ABSTRACT

Title:	IMPROVING SATELLITE LEAF AREA INDEX ESTIMATION BASED ON VARIOUS INTEGRATION METHODS

Dongdong Wang, Doctor of Philosophy, 2009

Directed by:	Dr. Shunlin Liang, Professor Department of Geography

Leaf Area Index (LAI) is an important land surface biophysical variable that is used to characterize vegetation amount and activity. Current satellite LAI products, however, do not satisfy the requirements of the modeling community due to their large uncertainties and frequent missing values. Each LAI product is currently generated from only one satellite sensor data. There is an urgent need for advanced methods to integrate multiple LAI products to improve the product's accuracy and integrality for various applications. To meet this need, this study proposes four methods, including the Optimal Interpolation (OI), Bayesian Maximum Entropy (BME), Multi-Resolution Tree (MRT) and Empirical Orthogonal Function (EOF), to integrate multiple LAI products. Three LAI products have been considered in this study: Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle Imaging SpectroRadiometer (MISR) and Carbon cYcle and Change in Land Observational Products from an Ensemble of Satellites (CYCLOPES) LAI.

As the basis of data integration, this dissertation first validates and intercompares

MODIS and CYCLOPES LAI products and also evaluates their geometric accuracies. The CYCLOPES LAI product has smoother temporal profiles and fewer spatial variations, but tends to produce spurious large errors in winter. The Locally Adjusted Cubic-spline Capping algorithm is revised to smooth multiple years' average and variance.

Although OI, BME and MRT based methods have been used in other fields, this is the first research to employ them in integrating multiple LAI products. This dissertation also presents a new integration method based on EOF to solve the problem of large data volume and inconsistent temporal resolution of different datasets. High resolution LAI reference maps generated with ground measurements are used to validate these algorithms. Validation results show that all of these four methods can fill data gaps and reduce the errors of the existing LAI products. The data gaps are filled with information from adjacent pixels and background. These algorithms remove the spurious large temporal and spatial variation of the original LAI products. The combination of multiple satellite products significantly reduces bias. OI and BME can reduce the RMSE from 1.0 (MODIS) to 0.7 and reduce the bias from +0.3 (MODIS) and -0.2 (CYCLOPES) to -0.1. MRT can produce similar results with OI but with significantly improved efficiency. EOF also generates the results with the RMSE of 0.7 but zero bias.

Limited ground measurement data hardly prove which methods outperform the others. OI and BME theoretically produce statistically optimal results. BME relaxes OI's linear and Gaussian assumption and explicitly considers data error, but bears a much higher computational burden. MRT has improved efficiency but needs strict assumptions on the scale transfer function. EOF requires simpler model identification, while it is more "empirical" than "statistical".

The original contributions of this study mainly include: 1) a new application of several different integration methods to incorporate multiple satellite LAI products to reduce uncertainties and improve integrality, 2) an enhancement of the Locally Adjusted Cubic-spline Capping by revising the end condition, 3) a novel comprehensive comparison of MODIS C5 LAI product with other satellite products, 4) the development of a new LAI normalization scheme by assuming the linear relationship between measurement error and LAI natural variance to account for the inconsistency between products, and finally, 5) the creation of a new data integration method based on EOF.

IMPROVING SATELLITE LEAF AREA INDEX ESTIMATION BASED ON
VARIOUS INTEGRATION METHODS

By

Dongdong Wang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Shunlin Liang, Chair
Professor Hanan Samet
Dr. Guoqing Sun
Professor John Townshend
Professor Naijun Zhou

DEDICATION

To my parents.

ACKNOWLEDGEMENTS

I would like to thank my primary advisor Professor Shunlin Liang. This dissertation could not have been completed without his continuous support and guidance. It has been a great experience to work with Professor Liang. Besides his profound insights, it is also his encouragement that helped me overcome difficulties I met in research.

I am also deeply grateful to other members of my dissertation committee: Professor Hanan Samet from the Department of Computer Science, Dr. Guoqing Sun, Professor John Townshend and Professor Naijun Zhou. They are always there and helpful when I need to discuss my research. Their valuable suggestions help improving the quality of my dissertation work.

Our research group is a big family and a vigorous team. I benefit a lot from the discussion with my colleagues and friends. I would like to say thanks to all of them. I also thank Ms. Elizabeth Hoy and Mr. Tom Otwell for proofreading the draft of my dissertation.

This dissertation uses some libraries kindly provided by other researchers. I would like to extend my gratitude to Professor George Christakos from San Diego State University for the BME library, Professor Jing M. Chen from the University of Toronto for the LACC code and Professor Paul Fieguth from the University of Waterloo for the multiscale software. I also thank Oak Ridge National Laboratory Distributed Active Archive Center and the investigators of the validation sites for providing the reference LAI data.

Finally, I would like to thank my parents and my fiancee. Their selfless and endless love makes all this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

Leaf Area Index (LAI), defined as the one-sided leaf area per unit of ground area (Chen and Black 1992), is a very important land surface biophysical variable used to describe vegetation amount and activity. Since vegetation plays a significant role in controlling terrestrial productivity, hydrological cycling, and surface energy balance, LAI is required as a key input for various ecosystem productivity models and land process models, especially under the changing climate. Moreover, vegetation is also affected by the anthropogenic activities and climate change. The change of LAI is one of the most direct indicators of how terrestrial ecosystems respond to climate change. This topic has received increasing attentions recently (Myneni *et al.* 1998; Yang *et al.* 1998; Badeck *et al.* 2004; Zhang *et al.* 2004a; Zhang *et al.* 2004b; Pettorelli *et al.* 2005; Notaro *et al.* 2006; Piao *et al.* 2006a; Piao *et al.* 2006b; Myneni *et al.* 2007). A reliable, consistent and accurate dataset of LAI would be the basis of such investigations.

Remote sensing is the ideal way to map biophysical variables globally on a regular basis. The historical Advanced Very High Resolution Radiometer (AVHRR) Vegetation Index (VI) products date from the early 1980s. However, due to problems

such as atmospheric correction, onboard calibration and orbit drift, the quality of

AVHRR data are in question (Vermote and Kaufman 1995; Gutman 1999; Kaufmann

et al. 2000). The new generation of moderate resolution optical sensors, such as

Moderate Resolution Imaging Spectroradiometer (MODIS), and VEGETATION,

provides finer spatial resolutions, more spectral channels, and better bandwidths and

radiometric calibration (Townshend and Justice 2002). Multi-angular sensors (e.g.

Polarization and Directionality of Earth Reflectances (POLDER), Multi-angle

Imaging SpectroRadiometer (MISR)) observe the Earth from several viewing angles

and are thought to improve the accuracy of surface characterization. With the launch

of the new generation of these instruments, we have entered a new era to generate

LAI and other biophysical products. Various sophisticated algorithms have been

developed to produce LAI products from several sensors or projects, for example,

MODIS, Carbon cYcle and Change in Land Observational Products from an

Ensemble of Satellites (CYCLOPES) and MISR LAI. However, many issues

associated with the current LAI products limit their applications in understanding the

dynamic environment. Specifically, the problems include the following:

First, there are many gaps or missing data due to instrumental malfunction, cloud

contamination and other factors (Fang *et al.* 2008). For instance, the gaps in MODIS

LAI data can be as high as 80% in the winter at high latitudes.

Second, LAI estimation from satellite imagery has large uncertainties. For the given

time and location, different products often have significantly different LAI values.

Figure 1-1 shows the discrepancy between CYCLOPES LAI, MODIS/Terra and

MODIS/Aqua LAI products at temperate forest. Although MODIS/Terra and

MODIS/Aqua LAI products are derived from the same type of sensors and using the

same inversion algorithm, significant differences exist between them. Phenology

determination from LAI time series data is difficult due to large uncertainties. Brut *et*

al. (2009) showed that there are also significant discrepancies between satellite data

and model simulated LAI both in absolute values and in derived phenological

parameters.
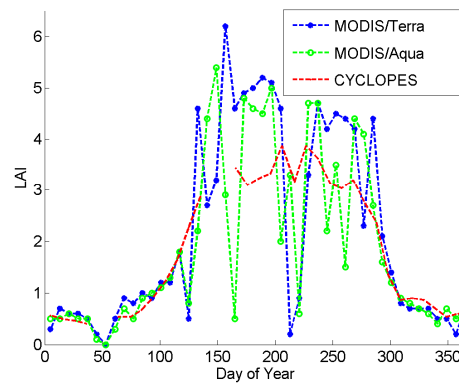
Figure 1-1 Time series of MODIS/Terra, MODIS/Aqua and CYCLOPES LAI
products in 2003 at a temperate forest pixel (39.04 N, 79.86W)

Third, multiple LAI products are typically incompatible in spatial and temporal

resolution, map projection and spatial ground coverage. For example, MODIS has a

swath of more than 2000 km while MISR's swath is as narrow as 400 km. The

difference between geometric and orbital properties among various sensors increases

the difficulty in applying and incorporating multiple-source LAI data.

The accuracy of current LAI products does not satisfy the systematic requirements for climate study and other applications. GCOS requires an accuracy of 0.5 and a spatial resolution of 1 km and a temporal resolution of 1 day for climate study (WMO 2006). Even more accurate estimation of LAI is required for other purposes. According to Francois *et al.* (1997), estimating leaf temperature at the accuracy of 0.5 requires less than 10% error of LAI and accurate retrieval of soil temperature needs a higher accuracy of LAI within 5% error. However, MODIS LAI is a 8-day composite product and has a 12% overestimation and RMSE of 0.66 (Yang *et al.* 2006), while CYCYLOPES LAI has a 10-day temporal resolution and RMSE of 0.84 (Weiss *et al.* 2007). These uncertainties may produce substantial errors when driving various numerical models.

These problems must be resolved. Data assimilation that combines the remotely sensed observations with dynamic models may be an effective way to address these issues. However, considerable errors exist in models prognosticating the dynamic leaf, which usually link the vegetation growth to climate variability and nutrient availability. The errors come both from the imperfect parameterization of vegetation growth and the uncertainties with the forcing data. Land data assimilation is an

emerging field and many issues remain (Liang 2004). Alternative solutions must be explored. In this study, I focus on developing algorithms based on integration methods to incorporate multiple LAI products and other ancillary data to improve the quality of existing products.

1.1 Literature review

The discontinuity or inconsistency of scientific data records is a universal phenomenon, in both *in situ* measurements (Falge *et al.* 2001; Ooba *et al.* 2006) and satellite observations (Moody *et al.* 2005; Fang *et al.* 2008), such as land surface datasets (Moody *et al.* 2005; Fang *et al.* 2008), atmospheric products (Zhang *et al.* 2007), and oceanic data archives (Pottier *et al.* 2008). Spatially, this discontinuity or inconsistency prevents forming an integrated map and conducting spatial analysis; temporally, it limits the ability to make time series analysis and obtain trend and change information. There are numerous investigations on developing algorithms to build spatially and temporally continuous scientific datasets and to improve the quality of these products (Sellers *et al.* 1994; Gregg and Conkright 2001; Buermann *et al.* 2002; Beckers and Rixen 2003; Kwiatkowska and Fargion 2003; Chen *et al.* 2004; Chen *et al.* 2006; Gu *et al.* 2006; Fang *et al.* 2008). However in terms of improving satellite vegetation products, temporal curve fitting is still the most common method. Besides, there are also several attempts to use spatiotemporal

statistical methods to incorporate more information.

Temporal curve fitting methods have been extensively used to smooth LAI and other biophysical variables. Due to the phenological cycles of terrestrial ecosystems, most biophysical variables derived from satellite imagery reveal some types of annual patterns. Different methods have been proposed, which vary in the selection of curve shape and fitting algorithms (See Table 1-1). Because atmospheric contamination usually leads to underestimating these biophysical variables, most algorithms adapt to the upper envelopes of annual curves. In order to achieve the upper envelopes, a recursive algorithm is usually used. The observations are replaced by filtered values to form new time series if they are smaller than filtered values. The filter is then carried out on the newly constructed time series until convergence. In addition to the use of least squares regression, the Kalman filter is used to estimate the parameters of the structural time series models (Young *et al.* 1999). The parameters of this type of time series model are also time-dependent variables, but change slower compared with the model outputs. Parameters are treated as state variables and model outputs are taken as measurements in the Kalman filter. The Kalman filter is used to estimate the state variables and so correct measurements. Alavi *et al.* (2006) applied this method to fill the gaps of evapotranspiration data. Jonsson and Eklundh (2004) developed a program called TIMESAT to fill gaps in satellite data and extract phenology information from time series analysis, which is used by Gao *et al.*(2008) to

produce continuous MODIS LAI data.

In temporal methods, additional information other than observations of the variable of interest itself may be included as well. In filling gaps of Normalized Difference Vegetation Index (NDVI) data, the data quality of NDVI (cloud mask etc) was used to group NDVI observations and assign them different weights in regression (Chen *et al.* 2004). Alavi *et al.*(2006) incorporated the relationship between evapotranspiration and other variables such as latent heat flux, available energy, and vapor pressure deficit to estimate evapotranspiration when no evapotranspiration observations are available.

Table 1-1 Major curve fitting methods used for filtering satellite vegetation products

Methods	References
Trigonometric function	Sellers *et al.*(1994)
Gaussian-type function	Jonsson and Eklundh (2002)
Logistic curve	Zhang et al. (2003)
Savitzky-Glay filter	Chen *et al.* (2004)
Cubic spline	Chen *et al.* (2006)
Wavelet	Lu *et al.*(2007)

Temporal methods are able to analyze the time series at the mean time filling gaps, and would be good choices for investigating the trend or phenology of vegetation activities (Zhang *et al.* 2003; Sakamoto *et al.* 2005; Piao *et al.* 2006b; Zhang *et al.* 2006). However, temporal methods usually do not utilize the spatial information.

When continuous gaps exist, they may not produce reliable results. Borak and Jasinski (2009) compared several methods of interpolating LAI and found the incorporation of spatial information would improve the results. Borak and Jasinski (2009) expanded Kang *et al.* (2005)'s method of utilizing spatial information, where the average of adjacent pixels with the same land cover was used to fill the missing data. Moody *et al.* (2005) used a similar method to fill gaps in MODIS albedo product by adjusting the ecosystem dependent phenology curve according to pixel to pixel variation. These methods utilize temporal and spatial information in a simple way, and the results are not statistically optimized.

The spatiotemporal statistical methods have also been employed to improve the consistency and accuracy of existing LAI and other vegetation products to some extent (Gu *et al.* 2006; Fang *et al.* 2008; Gu *et al.* 2009). Gu *et al.* (2006) averaged the observational LAI value with the multiple year mean according to their relative variance. Their method cannot be used for interpolation because it doesn't incorporate spatial or temporal covariance. Gu *et al.* (2009) used a similar but even more simplified method to reconstruct MODIS NDVI. In their study, the weights were given empirically according to the quality control information of NDVI products. Fang *et al.* (2008) designed a temporal spatial filter to replace missing or low quality values of MODIS LAI products. In Fang *et al.*'s method, data from temporally adjacent images were used to interpolate when data are missing. The weights were

given empirically according to the time distance between the points. Fang *et al.*'s method is not statistically optimal in a strict way. Fang et al.'s approach is based on Cressman (1959)'s method. Chen *et al.* (2008) has shown Cressman's method performs consistently worse than the geostatistics method Optimal Interpolation (OI) for the case of interpolating gauge rainfall data. Furthermore, all these methods have some generic problems: 1) limited information is incorporated; 2) only one product is used; and 3) errors of observations are not rigorously considered or fully utilized.

Besides geostatistical methods, the Empirical orthogonal function (EOF) based method is another choice to fill gaps and reduce noises. In EOF, the leading eigenvector components of the covariance matrix are used to reconstruct the original matrix. Because the leading components are supposed to contain most of the information, EOF is expected to reduce errors (Hannachi *et al.* 2007). We employed an EOF based method to improve MODIS LAI product. However, only one source was used in our previous investigation (Wang and Liang 2008) and this method is not suitable for large datasets due to storage and computational limitations.

1.2 Objective of this dissertation

The objective of this dissertation is to improve the accuracy and integrality of the LAI products by developing advanced methods to integrate all the available information

from the incompatible sources of satellite LAI products and any prior knowledge. Several issues must be solved in developing the integration methods:

- Multiple products have irregular spatiotemporal grids because they are using different map projections and temporal composite periods.

- Different products have variable support size and they are usually incompatible in spatial and/or temporal resolutions.

- Different products have different biases and the systematic inconsistency must be taken into account in data integration.

- Little ground truth data are available to model measurement error, correct systematic bias and validate working algorithm

- Moderate resolution LAI products have huge data volume. The designed algorithm needs to be computationally efficient.

In order to achieve this objective, four methods including OI, Bayesian Maximum Entropy (BME), Multi-Resolution Tree (MRT) and EOF, are presented to integrate MODIS, CYCLOPES and MISR LAI products in this dissertation. OI and EOF are selected, because they have been extensively used to solve missing data problem in meteorology.

OI is able to take irregular inputs and employs spatiotemporal covariance to interpolate variables at non-measured points and reduce errors at measured points. OI

will give the best estimation for the Gaussian process. Gu *et al.* (2006) tried OI to improve MODIS LAI. This dissertation improves their method by using data at adjacent points and combing MODIS with CYCLOPES.

Compared with the geostatistics method OI, EOF requires little input on measurement error and covariance structure. EOF methods calculate the covariance matrix of the datasets and use their leading components to reconstruct noisy and gap-prone data. Aimed to process large datasets, I develop a hierarchical EOF approach to both reduce high dimension matrix to multiple smaller ones and incorporate data information on higher levels. I also develop a scheme to incorporate multiple products with different temporal resolutions.

Besides the OI approach, LAI integration based on modern geostatistics BME is also proposed. Different from the linear estimator OI, BME doesn't have the Gaussian assumption but explicitly incorporates the measurement error in a strict way. BME has been used to improve satellite ozone products and other limited applications, but has never been applied to LAI products. LAI integration based on BME is proposed in this dissertation to combine MODIS and CYCLOPES data.

Both OI and BME need to inverse a matrix, which is computationally expensive. In addition, BME involves high dimensional integral. The computational cost limits

their applications for large datasets. MRT efficiently employs the Kalman filter on a tree structured data. MRT has been applied to analyze many geophysical data. This dissertation presents a method based on MRT to integrate MODIS and MISR L3 LAI products.

To my best knowledge, curve fitting or a one-dimensional (1D) temporal filter is still the most common method of improving satellite vegetation products. There are few attempts to employ spatiotemporal statistics, which are based on "traditional" geostatistics, with only limited information. No attempts to combine more than one source of LAI datasets have been made. This dissertation addresses these problems of current LAI products described in the previous section. It is the first attempt to apply sophisticated integration methods to combine multiple LAI products with variable resolutions and uncertainties. The newly developed EOF integration method is able to handle large datasets and multiple incompatible data. The improved mapping of LAI will improve modeling of vegetation dynamics and evaluating terrestrial productivity with higher accuracy and more integrality. The proposed methods can also be applied in other satellite land products with similar problems.

1.3 Structure of this dissertation

Chapter 2 summaries the datasets used in this dissertation, including three satellite

products and *in situ* measurements collected at Bigfoot and other programs. Besides the direct validation, the relative geometric accuracy of MODIS and CYCLOPES LAI products is also evaluated. Multiple years' mean and variance are calculated as the background of data integration. A periodic locally adjusted cubic-spline capping algorithm is improved to filter the multiple years' mean and variance data.

Chapter 3 presents two geostatistics based methods: OI and BME, and analyzes the results of integrating MODIS and CYCLOPES data. A data normalization scheme is developed to take into account the systematic inconsistency between the two products. Improvement is shown on points, 2D image and time series.

Chapter 4 applies the method based on MRT to integrate MODIS and MISR L3 LAI, which have different spatial resolutions. OI is also carried out at the same experiment area to compare with MRT. MRT shows the similar results but with significantly improved computational speed.

Chapter 5 develops a new integration method based on EOF to solve the problem of large data volume and inconsistent temporal resolution of different datasets, and applies this method to integrate MODIS and CYCLOPES LAI data.

Chapter 6 gives a general conclusion. The merits and shortcomings of each method

are discussed and the main original contributions of this dissertation study are summarized and remaining problems are identified for future study.

Chapter 2 Data and preliminary analysis

This chapter first describes the main datasets used in this study and then calculates LAI climatology (multiple years' mean and variance), which are important background knowledge for data integration. A revised version of locally adjusted cubic spline capping algorithm is developed to smooth satellite LAI climatology. Preliminary analysis results on geometric errors, measurement uncertainties and gaps distribution are also presented in this chapter.

