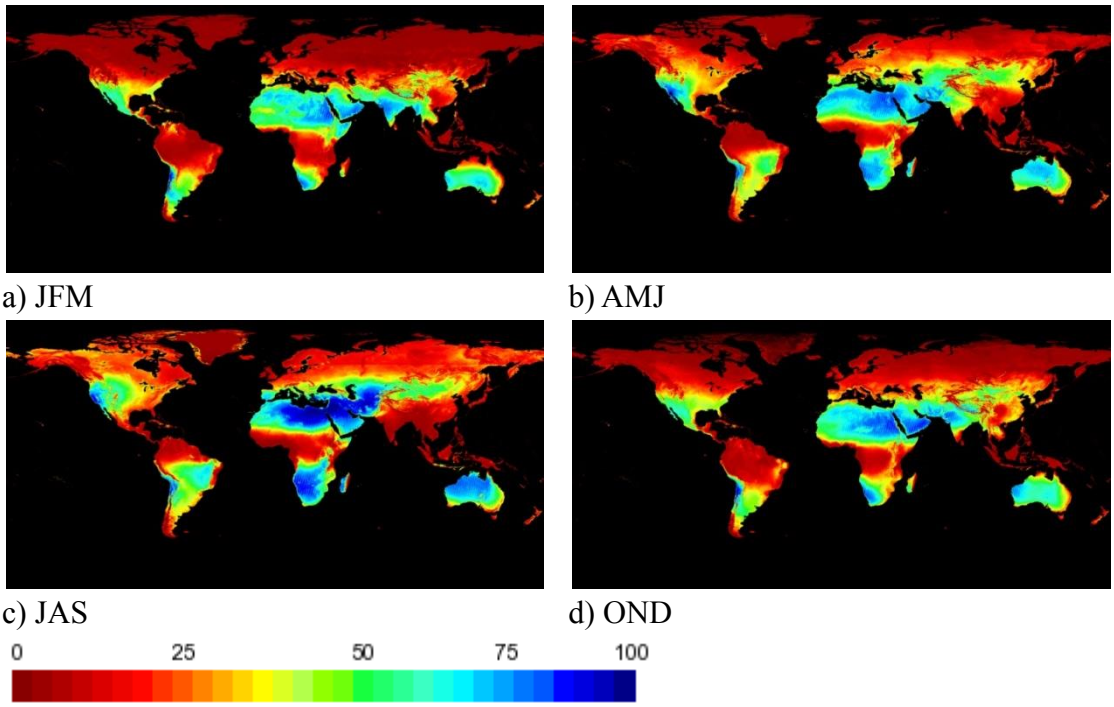


Fig 22. Terra MODIS Terra seasonal distribution of data reliability.

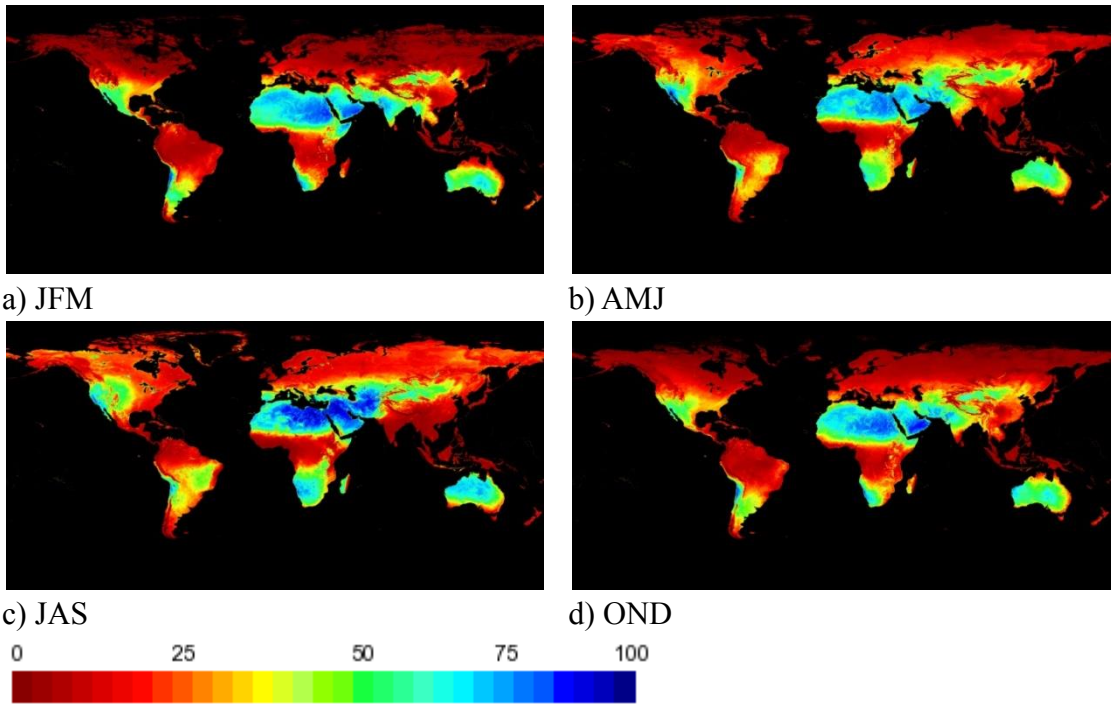


Fig 23. AVHRR and Terra MODIS seasonal distribution of data reliability.

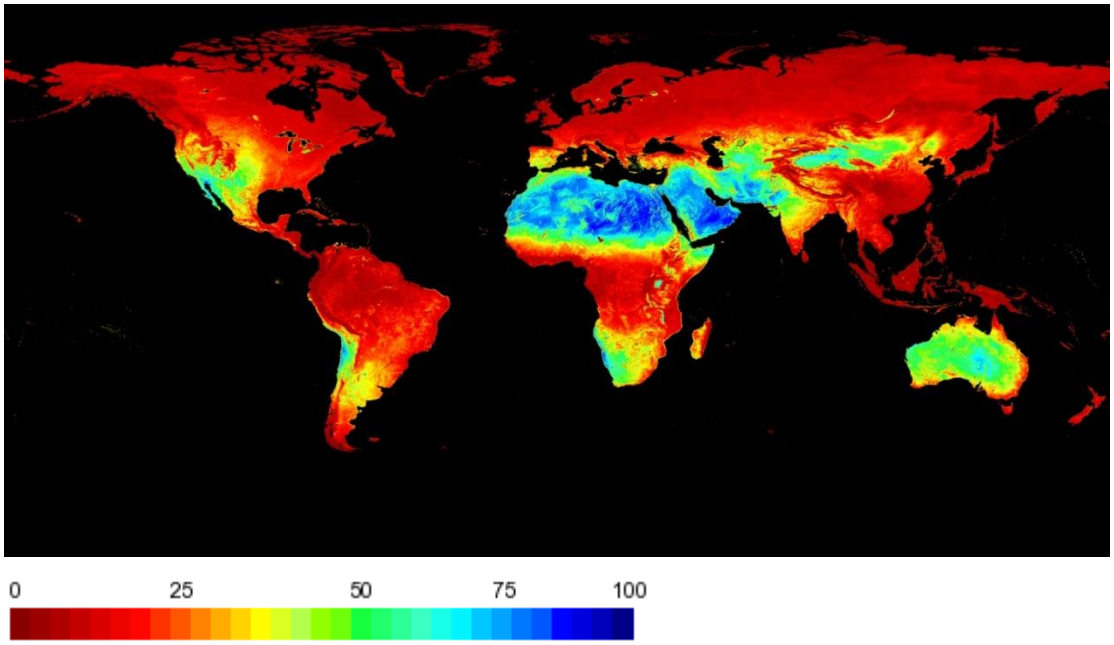


Fig 24. AVHRR annual distribution of data reliability.

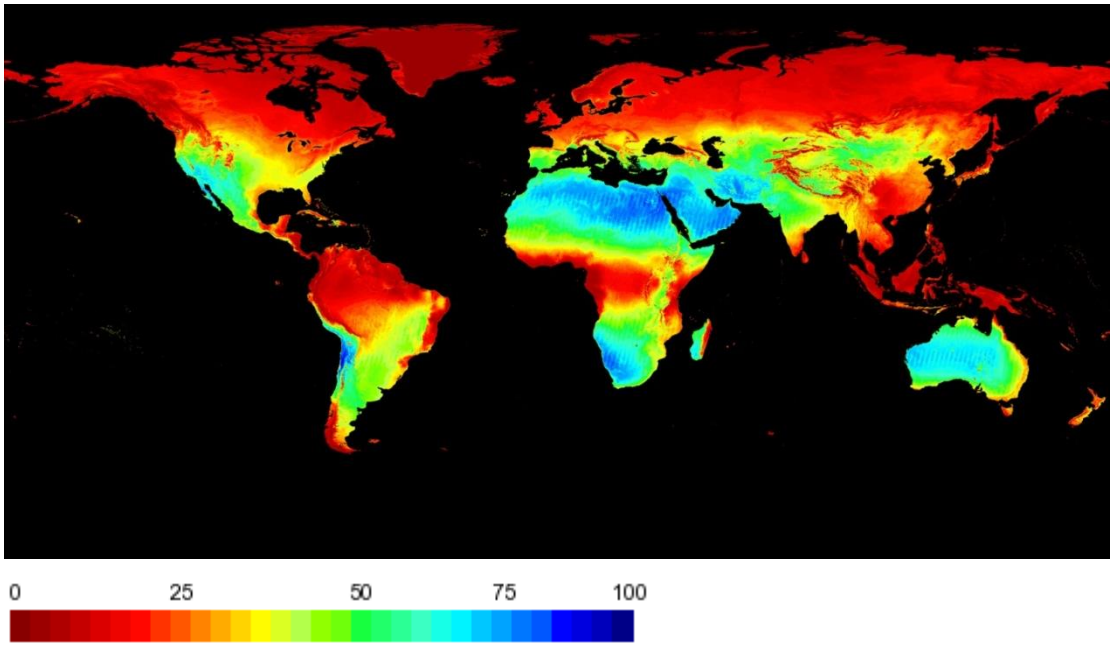


Fig 25. Terra MODIS annual distribution of data reliability.

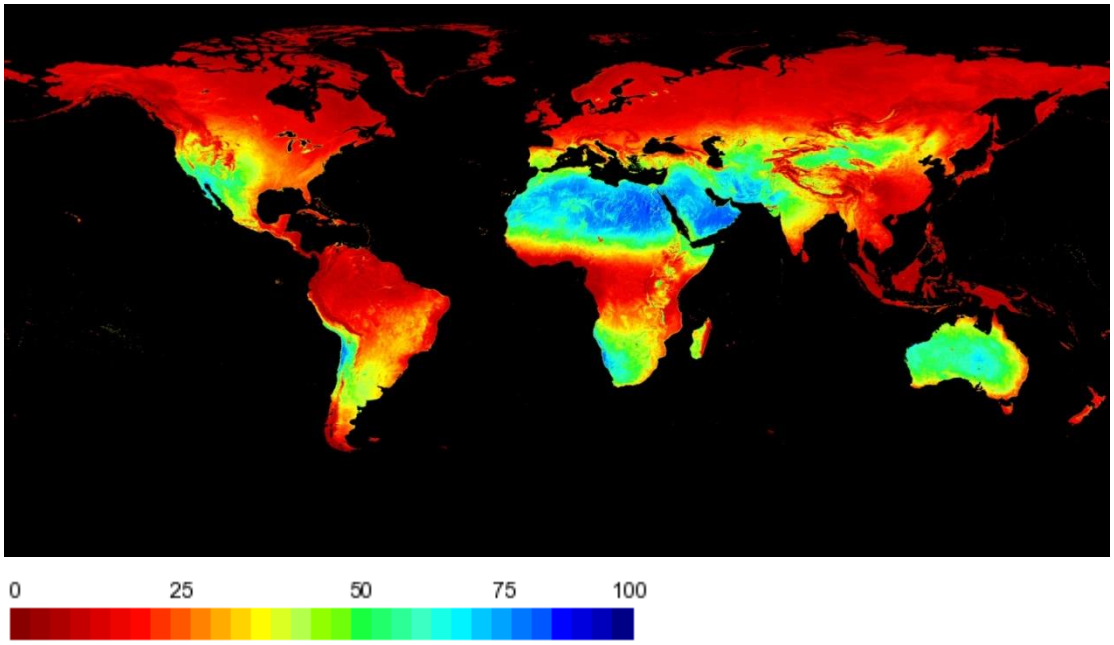


Fig 26. AVHRR and Terra MODIS annual distribution of data reliability.

From these global maps, areas with the highest yield of good quality data are located in the least vegetated areas (North Africa, South West United States and Northern part of Mexico, South Africa and Australia). And areas with the lowest yield correspond to some of the most important biomes like the rainforest in the Amazonas, and the evergreen and deciduous forest in Central Africa, and North America. This is mainly because of the presence of clouds in those areas.