2.1 Datasets

Both satellite and *in situ* LAI data are used in this study. Satellite data are the object of this study and the input of proposed data integration methods. *In situ* LAI measurements are used to evaluate satellite products and validate integration algorithms. Two moderate resolution satellite LAI data, MODIS and CYCLOPES LAI products and one coarse LAI data, MISR L3 LAI product will be used to illustrate the algorithms' ability of integrating multiple data with varied accuracy and spatiotemporal resolutions. High resolution LAI maps calibrated with *in situ* LAI measurements are employed to take into account the change of support problem in

validating satellite LAI products.

2.1.1 Satellite LAI products

Global LAI products have been routinely produced from different sensors, such as
MODIS (Myneni *et al.* 2002), CYCLOPES (Baret *et al.* 2007), GLOBCARBON
(Deng *et al.* 2006), MISR (Hu *et al.* 2007), POLDER (Bicheron *et al.* 1998) (See
Table 2-1). Among them, MODIS, CYCLOPES and MISR LAI are produced over
multiple years and easy to access. MODIS twin sensors aboard Terra and Aqua
satellites were launched in 1999 and 2002 respectively. MODIS has 36 spectral bands,
among which red and NIR bands are the input of MODIS LAI algorithm. This
algorithm is biome dependent, using the six biome classification system. The
vegetation is classified as Grasses/Cereal crops, Broadleaf crops, Shrubs, Savannah,
Broadleaf forest and Needleleaf forest (Friedl *et al.* 2002). A three dimensional
radiative transfer (RT) model was run offline to obtain look-up tables (LUT) to
improve computation efficiency (Tian et al. 2000; Myneni et al. 2002). An empirical
algorithm based on the relationship between LAI and NDVI is used as backup when
the main RT algorithm fails. To get sufficient clear data, MODIS LAI products are
composed at 8-day time step. For each composite period, the LAI value where the
Fraction of Absorbed Photosynthetically Active Radiation (fPAR) is largest is chosen
as the LAI for that period. MODIS LAI products are produced at 1 km spatial

resolution in the MODIS sinusoidal projection and the latest version of MODIS LAI

is collection 5, available from February 18, 2000 to present. Compared with the

previous versions, Collection 5 data improve the retrieval percentage at woody area

and correct the overestimation problem over the needle forests (Shabanov *et al.*

2007) . LAI products are produced using both Terra MODIS (MOD15A2) and Aqua

MODIS (MYD15A2). Started from collection 5, a new Terra and Aqua combined LAI

product is available (MCD15A2). Similar to other MODIS products, MODIS LAI

provides detailed Quality Control (QC) information with the LAI values. The QC data

of LAI Collection 5 contains two bytes, "FparLai_QC" and "FparExtra_QC".

FparLai_QC mainly gives the retrieval path, while FparExtra_QC contains the

information of acquisition conditions.

CYCLOPES LAI uses three bands (red, NIR and shortwave infrared) of

VEGETATION, a sensor onboard the European SPOT 4 and 5 satellites, as the input

data. Instead of the radiative transfer model inversion method, CYCLOPES algorithm

uses the nonparametric regression based on neural network (Bacour et al. 2003; Baret

et al. 2007). The training data are obtained through simulation of the PROSPECT

(Jacquemoud and Baret 1990) and Scattering by Arbitrarily Inclined Leave (SAIL)

(Verhoef 1984) models. All reflectance values are normalized to nadir geometry using

the Roujean BRDF model (Roujean et al. 1992). CYCLOPES LAI is produced at 10

day temporal sampling intervals, however, cloud free reflectance within 30 days may

be used to fit the BRDF model. CYCLOPES LAI data is generated at a spatial

resolution of 1/112 degrees in the plate carrée map projection. The latest version is

3.1, covering from year 1999 to 2003.

Table 2-1 Characteristics of major current satellite LAI products

Product name	Spatial resolution	Temporal resolution	Composite method	Map projection	Availability
MODIS	1km	8 days, starting from the data day	Maximal value of fPAR	Sinusoidal	February 2000-present (Terra) July 2002 –present (Aqua)
CYCLOPES	1/112 degree	10 days, centered at the data day	Temporal smoothing of reflectance	Plate carrée WGS-84	1999-2003
GLOBCAR-BON	1km, 10km, 0.25 degree, 0.5 degree	10 days or monthly	Temporal smoothing using cubic spline	Plate carrée WGS-84	1998-2007
MISR L2	1 km	Every 2 days at polar region and every 9 days at equator	Not applicable	SOM projection (Space Oblique Mercator)	February 2000-present
MISR L3	0.5 degree	Monthly	Simple average	Geographic	February 2000-present
POLDER	1/18 degree	10 days	Gaussian temporal averaging	Sinusoidal	November 1996 – June 1997 April- October 2003

MISR is a multi-angular radiometer onboard Terra, observing the entire Earth from

nine angles at four bands every 16 days. The MISR LAI algorithm chooses the most

probable LAI from all possible values by comparing modeled reflectance and MISR

directional reflectance (Knyazikhin *et al.* 1998). MISR LAI products are distributed at

two different resolutions, L2 1.1km moderate resolution and L3 0.5 degree coarse

resolution. The L3 product will be used in this study to facilitate the development of the algorithm for integrating LAI data of different resolutions. MISR LAI will be only used in Chapter 4, and the following analysis is mainly on the two moderate resolution LAI products, MODIS and CYCLOPES.

Both MODIS and CYCLOPES LAI have been validated (Yang *et al.* 2006; Weiss *et al.* 2007). A summary of validating MODIS LAI can be found in Yang *et al.* (2006). They are also intercompared (Weiss *et al.* 2007; Garrigues *et al.* 2008; Verger *et al.* 2008). The general conclusion is that both products can represent reasonable temporal and spatial variations when compared with *in situ* measurement or an LAI reference map. Compared with MODIS collection 4 LAI, CYCYLOPES LAI appears more accurate, with a smaller RMSE. CYCYLOPES LAI shows a smoother temporal and spatial profile. However, the highest value of CYCYLOPES LAI is 6, which cannot characterize the highly vegetated area like dense forests.

All five years (1999-2003) of CYCLOPES LAI, 2000-2008 MODIS LAI over North America and 2000-2008 MISR L3 monthly data are downloaded from their respective data centers. Forty MODIS tiles and 73 CYCLOPES tiles between 0 N and 80 N covering the North America continent are used. (Figure 2-1).
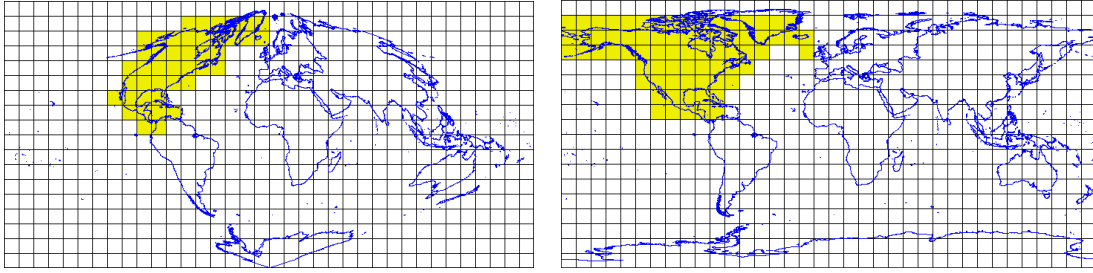
Figure 2-1 Tile systems used for a) MODIS LAI product b) CYCLOPES LAI product. The yellow tiles covering North America are used in this dissertation.

2.1.2 In situ LAI measurements

Besides the satellite LAI products, field measurement of LAI is also needed to evaluate the satellite products and validate the algorithms. Although more and more global satellite LAI products are becoming available, LAI is not frequently and regularly measured at the field experimental sites. Nevertheless, in order to validate or cross compare specific products or provide ancillary data for other research, several LAI validation campaigns or programs have been established (Chen *et al.* 2002; Morisette *et al.* 2002; Cohen *et al.* 2003; Morisette *et al.* 2006). The spatial resolution of moderate resolution satellite products is around one kilometer, while the *in situ* measurements represent only a much smaller area. Due to the heterogeneous of vegetation cover, the problem of change of support must be considered in the process of using field measurements. Well designed field sampling strategies can be used to down scale field measurement so that the scale of field measurement can match the pixel size of high resolution of satellite data (Burrows *et al.* 2002; Berterretche *et al.*

2005). Then in order to account for the difference of scale between field measurement and moderate or coarse resolution products, validation using LAI reference maps derived from high resolution satellite imagery and calibrated with *in situ* measurement is suggested (Chen *et al.* 2002; Cohen *et al.* 2003; Cohen *et al.* 2006). Twenty-eight high resolution LAI reference maps at 12 sites are collected to validate original LAI data and the integration algorithm (Figure 2-2). Four of these maps have a large spatial extent and more than one point on each map is chosen.
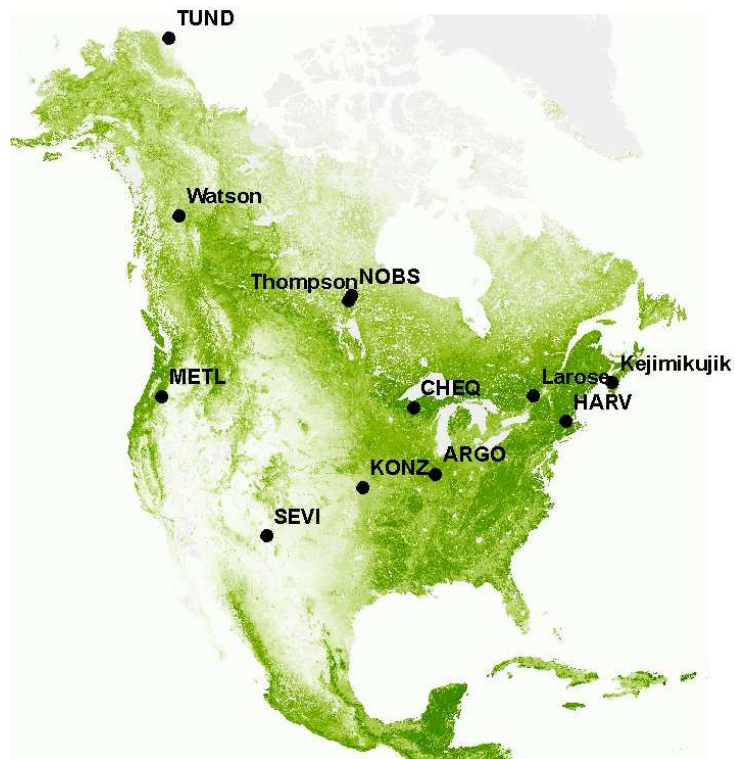
Figure 2-2 Location of 12 LAI field measurement campaign sites over North America

Table 2-2 Brief description of the collected LAI validation data

Site Name	Images available	Vegetation	Network	Reference
ARGO	2	cropland	Bigfoot	(Cohen *et al.* 2006)
CHEQ	1	mixed forest	Bigfoot	(Cohen *et al.* 2006)
HARV	4	mixed forest	Bigfoot	(Cohen *et al.* 2006)
KONZ	4	tall grass	Bigfoot	(Cohen *et al.* 2006)
METL	1	ponderosa pine	Bigfoot	(Cohen *et al.* 2006)
NOBS	3	black spruce	Bigfoot	(Cohen *et al.* 2006)
SEVI	8	short grass	Bigfoot	(Cohen *et al.* 2006)
TUND	1	tundra	Bigfoot	(Cohen *et al.* 2006)
Larose	1	mixed forest	VALERI	(Fernandes *et al.* 2003)
Watson Lake	1	mixed forest	Canada	(Fernandes *et al.* 2003)
Kejimikujik	1	mixed forest	Canada	(Fernandes *et al.* 2003)
Thompson	1	mixed forest	Canada	(Fernandes *et al.* 2003)

2.2 Climatology and temporal filter

Multiple years' mean and variance of the LAI products are needed as the background knowledge for data integration. Actually, multiple years' mean may be more accurate than satellite retrieval for the case where inter-annual variation of LAI is smaller than the measurement error of satellite LAI products. MISR L3 data at one day is used only by the MRT method and no temporal information is incorporated. So, temporal filter is only carried out on MODIS and CYCLOPES data.

Even after multiple years' data are used, it is still possible in some cases that there are no enough data to compute mean or variance. The time series of mean and variance also tend to show spurious fluctuation due to errors in original data. It is necessary to

apply some smoothing technique to filter the mean and variance to obtain more plausible climatology. Chen *et al.* (2006) proposed to use cubic spline to smooth MODIS NDVI products. In their paper, a left free end condition is used to keep the time series linear at the starting and ending point. Such cubic spline is called "natural" spline. For the time series of LAI, it is not necessarily left free. On the contrary, the time series of climatology data should have the periodic constraints since the climatology repeat itself from year to year. Based on the method developed by Chen *et al.* (2006), the locally adjusted cubic spline capping method with periodic end conditions is developed to assure the smoothed climatology has the following properties:

- The spurious low value is removed.

- The seasonal curve of vegetation is kept.

- The periodicity of LAI phenology is considered.

- The method is able to interpolate at no observation time in order to match two products with different temporal resolutions.

The same set of notations from Chen *et al.*(2006)'s paper is used here. For one pixel, there are N original LAI climatology values $L(t), t = 1, 2, ..N$ (N = 47 for MODIS and N=37 for CYCLOPES). The length of time series of LAI used here is one time span longer than the annual climatology. The first data is added after the last data to

keep the time series periodic. A piecewise cubic polynomial curve is used to fit the

LAI climatology.

$$L_i(t) = a_i(t - t_i)^3 + b_i(t - t_i)^2 + c_i(t - t_i) + d_i, i = 1,2,..N \qquad \text{2-1}$$

The continuous conditions at the boundaries are:

$$L_i(t) = L_{i+1}(t), L_i'(t) = L_{i+1}'(t), L_i''(t) = L_{i+1}''(t), i = 1,2,..N-1 \qquad \text{2-2}$$

$$L_N(t) = L_1(t), L_N'(t) = L_1'(t), L_N''(t) = L_1''(t) \qquad \text{2-3}$$

The equations 2-3 make sure the curve is periodic. By applying these conditions, one

can get a group of linear equations

$$(M + \mu Q^T \Gamma Q)b = Q^T L \qquad \text{2-4}$$

where

$$M = \begin{pmatrix}
p_1 & h_1 & 0 & \cdots & 0 & h_0 \\
h_1 & p_2 & h_2 & \cdots & 0 & 0 \\
0 & h_2 & p_3 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & p_{N-1} & h_{N-1} \\
h_N & 0 & 0 & \cdots & h_{N-1} & p_N
\end{pmatrix} \qquad \text{2-5}$$

$$Q^T = \begin{pmatrix}
f_1 & r_1 & 0 & \cdots & 0 & r_0 \\
r_1 & f_2 & r_2 & \cdots & 0 & 0 \\
0 & r_2 & f_3 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & f_{N-1} & r_{N-1} \\
r_N & 0 & 0 & \cdots & r_{N-1} & f_N
\end{pmatrix} \qquad \text{2-6}$$

$$\Gamma = \begin{pmatrix}
\gamma_{11} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \gamma_{nn}
\end{pmatrix} \qquad \text{2-7}$$

b and L is the vector of b_i and L_i, respectively. μ and γ_{ii} is respectively used to adjust the smoothness of curve globally and locally.

$$h_i = t_{i+1} - t_i,\ i = 1,2,..N \qquad \text{2-8}$$

The periodicity of the time series leads to $h_i = h_{k \bullet N + i}$ and k is a natural number. Specifically, there exists $h_0 = h_N$

$$p_i = 2(h_{i-1} + h_i) \qquad \text{2-9}$$

$$f_i = -\left(\frac{3}{h_{i-1}} + \frac{3}{h_i}\right) \qquad \text{2-10}$$

$$r_i = -\frac{3}{h_{i-1}} \qquad \text{2-11}$$

The coefficient b is obtained by solve the linear equations.

$$b = (M + \mu Q^T \Gamma Q)^{-1} Q^T L \qquad \text{2-12}$$

M, Q and Γ are all symmetric and cyclic matrices, which makes the coefficient b cyclic and the time series periodic. Then, a, c and d are obtained by inserting b into the continuous conditions.

2.2.1 Results

Figure 2-3 shows the LAI climatology on July 20-27 calculated from MODIS/Terra and MODIS/Aqua LAI products. The images before and after temporal smoothing do not show a large difference. Nevertheless, the image after cubic spline filter is smoother if zoomed in. This can be better noticed from the time series at one selected

point (Figure 2-4). The random error is reduced through the smoothing procedure and the smoothed climatology is more likely to represent the natural phenology curve. The climatology calculated from CYCLOPES data is also shown in Figure 2-5. In order to compare the two data, I interpolate CYCLOPES climatology to MODIS observation time at the same time of smoothing using cubic spline. The difference between the two products is obvious. This difference strongly correlates with land cover type (Figure 2-6). The difference is larger in highly vegetated forest than grassland or shrub. The evergreen broadleaf forest has the largest discrepancy.
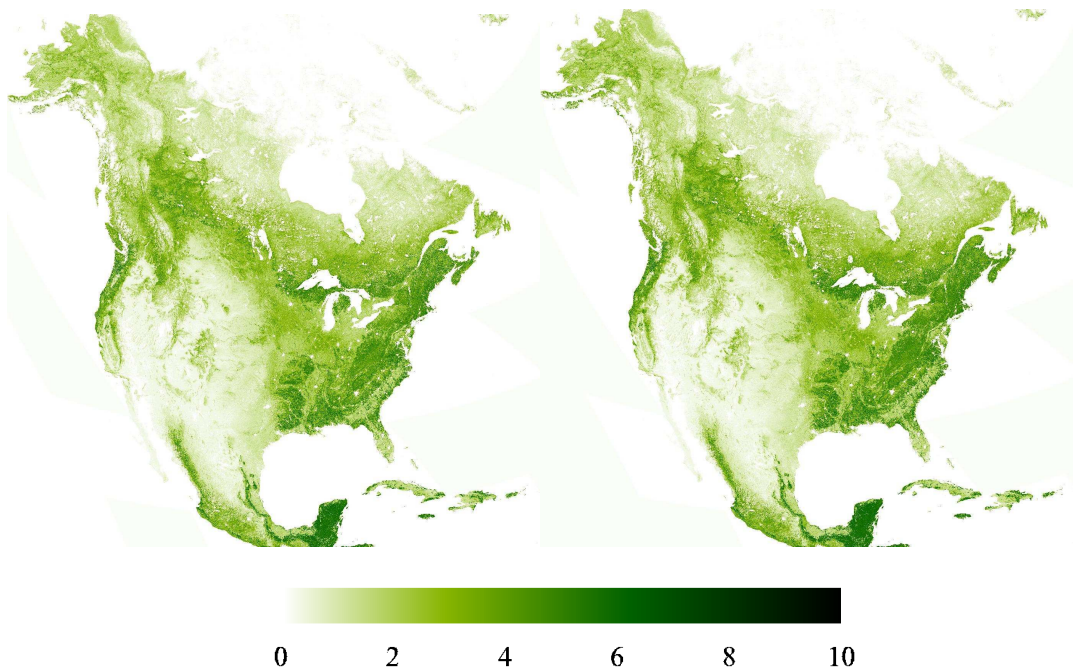
Figure 2-3 North America LAI climatology calculated from multiple years' MODIS data on July 20-27 a) original b) after temporal smoothing.

2.3 Geometric accuracy

Both MODIS and CYCLOPES LAI products are believed to achieve very high geometric accuracy. Through sophisticatedly designed orientation systems and a

global network of ground control points, MODIS observations are expected to have sub-pixel accuracies (Wolfe et al. 2002). The geometric error of CYCLOPES could be less than one pixel, since the VEGETATION data maintain a very accurate record. In terms of multiple scenes registration, VEGETATION1 has 95% points with error smaller than 450m, and the error for VEGETATION2 is 320 m (Sylvander *et al.* 2003). The high coregistration accuracy is the prerequisite for data integration. Nevertheless, there is little work on investigating the relative geometric error of the two products.
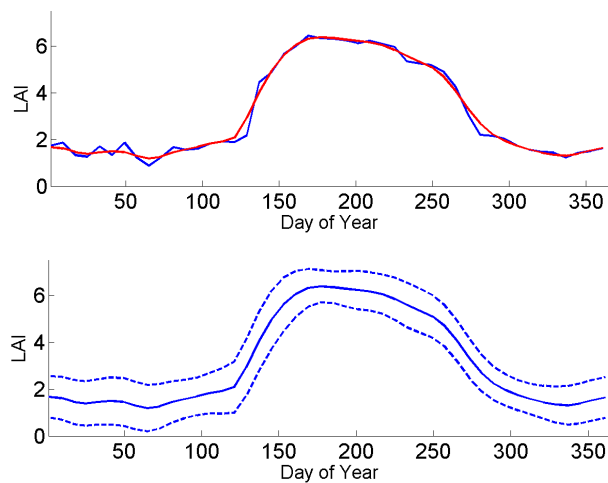
Figure 2-4 MODIS climatology at CHEQ before and after cubic spline filter.

The common way to evaluate the geometric accuracy is to compare the pairs of the ground control points. It is highly subjective and contains large error to manually pick up ground control points from these kinds of coarse resolution satellite products, especially when the large errors of missing values and large errors of LAI estimations exist. Both products provide water mask information. The water mask data are used to evaluate the relative geometric accuracy of the two products.
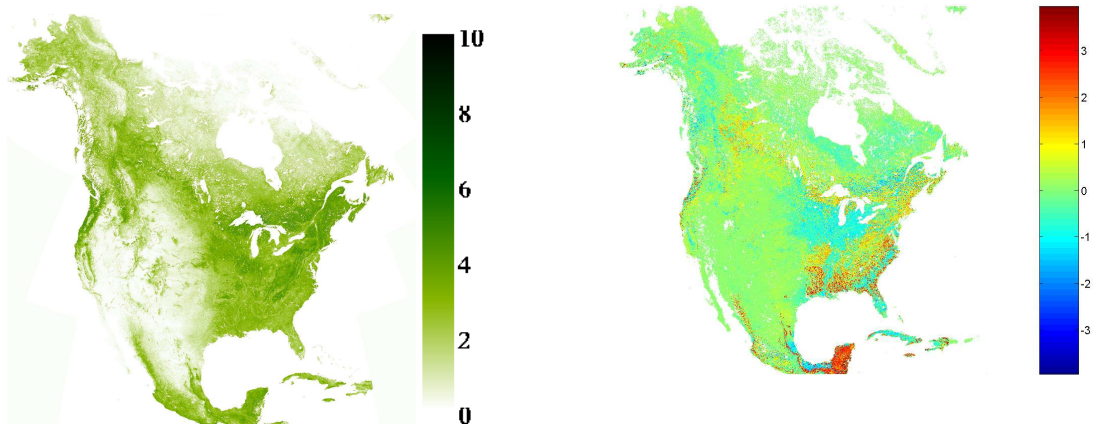
Figure 2-5　Smoothed CYCLOPES LAI climatology interpolated to MODIS temporal resolution on July 20-27 and its difference from MODIS climatology

Land Cover Type

■ water	■ deciduous needleleaf tree	■ grass	■ urban and bulit-up
■ evergreen needleleaf tree	■ deciduous broadleaf tree	■ cereal crop	■ snow and ice
■ evergreen broadleaf tree	■ shrub	■ broadleaf crop	■ barren or sparsely vegetated
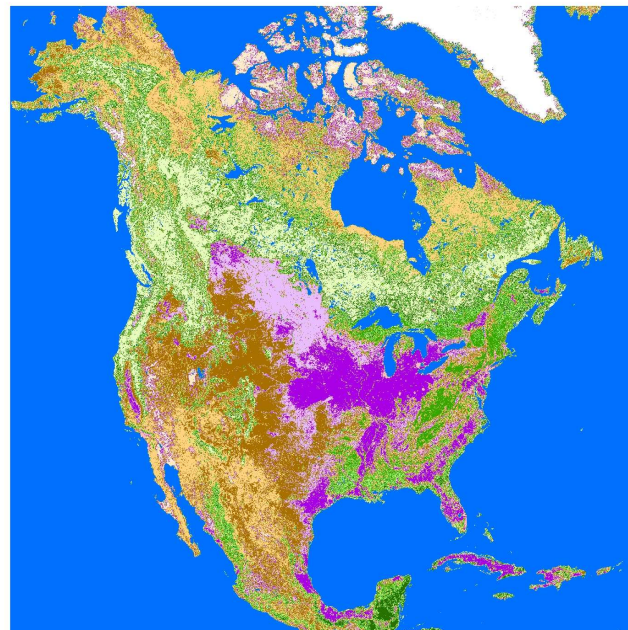
Figure 2-6 Land cover map over North America derived from the MODIS land cover product (MOD12Q1)

2.3.1 Geometric registration

There are two types of image registration methods: correlation based and feature

based. Since it is computationally burdensome to calculate the correlation of such large images, feature based method is selected. Both the point feature (centroids of small islands) and line feature (shorelines) are extracted and used for evaluating the geometric accuracy.

Method 1:Semi-automated coregistration of centroids of small islands

a. Select pairs of islands from the two images and mark one point on each island (as shown in Figure 2-7).

b. Use these points as seeds to segment images using the region growth algorithm and obtain all points of each island.

c. Compute the centroid's coordinate of each island.

d. Project the coordinates of the CYCLOPES image to the Sinusoidal projection.

e. Calculate the RMSE of the pairs of centroid coordinates.
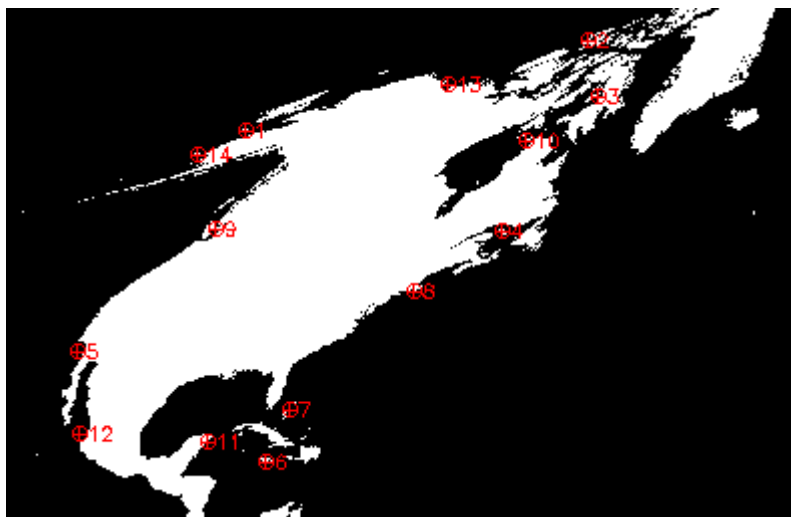
Figure 2-7 Seeds of small islands, used to segment islands from the water mask of satellite LAI products

The results are shown in Table 2-3. Overall the two products match well. The errors between most point pairs are less than one pixel and the mean error is 0.92 pixels. Besides the geometric error of each product, another possible reason for the mismatch is that the two products use different water masks, which have variable mapping accuracy for water/land boundary.

Table 2-3 Coordinate of small islands' centroids in MODIS and CYCLOPES images

Island No	MODIS X	MODIS Y	CYCLOPES X	CYCLOPES Y	Error
1	4045.30	1993.37	4046.17	1993.26	0.88
2	9713.36	505.58	9714.05	505.13	0.83
3	9859.49	1449.52	9859.82	1448.96	0.65
4	8289.11	3662.27	8289.75	3661.48	1.02
5	1262.58	5652.30	1262.98	5652.09	0.45
6	4385.19	7421.99	4385.11	7420.83	1.16
7	4785.13	6596.38	4784.74	6595.91	0.61
8	6841.74	4633.80	6841.61	4633.03	0.78
9	3536.20	3642.26	3537.89	3641.35	1.92
10	8710.54	2156.81	8711.09	2156.05	0.93
11	3429.13	7147.46	3428.84	7146.61	0.90
12	1311.21	7005.21	1311.51	7004.23	1.02
13	7378.74	1250.31	7379.38	1249.85	0.79
14	3247.64	2388.51	3248.37	2388.01	0.88

Method 2: Shoreline matching based on Chamfer matching

1. Extract shorelines from both images.

2. Calculate the Euclidean distance transform of one shoreline image (Figure 2-8).

3. Chamfer match another image to this distance transform, supposed that only translation error is present.

4. The best match position which minimizes the distance error represents the

coregistration error of the two images.
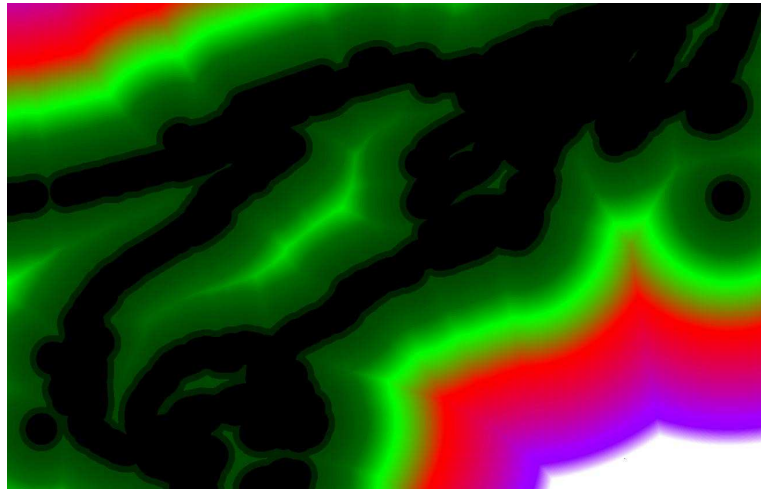
Figure 2-8 Euclidean distance transform of shoreline extracted from CYCLOPES

The water mask of CYCLOPES data doesn't contain the inland water. The shorelines extracted from the two products cannot exactly match. The chamfer distance is minimal when the offset is one pixel (Figure 2-9).
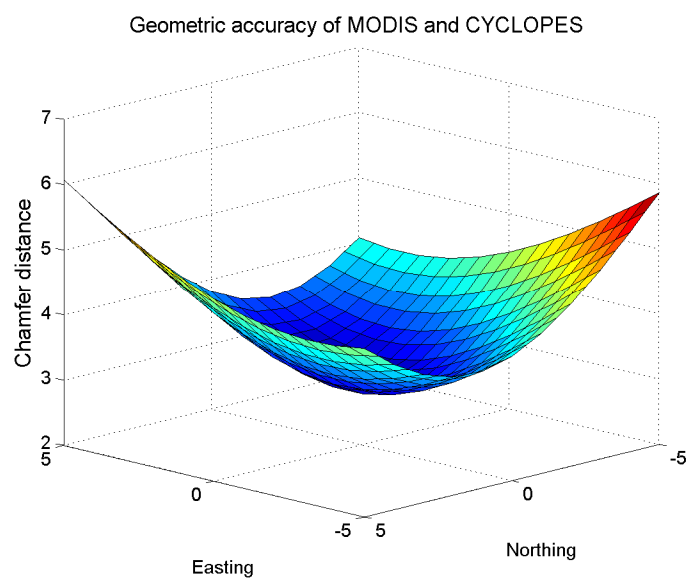
Figure 2-9 Chamfer distance as a function of easting and northing offsets

2.3.2 Influence of geometric error

Both methods show that the relative geometric error is around one pixel. The following will show how the one pixel error may influence the integration results. The CYCLOPES LAI product is used since it is believed to have higher accuracy. The LAI values at the origin and the mean of LAI values at this point and at one pixel lag are compared. The data from the growth season are chosen when LAI have maximal values. Because the original product contains large uncertainties and gaps, the multiple years' mean values are used. The spatial variation apparently determines the impact of geometric errors on LAI accuracy. The influence is calculated at different land cover types. The results on four typical vegetation regions are shown in Figure 2-10.

Temperate forest has the largest error, with the RMSE of 0.13. Grassland has the smallest error, where the overall LAI is smaller. For all the land cover types, the error due to misregistration is small, compared with the uncertainties of the retrieval algorithms. In the following data integration, the data in their original geographic coordinate are used to avoid the new errors introduced by inaccurate registration parameters.
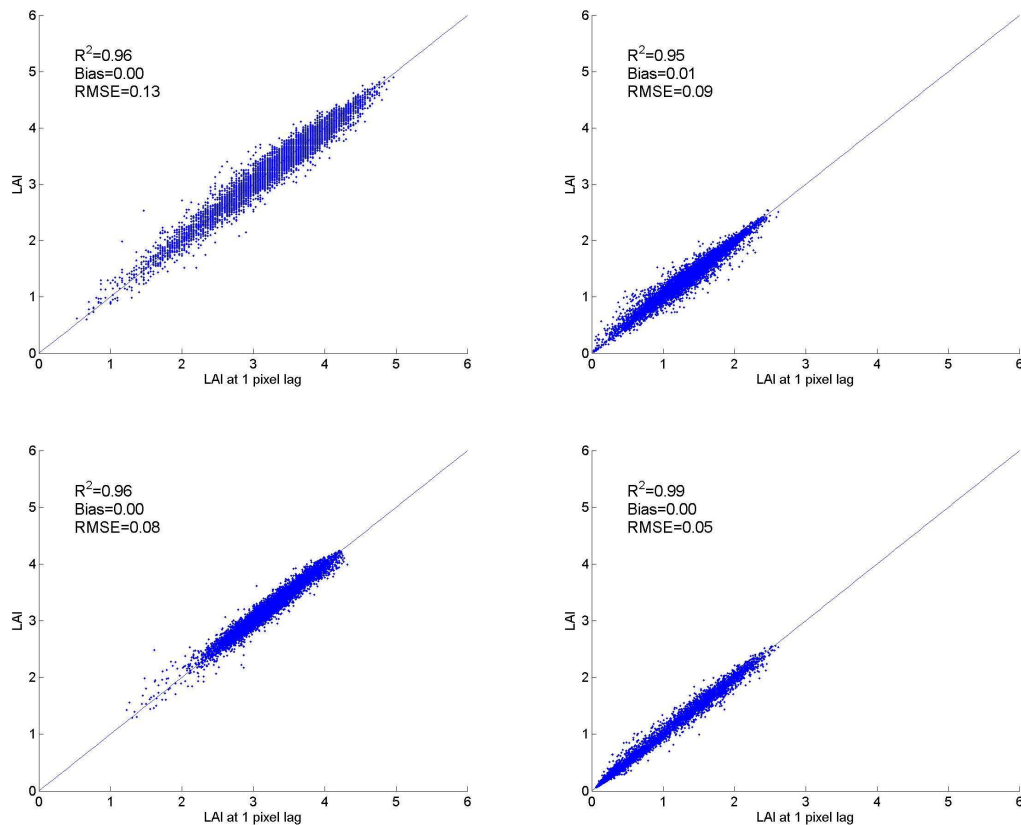
Figure 2-10 Impact of geometric errors on integration accuracy

a) temperate forest b)boreal c)tropical forest d)grassland

2.4 Impact of atmospheric condition on data quality

Error in satellite LAI products can be twofold: one is systematic, which is related to the shortcoming of retrieval algorithm, the inaccurate prior knowledge, i.e. the incorrect land classification input; the other is random, which comes from random input error such as atmospheric correction or BRDF effects. The systematic error

cannot be reduced through statistics without bias correction. I mainly focus on the random error. For retrieval of LAI, the atmospheric condition is an important factor determining the data quality. Comprehensive field measurements are unavailable to examine the impact of atmospheric condition on data quality. Here I take advantage of the twin sensors of MODIS to evaluate this issue. The actual LAI values in the morning and afternoon are similar. The difference between MODIS/Terra and MODIS/Aqua LAI should come mainly from the random error of LAI retrieval. The best LAI retrievals from one of the two sensors with the highest quality are used to evaluate the LAI from the other sensor. Here the best data are referred to those from the radiative transfer method and not contaminated by cloud, aerosol, snow, cirrus or cloud shadow. Limiting the atmospheric condition of the second retrieval to different categories, one can see their impact on LAI data accuracy (Figure 2-11).

Clouds reduce the data quality most seriously. The data under cloud contamination have a significant negative bias and a very large variance (1.57). The error distribution is far from normal. The data under cloud shadow also has a non-normally distributed error and with large variance (1.03). LAI retrievals when significant aerosol is present have almost no bias. The accuracy is also much better than that of cloud shadow. The data from empirical backup algorithm also have an unbiased error and a slightly higher variance than data under aerosol contamination.
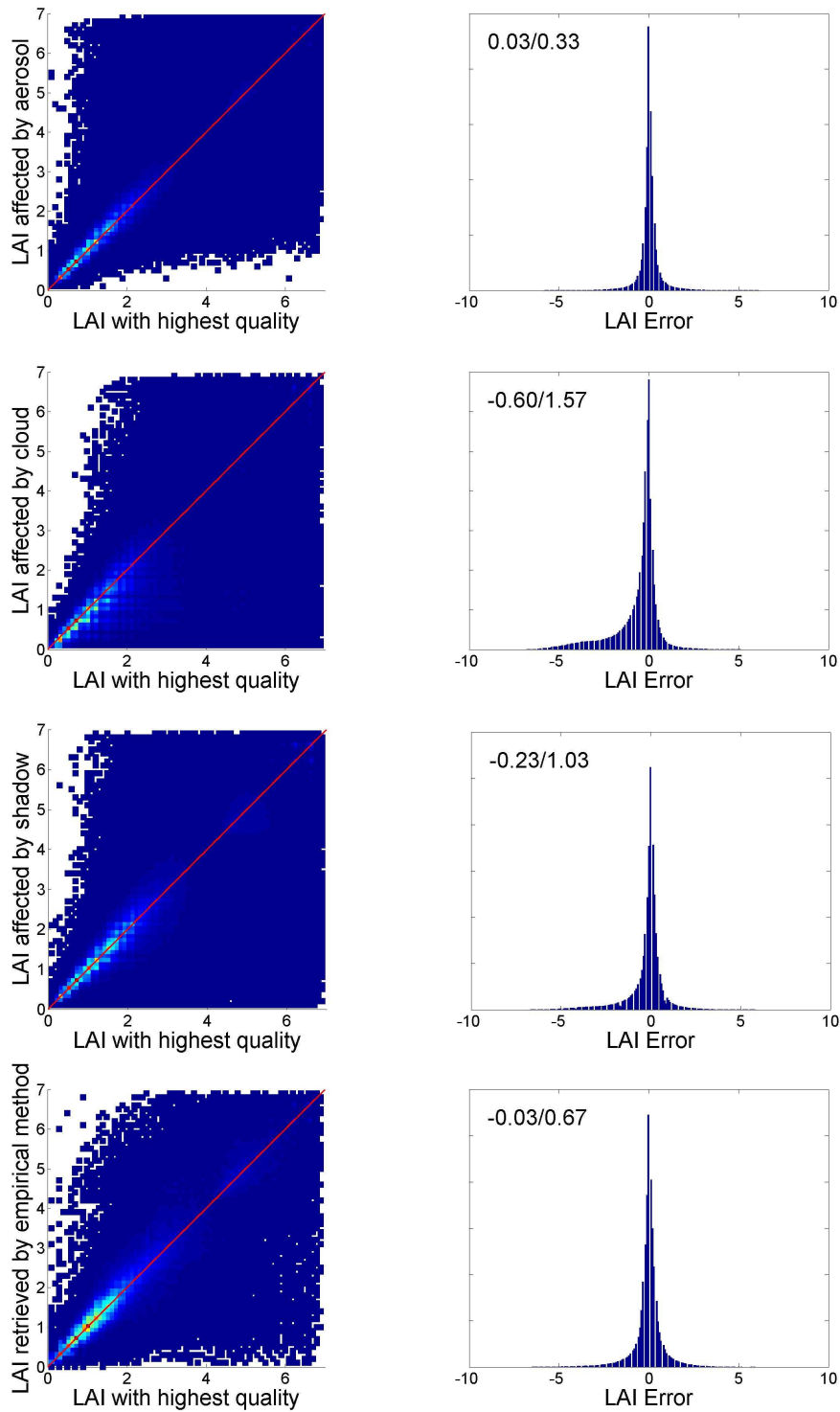
Figure 2-11 Impact of atmospheric condition on LAI data quality. The second column shows the histograms of LAI error distribution, and the first number is mean and the second number is variance.

2.5 Intercomparison of MODIS and CYCLOPES LAI products

Data quality such as measurement error is a critical input of data integration algorithms. This section compares two moderate resolution products: MODIS and CYCLOPES LAI. The exploratory analysis of the relative data quality will serve as the basis of data integration. The two products are evaluated in terms of retrieval rates, absolute accuracy and relative bias. MODIS has higher retrieval rates in winter and smaller in summer than CYCLOPES. Intercomparison shows systematic discrepancy exists between the two products. CYCLOPES has higher accuracy but may generate spurious large values in winter.