In general, when comparing the reliability map from AVHRR and MODIS for their respective periods there is 10-15% difference. AVHRR reliability maps report higher values in the Sahara desert and lower values in areas with dense forest vegetation. This is a direct result of the AVHRR cloud mask algorithm performance.

Conclusions

More than 60% of the Earth is covered by clouds, the presence of clouds reduces the visibility of the land observed by space-borne satellite platforms and renders their observation of limited to no use for land based research and applications. Clouds are the most severe inhibitor to synoptic remote sensing data. Although aerosols are problematic also, they can for the most part be measured and corrected by fairly well tested and validated algorithms, especially for low and average aerosol loads.

The low performance of the cloud masking of AVHRR resulted in higher rates of cloud commission when compared to MODIS. In addition since there is no aerosol retrieval or correction for AVHRR our results were limited to MODIS.

The probability of obtaining cloud free data for a particular geographic location can be assessed and even predicted based on this work. Because of these issues studying biomes like the tropical rainforests and boreal forests with satellite imagery may be very challenging and would require accurate preprocessing to eliminate the pitfalls of using poor quality data. It is interesting to note that the areas with lowest vegetation cover (deserts, barren soil) are the least impacted by these problems.

When comparing Terra MODIS versus Aqua MODIS, we found out that in general the PM overpass is subject to higher cloud cover in the range of 5%. When looking at the spatial distribution, the differences are larger in arid and snow/ice regions where it can range from 15-45% higher values than the AM overpass.

We developed a simple procedure to re-create the snow/ice and aerosol information flags for the AVHRR dataset by using long term information from MODIS. This information was then used to complement the ranking pixel algorithm and develop the reliability maps that include all the quality information needed to characterize the pixel quality status, instead of only using clouds in the case of AVHRR.

While compositing helps address some of these issues and limitations the temporal repeat is lost due to the reduction of data frequency since the compositing algorithms have to wait for cloud and problem free observations. For phenological or change studies where variability from day to day is important other approaches may need to be considered. For example, a dynamic compositing period could be considered as a hybrid approach solution. This work also showed that the PM overpass was slightly more problematic. Synoptic remote sensing is for the most part the corner stone of global change research, however, large amounts of data are of little to no use due to these problems and need to be discarded, a more effective solution could be the adoption of geostationary satellite that could opportunistically wait for the clouds to move and image the exposed land surface. Additionally our work points to the problematic nature of observing high latitude due to cloud, snow/ice, aerosol, and limited times where passive remote sensing is used.

VI. References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., & Gumley, L. E. (1998). Discriminating clear sky from clouds with MODIS. Journal of Geophysical Research: Atmospheres (1984–2012), 103(D24), 32141-32157.

Chou, M. D., Childs, J., & Dorian, P. (1986). Cloud Cover Estimation Using Bispectral Satellite Measurements. Journal of Applied Meteorology, 25, 1280-1292.

Derrien, M., Farki, B., Harang, L., LeGleau, H., Noyalet, A., Pochic, D., & Sairouni, A. (1993). Automatic cloud detection applied to NOAA-11/AVHRR imagery. Remote Sensing of Environment, 46(3), 246-267.

Didan, K., & Huete, A. (2006). MODIS vegetation index product series collection 5 change summary. MODIS VI C5 changes, 17. http://ncar.datamirror.csdb.cn/modis/resource/doc/MOD13_C005_Change.pdf

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment, 114(1), 168-182.

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. Journal of Geophysical Research: Atmospheres (1984–2012), 102(D6), 6831-6864.

Holben, B., Vermote, E., Kaufman, Y. J., Tanre, D., & Kalb, V. (1992). Aerosol retrieval over land from AVHRR data-application for atmospheric correction. Geoscience and Remote Sensing, IEEE Transactions on, 30(2), 212-222.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote sensing of environment, 83(1), 195-213.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., ... & Morisette, J. T. (2002). An overview of MODIS Land data processing and product status. Remote Sensing of Environment, 83(1), 3-15.

Justice, C. O., & Townshend, J. R. (1994). Data sets for global remote sensing: lessons learnt. International Journal of Remote Sensing, 15(17), 3621-3639.

Kaufman, Y. J., Remer, L. A., Tanré, D., Li, R. R., Kleidman, R., Mattoo, S., ... & Koren, I. (2005). A critical examination of the residual cloud contamination and diurnal sampling effects on MODIS estimates of aerosol over ocean. Geoscience and Remote Sensing, IEEE Transactions on, 43(12), 2886-2897.

Kaufman, Y. J., Tanré, D., Dubovik, O., Karnieli, A., & Remer, L. A. (2001). Absorption of sunlight by dust as inferred from satellite and ground-based remote sensing. Geophysical Research Letters, 28(8), 1479-1482.

Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. Journal of Geophysical Research, 102(D14), 17051-17.

Kawamoto, K., Nakajima, T., & Nakajima, T. Y. (2001). A global determination of cloud microphysics with AVHRR remote sensing. Journal of Climate, 14(9), 2054-2068.

Kishcha, P., Starobinets, B., Kalashnikova, O., Long, C. N., & Alpert, P. (2009). Variations of meridional aerosol distribution and solar dimming. Journal of Geophysical Research: Atmospheres (1984–2012), 114(D10).

Koslowsky, D. (1997). A cloud screening algorithm for daytime AVHRR data using dynamic thresholds. Advances in Space Research, 19(3), 533-536.

Oreopoulos, L., Wilson, M. J., & Várnai, T. (2011). Implementation on Landsat data of a simple cloud-mask algorithm developed for MODIS land bands. Geoscience and Remote Sensing Letters, IEEE, 8(4), 597-601.

Levy, R. C., Remer, L. A., Martins, J. V., Kaufman, Y. J., Plana-Fattori, A., Redemann, J., & Wenny, B. (2005). Evaluation of the MODIS aerosol retrievals over ocean and land during CLAMS. Journal of the Atmospheric Sciences, 62(4), 974-992.

Martins, J. V., Tanré, D., Remer, L., Kaufman, Y., Mattoo, S., & Levy, R. (2002). MODIS cloud screening for remote sensing of aerosols over oceans using spatial variability. Geophysical Research Letters, 29(12), 8009.

NASA, (1993). NASA EOS Reference Handbook, NASA Publication 202NASA Earth Science Support Office, Washington, DC (1993), p. 145

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J., ... & Holben, B. N. (2005). The MODIS aerosol algorithm, products, and validation. Journal of the Atmospheric Sciences, 62(4), 947-973.

Running, S. W., Baldocchi, D. D., Turner, D. P., Gower, S. T., Bakwin, P. S., & Hibbard, K. A. (1999). A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. Remote Sensing of Environment, 70(1), 108-127.

Stowe, L. L., Davis, P. A., & McClain, E. P. (1999). Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the Advanced Very High Resolution Radiometer. Journal of Atmospheric and Oceanic Technology, 16(6), 656-681.

Stowe, L. L., McClain, E. P., Carey, R., Pellegrino, P., Gutman, G. G., Davis, P., ... & Hart, S. (1991). Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. Advances in Space Research, 11(3), 51-54.

Vermote, E., & Kaufman, Y. J. (1995). Absolute calibration of AVHRR visible and near-infrared channels using ocean and cloud views. International Journal of Remote Sensing, 16(13), 2317-2340.

Vermote, E. F., El Saleous, N. Z., & Justice, C. O. (2002). Atmospheric correction of MODIS data in the visible to middle infrared: first results. Remote Sensing of Environment, 83(1), 97-111.

Vermote, E. F., Kotchenova, S. Y., & Ray, J. P. (1999). MODIS surface reflectance user's guide. MODIS Land Surface Reflectance Science Computing Facility, version, 1.

Vermote, E., Saleous, N. E., Kaufman, Y. J., & Dutton, E. (1997). Data pre-processing: Stratospheric aerosol perturbing effect on the remote sensing of vegetation: Correction method for the composite NDVI after the Pinatubo eruption. Remote Sensing Reviews, 15(1-4), 7-21.

APPENDIX B: VIP DATA EXPLORER - AN ONLINE SYSTEM SERVING 30 YEARS OF REMOTE SENSING VEGETATION INDEX AND PHENOLOGY OBSERVATIONS

Armando Barreto[1], Kamel Didan[1,2], Muluneh Yitayew[1,]
Tomoaki Miura[3], Javzandulam Tsendayush[3]

1. Agricultural and Biosystems Department, The University of Arizona, Tucson AZ.
2. Electrical and Computer Engineering Department & Institute of Environment. The University of Arizona. Tucson, AZ.
3. Department of Natural Resources and Environmental Management. University of Hawaii, Manoa, HI.

Paper was prepared to submit to journal of Remote Sensing of Environment, Elsevier B. V.

Abstract

The Vegetation Index and Phenology Lab at The University of Arizona produced a over 30-year seamless and sensor independent record of land surface vegetation indices and phenology metrics. These Earth Science Data Records were generated globally at 0.05-degree resolution and at daily, 7-day, 15-day and monthly temporal frequency. Remote sensed based products have been normally made available for distribution through the internet by data centers like the Land Processes Distributed Active Archive Center, Level-1 and Atmosphere Archive and Distribution System with simple interfaces for data browsing and acquisition. In this work, an online tool, the VIP Data Explorer, was developed to support the visualization, exploration and distribution of these ESDRs. The system is based on a client/server architecture that allows for the direct access, browsing, manipulation and download of these data records with minimum computing needs at the user side, while providing a set of online capabilities for the visualization, qualitative and quantitative exploration of these data records. This web-based tool makes it possible for users of these products to explore and evaluate this data before download and use.

Introduction

The Internet provides an easy channel of data distribution through its worldwide reach (Abel et al, 1998; Hu Shunfu, 2008) to millions of users. In recent years there has been an increase in its use for distribution of remote sensed data. With the emergence of new web technologies like JavaScript, Extensive Markup Language (XML), HyperText Markup Language (HTML5) and Asynchronous communication and processing, it is now possible to make large amounts of geospatial resources available on the net. This has resulted in the development of very powerful and interactive mapping applications and services (Standart et al., 2011) like Google maps (http://maps.google.com), Yahoo maps (http://maps.yahoo.com), Bing maps (http://bing.com/maps/), Nokia maps (http://m.here.com), Mapquest (http://www.mapquest.com) and the community maintained Openstreetmaps (http://www.openstreetmap.org). These web applications display high-resolution satellite imagery with a global coverage and present visual aid to routing traffic and driving directions, while providing access to actual data supportive of research.

Until recently, the most important use of the internet by the remote sensing community has been the distribution of large dataset archives of satellite images and derived products. However the evolution of web applications and wide adoption of the internet technologies by the remote sensing community and the relevance and free availability of virtual globes like Google Earth (http://www.google.com/earth/), NASA World Wind (http://worldwind.arc.nasa.gov/java/) and ArcGIS Explorer (http://www.esri.com/software/arcgis/explorer) has encouraged the creation of similar applications (Goodchild, 2007) and the use of these applications for the dissemination of satellite imagery.

Remote sensed data usually requires the use of specialized imagery processing software, huge amounts of data storage and high capacity processing power computers (Zhao et al., 2012). With the use of new web technologies these requirements can be minimized. The

simplicity of the client/server model allows processing scripts and applications, databases operations and imagery files storage to be handled by the server while the client interacts with the user (Abel et al., 1998; Chen et al., 2012) trough a web browser on a computer with minimum computing power.

The Land Processes Distributed Active Archive Center (LP DAAC), Level-1 and Atmosphere Archive and Distribution System (LAADS) distribute remote sensed data through web applications where users can visualize data and submit orders for posterior data download. Examples of these are: The Reverb (http://reverb.echo.nasa.gov/reverb/) system provides means for discovering, accessing and invoking data products and services. The USGS EarthExplorer (http://earthexplorer.usgs.gov/) allows the query and order of different cartographic products, aerial photographs, and the Landsat and MODIS products. The Global Data Explorer (http://gdex.cr.usgs.gov/gdex/) provides access to the ASTER Global digital elevation model data. The web-enabled landsat data project application (http://landsat.usgs.gov/WELD.php) provides 30-meter Landsat 7 image mosaics at different composites periods. The United States Geological Survey (USGS) Global Visualization Viewer (http://glovis.usgs.gov/) provides searching and data ordering capabilities of user selected geographical regions for different sensors.