2.5.1 Retrieval rate

Many factors may cause failure of retrieving LAI. Surface snow cover and high cloud frequency within the observation period are the two leading causes. Adverse atmospheric conditions, unfavorable observation geometry, instrumental failure and NIR reflectance saturation induced by dense vegetation, among other reasons, may also lead to missing data or low retrieval quality. The retrieval rate, which is the ratio between the number of successfully retrieved data and the number of expected data, is used to evaluate the distribution of missing data. All data downloaded are used to calculate this rate in terms of different land cover types and different seasons. The

results are shown in Figure 2-12.
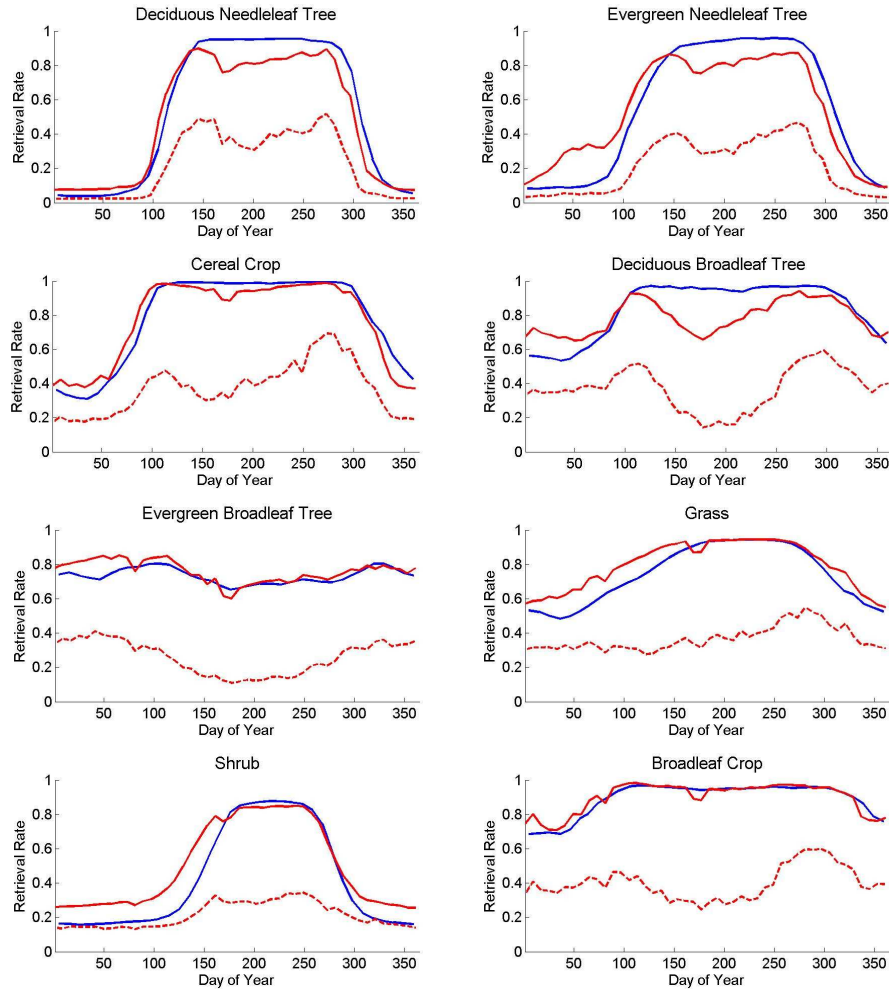
Figure 2-12   Retrieval rates of CYCLOPES and MODIS LAI products. Blue lines represent CYCLOPES, solid red lines are MODIS and dashed red lines mean MODIS with the highest data quality.

Generally, MODIS has higher retrieval rates in winter and smaller in summer than CYCLOPES. Both of the satellite data have extremely low retrieval rate for needleleaf tree in winter season, which is distributed mainly at high latitudes. The retrieval rates for evergreen broadleaf tree are also low across the whole year and

especially lower in summer due to the high atmospheric moisture. For all the vegetation type and seasons, the percentage of MODIS data with highest quality is low, generally smaller than 50%. The rates for broadleaf trees are extremely low in the growth season due to saturation and high atmospheric moisture.

2.5.2 Direct validation

Direct comparison of the two satellite LAI products with high resolution LAI maps has been done by Garrigues *et al.*(2008). However, they are not using the latest version of MODIS and their results are based on a degraded temporal and spatial resolution. Here, I use the latest version of MODIS LAI data and evaluate the two satellite products at their original resolution. The LAI reference maps are aggregated to 1km to match the satellite products. According to the time and location of the aggregated maps, a single closest pixel of the satellite products is chosen to compare with the reference data. Figure 2-13 shows the CYCLOPES LAI performs a little better in terms of R square and RMSE. MODIS overestimates LAI with a positive bias of 0.28, while CYCLOPES has a negative bias 0.20. The overall data quality of these two satellite LAI products is expected to be worse than the shown results. The availability of a high resolution LAI reference map indicates the atmospheric condition is ideal, which means possible high quality of satellite LAI products. Twenty-two out of 31 available MODIS observations have the QC value of zero,

much higher than the overall MODIS LAI product quality.
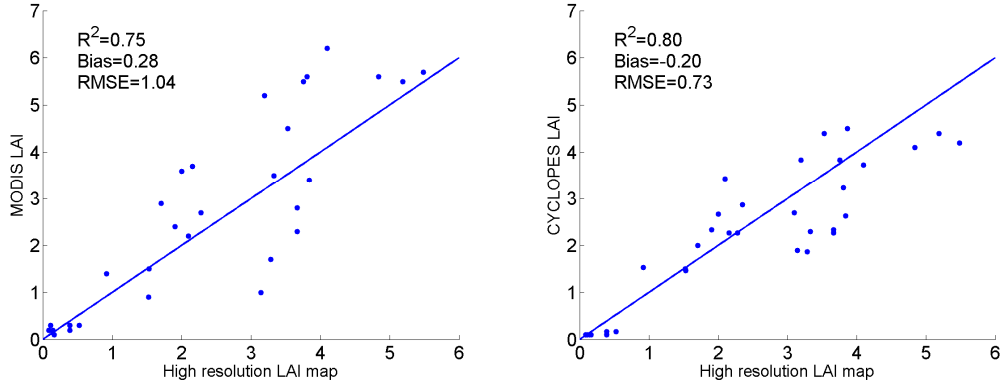
Figure 2-13 Direct validation of MODIS and CYCLOPES LAI products using high resolution LAI reference data

The relation between data quality and QC is further investigated. When applying the QC of MODIS, data with better quality control do not always mean higher accuracy (Table 2-4). The bias and RMSE decreases from low quality to high quality. However, excluding data from the backup algorithm and data contaminated by cloud fails to improve R square and relative RMSE. This indicates the retrieved LAI with lower quality may contain large systematic errors. The random errors are degrading the data quality seriously even for retrieved data with high confidence.

Table 2-4 Validation results of MODIS LAI when applying different quality control

	R2	Bias	RMSE	Relative RMSE
All data	0.75	0.28	1.04	44%
Data from RT	0.65	0.08	0.96	44%
Data from RT without cloud	0.72	-0.03	0.44	51%

Visual examination shows CYCLOPES has smoother temporal profiles and CYCLOPES LAI maps appear smoother than MODIS images. In order to evaluate which product better expresses the spatial heterogeneity of LAI, I calculate the coefficient of variation within selected sites (7*7km). The reference maps are aggregated to 1km resolution, then used to derive the coefficient. Because the original satellite products may contain too many random errors, the coefficients of variation are also computed on multiple years' average of LAI products (Table 2-5). The spatial variations of MODIS data are closer to reference maps, while CYCLOPES has overall smaller values. The multiple years' mean of CYLOPES has similar results with the original CYCLOPES data, which indicate the CYCLOPES LAI is stable over years. The random errors in MODIS original data are reduced to make MODIS climatology have much smaller variation.

Table 2-5 Coefficients of variation within a 7*7 km region at validation sites

Site	Date	Reference Map	Terra	Modis Climatology	CYCLOPES	Cyclopes Climatology
ARGO	2000 07 04	0.21	0.19	0.06	0.03	0.03
CHEQ	2002 08 01	0.21	0.15	0.17	0.00	0.01
HARV	2002 08 24	0.08	0.09	0.06	0.04	0.01
KONZ	2000 06 07	0.09	0.14	0.08	0.01	0.00
METL	2002 09 24	0.16	0.25	0.07	0.04	0.04
NOBS	2000 07 14	0.17	0.10	0.06	0.03	0.02
SEVI	2002 07 26	0.01	0.00	0.00	0.00	0.00

2.5.3 Intercomparison

Direct comparison of the two products shows large scattering. Even after land cover

types are considered, the relationship between CYCLOPES and MODIS LAI is not significantly improved (Figure 2-14). The large scattering may come from both the systematic bias of the two products and their respective random errors. In order to further investigate the consistency between the two products, multiple years' averages for each land cover type are compared.
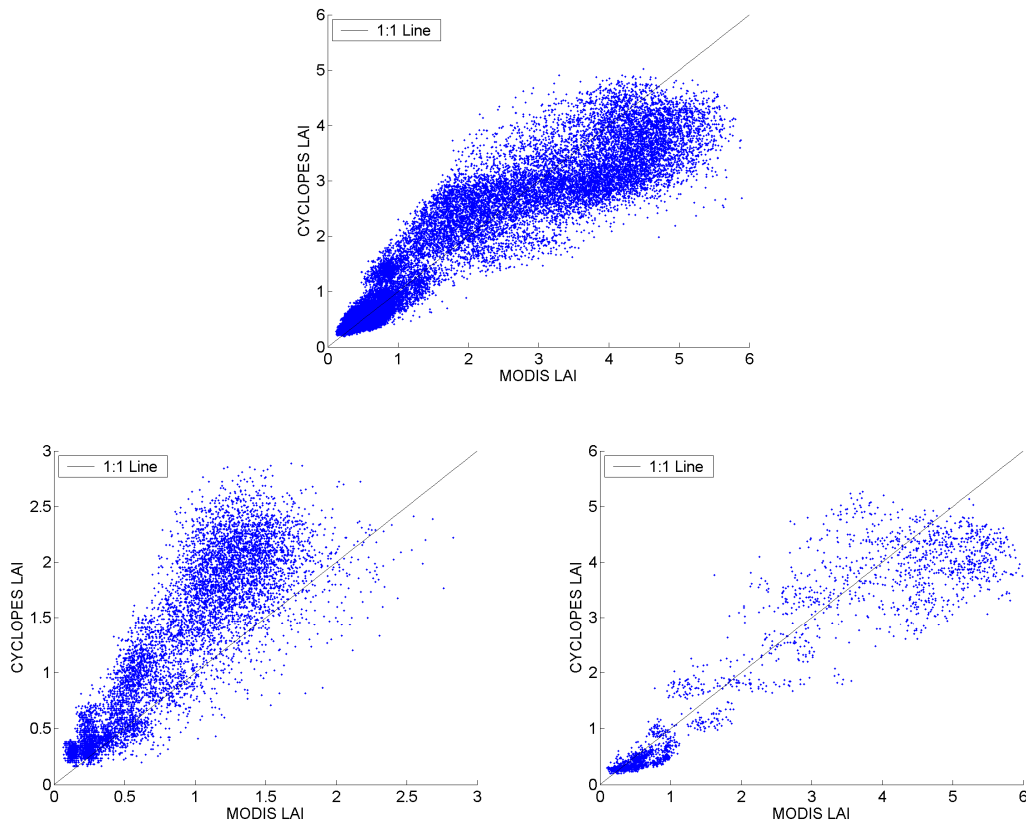
Figure 2-14 Relationship between CYCLOPES LAI and MODIS LAI at a mid-latitude location near the Beltsville Agriculture Research Center, MD a) all pixels b) broadleaf crop c) deciduous broadleaf tree

The two products agree more closely because the random errors are reduced greatly through multiple years' averaging and averaging among multiple pixels with the same land cover type. Especially for grass, shrub, cereal crop, deciduous needleleaf tree

and evergreen needleleaf tree, the two products have almost the same results. For broadleaf tree and sparsely vegetated area, MODIS produces much larger values. There is only one case, broadleaf crop, where MODIS is smaller than CYCLOPES. The MODIS LAI is almost twice that of the CYCLOPES LAI for evergreen broadleaf tree. There are two possible reasons for this according to Weiss *et al.* (2007): LAI saturation is worse for CYCLOPES and MODIS data are corrected for some clumping effects. This systematic mismatch between CYCLOPES and MODIS must be addressed in the integration of the two products. Another problem worthy of notice is that CYLCOPES has some spurious high LAI values in winter (Figure 2-16). One possible explanation is that CYLCOPES has bad input of reflectance under snow contamination.
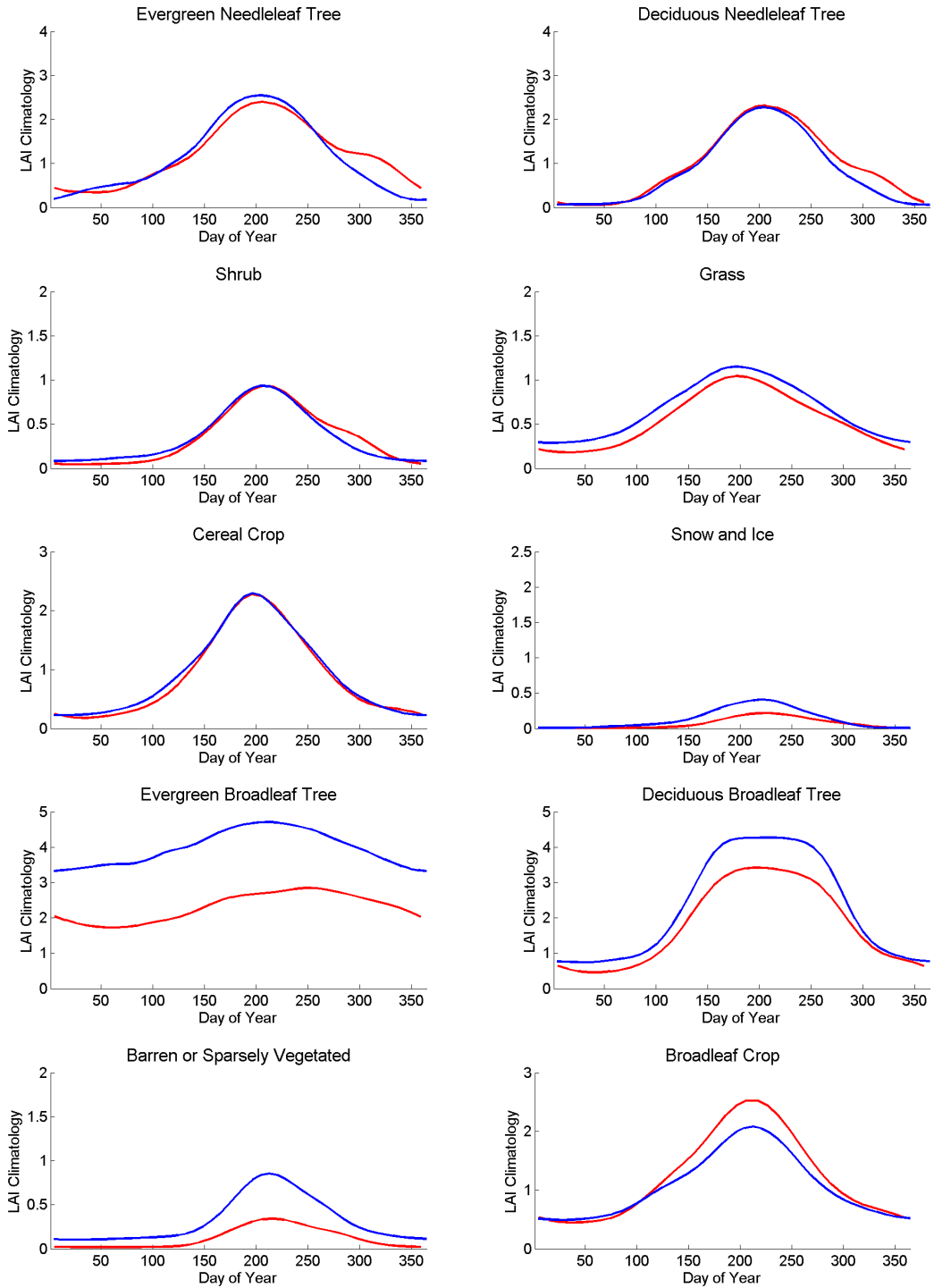
Figure 2-15 Comparison of multiple years mean of MODIS and CYCLOPES
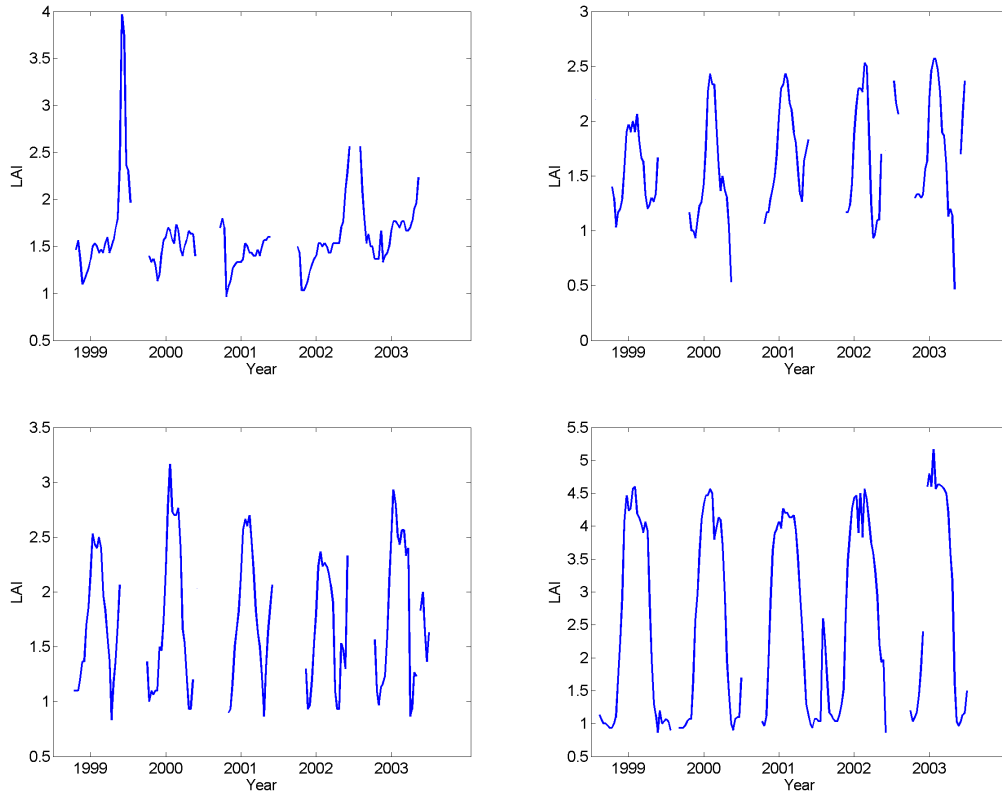
Figure 2-16 Spurious large values of CYLCOPES LAI in winter. Examples are shown at a) METL, b) NOBS c) Thompson d)HARV

Chapter 3 Geostatistical methods for integrating MODIS and CYCLOPES LAI products

This chapter applies both "traditional" geostatisical method OI and modern geostatisical method BME to integrate MODIS and CYCLOPES LAI products. After a brief introduction to the two methods, a normalization scheme to account for the systematic biases of the two satellite products is developed. The integrated results are validated and analyzed in both time series and two dimensional images.

3.1 Methodology

This section first gives the basic concepts of the space time stochastic process. Then, two geostatistical mapping methods and their application are reviewed. OI will give the best linear unbiased prediction for the Gaussian process, while BME relaxes the linear and Gaussian assumption of OI.

3.1.1 Introduction to space time stochastic process

In the context of spatiotemporal statistics, LAI is treated as a space-time stochastic process, a collection of random variables $L(s,t)$ over probability space (Ω, F, P).