The Vegetation Index and Phenology Laboratory (VIPLab, http://vip.arizona.edu) at The University of Arizona in collaboration with the University of Hawaii, Boston University, National Oceanic and Atmospheric Administration (NOAA), and the USGS-EROS LP-DAAC, was supported by NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) project to generate a over 30-years multi-sensor global data record of vegetation index (VI) and land surface phenology (LSP). This effort created a global long-term seamless and sensor-independent data records by combining data from the Advanced Very High Resolution Radiometer (AVHRR) aboard the NOAA polar-orbiting satellites and data from the Moderate Resolution Imaging Spectrometer (MODIS) aboard both Terra and Aqua satellites. This record covers the period 1981 to

2012. Temporal gaps in the record were addressed using the European SPOT-VEGETATION data (Tucker et al., 2005; Brown et al., 2006). These VIPLab data products consist of daily global images at the climatic modeling grid resolution (CMG, ~5.6km). Using the long-term data record (LTDR) AVHRR09 version 3 (Vermote et al., 2010; Pedelty et al., 2007) for the period 1981 to 1999 and the MODIS Collection 5 data from 2000-2012 (Huete et al., 2002; Justice et al., 2002) as the input for these records. Two vegetation indices were generated, the Normalized Difference Vegetation Index (NDVI)(Tucker, 1979; Holben,1986; Huete et al., 2002; Tucker et al., 2005) and the 2-band Enhanced Vegetation Index (EVI2) (Huete et al., 2006; Jiang et al., 2008). Using the per-pixel Quality Assurance flags each input pixel was ranked and poor quality pixels discarded. An across sensor continuity algorithm (Miura et al., 2002; Yoshioka et al., 2005; Murphy et al., 2001; Tsend-Ayush et al., 2011; Yoshioka et al., 2012) to address the sensors' radiometric, spectral, and spatial differences was used to normalize the AVHRR with that of MODIS. In addition to daily images, composited products (Holben, 1986; Van Leeuwen et al., 1999; Huete et al., 2002) at 7 days, 15 days, and monthly were also generated.

Using the standard MODIS algorithm for vegetation dynamic (Zhang et al., 2003; Zhang et al., 2005) the daily vegetation index data were used to generate a global record of land surface phenology metrics. Phenology measures the land surface vegetation change over time and captures the growing season, it is an excellent indicator of global, environmental, and climate change impacts on vegetation (Braswel et al., 1997; Zhang et al., 2003). The growing season in a remote sensing context is defined by a set of parameters that indicate when the season started, ended, and when certain biophysical events took place.

Additionally, the algorithm provides information about the integration of the VI signal, a strong proxy of vegetation biomass production (Boelman et al., 2003; Running et al., 2002). To address the bimodal growing season of certain cropped areas and areas subject

to multiple wet seasons (Ethiopian high plateau, Himalayan foothills) the land surface phenology algorithm was designed to look for multiple growing seasons (Didan & Huete, 2004; Zhang et al., 2005).

The first objective of this project was to generate the 30-years seamless sensor independent data record by designing, evaluating and executing the filtering compositing and gap filling algorithms. Also, to implement and execute the continuity and phenology metrics algorithms designed by the team members of the VIPLaboratory. An additional objective was to create a web-based tool that allows for the direct online exploration of these records and provides means for data distribution. Thus, in this manuscript we present a set of community vetted algorithms developed by the VIPLab and used to derive these long term, seamless, and sensor independent data records, and an online system to preview, explore, manipulate, and acquire these records.

Multi-sensor data record generating algorithms

Thirty years of daily global surface reflectance data in the Hierarchical Data Format - Earth Observation System (HDF-EOS) in the latitude/longitude projection at a spatial resolution of 0.05 degrees in a grid of 3600 rows and 7200 columns were acquired from AVHRR Long Term Record and MODIS. The AVHRR dataset was downloaded from the LTDR (AVHR09 V3) project website (http://ltdr.nascom.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi) covering the 1981-1999 across 4 different platforms N07 (1981-1985), N09 (1985-1988), N11 (1988-1994) and N14 (1994-1999). Every file consisted of 10 layers (Table 1) of data stored in 16 bits integers and is approximately 160 MB.

Table 1. AVHRR09 version 3 data layer specifications

Science Data Sets (HDF Layers)	UNITS	BIT TYPE	FILL	VALID RANGE	SCALE
Surface Reflectance 640 nm (SREFL_CH1)	Reflectance	16-bit signed integer	-9999	-1000 – 10000	0.0001
Surface Reflectance 860 nm (SRFEL_CH2)	Reflectance	16-bit signed integer	-9999	-1000 – 10000	0.0001
Brightness Temperature 3.75 microns. (BT_CH3)	Kelvin	16-bit signed integer	-9999	0 - 1000	0.1
Brightness Temperature 11.0 microns. (BT_CH4)	Kelvin	16-bit signed integer	-9999	0 – 1000	0.1
Brightness Temperature 12.0 microns. (BT_CH4)	Kelvin	16-bit signed integer	-9999	0 – 1000	0.1
Surface Reflectance 3.75 microns (SREFL_CH3)	Reflectance	16-bit signed integer	-9999	-1000 – 10000	0.0001
Solar Zenith Angle (SZEN)	Degree	16-bit signed integer	-9999	-9000 – 9000	0.01
View Zenith Angle (VZEN)	Degree	16-bit signed integer	-9999	-7000 – 7000	0.01
Relative Azimuth (RELAZ)	Degree	16-bit signed integer	-9999	-18000 – 180000	0.01
Quality Assurance (QA)	Bit Field	16-bit signed integer	na		na

The MODIS data consisted of the Terra MOD09CMG collection 5 product from February 2000 to December 2012. It was downloaded from the MODIS data pool (ftp://e4ftl01u.ecs.nasa.gov/). All data were stored on a dedicated storage system at the VIP Lab at the University of Arizona. Similar to the AVHRR dataset, the MODIS data is daily global, gridded on a simple 0.05 degree geographic projection. Every file consisted of 19 layers (2000-2001) and 21 layers (2002-2012) in a 3600 rows and 7200 columns grid system and approximately 985 MB in size. The 21-layers are listed on Table 2, however files for Year 2000 and 2001 do not contain the layers Coarse Resolution Band 3 Path Radiance or Coarse Resolution State QA. The Coarse Resolution State QA layer contains the Pixel State Quality Assessment and when missing the Coarse Resolution Internal CM layer was used instead, however this data layer was not as accurate.

Table 2. Terra MODIS collection 5 data layer specifications.

Science Data Sets (HDF Layers)	UNITS	BIT TYPE	FILL	VALID RANGE	SCALE
Coarse Resolution Surface Reflectance Band 1 (620–670 nm)	Reflectance	16-bit signed integer	-28672	-100 – 16000*	0.0001
Coarse Resolution Surface Reflectance Band 2 (841–876 nm)	Reflectance	16-bit signed integer	-28672	-100 – 16000	0.0001
Coarse Resolution Surface Reflectance Band 3 (459–479 nm)	Reflectance	16-bit signed integer	-28672	-100 – 16000	0.0001
Coarse Resolution Surface Reflectance Band 4 (545-565 nm)	Reflectance	16-bit signed integer	-28672	-100 - 16000	0.0001
Coarse Resolution Surface Reflectance Band 5 (1230–1250 nm)	Reflectance	16-bit signed integer	-28672	-100 – 16000	0.0001
Coarse Resolution Surface Reflectance Band 6 (1628–1652 nm)	Reflectance	16-bit signed integer	-28672	-100 – 16000	0.0001
Coarse Resolution Surface Reflectance Band 7 (2105–2155 nm)	Reflectance	16-bit signed integer	-28672	-100 – 16000	0.0001
Coarse Resolution Solar	Degree	16-bit	0	0 –	0.01

Zenith Angle		signed integer		18000	
Coarse Resolution View Zenith Angle	Degree	16-bit signed integer	0	0 – 18000	0.01
Coarse Resolution Relative Azimuth Angle	Degree	16-bit signed integer	0	-18000 – 180000	0.01
Coarse Resolution Ozone	cm atm	8-bit unsigned integer	0	0 –255	0.04
Coarse Resolution Brightness Temperature Band 20	Kelvin	16-bit unsigned integer	0	0 – 40000	0.01
Coarse Resolution Brightness Temperature Band 21	Kelvin	16-bit unsigned integer	0	0 – 40000	0.01
Coarse Resolution Brightness Temperature Band 31	Kelvin	16-bit unsigned integer	0	0 – 40000	0.01
Coarse Resolution Brightness Temperature Band 32	Kelvin	16-bit unsigned integer	0	0 – 40000	0.01
Coarse Resolution Granule Time	HHMM	16-bit signed integer	0	0 –2355	1
Coarse Resolution Band 3 Path Radiance	Reflectance	16-bit signed integer	-28672	-100 – 16000	0.0001
Coarse Resolution QA	Bit Field	32-bit unsigned integer	0	na	na
Coarse Resolution Internal CM	Bit Field	16-bit unsigned integer	0	0 – 8191	na
Coarse Resolution State QA	Bit Field	16-bit unsigned integer	0	0 – 65535	na
N pixels averaged	none	8-bit unsigned integer	0	0 – 40	na

*: Theoretically a surface reflectance data should be between 0 and 1.0, the additional range is used in MODIS to capture atmosphere correction performance issues and data outside the normal range (0-1.0) should be discarded.

The remaining AVHRR gap in 1999 to MODIS 2000 was bridged using SPOT-VEGETATION data at 1km, which was downscaled and resampled to the Climatic Modeling Grid CMG resolution (Barreto et al., 2013).

The system processed these records into daily and multi-day products of vegetation index and phenology. A series of algorithms (Fig 1) were used to filter and remove poor quality data, then generate the vegetation product, the continuity algorithm is then applied to the AVHRR record. The resulting spatial gaps are filled using a simple Inverse Distance Weighting interpolation method (Riveracamacho et al., 2011), which assigns more weight to the neighboring temporal observations.

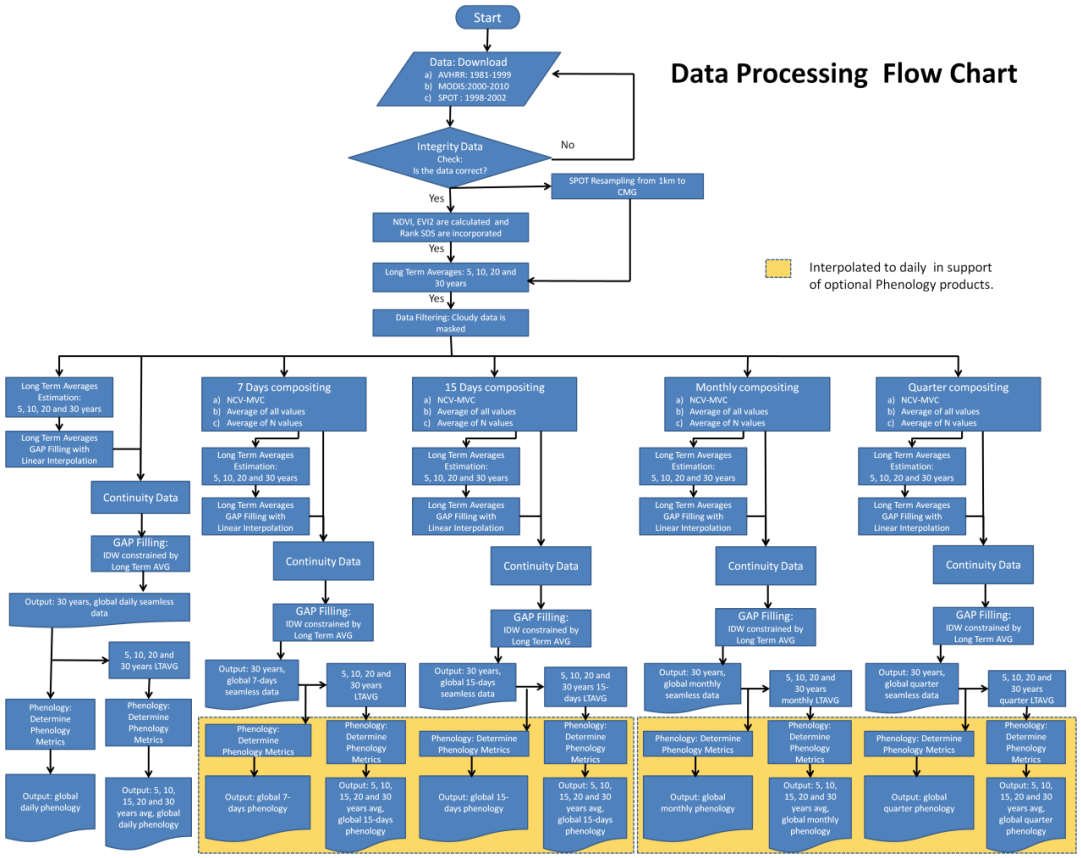


Figure 1: The VIP data processing plan.

A large effort was placed on data filtering to remove questionable data, particularly cloudy and poor quality that may compromise research and applications. An end to end reprocessing of the full data record takes between 1-2 months using an off the shelf 2 Linux Centos and 3 Red Hat systems cluster with a total of 104 Intel Xeon CPUS @ 2.93GHz, and with an online storage capacity of 200 TB and 152 GB of RAM.

Data preprocessing

The data processing follows a loose sequential approach (Fig. 1), however, there are processing nodes that could be initiated in parallel to speed processing. In the following sections we describe each step in detail.

a) Vegetation Indices and Ranking

The full input record (1981-2011) was preprocessed at the VIP Lab. Preprocessing consisted of calculating NDVI, EVI2, and ranking each observation based on its input QA. The current record stops at 2011 because adding the newer year takes some time. Starting Version 3 newer years (2102+) will be added to the record.

Pixel values from the RED and NIR layers were checked to be in the valid range (according to Table 1 and 2). If so, the NDVI and EVI2 values were computed, otherwise a FILL_VI value (-15000) was assigned to the pixel. Only land pixels were considered for processing.

b) Ranking

Each pixel was ranked based on the QA layer information (Quality Assurance layer for AVHRR and Coarse Resolution State QA for MODIS), View Zenith Angle and Sun Zenith Angle. A pixel quality category was assigned (Table 3) to each pixel using the decision tree in Fig 2. With this information, a Pixel Reliability was estimated to indicate how likely a pixel would be free of problems. A layer was created to store this info for each daily global image.

Table 3 Pixel ranking classes and description.

Rank	Description	Rank	Description
0	Excellent	6	Cloud
1	Good	7	Estimated
2	Marginal	-1	NO DATA
3	Questionable	-2	NO DATA / High Latitude
4	Cloud Shadow	-4	Antarctica
5	Ice/Snow	-5	Water/Ocean

Since AVHRR data neither reports aerosol information nor snow/ice, low aerosol was assumed for non-cloudy pixels and snow/ice was not considered for this step. Similarly data from MODIS years 2000 and 2001 aerosol information was not reported, and low

aerosol was assumed for cloud-free pixels. Pixels were flagged properly in the QA to account for these assumptions.

The NO-DATA class was assigned to pixels over land where data was a) not reported due to orbit not-overlapping and/or data not reported, b) surface reflectance values out of range (according to Table 1 and 2) and c) the NDVI and/or EVI2 fell out of range. NO-DATA High latitude class was used for pixels over land during the winter months when the sun angle is $\geq 85^{o}$.

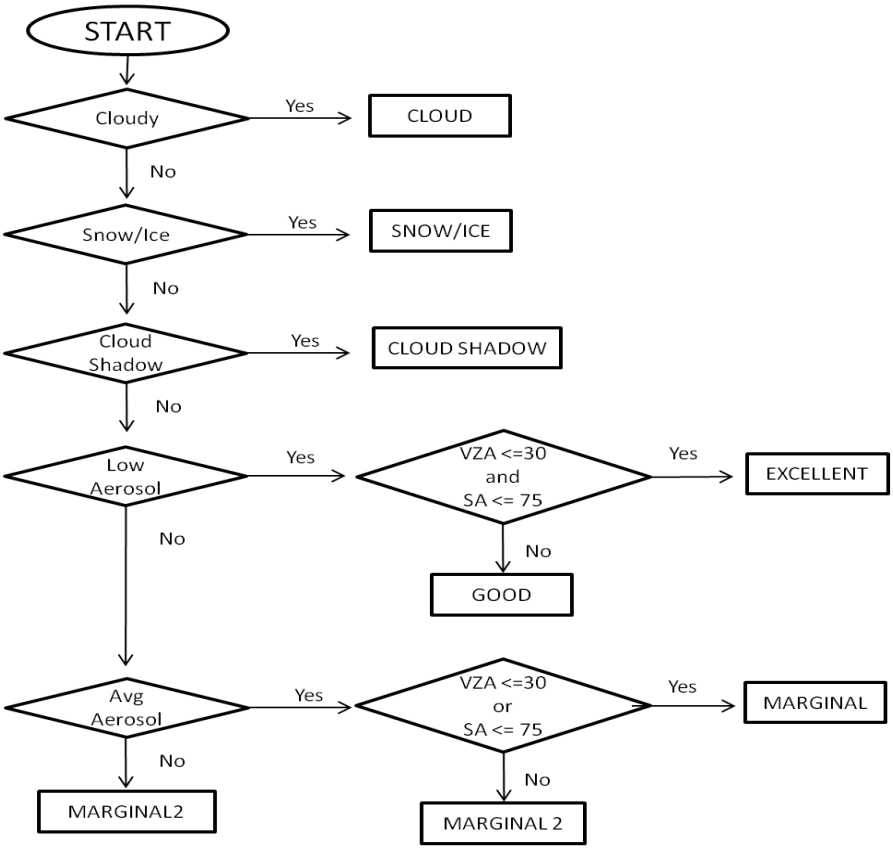


Figure 2. Pixel ranking approach.

Given the different QA bits descriptions from each sensor (AVHRR, MODIS 2000-2001 and MODIS 2002-2012), a special VIP QA Layer was created to standardize this QA information (Table 4). This new QA is used for subsequent processing steps and in

support of consistency for this 30-year data record. This QA layer is 16-bit unsigned integer with bits listed from the most significant bit (MSB) 15, to the least significant bit (LSB) 0.

Table 4. VIP product suite specific per-pixel QA layer description.

Bit No.	Parameter Name	Bit Value	State QA
13-15	Land/water flag	000	Shallow ocean
		001	Land
		010	Ocean and lake shorelines
		011	Shallow inland water
		100	Ephemeral water
		101	Deep inland water
		110	Continental/moderate ocean
		111	Deep ocean
12	View angle > 30	0	No
		1	Yes
11	Sun zenith angle > 85	0	No
		1	Yes
10	Sun zenith angle > 75	0	No
		1	Yes
8-9	Gap fill status	00	Not gap filled
		01	Interpolated
		10	Long Term average
		11	Not set
7	Snow/Ice estimated	0	No
		1	Yes
6	Snow/Ice flag	0	No
		1	Yes
5	Aerosol estimated	0	No
		1	Yes
3-4	Aerosol quantity	00	Climatology
		01	Low
		10	Average
		11	High
2	Cloud shadow	0	No
		1	Yes
0-1	Cloud state	00	Clear
		01	Cloudy
		10	Mixed
		11	Not set, assumed clear

This new dataset containing NDVI, EVI2, VIPQA, Pixel reliability layer, RED, NIR, (BLUE and MIR layer for MODIS) View zenith angle, Sun zenith angle and Relative azimuth was labeled as VIP01P1 product, where the 01 represents daily data (i.e 07 for 7-days and 15 for 15-days composite) and P1 indicates the first processing step

c) Long Term Averages

Because prior data analysis indicated that the pixel QA information is not 100% accurate and that cloud-free pixels may actually be cloudy and vice versa, we decided to calculate a long term average data record from all the cloud free pixels NDVI and EVI2 in order to use to further filter and rank the data. If not disturbed, a pixel should usually follow the long term average profile plus noise related to inter-annual change. Any large deviation ($\geq 1.5\ \sigma$) is considered the result of poor quality that was not captured by QA and is subsequently rejected. This approach usually works fine except in case of disturbance (fire, infestation, etc.) or when natural change is greater than σ (ex: land use change). However, the combination of QA filtering and this long term constrainer addresses this rather well and only very few pixels are wrongly labeled. Even if we mislabel a pixel using this approach, it is usually acceptable to be conservative in this context, as throwing a good observation is more tolerable than using a poor quality one with potential for biasing results. Overall, this approach performed extremely well at addressing poor quality data and helped improve the product record fidelity (Didan et al., 2012; Barreto et al., 2013)

Long term average NDVI and EVI2 values were calculated from daily images, excluding the questionable data (Rank ≥ 3). 5, 10, and 20 year Long term averages were computed to promote the use of most recent data to be temporally close to the pixel being analyzed if possible. In this context, the 5 year is used first, then the 10, and finally the 20-year is used last. Additionally this aids in processing by not mixing data from more than one platform to avoid bias.

d) Aerosol and Snow/Ice Mask

Using the Terra MODIS daily dataset (2000-2012) a long term average aerosol and snow/ice mask were derived. These masks consist of the frequency of occurrence of snow/ice, and aerosols and numbers range from 0 to 100%.

The MODIS Coarse Resolution State QA and the Coarse Resolution Internal CM layer were used for this analysis. Cloud, aerosol, snow/ice and water/land mask information were extracted and analyzed and only land pixels were processed. Aerosol was separated into the three classes as described in the QA (low, average and high). This results in 365 daily files that indicate how likely a pixel could be cloudy, ice/snow covered, or aerosol covered, which are averages from analyzing the full MODIS record (12 years). This record can serve as a benchmark for assessing the ice/snow and aerosol quality for each of the AVHRR pixels since no information is available. We realize the inaccuracies of such process, but due to snow/ice dynamics and bulk of aerosol due to smoke resulting from the slash and burn practice in the rainforest which is seasonal and quite repetitive at about the same time period. This data record can then serve as an additional info to screen and filter the AVHRR data which lacks this information.

Filtering

Using the calculated Long Term Average values of NDVI and EVI2, the estimated snow/ice and aerosols masks, the 30 years VIP01P1 daily products were reprocessed into level-2 VIP01P2 products.

This consisted of reevaluating the ranking system using the additional historical information. For every pixel the NDVI and EVI2 values were compared against the long-term value for the 5 years period containing the processing year. Using the long-term average VIs and the long term standard deviation a min/max range was created using 1.5

standard deviations to help identify outliers. If a pixel falls outside this margin and was previously ranked as cloud free or otherwise good it was re-ranked as cloudy pixel.

In addition, for AVHRR, it was possible to incorporate aerosol quality information and snow/ice status in the rank and QA layers. If for a given AVHRR pixel, the estimated snow/ice mask indicated that the pixel was always (2000-2012) snow/ice, then it was re-ranked as snow/ice and the snow/ice estimated flag was turned to YES to indicate it was estimated and not real. If the pixel was labeled as cloudy in the input, but the NDVI value was within the min-max boundaries it was re-ranked as snow/ice if the probability of snow/ice is above 80%.

Because aerosols attenuate the VI signal, the higher the concentration of aerosol was associated with the lower the signal. If VI values were above the standard deviation of the long term, low aerosol or average aerosol category was assigned to the pixel. If the VI value was below the average VI but above the min value, then average aerosol or high aerosol was assumed. With this new re-ranking process the QA layer was updated to reflect these changes and permit the proper post-processing. Using this information a filtered and extremely high fidelity dataset can easy generated.

Continuity

In order to create a seamless and sensor-independent dataset for the 30-year span, two continuity methods were used. A simple single global regression equation (Tsend-Ayush et at., 2011; Yoshioka et al., 2012; Miura et al., 2002) and equations by land cover (Tsend-Ayush et at., 2011). A per pixel regression (Didan et al., 2013) is now being designed and currently being tested for a future data release. These equations permit the conversion of any 1981-1999 AVHRR NDVI or EVI2 value (y variable) to an AVHRR-MODIS equivalent. Translation equations were applied only to cloud free high quality

data. All other pixels were omitted and considered missing to be replaced by the gap filling algorithm.

a) Version 1: Single Continuity Equation

For Version 1.0 we used the continuity equations in Table 5 (Tsend-Ayush et at, 2011) which account for biases across platforms and sensors, a single global equation was used per AVHRR platform for NDVI and EVI2. This translation equation was applied pixel by pixel to cloud-free data over land to derive the MODIS equivalent AVHRR daily data from 1981 to 1999.

Table 5. AVHRR to MODIS continuity translation equations.

Sensor	NDVI Equation	Uncertainty (95% PI)
N07	$Y=-0.0646111+1.2409713x-0.0304219x^2$	± 0.138
N09	$Y=-0.0621082+1.2487272x-0.0307315x^2$	± 0.138
N11	$Y-0.0606805+1.2456808x-0.0335204x^2$	± 0.138
N14	$Y=-0.0571829+1.2372178x$	± 0.138
	EVI2 Equation	
N07	$Y=-0.0403338+1.2400319x$	± 0.088
N09	$Y=-0.0403338+1.2400319x$	± 0.088
N11	$Y=-0.0403338+1.2400319x$	± 0.088
N14	$Y=-0.0403338+1.2400319x$	± 0.088

b) Version 2: Continuity Equations by Land Cover

A single global equation maybe simpler to use and more consistent, however issues were noticed and a strong land cover dependency was identified. To address this, a set of land cover specific equations were used to minimize these biases (Tsend Ayush et al., 2012). These equations (Table 6) were normalized by the 2001 MODIS Land Cover map (MOD12C1) with 16 different land cover types. The year 2001 was selected as a mid-point between the two, AVHRR and MODIS, records.

Table 6. AVHRR to MODIS translation equations by land cover.