The index here contains location s and time t:

$$\{s \in R^2, t \in R\} \qquad \text{3-1}$$

For every realization of $L(s,t)$, its probability can be denoted using the cumulative distribution function CDF F:

$$F_{L(s,t)}(x) = Probability(L(s,t) < x) \qquad \text{3-2}$$

The derivative of CDF defines another function of $L(s,t)$ probability density function (*pdf*) f:

$$f_{L(s,t)}(x) = \frac{dF_{L(s,t)}(x)}{dx} \qquad \text{3-3}$$

Given *pdf*, the mean of this stochastic process can be calculated by

$$\mu_L(s,t) = \int x \cdot f_{L(s,t)}(x) dx \qquad \text{3-4}$$

For two points (s,t) and (s',t'), their covariance is defined by

$$C_L(s,t,s',t') = \int (L(s,t) - \mu_L(s,t)) \cdot (L(s',t') - \mu_L(s',t')) f_{L(s,t),L(s',t')}(x,x') dx dx' \qquad \text{3-5}$$

where $f_{L(s,t),L(s',t')}(x,x')$ is multivariate *pdf* of random variable $L(s,t)$ and $L(s',t')$. The covariance function must be positively definite to keep it valid. For a covariance function to be positively definite, it must satisfy:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C_L(s_i,t_i,s_j,t_j) > 0 \text{ given any } a \qquad \text{3-6}$$

For the sake of simplicity, the random field usually is assumed to have the property of stationarity. A strict stationarity is defined by the multivariate CDF. The CDF is invariant with the temporal lag and spatial lag. In most cases, only weak stationary (wide sense stationary) is required, which has the following properties:

$$\mu_L(s,t) = c \hspace{4cm} \text{3-7}$$

$$C_L(s,t,s',t') = C_L(s-s',t-t') \hspace{3cm} \text{3-8}$$

This means the mean is a constant and the covariance is only the function of distance in space-time domain. According to its definition, a stationary random field cannot contain non-constant space-time trend. One important task of geostatistics is to remove large-scale trend T and get homogenous zero-mean small scale variation V:

$$L(s,t) = T(s,t) + V(s,t) \hspace{3cm} \text{3-9}$$

T is assumed to be deterministic and V is space-time stochastic field. V is the real target of spatiotemporal geostatistics.

I start the data integration theory with a very simple case: the point to be estimated coincides with the point measured. Without loss of generality, supposing that there are two LAI observations L_1, L_2 from two satellite products at the same space-time point with errors e_1 and e_2, and the real LAI value at this space-time coordinate is L_t, there are two measurement equations:

$$L_1 = L_t + e_1 \hspace{3cm} \text{3-10}$$

$$L_2 = L_t + e_2 \hspace{3cm} \text{3-11}$$

Assume both of the satellite products are unbiased, that is

$$E(L_1) = E(L_2) = L_t \hspace{3cm} \text{3-12}$$

Because the two observations are independently retrieved from different sensors using different algorithms, the errors of the two products are assumed to be

uncorrelated, that is

$$Cov(e_1, e_2) = 0 \qquad\qquad 3\text{-}13$$

The variance of the two observations is $\sigma_1{}^2$ and $\sigma_2{}^2$. Under such assumption, an estimation of L_t can be expected in the linear form:

$$L_t = a_1 L_1 + a_2 L_2 \qquad\qquad 3\text{-}14$$

Whether in terms of the least square, variational method or Bayesian theorem, the best estimator has the same parameter:

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \qquad\qquad 3\text{-}15$$

So, starting from either theory, for a Gaussian process, the final result will be the same (Lorenc 1986; Wikle and Berliner 2007). Gu *et al.*(2006)'s method of improving MODIS LAI is based on this simple one point optimal interpolation. However, in data integration of LAI, usually there is no reliable observation available in the same spatiotemporal point of estimation. Information from temporally and spatially adjacent points is needed to improve the estimation. Accordingly, temporal or spatial methods with the ability of interpolating are needed.

3.1.2 Optimal interpolation

OI, also known as objective analysis, is the statistical method of spatial interpolation on the Gaussian process. OI was invented in the field of numerical weather prediction

by Gandin (1965). In geostatistics, a very similar set of theories, Kriging, was independently developed almost at the same period (Matheron 1963). OI is designed for spatial interpolation, however it is natural to add time as another dimension of space and extend OI to the spatiotemporal domain. For some variables, this extension is plausible. These variables evolve in both spatial and temporal domains and have both spatial and temporal dependency. For example, insolation can be modeled through a spatiotemporal covariance matrix and predicted using Kriging in the spatiotemporal domain (Huang *et al.* 2007). However, some variables show significantly variable properties in spatial and temporal domains. Uz and Yoder (2004) found temporal correlation barely exists in oceanic chlorophyll concentration although spatial correlations were found, so Pottier *et al.* (2006) used only spatial dependency in merging MODIS and SeaWIFS chlorophyll products. Without loss of generality, LAI is treated as both spatially and temporally indexed.

Given the mean and covariance function of this Gaussian process, we have observations $L(s_i, t_i)$ of LAI at a set of space time points (s_i, t_i), and we want to predict the value of LAI at a point (s_0, t_0). OI pursues the linear predictor $\hat{L}(s_0, t_0)$ of $L(s_0, t_0)$ in the form of:

$$\hat{L}(s_0, t_0) = \mu(s_0, t_0) + \sum \lambda_i \cdot [L(s_i, t_i) - \mu(s_i, t_i)] \qquad \text{3-16}$$

through minimizing the error $[\hat{L}(s_0, t_0) - L(s_0, t_0)]^2$. To minimize the cost function, we

take the first order condition and get the parameters λ_i in the form (Bretherton *et al.* 1976) :

$$\lambda_i = \sum_{i=1}^{N}\sum_{j=1}^{N} C(s_i,t_i,s_j,t_j)^{-1} C(s_0,t_0,s_i,t_i) \qquad \text{3-17}$$

Recently, OI was extensively used to reconstruct satellite products. Reynolds and Smith (1994) used OI to fuse *in situ* measurement and AVHRR data and produce a gridded sea surface temperature product at 1 degree. Le Traon *et al.* (1998) improved sea level anomaly mapping by integrating TOPEX/Poseidon and ERS-1 altimeter data. Pottier et al. (2006) combined SeaWIFS and MODIS observations together to enhance the mapping area of chlorophyll and reduce measurement error. Sapiano *et al.* (2008) blended Special Sensor Microwave/Imager (SSM/I) data with ERA-40 reanalysis to generate 2.5 degree global precipitation data. Chao *et al.* (2009) used an OI equivalent two-dimensional variational data assimilation to merge five satellite SST data. In reconstructing LAI products, both the Gu *et al.* (2006) and Fang *et al.* (2008) methods used the concept of OI in a simplified form. Gu *et al.* (2006) used only the data at one point and didn't consider spatiotemporal covariance among different points. Fang *et al.* (2008) empirically calculated the weights of adjacent observations instead of optimally inversing covariance matrix.

OI is a linear predictor on the Gaussian process. A new epistemological

spatiotemporal mapping method, BME, has been recently developed. Uncertain observations are treated as soft data in BME and the errors are rigorously considered. Unlike the "traditional" geostastitics, Kriging, BME makes no restrictive modeling assumption such as linearity and normality (Serre and Christakos 1999). BME is different from the linear interpolation used in Kriging in that a flexible form is incorporated. Under some scenarios, BME is simplified to Kriging.

3.1.3 Bayesian Maximum Entropy

Christakos's (2000) monograph systematically describes the BME theory. BME treats spatiotemporal stochastic processes in a different way from traditional geostatistics (e.g. Kriging) by incorporating physical knowledge into the spatiotemporal analysis instead of using a "pure inductive" framework (Christakos 1990; Serre and Christakos 1999). Under the BME's framework, "traditional" geostatistical methods are a special case of BME.

BME has been successfully applied to solve many problems. For example, Christakos *et al.* (2004) employed a nonlinear estimator based on BME to combine both the satellite ozone product and the empirical relationships between ozone and tropopause pressure to produce high spatial resolution ozone products with greater accuracy. Douaik *et al.*(2005) used BME to map soil salinity and found that the BME approach

produced less biased and more accurate predictions than the traditional Kriging approach. Kolovos *et al.* (2002) combined site-specific observations and stochastic partial differential equations into an assimilation of the advection-reaction process.

BME is capable of incorporating both general knowledge and site-specific information. The general knowledge could be either physical laws or statistical moments. Both are expressed in the form of teleologic equations. There are generally three stages for applying BME: *prior*, *meta-prior* and *posterior* (Serre and Christakos 1999; Christakos 2000):

In the *prior stage*, the general knowledge, related with a g_α function, is expressed by the representation of a *G*-operator:

$$\int d\chi_{map} G(g_\alpha) f_G(\chi_{map}) = 0 \quad (\alpha = 1,...,N) \qquad 3\text{-}18$$

The prior *pdf* $f_G(\chi_{map})$ is obtained by maximizing the informative entropy constrained by this equation. χ_{map} represents all the data points, including observed data χ_{data} and the data χ_k to be predicted at unobserved points *k*. In the *meta-prior stage*, the available site-specific data can be divided into true (hard) data χ_{hard} or uncertain (soft) data χ_{soft}. The uncertainties with soft data will be considered explicitly in a rigorous way at this stage. In the *posterior stage*, the posterior *pdf* $f_K(\chi_k)$ of predicted points χ_k is given by:

$$f_K(\chi_k) = A^{-1} \int_D d\chi_{soft} f_G(\chi_{map}) \qquad 3\text{-}19$$

where $A = \int_D d\chi_{data} f_G(\chi_{data})$ is the normalization coefficient. When the physical knowledge is the statistical moments of underlying stochastic process, the *G*-operator of the prior stage takes the form of multivariate Gaussian function and the posterior *pdf* leads to (Serre and Christakos 1999):

$$f_K(x_k) = A^{-1}\phi(x_k; B_{k|h}x_{hard}, c_{k|h})\int dx_{soft} f_S(x_{soft})\phi(x_{soft}; B_{s|kh}x_{kh}, c_{s|kh})\qquad\text{3-20}$$

where, $\phi(x, \bar{x}, c)$ is the multivariate Gaussian function of variable x with the mean $\bar{x}$ and covariance c; $c_{k|h}$ is the covariance matrix of predicted points conditional to hard data; $c_{s|kh}$ is the covariance matrix of soft data points conditional to predicted points and hard data points; $B_{k|h}$ is defined as $c_{k,h}c_{h,h}^{-1}$ and $B_{s|kh}$ is defined as $c_{s,kh}c_{kh,kh}^{-1}$.

Based on the posterior *pdf*, the estimates $\bar{x}_k$ and the errors of estimation σ_k^2 can be expressed as:

$$\bar{x}_k = \int x_k f_K(x_k)dx_k\qquad\text{3-21}$$

$$\sigma_k^2 = \int (x_k - \bar{x}_k)^2 f_K(x_k)dx_k\qquad\text{3-22}$$

Under this framework, all the information is integrated into the Bayesian inference to maximize the information. The results are expected to be more informative and accurate. These properties of BME can address the problems the land remote sensing community is faced with and can make BME ideally suitable for data integration of satellite LAI products.

3.2 Normalizing LAI anomaly

Theoretically, both BME and OI can take irregular observations as input. Consequently, the mismatched spatiotemporal grids of the two LAI products will not be a problem and the proposed methods should work well for the case of two products. However, the two products are found to be significantly different. Besides the random measurement error, the bias and other systematic measurement error also need to be considered. Comprehensive field measurements will be the ideal way to remove the systematic bias or mismatch between the two products. However, such an observation network of LAI is not available. I propose to remove the mismatch by working on normalized anomalies of the two satellite products.

Assume the error of satellite LAI products could be expressed in a linear model:

$$L_M = a_M \cdot L_t + b_M + e_M \qquad \qquad \text{3-23}$$

L_M is MODIS estimated LAI, L_t is the true LAI, a_M and b_M are constants, dependent on the location and time of year. a_M and b_M together with random error e_M can represent different sources of measurement errors. a_M represents the ratio of dynamic range between the estimation and actual LAI values, b_M depicts the estimation bias, and e_M is the normally distributed random error with variance σ_{Me}^2. If MODIS LAI has the same dynamic range and no estimation bias, this

measurement error model will reduce to the simple form:

$$L_M = L_t + e_M \qquad\qquad 3\text{-}24$$

Since current satellite LAI products have large uncertainties, the dynamic error and bias cannot be ignored here. For the CYCLOPES product, there exists the similar equation:

$$L_C = a_C \cdot L_t + b_C + e_C \qquad\qquad 3\text{-}25$$

If the dynamic error and bias of these two products is known, the satellite retrieved values can be corrected before these two products are merged to obtain an improved product without bias and with consistency. The sparse *in situ* measurements make this approach infeasible. Although it is hard to remove the bias without independent LAI measurements, I need to at least try and develop a mechanism to make sure the results have consistent bias and dynamic range.

In dealing with more than one variable with different units, normalization is crucial to make variable commensurable. A straightforward way to standardize the LAI data is

$$Ls = \frac{L_M - \mu_M}{\sigma_M} \qquad\qquad 3\text{-}26$$

$\sigma_M{}^2$ and μ_M are the product variance and mean calculated directly from data, as illustrated in Chapter 2. However, through this direct standardization, one cannot obtain a variable with the property of bias free due to the existence of measurement errors. Given the existence of measurement errors, the calculated product variance

contains two parts:

$$\sigma_M{}^2 = a_M{}^2 \sigma_t{}^2 + \sigma_{Me}{}^2 \qquad\qquad 3\text{-}27$$

where $\sigma_t{}^2$ is the variance of true LAI. Instead of standardized LAI using the calculated product variance, I need to construct a new variable $(a_M \sigma_t)^2$ using the product variance without measurement error:

$$Ls' = \frac{L_M - \mu_M}{a_M \sigma_t} \qquad\qquad 3\text{-}28$$

The normalized anomaly Ls' will be dynamic error and bias free:

$$Ls'(MODIS) = Ls'(CYCLOPES) = \frac{L_t - \mu_t}{\sigma_t} \qquad\qquad 3\text{-}29$$

In order to obtain the product variance without the measurement error from the product variance, I need to make some simplifications. I assume the variance of measurement errors is proportional to the variance of LAI. This assumption is reasonable because larger retrieval error is expected when LAI has larger variation:

$$\sigma_{Me} = k_M a_M \sigma_t \qquad\qquad 3\text{-}30$$

Under this assumption, the normalized anomaly has a simple relationship with the standardized anomaly

$$Ls' = \sqrt{1 + k_M{}^2}\, Ls \qquad\qquad 3\text{-}31$$

The derivation of k will be given in the next section. Thus, the normalized anomaly will be free of dynamic error and bias. OI or BME should be carried out on the normalized anomaly Ls' instead of LAI observation itself. Given the input of unbiased normalized anomalies, the output of data integration is expected to also be

bias-free:

$$\hat{L}s' = f(Ls_i')$$
3-32

The true LAI variance σ_t^2 and mean μ_t of LAI are needed to reconstruct LAI from

the integrated result $\hat{L}s'$.

$$\hat{L} = \hat{L}s' * \sigma_t + \mu_t$$
3-33

They can be estimated by averaging the product variance without measurement error

and the multiple years' mean of MODIS and CYCLOPES:

$$\hat{\mu}_t = \frac{\mu_M + \mu_C}{2}$$
3-34

$$\hat{\sigma}_t = \frac{\sigma_M / \sqrt{1 + k_M^2} + \sigma_C / \sqrt{1 + k_C^2}}{2}$$
3-35

3.3 Analysis of results

3.3.1 Modeling spatiotemporal dependency

The assumption of second order stationarity is hardly valid throughout the earth.

However, the normalized anomaly of LAI can be thought to be stationary within a

small region. Smaller window size can better characterize vegetation heterogeneity,

but can not model large scale dependency. I balance this by setting the window size to

300km and model spatiotemporal dependency in each 300 by 300 window. It is still

impractical to compute the spatiotemporal covariance using all available data within

the window, although I segment the data to 300 by 300 km windows. A stratified

sampling strategy is designed to obtain both small scale and large scale dependency within the window (Figure 3-1).
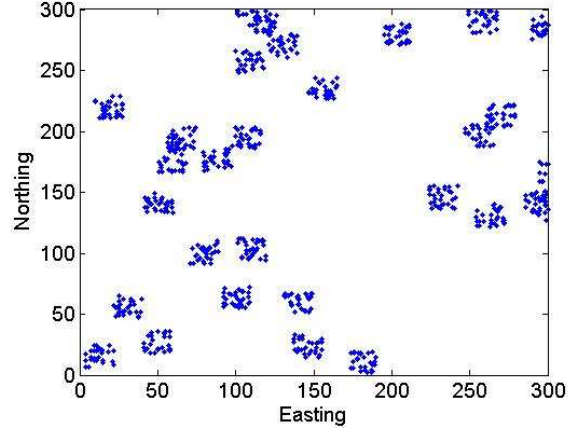
Figure 3-1 Points used to compute covariance within a 300 by 300 window

Exponential models are frequently used (Christakos *et al.* 2004; Douaik *et al.* 2005). I use them to fit the temporal covariance. To model the spatial dependency at both short distance and long distance, a nested exponential covariance model is chosen. Due to the existence of measurement error, the model contains a nugget effect.

$$C(s,t) = [c_1 \exp(-\frac{3s}{a_{s1}}) + c_2 \exp(-\frac{3s}{a_{s2}})]\exp(-\frac{3t}{a_t}) \qquad\qquad s > 0, t > 0 \qquad \text{3-36}$$

$$C(s,t) = c_{Nugget} \qquad\qquad s = 0, t = 0$$

where c_1 , c_2 and c_{Nugget} are coefficients, representing short-distance, long-distance spatial covariance and covariance at the zero-lag point; a_{s1} and a_{s2} are spatial range while a_t is temporal range. Assume the covariance function of the underlying field is continuous. Thus, the discontinuity at the zero lag point is solely from the measurement error. The variance of measurement error would be:

$$\sigma_e{}^2 = c_{Nugget} - (c_1 + c_2) \qquad 3\text{-}37$$

Recalling the ratio k between the variance of measurement error and the product variance without measurement error defined in the previous section, we can calculate k by:

$$k = \sqrt{\sigma_e{}^2 / (c_1 + c_2)}$$

Figure 3-2 shows the fitted and calculated covariance at the Beltsville Agriculture Research Center (BARC). Figure 3-3 shows the fitted spatiotemporal covariance at all validation sites. At most sites, correlation in spatial domain is stronger than the temporal domain, which means spatially adjacent pixels will have large influence on integrated results than temporally adjacent pixels.
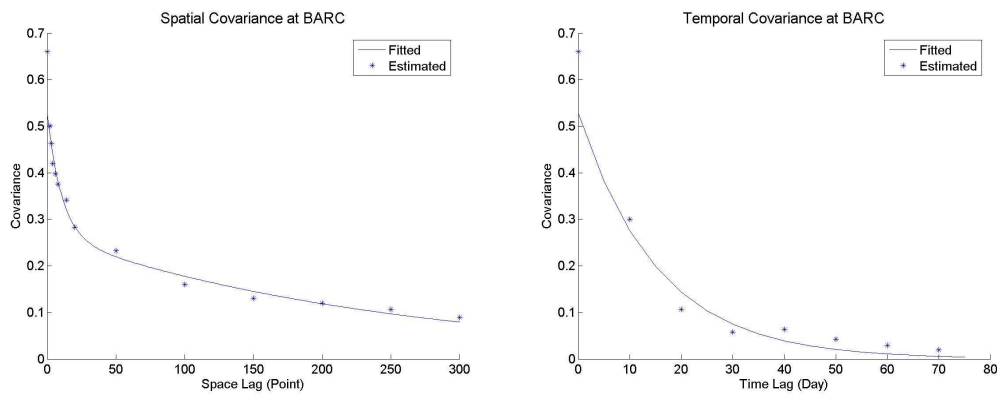

Figure 3-2 Estimated and fitted covariance at the Beltsville Agriculture Research Center. a) spatial covariance b) temporal covariance
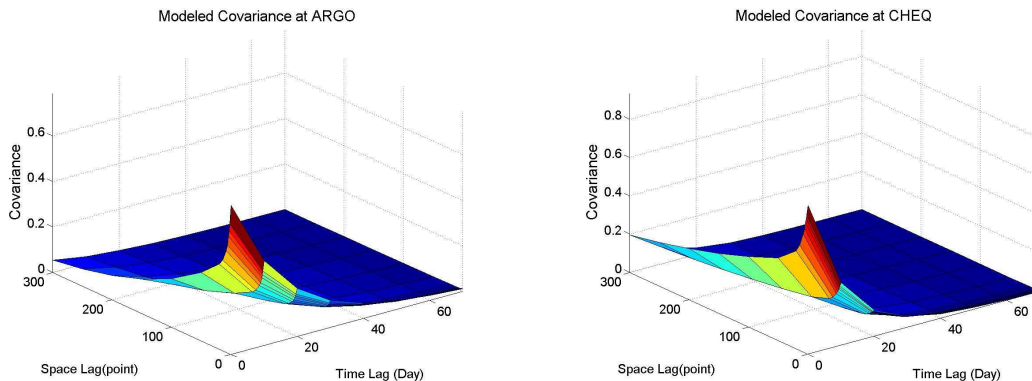
Modeled Covariance at HARV
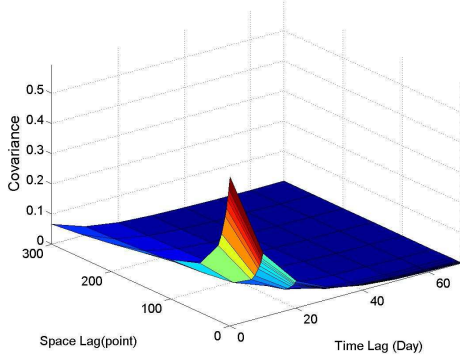
Modeled Covariance at Kejimikujik
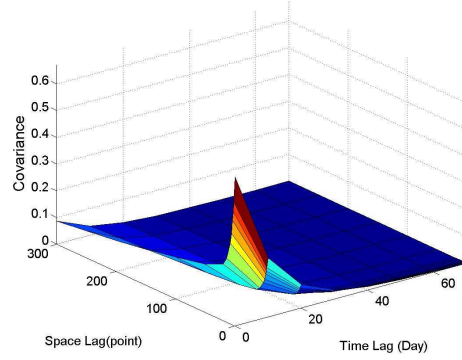
Modeled Covariance at KONZ
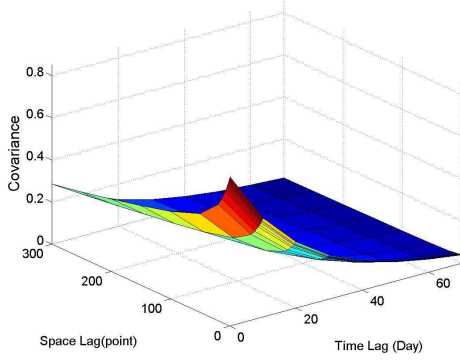
Modeled Covariance at Larose
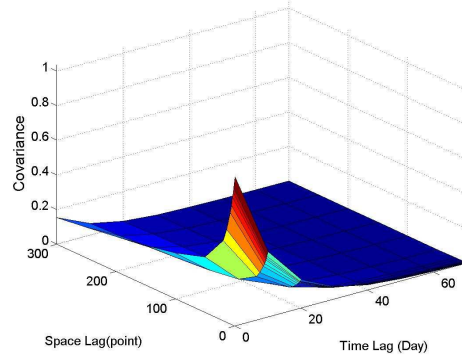
Modeled Covariance at METL
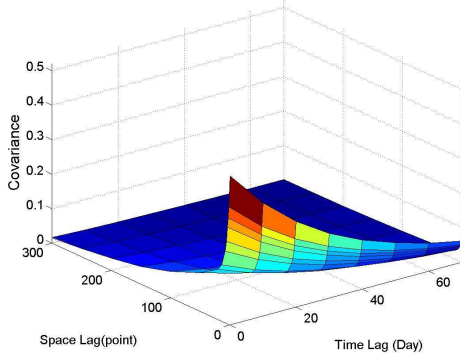
Modeled Covariance at NOBS
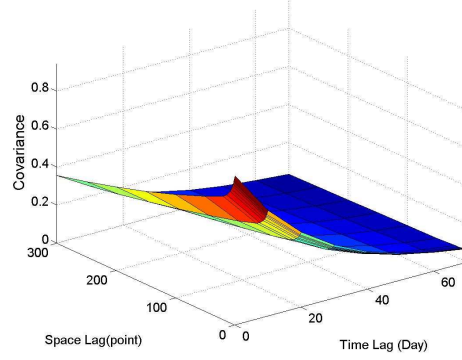
Modeled Covariance at SEVI
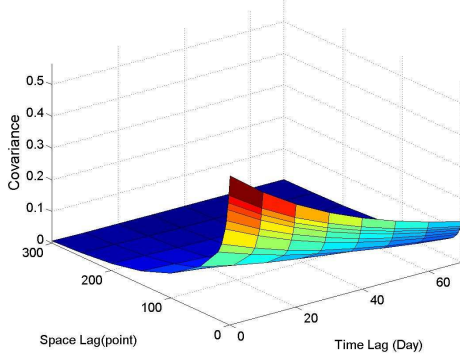
Modeled Covariance at Thompson
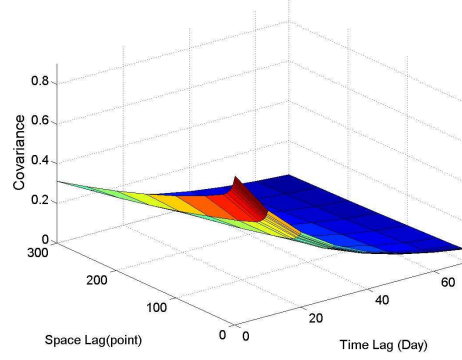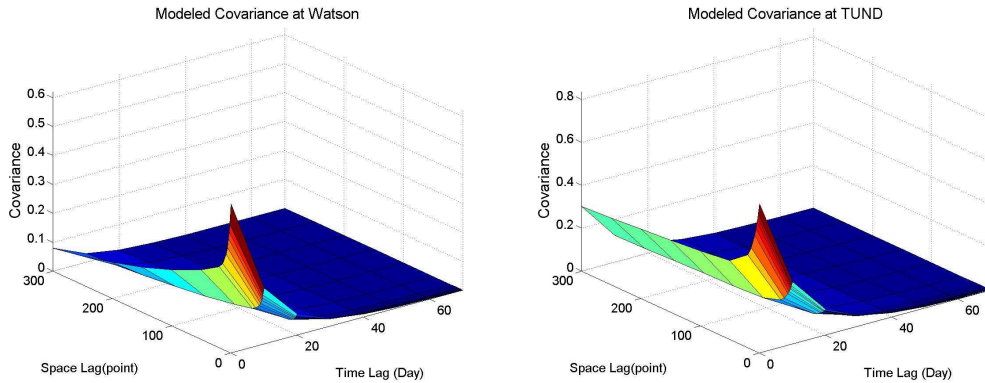
Figure 3-3 Modeled spatiotemporal covariance at all validation sites

3.3.2 Considering measurement error

OI and BME treat measurement errors in different ways. Measurement errors change the weight of the linear estimator through changing covariance between observations in OI. The measurement errors are represented by their variance. The measurement error in OI is usually assumed to follow normal distribution, so measurement error can be completely characterized by its variance and mean (always zero). BME has two ways of incorporating measurement error. One is to use the interval data and the other is to use the *pdf* of measurement error as the soft data. The *pdf* type soft data are chosen here. Given the variance, the *pdf* can be constructed by assuming measurement errors follow the normal distribution.

Variance of measurement error for CYCLOPES and MODIS data with the best quality is set as the value of $\sigma_e^{\,2}$, given in the previous section. The uncertainties of MODIS good data are estimated from the relative accuracy between best and good

data. The RMSE of Terra and Aqua good data is 4.25 and the RMSE of the best data is 2.56 (Figure 3-4), indicating the relative accuracy of good and best data. Accordingly, the variance of MODIS good data is $(1.7\sigma_e)^2$.

Some outliers may still exist, although ancillary data quality information is used to exclude data with large uncertainties. The data quality will be further controlled in preparing soft data. The data with absolute values larger than three times standard deviation will not be used as input.
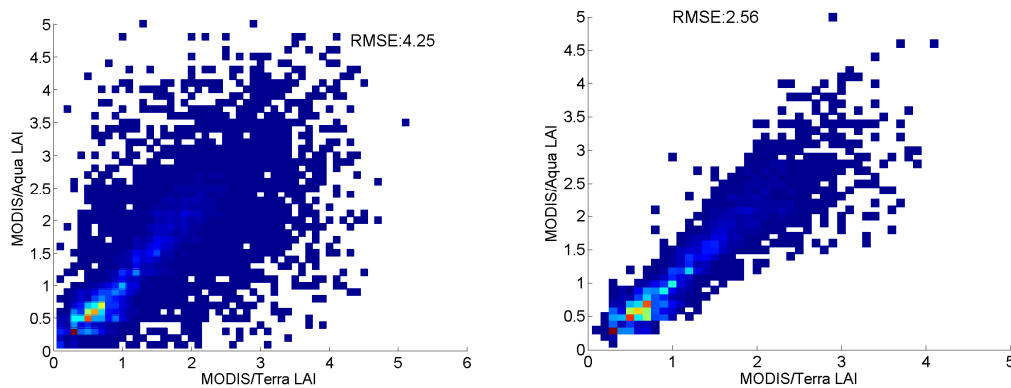
Figure 3-4 Scatter plots of MODIS LAI good data and best data.

3.3.3 Integration results

The proposed method is carried out at 12 reference sites. The integrated results are validated using reference maps (Figure 3-5). Both BME and OI improve the original data significantly. The integrated LAI has stronger correlation with reference LAI and the bias is reduced. The improvement of RMSE and R^2 from CYCLOPES LAI to the integrated one is slight, but very significant for MODIS LAI.
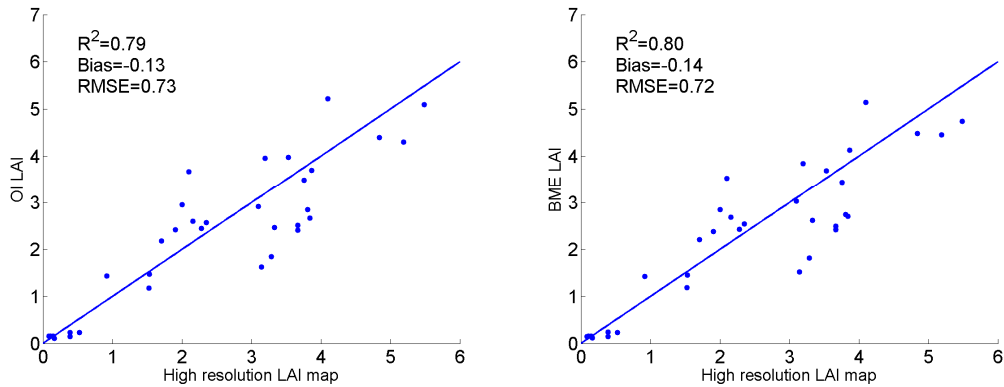
Figure 3-5 Validation results of BME and OI integrated LAI using high resolution
LAI reference maps at collected sites

However, the computation cost of BME and OI would be an obstacle for applying them to large datasets. Small neighbors are used to avoid this problem, since the pixels far away from the pixel to be estimated have little influence. This cost for BME is especially heavy, which needs multidimensional integral in addition to matrix inversion. Therefore, only OI will be used to show the results of geostatistical methods on time series and two dimensional images.

Figure 3-6 shows two examples of integrated results in time series. The time series at ARGO is very smooth since the original data has reliable data quality. Results at BARC are slightly worse. The integrated time series still has some unexpected fluctuations, but are significantly smoother than the original data. Although the two source data have apparent biases, the integrated results are systematically consistent. The integrated results are smaller than CYCLOPES and larger than MODIS,

mitigating their respective underestimation and overestimation problems. Integration

methods based on geostatistics can also generate estimation errors together with the

absolute estimated values. Figure 3-7 indicates the error bar at the ARGO site. In the

growth season, one has large errors because of the large uncertainties of original

retrieval. Only CYCLOPES data is produced in 1999 and MODIS/Aqua is not

available until 2002. The error of 2002 and 2003 is smaller than the previous years

because more observations are available from MODIS/Aqua.
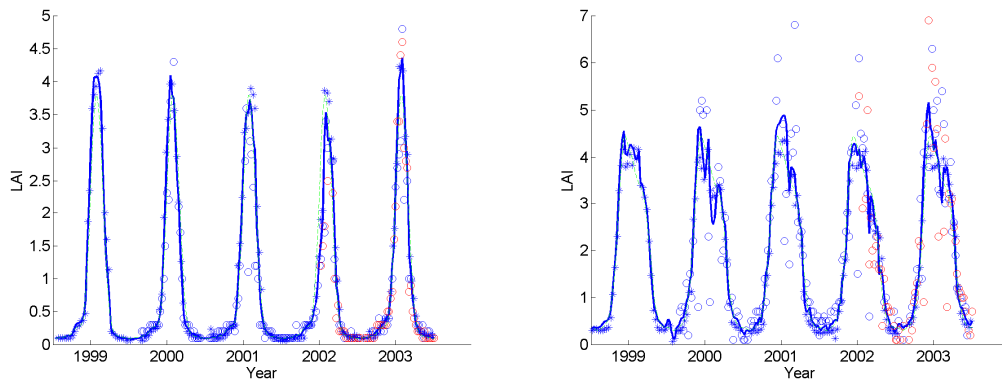
Figure 3-6 Five years' LAI time series at a) ARGO b)BARC. The solid lines:
integrated results by OI, dashed lines: climatology, stars: CYCLOPES, green circles:
MODIS/Terra, red circles: MODIS/Aqua.

To illustrate the approach at a regional scale, the OI based integration is also applied

to the greater Washington DC area. The high spatial resolution of ETM+ LAI on

August 5 2001 is estimated using the hybrid algorithm and calibrated with field

measurements (Fang and Liang 2005). Compared with ETM+ LAI, MODIS LAI has

two main problems. MODIS overestimates LAI, especially in woody area. Data gaps

pose the second problem. If zooming in the MODIS map, one can see a lot of gaps.

There are no gaps in the integrated map. The integration algorithm excludes most spurious high values; therefore, the histogram of the integrated map is closer to the histogram of the ETM+ map.
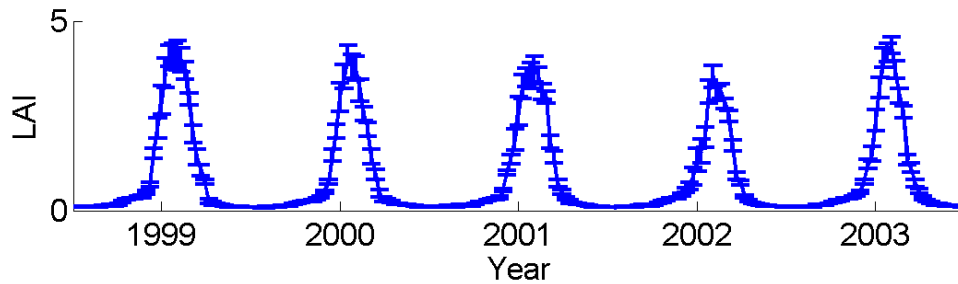
Figure 3-7 Integrated LAI and its error at ARGO

3.4 Summary

This chapter applies two geostatistical methods OI and BME to integrate MODIS and CYCLOPES LAI products. These two methods cannot directly operate on original LAI values because of their large discrepancies. I develop a framework to normalize anomaly and solve the problem that the two products have inconsistent bias and dynamic range. The anomaly can be normalized by assuming the linear relationship between the variance of measurement errors and the variance of true LAI. The integrated result has improved accuracy and no missing values. LAI time series is smoother and more reliable. It is hard to remove systematic error when there is no sufficient ground true data to evaluate the performance of retrieval algorithms. I make the integrated results consistent through the proposed algorithm.
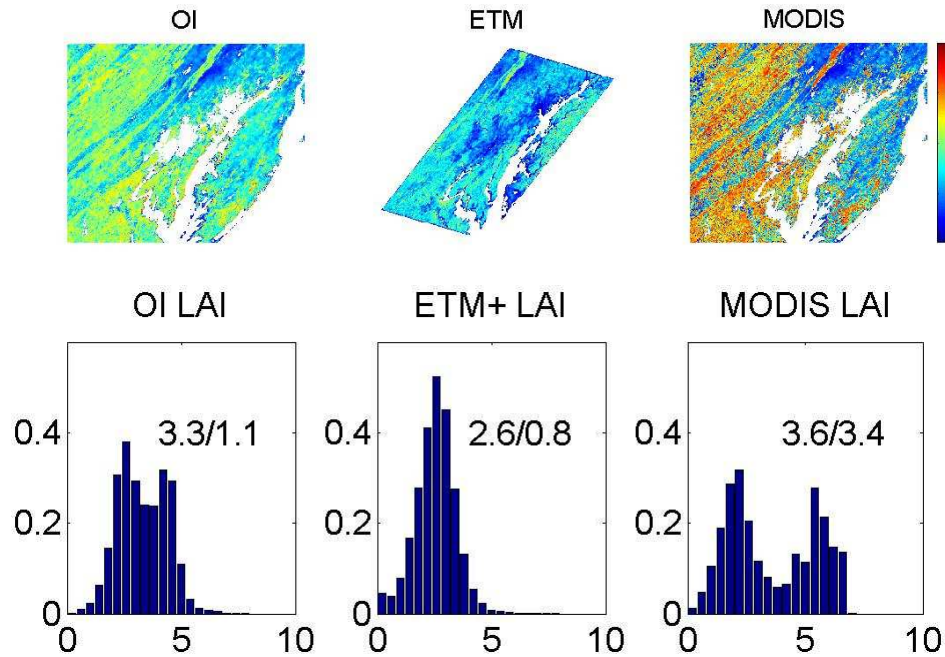
Figure 3-8 OI integrated result on a region near BARC on August 5 2001

Besides the statistically optimal integrated values, the geostatistical methods are also able to produce the estimation error. However, as mentioned earlier, one shortcoming of geostatistics based methods is their heavy computational cost. There are some solutions on improving the efficiency, for example: 1) using the Kalman filter on multiresolution tree (Fieguth *et al.* 1995; Huang *et al.* 2002; Johannesson *et al.* 2007) 2) introducing the predictive process model, which has a lower dimension (Banerjee et al. 2008) 3) approximating covariance matrix, through sparse matrix technique (Barry and Pace 1997), tapering (Furrer et al. 2006), spectral domain(Fuentes 2007) or wavelet basis functions (Nychka et al. 2002) 4) using the random effect model, expressing covariance matrix in the forms of limited basic function with much smaller dimension (Cressie and Johannesson 2008). I will show the results using the multiresolution efficient interpolation algorithm in the next chapter.

Chapter 4   Multiresolution tree method for integrating MODIS and MISR L3 LAI products

Optimal interpolation or BME cannot be applied on a large dataset due to the inefficient inversion of large covariance functions. One way to solve this problem is through a scale recursive filter of MRT based data structure. This method was first introduced by Chou (1991) and has been extensively used to interpolate and filter satellite altimetry data (Fieguth *et al.* 1995; Fieguth *et al.* 1998), temperature (Menemenlis *et al.* 1997), soil moisture (Parada and Liang 2004), aerosol (Huang et al. 2002), topography (Slatton *et al.* 2001) and so on. MRT not only carries out spatial interpolation efficiently but also is able to take measurements with different spatial resolutions. For example, de Vyver and Roulin (2009) used this approach to fuse two remotely sensed precipitation datasets with different spatial resolutions. In order to show its ability to integrate data with various spatial resolutions, MRT will be explored to integrate MISR L3 LAI and MODIS LAI in this chapter.