Land Cover Type	NDVI Equation
Evergreen needle leaf forest	y=-0.04178518+1.252120236x
Evergreen broadleaf forest	y=0.063729956+1.101413471x
Deciduous needle leaf forest	y=-0.079968156+1.308736967x
Deciduous broadleaf forest	y=0.049592289+1.100792188x
Mixed forest	y=0.003659916+1.158981715x
Closed shrubland	y=-0.033027165+1.187927641x
Open shrubland	y=-0.059602521+1.234259184x
Woody savannas	y=-0.017827492+1.195172775x
Savannas	y=-0.031046917+1.193979443x
Grasslands	y= -0.041965171+ 1.17542088x
Permanent wetlands	y=-0.057182852+1.237217846x
Croplands	y=0.002047676+1.12516719x
Urban and built up	y=-0.057182852+1.237217846x
Cropland/nature veg	y=0.005918222+1.129474142x
Snow and ice	y=-0.057182852+1.237217846x
Barren or sparsely vegetated	y=-0.007130528+0.865844462x
	EVI2 Equation
Evergreen needle leaf forest	y=-0.003041271+1.160234372x
Evergreen broadleaf forest	y=0.075359038+1.087815468x
Deciduous needle leaf forest	y=-0.015021464+1.171831062x
Deciduous broadleaf forest	y=0.011102573+1.16520626x
Mixed forest	y=-0.001067469+1.171690446x
Closed shrubland	y=-0.009871674+1.112843881x
Open shrubland	y=-0.024113998+1.120142751x
Woody savannas	y=0.00748675+1.133180482x
Savannas	y=-0.014023451+1.175724903x
Grasslands	y=-0.02039079+ 1.121235191x
Permanent wetlands	y=-0.040333833+1.240031857x
Croplands	y=0.001648462+1.101353053x
Urban and built up	y=-0.040333833+1.240031857x
Cropland/nature veg	y=0.002845267+1.136705983x
Snow and ice	y=-0.040333833+1.240031857x
Barren or sparsely vegetated	y=0.002196079+0.77907402x

These equations were applied to each pixel and each daily image for the AVHRR record

GAP Filling

Because only high fidelity data is considered in our processing, which results in large swaths of missing data, in particular areas prone to clouds and aerosols, we designed a gap filling algorithm to address this missing data. This gap filling approach is applied in two steps, an interpolation model to fill the missing values and a long term average value to replace hard to interpolate cases.

The inverse distance weighting (IDW) interpolation method was used because of its simplicity and superior performance than the traditional linear interpolation gap filling technique. The IDW method reconstructed the daily vegetation index time series for each year, using two search radiuses (Riveracamacho et al, 2011). First an attempt with a maximum search radius of 30 days (before and after the missing observation day) was used. This was done to limit the number of days contributing to the estimation of the missing pixel, which is beneficial for areas where cloud free days are easy to obtain in short periods. Given the nature of the method, the days closer to the day of the year in question, contribute more than the far away days, keeping the trends and growing season dynamic mostly intact with minimal perturbation.

A second search radius of 60 days (before and after) was applied when the first radius was not successful and no useful data was found to help reconstruct the missing observation. This is true for regions where clouds remained for long periods of time (ex: Rainforests like the Amazon). By using this extended search period, the likelihood of finding useful vegetation index values was higher. The interpolated value is further compared to the long term average value to insure an unbiased estimation. If the IDW model was not able to produce a value within that range, the long term average was used instead. First, the 5-year long term average is tried and if no value is found the 10-year, 20-year and 30-year were tried.

The gap fill status flag in the VIP QA layer is used to track real versus interpolated values. The flag indicates if the pixel was estimated (IDW method), filled from long term average or not filled, which happens in extreme cases in particular over high latitude areas.

Land Surface Phenology data record

Using this new long-term multi-sensor record of NDVI and EVI2 we generated an annual record of land surface phenology (White et al., 1997, 2009; Reed et al, 2003). To generate the land surface phenology product suite we used two approaches: (1) the standard MCD12Q2 vegetation dynamic algorithm (Zhang et al., 2003); and (2) the homogeneous cluster approach (White et al., 2009).

The MODIS land cover dynamics product algorithm is designed to estimate four transition dates about the vegetation growing season: (1) The greenup or date of onset of VI increase for the stable minimum (dormancy); (2) maturity or date of onset of VI maximum; (3) senescence or date of VI decrease start; and (4) dormancy or date when the VI drops down to its minimum. In addition, the algorithm integrates the VI signal. Because the standard MODIS algorithm looks at the landscape and is based on the analysis of the VI transition dates and not the specific plant phenology we specifically refer to transitions in VI (onset of increase in VI, etc.) rather than start or end of growing season. The actual MODIS land cover dynamics algorithm is described in detail in Zhang et al., (2003, 2006) and Ganguly et al., (2010). The algorithm looks at the periods with sustained VI increase and decrease to avoid the small decreases or increases caused by noise or changes unrelated to actual plant activities. The algorithm requires that the change in VI must be larger than 35% of the annual range in VI for the pixel and that the ratio of the local maximum VI to the annual maximum VI should be at least 0.7. This insures the filtering of short-term variations in VI that are usually due to things other than plant activities. The actual phenology (both rising and falling limbs of the profile) are modeled using a logistic function. The actual phenological transition dates are then

estimated by analyzing the change in curvature of these fitted logistic models for each pixel (Zhang et al., 2003).

Using the homogeneous cluster approach, the algorithm treats swaths of land that are covered by the same land cover type, under the same climate regime (determined by temperature and precipitation), soil, elevation, and latitude as area having the same growing season. The algorithm computes first these homogeneous clusters (Didan et al., 2010), then the Half-Max VI (White et al., 1997; White et al., 2009) phenology extraction algorithm is applied to each cluster. This same Half-Max approach could also be applied per-pixel to generate a per-pixel phenology product. This version of the algorithm estimates 10 growing season related parameters: (1) Start of Season, (2) End of Season, (3) Length of Season, (4) Green up rate, (5) Brown down rate, (6) Peak Date, (7) Peak Value, (8) Average season signal, (9) Seasonal integrated signal, and (10) number of seasons for areas with more than one defined season.

This overall process results in a high fidelity gap free NDVI, EVI2, and land surface phenology product suite that is seamless, sensor independent and well characterized. Table 7 lists the product suite.

Table 7. VIP version 2 standard and nonstandard datasets.

	Daily	7 Days	15Days	Monthly	Quarterly
Input Data	X				
Filtered Data	X	X	X	X	X
Continuity GAP Filled	X	X	X	X	X
Long Term Average VI	X	X	X	X	X
Pixel Phenology	X	X	X	X	X
Cluster Phenology	X	x	x	x	X

Online tool in support of the 30 Year VI and Phenology record

To support these products and provide the users with an online tool for their exploration and evaluation prior to download and use, the VIP Lab. designed the Data Explorer web tool. To design this tool, we first created an inventory of all the datasets available. A MySQL database was designed in order to quickly query and recover information about the different data files. The database stores and keeps track of the file properties, metadata content, reference and location. This database is the backbone of the online web tool query engine. Management of this database is done via a set of php scripts running on the Linux cluster where the processing and data are housed.

The main purpose of this web-based tool is to centralize the information, delivery and provide for a useful user experience. In doing so one of the main issues requiring special consideration was the Internet speed and bandwidth at the user end. Moving large amounts of information is usually prohibitive and counterproductive and for these reasons we've adopted a modular approach to data services and browse following the Google Earth tile data model. Only the images and/or data in the view are active and available for exploration.

To further improve the performance of this online tool, only the main and key layers associated with these data records are available for browsing and inspection, this includes the NDVI, EVI2, associated QA, False color composite (MNR) composite images, and land surface phenology metrics. The display uses the global full resolution (3600x7200) jpeg images that were made from the data records. Following the Google Earth model, these images are then tiled to 300x300 pixels at full resolution and then coarser resolution to support fast zooming in and out following a pyramid structure. Per-pixel Reliability index images are also available for exploration, as well as MNR false color images. Because AVHRR does not have an MIR channel we developed a simple approach to generating an AVHRR-equivalent MIR band using the relationship between the MODIS

MIR band and the NIR and Red channels. Using this relationship we could generate a false AVHRR equivalent MIR. These data are however available for viewing only.

Data Explorer development

The VIP Data Explorer online tool makes use of several web-based technologies and standalone programs and methods. The tool is organized conceptually in two major parts, the client and the server sides (Fig 3). At the client side the application was developed using HTML5, JavaScript, Cascading Style Sheets (CSS), Extensible Markup Language (XML), Keyhole Markup Language (KML) which is the Google Earth file format, and the Google Earth JavaScript API. The server side resides in a Linux environment and uses the MYSQL engine, with support from HyperText Preprocessing (PHP) and C programming languages. The HDF-EOS library was used to manipulate the data records that are stored in the HDF-EOS file format.

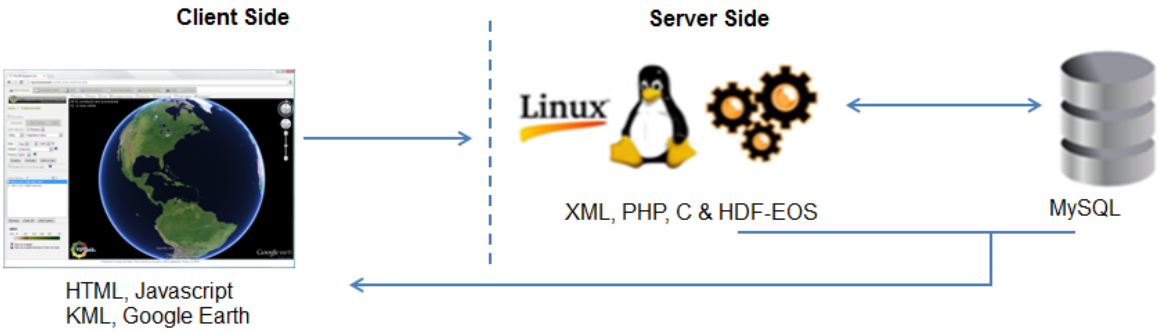


Fig 3. Architecture application flow-diagram

The Data Explorer server side consists of two core layer applications, a private VIP-API and the Data Explorer server processing engine powered by a LAMP open source web application software stack composed of a cluster of Red Hat Enterprise Linux and CentOS Servers, Apache Server, MySQL and PHP. Additionally the HDF-EOS library was installed to enable accessing the raw HDF files data on the fly.

The private VIP-API is responsible of all communications between the client and the server, ranging from serving image tiles from a specific dataset and geographic area to extracting data from the HDF files and scheduling time series processing and data orders. The private VIP-API is written in PHP (HyperText Preprocessing) and it interacts with the MYSQL Database and low level programs written in ANSI C and the HDF-EOS toolkit library to complete tasks and serve users requests.

This API includes a series of scripts that return preprocessed data to the client based on user selections and parameters. The client side waits for the response to be shown to the user. In addition the system has a set of scripts and programs designed for real time interaction with the hierarchical data format (hdf)files; however these tasks are kept to simple pixel query and other services that take little time to complete.

The front end client side is a web application with a dynamic graphical user interface developed using HTML5. The client permits users the selection of datasets, and request from the server to display images or results. The interaction between client and server is mostly done via ajax calls. In general, when the client side makes a request to the server, it sends the required parameters for the task to be completed through the respective API call. The server evaluates the parameters and runs the corresponding application to generate the output, which is returned back to the client to be visualized. At the server side, a set of scripts, using several computer and web languages manage the client's requests. These requests are grouped in two work categories: 1) Needing immediate server response, and 2) Scheduling tasks.

The immediate response requests are managed by scripts that return already preprocessed data to the client based on the user selections and date/time, and spatial parameters. This ranges from returning an image for a specific day and location, to returning a time series plot or a database query results. Some of the requests interact in real time with the hdf files, residing on the server, particularly when extracting a pixel value from the data

display or inspecting a pixel profile. The real time requests are kept to a minimum and are mostly simple tasks that do not take CPU time and in most cases users will not notice the delay.

To reduce latency, tasks that require the manipulation and processing of large amounts of data take place at the server side where the full data record resides. In doing so, the burden of data processing is kept to the server side.

VIP Data Explorer Operational Description

The conceptual design and programming of the several tools that make up the VIP Data Explorer are:

Data Visualization: Datasets are presented to the user as browse images of a user selected layer from a particular dataset on a mapviewer environment. The actual HDF data files are up to 1GB in size, which is impossible to manipulate via the internet, therefore, images were created from the different HDF data layers in advance. These consisted of full resolution global (3600 rows and 7200 columns) jpegs images. A tiling system is used to improve the image browse based on the Google Earth/Map pyramid approach. The VIP-API receives input parameters about the extent of the geographic area and the dataset (type, date, processing level) selected by the user. It returns the necessary image tiles to cover the view extent at the current zoom. This process is repeated every time the user zooms in and out, or changes to a different location view, but thanks to the tiling algorithm call occurs very fast even on slow internet connections. This action is triggered by a JavaScript listening event that occurs when the user moves the map to a different location, or zooms in/out. The Google Earth plugin calls the VIP-API to refresh the view content and it sends the current geographic extent coordinates. In returns it receives the http image resources.

The mapping is handled by the reuse of the Google Earth plugin. In addition it brings the added value of accessing Google-APIS for displaying high resolution imagery, roads, borders, places, terrain, names and search capabilities.

Data Animations: Animation allows users to visualize change over time. Users can select anytime frame and then the system will create an animation from the selected dataset. The full resolution jpg images were resampled to 10% size (360x720 pixels) with 90% compression rate to support fast animation and transmission through low internet connections. Any individual frame from the animation could be explored in detail in the mapviewer.

A JavaScript/DOM function calls the VIP API for retrieving the images to construct the movie and the resources to create a custom animation player with controls to allow play, stop, slow/fast animation and the option to open on the Data Explorer viewer a current frame. The images are animated by modifying the DOM trough JavaScript.

Pixel Data Extraction: This action is executed on the fly. Once the user selects a dataset from the menu options and provides a lat/long location by clicking on the map, the server API converts the geographic location to a row/column position and the name and location of the source file is queried from the database. This information is fed to a c/hdf program that retrieves the pixel location values for each layer on the HDF file. If the file contains a QA layer, the QA bits are decoded for easy understanding.

Transects visualization: A list of lat/long coordinates representing segments (minimum of one segment), is required by the system to extract the transect values. The coordinates are captured by clicking a path over the map viewer. The server side API, converts the array of geographic locations to a row/column positions list on a grid system. A c/hdf program extracts the pixel values that fall within a straight line between each pair of coordinates from the HDF source file. These values are plotted against accumulated

distance and saved as a jpg image. The image is returned on the fly to the client for visualization.

Time Series Viewer: Data Explorer allows the immediate exploration of series from the full 30+ years record. While user defined location time series processing takes time to complete, two sets of already made time series are available for viewing. Time series generated by other users and a global predefined 0.5 degree grid. This on demand service relieves the user from downloading data just for extracting time series. An action that takes hours or days. The time series are displayed in an interactive open source JavaScript charting platform (DyGraph, http://dygraphs.com/). This allows for the dynamic exploration of the chart. Comparison of data charts between different datasets and processes levels. Users can download the actual time series data as a comma separated values (cvs) file for use in other plotting and analysis applications.

On demand Time Series: The extraction of time series is a computer intensive process, especially when working with daily datasets for 30-year record like the VIP. The user selects a geographic location from the map viewer and it is sent to the server for processing. The VIP-API schedules the request on a queue table on the database and the Data Explorer processing engine will redirect the action to a c/hdf program for execution. While the process of extracting one pixel value from a single file is quite simple and requires no time, the repeating of this operation thousands of times (open/close files) becomes a resource intensive operation that require special handling. The system actually schedules the TS extraction and puts it in a special queue. When the extraction is complete the user is notified by email. The email contains a direct access link to the series viewer and instructions on how to access the results.

VIP Data Explorer Interface and Tools Description

The VIP Data Explorer is operating system (OS) independent and its performance will for the most part depend on the user computer and internet bandwidth. The application web address is http://vip.arizona.edu/viplab_data_explorer/ and will work with a minimum configuration.

Data Explorer Viewer

The Data Explorer interface (Fig 4) is simple and intuitive and builds on the Google Earth model. The UI is organized into five sections: 1) Main display area, where maps and images are viewed, 2) data record selection, 3) layers display management, 4) Display color legend, and 5) Data records support documentation and direct database access.

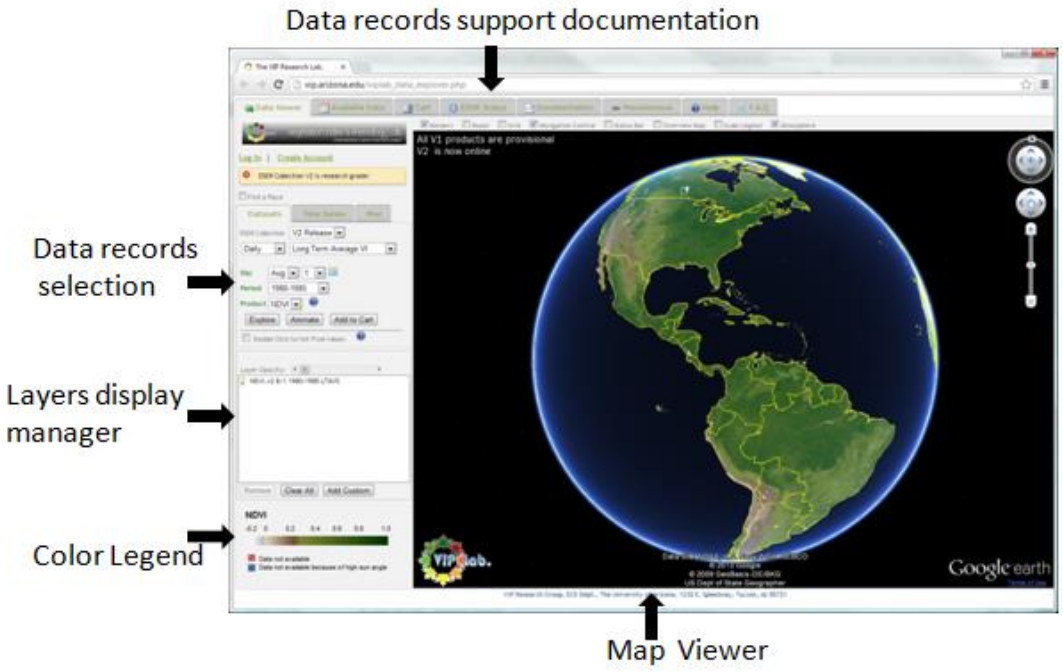


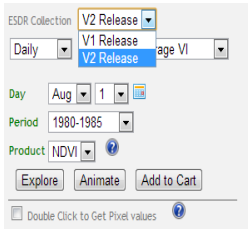
Figure 4. Data Explorer graphical user interface.

The map/image viewer occupies the majority of the screen display, and this is where the selected datasets are displayed. Using the standard Google Earth display manipulation

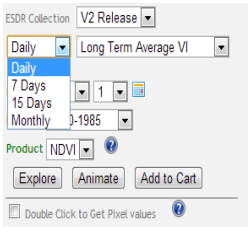
tools, the users can manipulate the display, pan the map and inspect the coordinate location. The UI uses the Google Earth plug-in which needs to be installed at the client side prior to using the Data Explorer. The system allows access to all high resolutions images and ancillary data provided by Google, like roads, country boundaries, place names, and other useful layers for reference purposes. Combined with the VIP datasets this interface provides a highly interactive geographic map viewer and explorer.

Data sets available through VIP Data Explorer

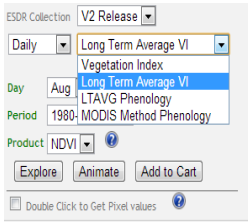
The online tool provides for exploring the full VIP data records (Table 7). A drilldown style menu system (Figure 5) is used for faster access and selection of data records for display. The submenus adjust depending on the data records being explored and the first entries are time selection and re-processing version.



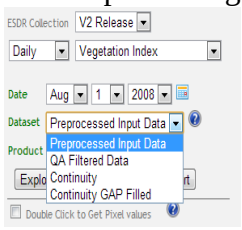
Data Reprocessing release



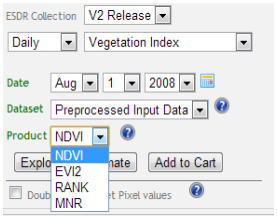
Data Temporal Frequency



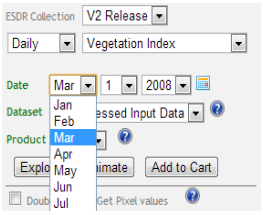
Data Product



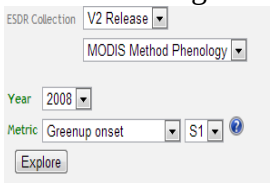
Data Processing level



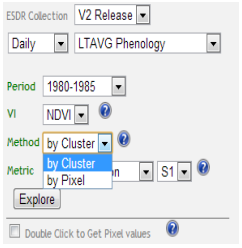
Data Product Layer



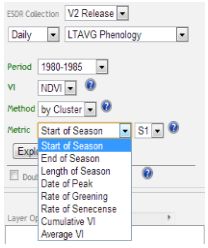
Date selection



Phenology Data selection



Phenology Algorithm



Phenology metrics

Figure 5. Data Explorer navigation menus.

Any single selection dataset or product can immediately be displayed on the Google Earth globe. Separately, longer time series can be animated through the animation window. All selected datasets can be added to the shopping cart for download. The Data Explorer allows for the stacking of multiple datasets using a data layer management control system (Figure 6). The layer stacking approach transforms the Data Explorer into a very powerful data manipulation and exploration tool by providing a fast method of data visual cross comparison across space and time. Additionally users have the option of adding their own custom datasets to these stacks for added versatility. The system allows for simple geographically projected images in JPG, PNG, or KLM format to be added to the Data Explorer display. User can change the opacity of any layer or switches them on/off to visualize the spatial and/or temporal relationships across these data records.

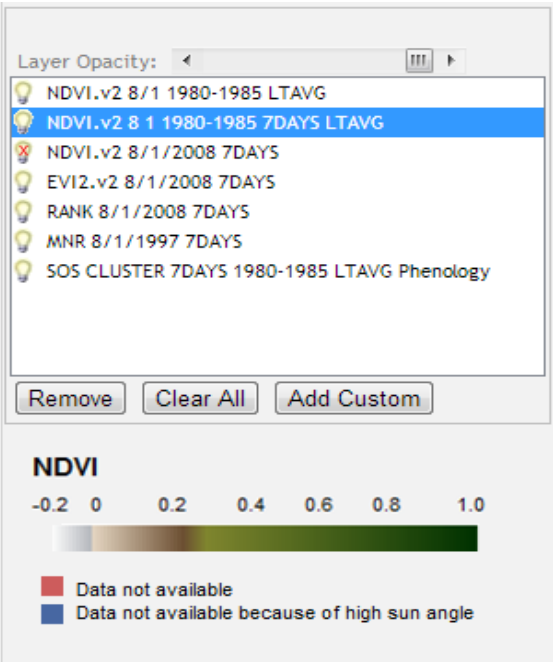


Figure 6. Data Explorer layers display management and control system.

The client side of the Data Explorer automatically assigns and keeps track of all active layers using their metadata or user provided information. A layer can be activated by clicking on its name, which sends a request to load the proper color legend for the display and allows for control of its opacity in the Google Earth globe. The display manager permits the turning on and off of any displayed data record or the complete removal of the records from display. User can upload their own data and images to be visualized and cross compared using the data display manager (Figure 6).

User custom data must be in the form of a jpg/png image or a Google Earth KML or KMZ format. These images must also be in the latitude/longitude coordinate system. User will be asked to provide the corners' coordinates of the image. Data Explorer will handle the display and georeferencing of these custom data sets through the Google Earth API. All custom images are volatile and are not stored in the system for later use and will expire when the session is closed. This service is available only to registered users.

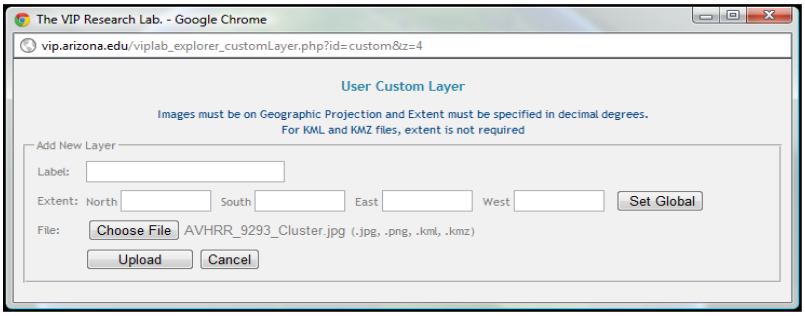
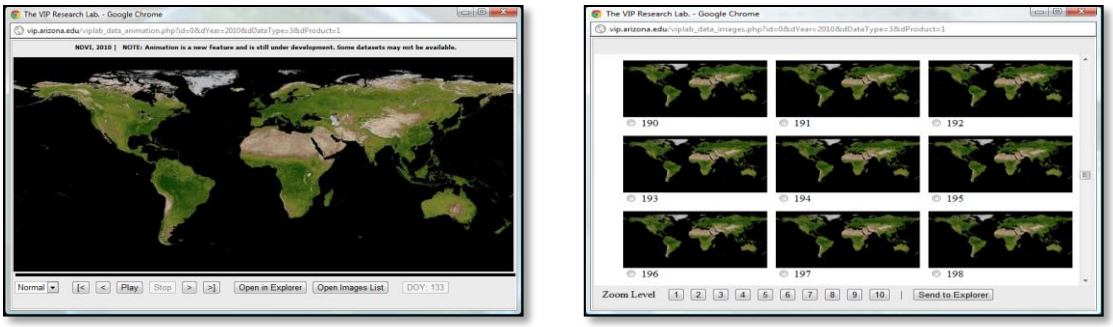


Figure 7. User custom data upload.

b) Data Explorer - Animations

While Data Explorer allows for the full spatial visual exploration and inspection of any single image via the display manager, displaying and exploring multiple images at once is impractical considering the length of our data record. To speed the visualization and exploration of a large stack of images the system provides a specialized animation tool optimized for the visualization of time series images. The Data Explorer allows for the visualization of multiple images over a user selected period of time (Figure 8) in a dynamic animation window, where speed, resolution and other display parameters can been adjusted. It is also possible to display the images time series in a mosaic tabular format and any image can be sent to the Date Explorer display manager for visualization.



a) Movie animation b) Images mosaic

Figure 8. Data Explorer images display and animation.