4.1 Methodology

The basic concept of multiresolution optimal interpolation is to organize data in a tree structure (Figure 4-1). Using similar notations with Luettgen (1993), the data between

different resolutions can be expressed in a linear state model:

$$x(s) = A(s)x(ps) + w(s) \qquad \text{4-1}$$

$x(s)$ and $x(ps)$ are the variable of interest respectively at scale s and its parent scale ps. $w(s)$ is the white noise, having a Gaussian distribution $N(0, Q(s))$. A is the state transition matrix from parent to children. Accordingly, the state transition matrix F from children to parent can be calculated by (Luettgen 1993):

$$F(s) = P(ps)A(s)P^{-1}(s) \qquad \text{4-2}$$

where P(s) is the variance at scale s. Thus, an up tree state equation can be obtained:

$$x(ps) = F(s)x(s) + w'(s) \qquad \text{4-3}$$

Besides these two state equations, there is an observation equation:

$$y(s) = Hx(s) + v(s) \qquad \text{4-4}$$

H is the observation matrix. Because the satellite product is the same as the variable of interest and the satellite data at the same location are used, H is taken as the identity matrix. $v(s)$ represents the measurement error $N(0, R(s))$.
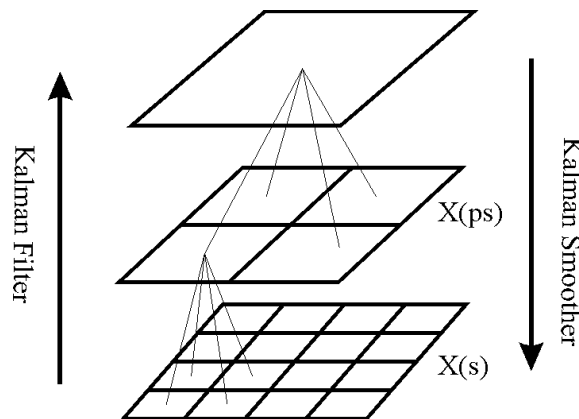
Figure 4-1 Framework of multiresolution tree method

This three equation system is similar to the state space model in time series, where the Kalman filter is proposed to incorporate the dynamic model and observational data. Similarly, the Kalman filter could also be applied on this tree-based model. From children to parents, the Kalman filter is employed to derive estimator at scale s1 $\hat{x}(s1 \mid s2)$ by incorporating observations up to scale s2. If there is no observation, the process evolves purely according to the state equation. When observations are available, the system is updated (Luettgen 1993):

$$\hat{x}(ps \mid ps) = \hat{x}(ps \mid s) + K(ps)(y(ps) - H\hat{x}(ps \mid s)) \qquad 4\text{-}5$$

where $K(s)$ is the Kalman gain and given by:

$$K(ps) = P(ps \mid s)HV^{-1}(ps) \qquad\qquad 4\text{-}6$$

where V(s) is the innovation covariance:

$$V(ps) = HP(ps \mid s)H^{T} + R(ps) \qquad 4\text{-}7$$

After the root of the tree is achieved by the upward loop, the Kalman smoother is applied from parent to children to incorporate the observations at all levels.

The critical step in using MRT is to assign the model parameters. Following the method used by Tzeng *et al.* (2005), I use identity matrix for the state transition matrix. The measurement error is dependent on sensor characteristics and the retrieval algorithm, which can be obtained through validation. However the acquisition of the state parameter is still a research topic. Huang *et al.*(2002) calculated the variance

parameter from the covariance function. Some use a $1/f^{\mu}$-like stochastic model, since many natural phenomena display a self-similar property (Fieguth *et al.* 1995; Fieguth and Willsky 1996; Fieguth *et al.* 1998). Kannan *et al.* (2000) used an expectation-maximization algorithm to estimate the parameters. de Vyver and Roulin (2009) directly calculated the parameters from the average radar measurement.

4.2 Data

In addition to having the high efficiency, MRT also has the advantage of integrating incompatible data with different spatial resolutions. MRT is applied here to integrate MODIS and MISR L3 LAI. Due to its narrow swath, MISR is less frequent to cover the entire Earth surface than MODIS. MISR LAI data are aggregated to 0.5 degree to generate L3 monthly data to enhance the surface coverage. MODIS Tile H10V5 is selected, most of which is over land. MISR L3 data is re-projected to the sinusoidal projection with the pixel size of 64 km. I use a tree with 6 levels and every node has sixteen children. MODIS LAI is the finest level and MISR takes the level 3. LAI anomaly instead of LAI is used as the input of MRT to achieve the goal of zero prior. Compared with the original values, the anomaly values of the two products match more closely (Figure 4-2). MISR L3 monthly LAI data has different temporal resolution from MODIS LAI. The temporal mismatching is simplified by assuming LAI anomaly doesn't change too much during the MISR averaging period. Day 217

2001 MODIS data and August 2001 MISR data is used.

From the direct validation in Section 2.5.2, the RMSE of MODIS LAI is 1.04. I use its square, 1.08, as the variance of the measurement error of MODIS data. MISR L3 has a resolution of 0.5 degree. The available high resolution LAI maps cannot cover the size of one MISR pixel. Instead, the aggregated MODIS LAI is used to validate MISR L3 LAI (Figure 4-3). The square of RMSE is used as the variance of MISR data.
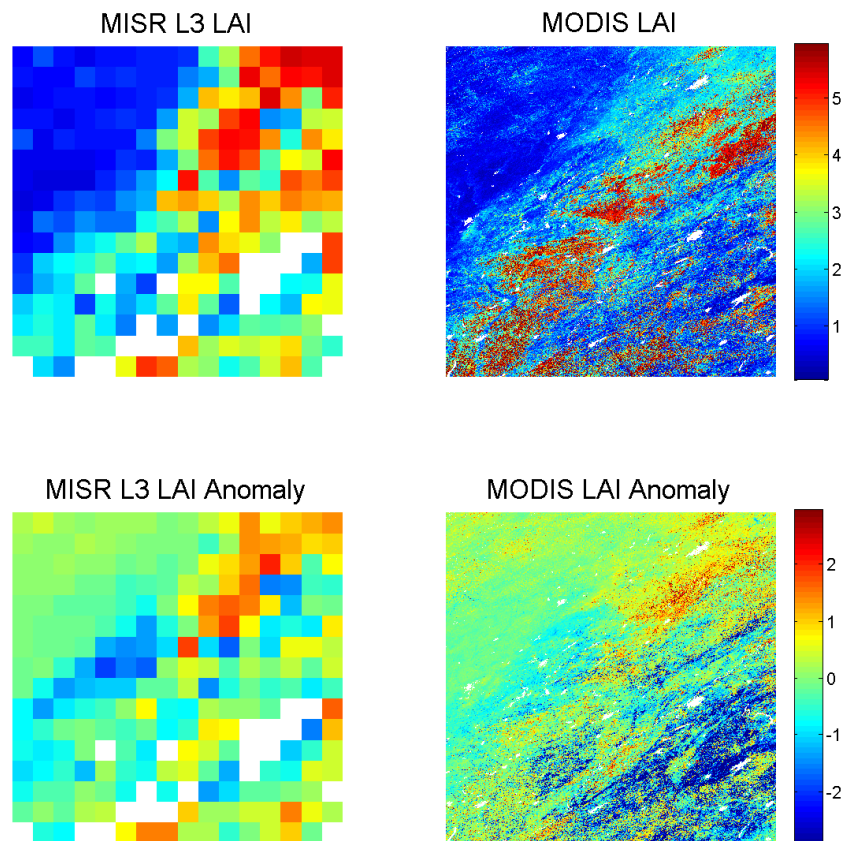
Figure 4-2 MISR, MODIS/Terra LAI and their anomaly of August 2001 at tile H10V5
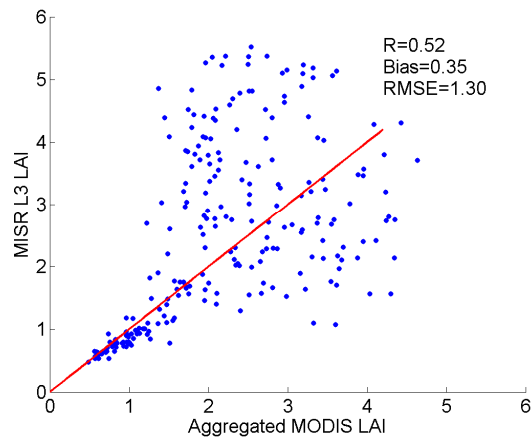
Figure 4-3 Validation results of MISR L3 LAI using aggregated MODIS LAI data

4.3 Results

In addition to the measurement error, another parameter is the variance when data are transferred from the parent to children. This parameter is empirically calculated using aggregated MODIS data at different levels (Table 4-1).

Table 4-1 Variance of LAI anomaly at different scales

Scale	1km	4km	16km	64km	256km
Variance	1.48	0.79	0.58	0.44	0.36

MRT integrated results are shown in Figure 4-4. The MRT integrated results are gap-free and smoother than the original anomaly map. The inconsistency among input data at different levels is mitigated through the data integration process. All the results

are consistent across the scales. The MRT errors are also shown, which depend mainly on the availability of input data and the accuracy of available data. Judging from MODIS QC, the left lower part of the image has lower data quality, and thus larger MRT error. However, the estimation errors are reduced through MRT, compared with the original measurement error.
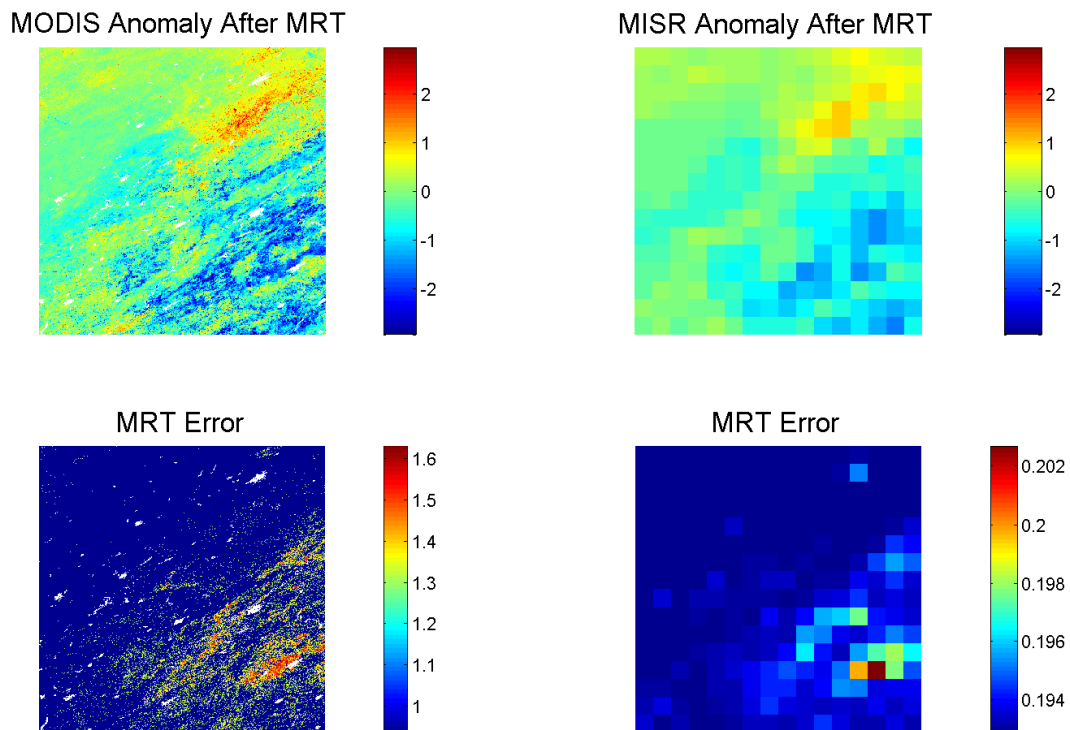
Figure 4-4 MRT integrated LAI anomaly and its error at MODIS and MISR scales

4.3.1 Comparison with OI

OI is also carried out over the same image to compare with the MRT method. The modeled spatial covariance is shown in Figure 4-5. Under the same computation facility (Pentium D 3.20GHz, 3.25Gb Memory), OI takes 4.5 hours when using an

11*11 small neighbor, while MRT needs only 6.5 seconds. MRT produces similar results with OI (Figure 4-6). Large differences appear in the left lower part of the map, where the original products have large errors.
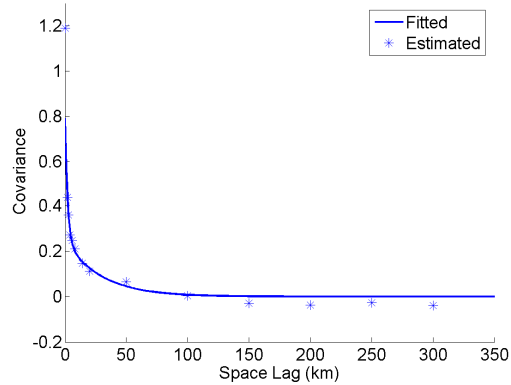
Figure 4-5 Modeled covariance at tile H10V5 using nested exponential covariance function
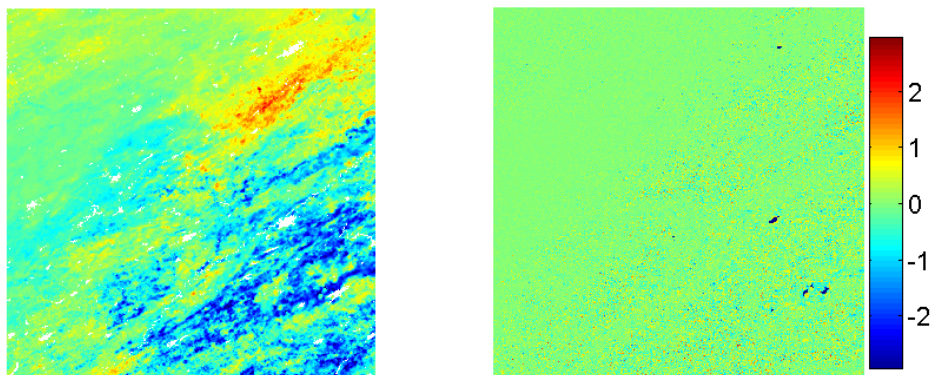
Figure 4-6 OI processed LAI anomaly and its difference with MRT results

Chapter 5   Empirical orthogonal function method for integrating MODIS and CYCLOPES LAI products

Empirical orthogonal function (EOF) methods deal with the matrix formed by the space-time variable, solving the eigenvector problem of data's covariance matrix to reduce data dimension and remove measurement noise. A new data integration method based on Kondrashov and Ghil (2006)'s iterative algorithm is developed in this chapter. Then this method is applied to integrate MODIS and CYCLOPES LAI products.

5.1 Introduction

EOF is one of the most extensively used methods in geosciences (Preisendorfer 1988; Hannachi *et al.* 2007). EOF sometimes is also called Principal Components Analysis (PCA). Singular Spectrum Analysis (SSA) also belongs to the EOF family, but deals with the temporal correlations of short and noisy time series (Vautard and Ghil 1989). Multi-channel SSA (MSSA) may handle multivariate time series and these different channels can be the same variable at different spatial locations, so MSSA has the ability of processing both temporal and spatial information. Similarly, an extended version of EOF deals with both spatial and temporal correlation by using a sliding

window to incorporate the lagged information in the time domain (Weare and Nasstrom 1982). In the atmospheric literature, this technique is called extended EOF (EEOF), which can be treated as the synonym of MSSA.

Through a regression and using information on spatial EOF of existing data, Smith *et al.* (1996) predicted data for places with no observations. Kondrashov and Ghil (2006) introduced an iterative algorithm based on SSA to fill missing data points and test it in multiple geophysical datasets. Liu *et al.* (2005) fit the *in situ* aerosol measurement with the leading EOF of satellite retrievals and model simulation in order to merge the information from three sources. Zhang *et al.*(2007) used EOF to improve poor quality insolation data to generate homogenous maps. Beckers and Rixen (2003) developed a "self-consistent" and "parameter free" Data Interpolating Empirical Orthogonal Functions (DINEOF). Alvera-Azcarate *et al.* (2005) and Alvera-Azcarate *et al.*(2007) applied this method in real oceanographic data for univariate and multivariate cases respectively. Ding *et al.* (2009) improved DINEOF by considering the influence of extreme values.

5.2 Methodology

The basic theories of all EOF based methods are similar: reconstruct noisy and gap-prone data using the leading components of data. Given a N*P matrix **L** formed

by the LAI retrievals $L_{t,s}$ at time t and location s, $t \in [1, N]$, $s \in [1, P]$, there are two ways to compute the EOFs of $\mathbf{L}$. One is to solve the eigenvalue problem of covariance matrix $\mathbf{C_L}$:

$$\mathbf{C_L} = \mathbf{L}^T \mathbf{L} \qquad 5\text{-}1$$

The other way is to apply singular value decomposition directly on matrix $\mathbf{L}$:

$$\mathbf{L} = \mathbf{U} \mathbf{S} \mathbf{V}^T \qquad 5\text{-}2$$

The diagonal matrix $\mathbf{S}$ contains the singular values sorted in descending orders. The left singular vectors are usually called EOFs (spatial domain), while the right singular vectors are named principal components (PCs, temporal domain). By this means, the original data field is divided into spatial and temporal components. The original data field could be reconstructed if all the EOFs and PCs are used. The leading components contain most of the information through which the patterns of complex spatiotemporal geophysical datasets can be analyzed and the original noisy data can be filtered. This is the basic idea of EOF methods. Instead of working directly on the original matrix $\mathbf{L}$, EOF could also be extended if the window of the length W is moved on the original matrix to emphasize temporal information. A new matrix $\mathbf{L}'$ with the dimension $(N - W + 1) \times WP$ can be formed by incorporating the lagged information (Hannachi et $al.$ 2007):

$$\mathbf{L}' = \begin{pmatrix} L_{1,1} & L_{2,1} & \cdots & L_{W,1} & L_{1,2} & \cdots & L_{W,P} \\ L_{2,1} & L_{3,1} & \cdots & L_{W+1,1} & L_{2,2} & \cdots & L_{W+1,P} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ L_{N-W+1,1} & L_{N-W+2,1} & \cdots & L_{N,1} & L_{N-W+1,2} & \cdots & L_{N,P} \end{pmatrix} \qquad 5\text{-}3$$

The temporal information is emphasized this way. This method has the name of EEOF or MSSA.

Usually, iterative algorithms are proposed when applying EOF methods to fill data gaps (Schoellhamer 2001; Beckers and Rixen 2003; Kondrashov and Ghil 2006; Zhang *et al.* 2007). Schoellhamer (2001) calculated the covariance using only existing data. This may lead to a non-positively definite matrix (Beckers and Rixen 2003). Therefore, the missing points are usually filled with mean before applying iteration. Zhang *et al.* (2007) used a simple single loop iteration. In each iteration, the same number of EOFs is used. Other authors (Beckers and Rixen 2003; Kondrashov and Ghil 2006) used a double loop. In the outer loop, the number of EOFs is increased from 1 to a preset number. The inner loop is similar with Zhang *et al*'s method where EOF decomposition and reconstruction are computed until convergence happens. I will use the latter strategy.

5.2.1 Hierarchical EOF

Two problems have to be solved before EOF can be used to integrate multiple LAI products. The first problem is computational cost and the second is that the two products to be integrated have different temporal and spatial resolutions. I propose to use two runs of Hierarchical EOF (HEOF) to solve these problems.

The usual solution for large datasets is to divide them into multiple small sub-datasets and apply EOF on each of them. However, the information from other parts cannot be used this way and there will be inconsistent edges between different sub-datasets. To overcome these problems, a hierarchical EOF algorithm is proposed. EOF is applied at two levels: coarse resolution aggregated data and multiple fine resolution sub-datasets, both of which have a small dimension feasible for EOF. The data are first aggregated to coarse resolution. EOF is then carried out on coarse resolution data. The poor quality block will be replaced during the EOF procedure. The improved coarse resolution data are used as the mean to center the fine resolution subsets. For the fine resolution subsets, two adjacent blocks are intentionally overlapped. After EOF is applied, the mean of the overlapping area is calculated as the final filtered values to reduce the "blocky effect" (See Figure 5-1).
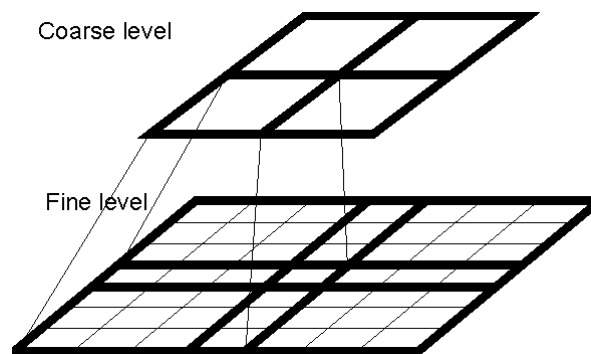
Figure 5-1 Two level hierarchical data structure for EOF. Overlapping is intentionally designed to reduce "blocky effect".

For each level, an iterative method revised from Kondrashov and Ghil (2006)'s method is used. Kondrashov and Ghil (2006) centered the original data by subtracting

the mean and filled the missing points with zero. My approach applies the same strategy at the coarse resolution, while the reconstructed values of coarse resolution are used to center the corresponding fine resolution data. Then, a double-loop iteration of EOF is applied on the filled matrix. The inner loop replaces the missing points using the leading components of the filled matrix to form a new matrix, on which new leading components are computed to repeat the loop until the results converge. The outer loop increases the number of leading components used for reconstruction from 1 to a predefined value and performs the inner loop. In Kondrashov and Ghil (2006)'s algorithm, the original data are treated as the true data and kept intact. In the case of LAI, some existing observations may contain large uncertainties. I revise Kondrashov and Ghil's method by replacing these data as well with the reconstructed values to filter out the noise and reduce uncertainties in the last set of inner loop.