The animation runs in a yearly loop at global scale. It can be stopped at any day of the year and the corresponding dataset sent to the Data Explorer for viewing and further inspection. This tool is especially useful when exploring daily datasets in a fast and efficient way.

In addition to visually exploring the data record, Data Explorer provides for the quantitative exploration of any displayed data. It allows for displaying the pixel value via the pixel exploration tool for plotting transects and the creation of time series of any of the data records.

c) Data Explorer - Pixel values retrieval

The Data Explorer can directly access any data record in the database and extract the value of any user selected pixel. When the user selects a location of interest the Data Explorer will request the extraction of the corresponding information from the HDF-EOS file and returns it for viewing. The pixel report contains the location in lat/long coordinates and row/column within the file, the date and type of data, and the value from each layer in the file. If the dataset contains a Quality Assurance layer, the system will extract and organize the bit fields into a readable format along with their description.

d) Data Explorer - Transects

In support of exploring geographic transects or profiles the Data Explorer provides a transect tool for the inspection of a range of locations without having to click every pixel. While images can show values and change over a given location, it still requires some visual accuracy on the part of the user to properly correlate colors and understand the change. The transect tool on the other hand allows for the plotting and display of multiple pixels along a set of straight lines all at once. The transect path can be a single segment or a group of segments (Figure 9), which allows for direction change and for the exploration of complex profiles. This tool is limited to only the exploration of NDVI and EVI2 data records.

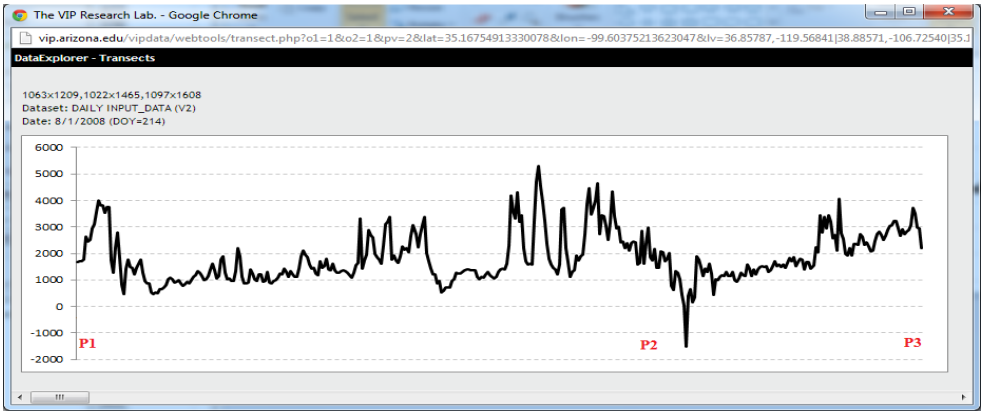


Figure 9. Data Explorer transect plotting.

e) Data Explorer - Time Series

A foundational aspect of remote sensing data analysis is rooted in time series analysis. Global change and all associated data records and analysis are mostly based on the evaluation of time series for trends and change, especially the ones related to climate and land use change. In this context, a trend is a change over time of a key land surface parameter, like a VI or a phenology metric. To support this basic remote sensing data analysis need, the Data Explorer provides a time series analysis tool. Users can either browser the thousands of already preprocessed locations by our systems, or select a new location. A 0.5^{o} regular grid was preprocessed globally and is made available via the Data Explorer. In addition the hundreds of users submitted locations are also available for

exploration. Once a location is selected the user is asked to fill a table to describe and define his site, and the system will start promptly in processing this request. All data from all the processing steps are available for time series visualization (Fig 10). The time series plot is interactive and user can manipulate the time frame and resolution of the display.

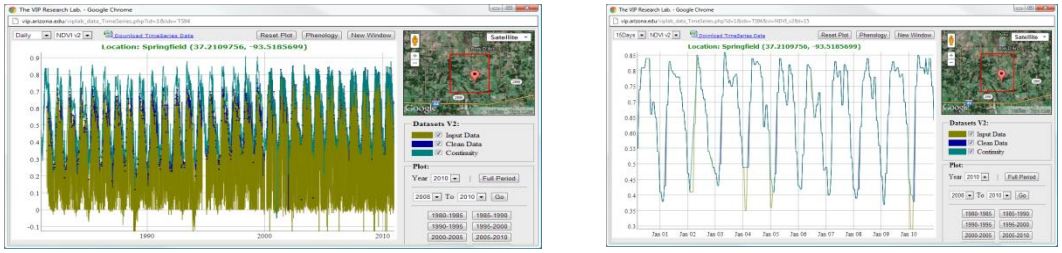
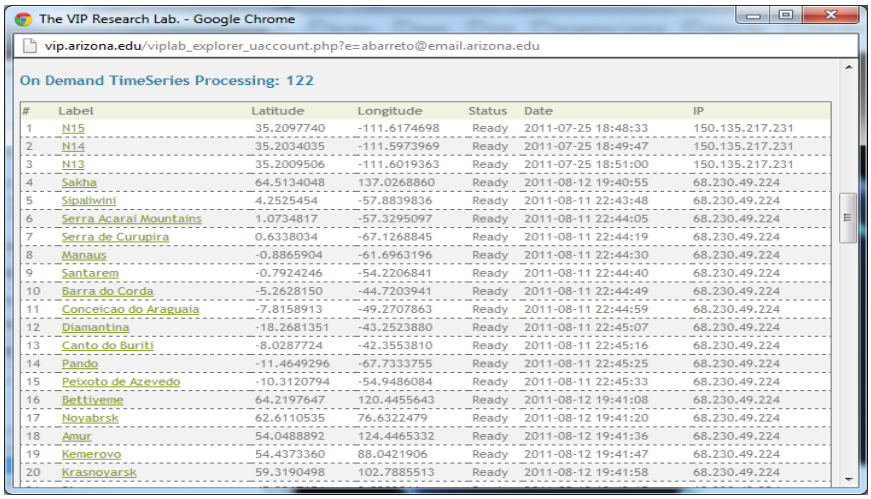


Figure 10. Data Explorer time series.

All daily, 7-day, 15-day, and monthly data are viewable, additionally the time series data can be downloaded as a comma separated values (csv) file.

When the requested time series does not exist the Data Explorer will submit the request to the server and the user will be notified by email when the time series has been processed. This usually takes anywhere between 1 hour to 8 hours depending on other outstanding requests and system resources. Registered users can follow the progress of their request by login to the system and checking their Data Explorer folder (Fig 11).



#	Label	Latitude	Longitude	Status	Date	IP
1	N15	35.2097740	-111.6174698	Ready	2011-07-25 18:48:33	150.135.217.231
2	N14	35.2034035	-111.5973969	Ready	2011-07-25 18:49:47	150.135.217.231
3	N13	35.2009506	-111.6019363	Ready	2011-07-25 18:51:00	150.135.217.231
4	Sakha	64.5134048	137.0268860	Ready	2011-08-12 19:40:55	68.230.49.224
5	Sipaliwini	4.2525454	-57.8839836	Ready	2011-08-11 22:43:48	68.230.49.224
6	Serra Acarai Mountains	1.0734817	-57.3295097	Ready	2011-08-11 22:44:05	68.230.49.224
7	Serra de Curupira	0.6338034	-67.1268845	Ready	2011-08-11 22:44:19	68.230.49.224
8	Manaus	-0.8865904	-61.6963196	Ready	2011-08-11 22:44:30	68.230.49.224
9	Santarem	-0.7924246	-54.2206841	Ready	2011-08-11 22:44:40	68.230.49.224
10	Barra do Corda	-5.2628150	-44.7203941	Ready	2011-08-11 22:44:49	68.230.49.224
11	Conceicao do Araguaia	-7.8158913	-49.2707863	Ready	2011-08-11 22:44:59	68.230.49.224
12	Diamantina	-18.2681351	-43.2523880	Ready	2011-08-11 22:45:07	68.230.49.224
13	Canto do Buriti	-8.0287724	-42.3553810	Ready	2011-08-11 22:45:16	68.230.49.224
14	Pando	-11.4649296	-67.7333755	Ready	2011-08-11 22:45:25	68.230.49.224
15	Peixoto de Azevedo	-10.3120794	-54.9486084	Ready	2011-08-11 22:45:33	68.230.49.224
16	Bettiyeme	64.2197647	120.4455643	Ready	2011-08-12 19:41:08	68.230.49.224
17	Novabrsk	62.6110535	76.6322479	Ready	2011-08-12 19:41:20	68.230.49.224
18	Amur	54.0488892	124.4465332	Ready	2011-08-12 19:41:36	68.230.49.224
19	Kemerovo	54.4373360	88.0421906	Ready	2011-08-12 19:41:47	68.230.49.224
20	Krasnoyarsk	59.3190498	102.7885513	Ready	2011-08-12 19:41:58	68.230.49.224

Figure 11. Data Explorer user activity manager.

f) Data search, order, download, and cart system

The core objective of Data Explorer is to facilitate data distribution and the enrichment of the user experience in interacting with the data prior to acquisition. The visualization tools described above serve to introduce the data and enable their exploration prior to order and acquisition, and although these tools are extremely useful and represent a new approach to interaction with large remote sensing data records, users are for the most part interested in acquiring data in support of their research and applications. The Data Explorer system provides a faster and more direct method for data search, order and download that bypasses the visualization and user interface. If the user is already familiar with the VIP Data products suite, or if he is only interested in acquiring specific data, the Data Explorer provides a simpler and a more direct access to the full data records database. User can use the database querying tool to search for the desired data record. This simple interface allows the selection across reprocessing versions, dates ranges, and dataset type. A query is then submitted and the Data Explorer returns the records that meet the user specified parameters with a direct link to the HDF file, access to quick look images of the data, and metadata display (Fig 12). Files can be downloaded right away via the provided live link created by the server. A complex order with multiple and different files can also be created by clicking the link next to the desired file names, which are added to a dynamic shopping cart. The user can then double check the contents of his cart (Fig 13) and adjust his order by adding or removing items before submitting the order.

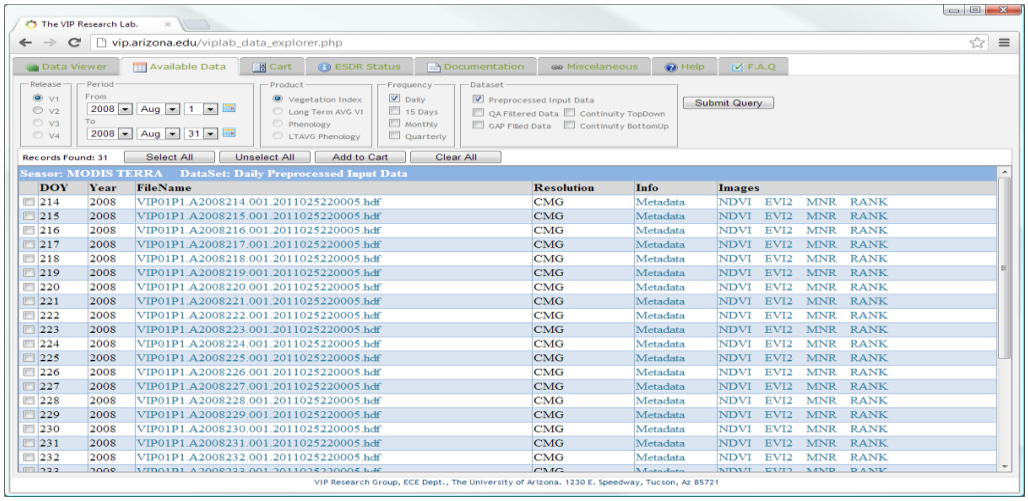


Figure 12. Data Explorer database query.

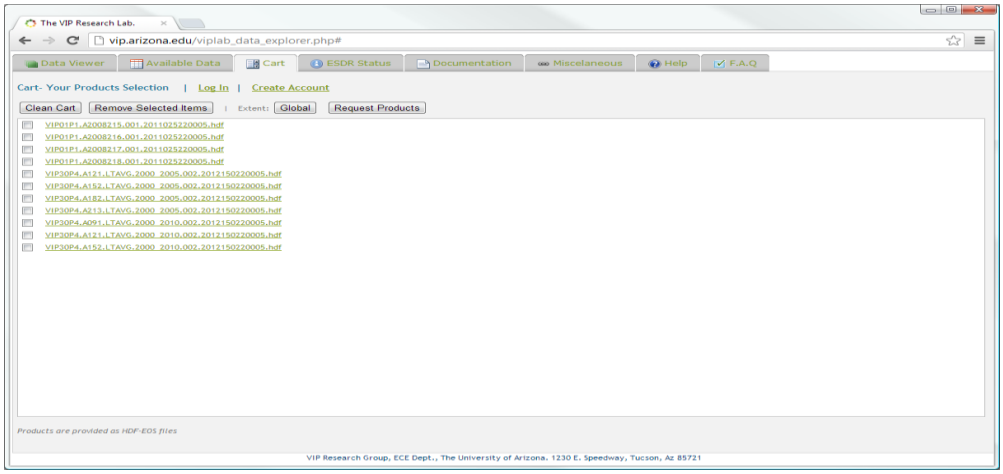


Figure 13. Data Explorer data search, order, and cart system.

Once submitted the Data Explorer server will stage the order and prepare the data for the user. Once ready, an order which usually takes few minutes, the system will send an email to the user with instructions on how to get the data. The email information will specify an expiration date and links to access and download the data (Figure 14).

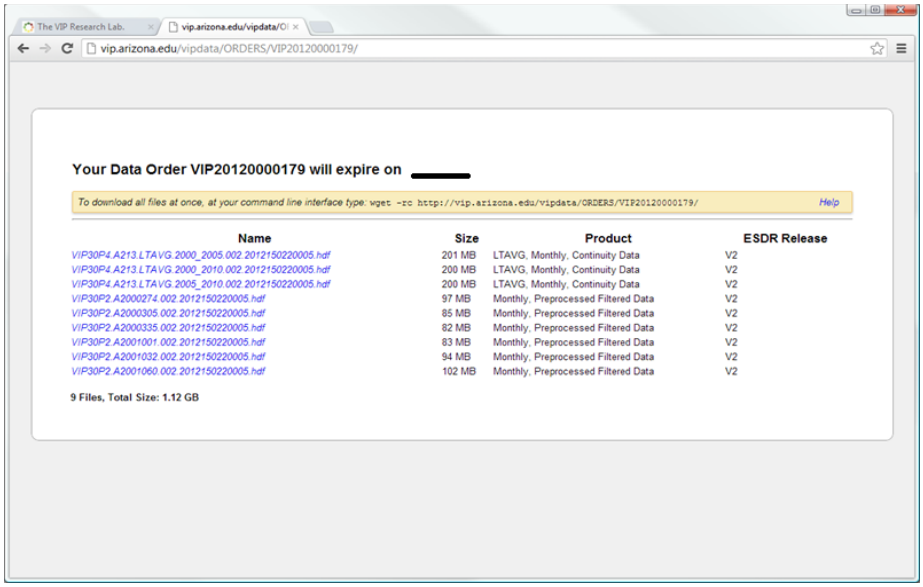


Figure 14. Data Explorer data order visualization and access.

Data Records Documents and Specification Files

Data Explorer is also home to the VIP Lab data production status and documentation center. It provides information about data production and system status. Additionally it contains the files dataset document descriptions and specifications and a detailed description of the various algorithms used to generate the data records. Users can learn the methodologies, algorithms and processes used for the creation of these data products.

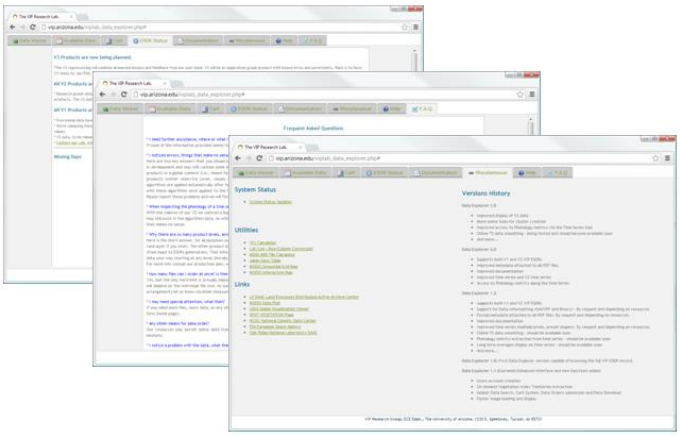


Figure 15. Data Explorer VIP ESDRs documentation center

Conclusions

Using a set of community maintained algorithms the Vegetation Index and Phenology Lab has produced three decades of global sensor independent and seamless long term data record of vegetation index and land surface phenology metrics. These data records, currently at V3, are provided at monthly, 15-day, 7-day, and daily time intervals and at 0.05 degrees spatial resolution. In support of these records we have designed and developed a highly interactive online tool to enable the users of these data records to explore, manipulate, search and order, and to perform basic data analysis operations on these records. The system has built in on demand processing capabilities and tools, especially the time series and pixel inspection tool.

This Data Explorer puts three decades of remote sensing data at the fingertips of the global long term land surface vegetation user community, while freeing the users from many preprocessing and data exploration tasks.

While the system is currently performing very well we plan to add more functionality and improve the users experience.

References

Abel, D. J., Taylor, K., Ackland, R., & Hungerford, S. (1998). An exploration of GIS architectures for Internet environments. Computers, Environment and Urban Systems, 22(1), 7-23.

Boelman, N. T., Stieglitz, M., Rueth, H. M., Sommerkorn, M., Griffin, K. L., Shaver, G. R., & Gamon, J. A. (2003). Response of NDVI, biomass, and ecosystem gas exchange to long-term warming and fertilization in wet sedge tundra. Oecologia, 135(3), 414-421.

Braswell, B. H., Schimel, D. S., Linder, E., & Moore, B. I. I. I. (1997). The response of global terrestrial ecosystems to interannual temperature variability. Science, 278(5339), 870-873.

Brown, M. E., Pinzón, J. E., Didan, K., Morisette, J. T., & Tucker, C. J. (2006). Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-Vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. Geoscience and Remote Sensing, IEEE Transactions on, 44(7), 1787-1793.

Didan, K., & Huete, A. (2006). MODIS vegetation index product series collection 5 change summary. MODIS VI C5 changes, 17.

Didan, K., & Huete, A. (2004, September). MODIS seasonal and inter-annual responses of semiarid ecosystems to drought in the Southwest USA. In Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International (Vol. 3, pp. 1538-1541). IEEE.

Didan, K., Barreto, A., Miura, T., Tsend-Ayush, J. (2012, December). A framework for capturing and characterizing the Error and Uncertainty in a Multi-Sensor Vegetation Index and Phenology Earth Science Data Record. Abstract IND13D-07 presented at 2012 Fall Meeting, AGU, San Francisco, Calif. 3-7 Dec.

Ganguly, S., Friedl, M. A., Tan, B., Zhang, X., & Verma, M. (2010). Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product. Remote Sensing of Environment, 114(8), 1805-1816.

Global Data Explorer. http://gdex.cr.usgs.gov/gdex/ (Accessed April 15, 2013)

GloVIS. The USGS global visualization viewer. http://glovis.usgs.gov/ (Accessed April 23, 2013)

Goodchild, M. F. (2007). in the World of Web 2.0. *International Journal*, *2*, 24-32.

Google Earth API Reference. https://developers.google.com/earth/documentation/reference/ (Accessed April 15, 2013)

HDF-EOS. http://hdfeos.org/ (Accessed April 15, 2013)

Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing, 7(11), 1417-1434.

Hu, S. (2008). Advancement of Web Standards and Techniques for Developing Hypermedia Maps on the Internet. In International Perspectives on Maps and the Internet (pp. 115-124). Springer Berlin Heidelberg.

Huete, A. R., Miura, T., Kim, Y., Didan, K., & Privette, J. (2006, August). Assessments of multisensor vegetation index dependencies with hyperspectral and tower flux data. In Proc. SPIE (Vol. 6298, p. 629814).

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote sensing of environment, 83(1), 195-213.

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment, 112(10), 3833-3845.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., ... & Morisette, J. T. (2002). An overview of MODIS Land data processing and product status. Remote Sensing of Environment, 83(1), 3-15.

LP DAAC. https://lpdaac.usgs.gov/ (Accessed April 15, 2013)

LTDR Data Product Description, version 3. 2010.

LTDR. Land Long Term Data Record. http://ltdr.nascom.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi (Accessed April 15, 2013)

Miura, T., Huete, A., Yoshioka, H., & Kim, H. J. (2002). An application of airborne hyperspectral and EO-1 Hyperion data for inter-sensor calibration of vegetation indices for regional-scale monitoring. In Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International (Vol. 6, pp. 3118-3120). IEEE.

Murphy, R. E., Barnes, W. L., Lyapustin, A. I., Privette, J., Welsch, C., DeLuccia, F., ... & Kealy, P. S. (2001). Using VIIRS to provide data continuity with MODIS. In Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International (Vol. 3, pp. 1212-1214). IEEE.

Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., ... & Pinheiro, A. (2007, July). Generating a long-term land data record from the AVHRR and MODIS instruments. In Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International (pp. 1021-1025). IEEE.

Reed, B. C., Schwartz, M. D., & Xiao, X. (2009). Remote Sensing Phenology: Status and the Way Forward. Phenology of ecosystem processes, 231.

Reed, B. C., White, M., & Brown, J. F. (2003). Remote sensing phenology. In Phenology: An Integrative Environmental Science (pp. 365-381). Springer Netherlands.

Reverb. http://reverb.echo.nasa.gov/reverb/ (Accessed April 15, 2013).

Riveracamacho, J. Didan, K. Barreto-Munoz , A. Yitayew, M. (2010, December) Abstract B51M-0597 presented at 2011 Fall Meeting, AGU, San Francisco, Calif., 5-9 Dec.

Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., ... & Zhang, C. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. Remote Sensing of Environment, 114(1), 35-49.

Standart, G. D., Stulken, K. R., Zhang, X., & Zong, Z. L. (2011). Geospatial visualization of global satellite images with Vis-EROS. Environmental Modelling & Software, 26(7), 980-982.

The HDF Group. http://www.hdfgroup.org/ (Accessed April 15, 2013)

Tsend-Ayush ,J., Miura, T., Didan, K., Barreto-munoz, A. (2011, December). Compatibility Analysis of Global Vegetation Index from Terra MODIS and SPOT-4 VEGETATION for Continuity Studies. American Geophysical Union. Fall Meeting 2011. #IN21A-1412

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., ... & El Saleous, N. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing, 26(20), 4485-4498.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote sensing of Environment, 8(2), 127-150

USGS EarthExplorer. http://earthexplorer.usgs.gov/ (Accessed April 15, 2013)

USGS Global Visualization Viewer. http://glovis.usgs.gov/ (Accessed April 15, 2013)

Vermote, E., Devadiga, S. (2010). LTDR Data Format Descriptions (version 3 release) http://ltdr.nascom.nasa.gov/ltdr/docs/LTDRDataFormatDescriptions_03_2010_30.pdf

Van Leeuwen, W. J., Huete, A. R., & Laing, T. W. (1999). MODIS vegetation index compositing approach: a prototype with AVHRR data. Remote Sensing of Environment, 69(3), 264-280.

Vermote, E., & Kaufman, Y. J. (1995). Absolute calibration of AVHRR visible and near-infrared channels using ocean and cloud views. International Journal of Remote Sensing, 16(13), 2317-2340.

WELD. Web-enabled Landsat data project. http://landsat.usgs.gov/WELD.php (Accessed April 22, 2013)

White, M. A., Brunsell, N., & Schwartz, M. D. (2003). Vegetation phenology in global change studies. In Phenology: An integrative environmental science (pp. 453-466). Springer Netherlands.

White, M. A., Beurs, D., Kirsten, M., Didan, K., Inouye, D. W., Richardson, A. D., ... & Lauenroth, W. K. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. Global Change Biology, 15(10), 2335-2359.

White, M. A., Thornton, P. E., & Running, S. W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. Global biogeochemical cycles, 11(2), 217-234

Yaheng, C., Junmei, Z., & Mingxin, M. (2012, August). WebGIS-based farmland ecosystem spatial decision support system. In Agro-Geoinformatics (Agro-Geoinformatics), 2012 First International Conference on (pp. 1-9). IEEE.

Yoshioka, H., Miura, T., & Obata, K. (2012). Derivation of relationships between spectral vegetation indices from multiple sensors based on vegetation isolines. Remote Sensing, 4(3), 583-597.

Yoshioka, H., Miura, T., & Yamamoto, H. (2005, January). Relationships of spectral vegetation indices for continuity and compatibility of satellite data products. In Fourth International Asia-Pacific Environmental Remote Sensing Symposium 2004: Remote Sensing of the Atmosphere, Ocean, Environment, and Space (pp. 233-240). International Society for Optics and Photonics.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., ... & Huete, A. (2003). Monitoring vegetation phenology using MODIS. Remote sensing of environment, 84(3), 471-475.

Zhang, X., Friedl, M. A., & Schaaf, C. B. (2006). Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. Journal of Geophysical Research: Biogeosciences (2005–2012), 111(G4).

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., & Liu, Z. (2005). Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments. Journal of Geophysical Research: Atmospheres (1984–2012), 110(D12).

Zhao, P., Foerster, T., & Yue, P. (2012). The geoprocessing web. Computers & Geosciences, 47, 3-12.