5.2.2 Multivariate HEOF

DINEOF is able to take more than one variables as inputs. Alvera-Azcarate *et al.* (2007) tried this to reconstruct the combination of sea surface temperature, chlorophyll a and wind data. Their validation results show combining other information improved the data quality of reconstructed SST. In the case of integrating LAI, multiple LAI products cannot be directly used as different variables of

multivariate EOF. Although EOF doesn't require all the variables to bear the same spatial location, their temporal resolution must be identical in order to form one matrix. In the case of integrating MODIS and CYCLOPES LAI, this prerequisite condition cannot be satisfied. Two runs of EOF are proposed to mitigate this problem.

In the first run, only MODIS LAI is filtered by EOF. Then, MODIS LAI is temporally smoothed to match with CYCLOPES's temporal knots using:

$$\widehat{L}(t_C) = \sum_{i=1}^{N} L(t_i) \frac{f(t_i)}{\sum_{j=1}^{N} f(t_j)} \qquad 5\text{-}4$$

$L(t_i)$ is the MODIS data after EOF. I choose N=4 and $f(t)$ to be Gaussian function, because similar parameters are used by Baret *et al.* (2007) to smooth VEGEATAION reflectance, which is ultimately used to generate the CYCLOPES LAI product. Then multivariate EOF is carried out on the smoothed MODIS data and CYCLOPES data. Before running multivariate EOF, two LAI anomalies will be normalized using their standard deviations over the whole dataset. After the multivariate EOF is done, the output MODIS and CYCLOPES data are averaged to obtain the final result.

5.3 Analysis of results

In the proposed EOF method, two parameters, the number of leading components and window length, need to be optimized. Zhang *et al.* (2007) used an empirical method,

simply choosing the EOF which could represent 80% of total variance. The explained

variance of LAI anomaly is shown in Figure 5-2. A total of 5 leading EOF modes

contain 80% of LAI covariance. In addition to this subjective method, cross

validation may also be used to derive the parameters. Kondrashov and Ghil (2006)

used cross-validation to determine the optimum parameters. The satellite LAI product

cannot be directly used to carry out the cross validation because the original product

contains large errors. Instead, the OI interpolated LAI anomaly is used as the true

value to carry out cross validation. OI is applied on a 40*40 window around BARC

for a whole year. EOF with varied parameters is used to reconstruct the same dataset.

The relative errors of the reconstructed datasets are then calculated to determine the

optimum parameters.
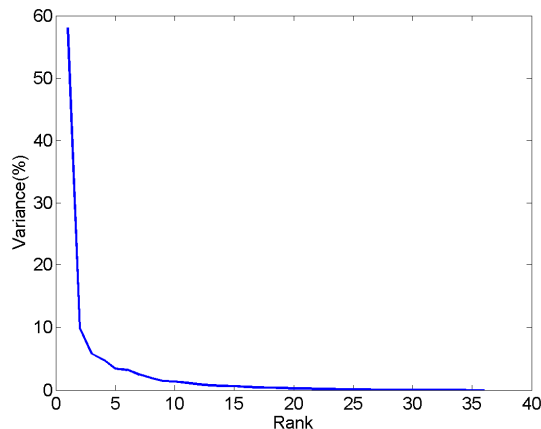
Figure 5-2 Variance explained by EOF modes of one year CYCLPOES LAI anomaly
data.

The results of cross validation are shown in Figure 5-3. When the window length is 1,

which means no temporal information is used and EEOF is simplified to simple EOF,

the relative error is very high. When the window length is increased to 2, significant improvement of the results can be noticed. After that, the change of window length has little influence on the relative errors. The estimation error first decreases with the increase in the number of EOF modes used for the reconstruction. However, continuous increase of EOFs leads to a large error of LAI estimation due to the large uncertainties of the satellite LAI products. The optimal number depends on the window length. The dimension of the working matrix $(N-W+1)\times WP$ is proportional to the window size. I set window size to 2 to reduce the computational time.
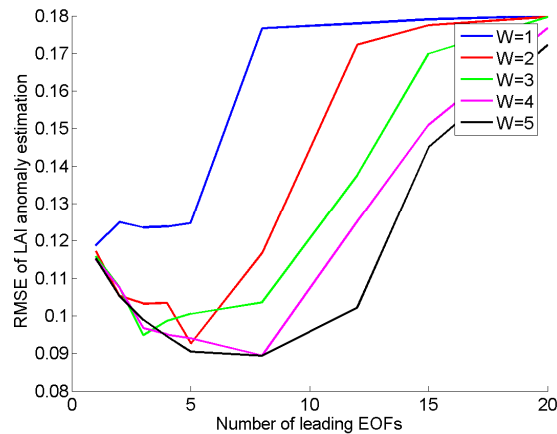
Figure 5-3 Relative errors of cross validation using CYCLOPES LAI anomaly, as the function of window length W and the number of leading components.

The proposed method is carried out at 12 reference sites. The integrated results are validated using reference maps (Figure 5-4). The quality of MODIS has been improved significantly in terms of R^2 and RMSE, while the reconstructed CYCLOPES data is slightly worse than the original ones. Nevertheless, the final

integrated results have much higher quality than the original products. The estimation

bias is significantly reduced by combing two data sources.
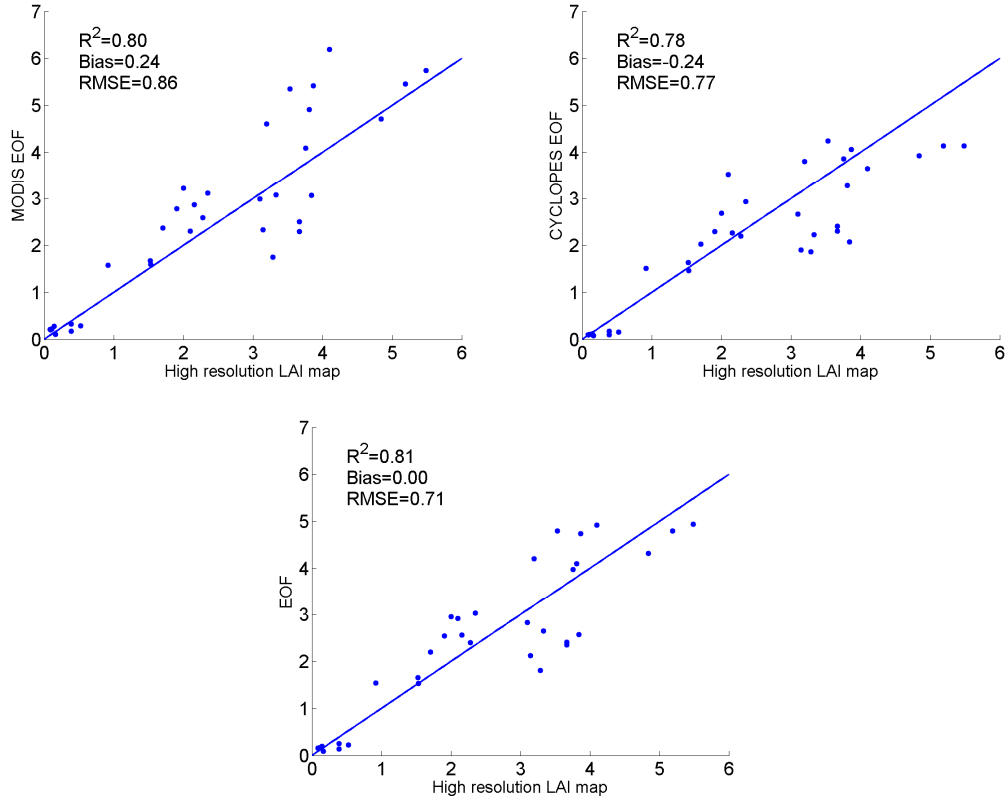
Figure 5-4 Direct validation of EOF methods at collected sites

The proposed method is also carried out at a region with 390*240 km around BARC.

For one year data, EOF cannot directly operate on such a large image under a 32 bit

Matlab® environmental due to the limitation of memory. By hierarchical EOF, one

can work on any size datasets. Figure 5-5 shows the results of three consecutive maps.

The missing data in the left upper corner of CYCLOPES data at Day 227 2001 are

filled with information from existing data. The EOF's ability of gap filling is further

illustrated in Figure 5-6. There are no MODIS data available in Day 169 and Day 177

2001. EOF generates continuous maps through data covariance calculated from existing data. The reconstructed MODIS LAI anomaly maps at missing days show similar patterns with the existing maps. The EOF reconstructed results appear smoother than the original one. Actually, the EOF integrated data also appear smoother in temporal domain (Figure 5-7). The spurious data values are removed through EOF. Thus the EOF integrated results show higher accuracy.
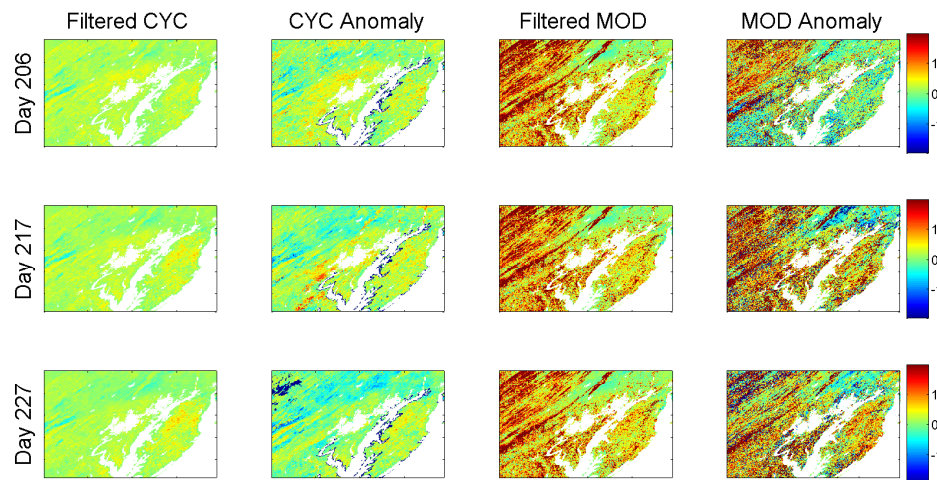
Figure 5-5 EOF results on three consecutive maps from July 25 to August 15 2001 around BARC. Each row shows the data at one day. The four columns respectively are the filtered and original CYCLOPES anomaly, the filtered and original MODIS anomaly.
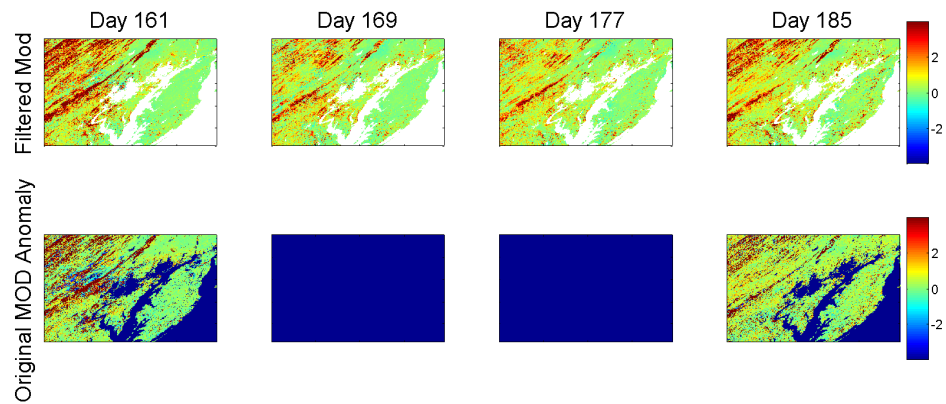
Figure 5-6 Four consecutive maps of MODIS LAI anomaly after and before EOF. Blue area means missing data.
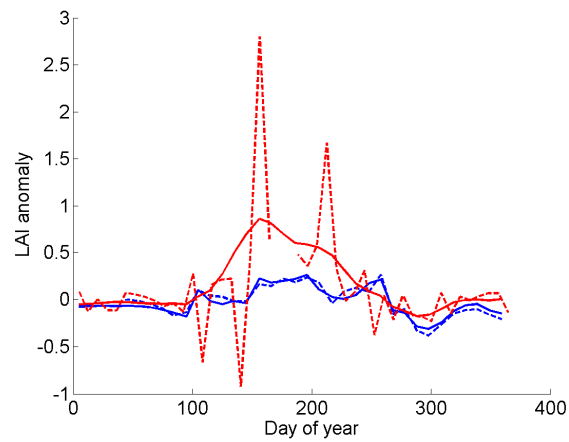
Figure 5-7 One year time series of LAI anomaly at one selected point. Blue lines mean CYCLOPES data and red ones represent MODIS data. The dashed lines are original anomalies and the solid lines are reconstructed results by EOF

Chapter 6   Conclusions

Accurate LAI products with high temporal and of different spatial resolutions are greatly needed to support climatological and ecological research efforts. None of the existing LAI datasets satisfies these requirements. Data assimilation may be a choice to address this issue. Nevertheless, independent observational data without incorporating physical dynamic models are essential to drive and validate all kinds of physical models. However, investigations on integration of high quality LAI from multiple existing LAI products are still not well established. This dissertation examined four methods based on OI, BME, MRT and EOF to integrate multiple LAI products and improve their accuracy.

6.1 Major findings

When validating and comparing MODIS and CYCLOPES LAI products that was the necessary step in the data integration analysis, I found that MODIS and CYCLOPES LAI have the relative geometric accuracy of around one pixel. Validation using LAI reference maps shows CYCLOPES underestimates LAI with a bias of -0.2 while MODIS overestimates LAI with a positive bias of 0.3. CYCLOPES may produce spurious large errors in winter. The direct comparison of the two datasets at pixel

level shows very large scattering. Multiple years' mean over each land cover type shows the two products have similar values over shrub, grass, cereal crop and needleleaf tree but great discrepancy on broadleaf tree, sparsely vegetated area and broadleaf crop. Besides these systematic biases, atmospheric condition affects the retrieval quality of LAI. Cloud contamination produces non-Gaussian negative error and cloud shadow may also seriously degrade data quality.

All of these four methods can fill gaps and reduce errors with existing data. After integration, all the data gaps are filled with information from adjacent pixels and prior knowledge. The integration process removes the spurious large temporal and spatial variation over the original data. Validation results indicate that the combination of two data sources reduces the bias and random error. The EOF method produces the integrated results with zero bias. The geostatistical methods reduce bias from +0.3 (MODIS) or -0.2 (CYCLOPES) to -0.1. The data quality has been improved through data integration. The integrated results' improvement from MODIS product is significant: R^2 increases from 0.75 to 0.8 and RMSE decreases from 1.0 to 0.7. Limited *in situ* measurements hardly prove which methods outperform the others. However, the four methods do have their own pros and cons.

Two geostatistics methods (both "traditional" OI and modern BME) have a solid theoretical basis. They generate statistically optimal results by incorporating both the

dependency among data and the data errors. In theory, BME will be the best choice for the integration of multiple data, since it doesn't assume the Gaussian process and linear summation of data. If the data obey the prerequisite conditions of the Gauss-Markov theory, OI is the best unbiased estimator. However, the computational cost of applying BME and OI would be an obstacle for handling large datasets. The dataset for the North America has the magnitude of 10^8. Inversion of the covariance matrix with such a large dimension is still impractical. This is even worse for BME which needs multidimensional integral in addition to matrix inversion.

MRT outperforms BME and OI in terms of algorithm efficiency. Besides, MRT also has the ability of integrating data with different spatial resolutions and generating consistent results at different resolutions. However, MRT usually assumes a simple linear state transfer function from coarse to fine resolution and vice versa and it is hard to verify if the real data satisfy such an equation.

A hierarchical two-run EOF is proposed to handle large datasets and integrate data with different temporal knots. The new algorithm divides a large dataset into a set of overlapping small subsets. The information from different subsets is used at the aggregated coarse resolution. Compared with the aforementioned three methods, EOF requires less model assumption and parameter identification. However, EOF lacks a mechanism to explicitly consider the measurement error.

6.2 Major contributions

This is the first study to apply several different integration methods to combine multiple satellite LAI products to reduce uncertainties and improve integrality. Existing methods (e.g., temporal filtering) usually work on one single product. Spatial temporal statistical based methods are hardly explored in previous investigations. Although BME, OI and MRT based methods have been used in other fields, this is the first time they have been applied to the integration of multiple LAI products.

This dissertation has enhanced the locally adjusted cubic-spline capping method by revising the end condition. The proposed periodic end condition is more reasonable to filter multiple years' LAI mean and variance.

This dissertation presents a new comprehensive comparison of MODIS C5 LAI product with other satellite products. Besides examining the LAI values, this work has also evaluated the relative geometric accuracy of MODIS and CYCLOPES LAI products. By taking advantage of the MODIS twin sensors, this dissertation has quantified the influence of atmospheric conditions on data quality.

In order to account for the inconsistency between products, a new LAI normalization scheme is developed by assuming the linear relationship between measurement error

and LAI natural variance. Although this scheme cannot remove systematic bias, it is able to generate LAI results with consistent bias.

A new data integration method based on Kondrashov and Ghil (2006)'s iterative EOF algorithm is developed. The new algorithm has the ability of handling large datasets by working on two different levels and integrating multiple data with different temporal resolutions. EOF is carried out on both coarse resolution aggregated data and fine resolution small sub-datasets. Results from coarse resolution are used as prior knowledge in filtering the fine resolution data. This approach is first run as a univariate case on MODIS data only. The filtered MODIS data is interpolated to match with CYCLOPES by a Gaussian function. Then, EOF is run as a multivariate case on MODIS and CYCLOPES. The intermediate data are averaged to obtain the final integration results. Besides filling gaps, validation shows this method significantly improves the data accuracy, reducing the bias to 0, improving R^2 to 0.81 and reducing RMSE to 0.71.

6.3 Suggestions for future study

This dissertation presents the initial efforts to use four different methods to integrate LAI products. Several issues need to be addressed in future studies. First of all, additional *in situ* measurements are required to evaluate existing LAI products and

correct the systematic bias with each product. The spatiotemporal correlation of LAI

and the distribution of its measurement error will be the basis of various data

integration methods. Comprehensive field measurements provide a way to better

characterize them. Furthermore, *in situ* measurements are also needed to validate the

integrated results and refine the developed algorithms.

A stationary spatiotemporal covariance function is currently used to model the

dependency of LAI anomaly data. More flexible models need to be developed to

account for the heterogeneity of LAI. Geostatistical methods are promising because

they have a strict theoretical basis. However, more research efforts are needed to

reduce their computational cost and make them suitable for large datasets.

Two products are integrated in each of the four methods used in this dissertation,

although these methods are able to employ more than two products. Future research

could focus on more LAI products. Attempts could also be made to incorporate

variables other than LAI to improve LAI estimation through multivariate methods,

such as Co-Kriging, multivariate BME or multivariate EOF.

In addition to LAI, other satellite products, such as albedo and insolation, suffer from

similar problems. For example, geostationary satellites have more observations but

are unable to map polar areas, while polar orbit satellites observe there frequently.

None of them can map the whole Earth at dense temporal intervals. The presented methods could be applied to integrate the geostationary GOES insolation product with polar orbit MODIS insolation information to solve this problem.

Reference

Alavi, N., Warland, J. S. and Berg, A. A. (2006). "Filling gaps in evapotranspiration measurements for water budget studies: Evaluation of a Kalman filtering approach." Agricultural and Forest Meteorology **141**(1): 57-66.

Alvera-Azcarate, A., Barth, A., Rixen, M. and Beckers, J. M. (2005). "Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature." Ocean Modelling **9**(4): 325-346.

Alvera-Azcarate, A., Barth, A., Beckers, J. M. and Weisberg, R. H. (2007). "Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields." Journal of Geophysical Research-Oceans **112**(C3).

Bacour, C., Baret, F. and Derive, G. (2003). CYCLOPES algorithmic development for estimating biophysical products from large swath sensors. Proceedings of IEEE International Geosciene and Remote Sensing Symposium 2003: 3902-3904

Badeck, F. W., Bondeau, A., Bottcher, K., Doktor, D., Lucht, W., Schaber, J. and Sitch, S. (2004). "Responses of spring phenology to climate change." New Phytologist **162**(2): 295-309.

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). "Stationary process approximation for the analysis of large spatial datasets." Journal of the Royal Statistical Society Series B-Statistical Methodology **70**: 825-848.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Nino, F., Weiss, M., Samain, O., Roujean, J. L. and Leroy, M. (2007). "LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION - Part 1: Principles of the algorithm." Remote Sensing of Environment **110**(3): 275-286.

Barry, R. P. and Pace, R. K. (1997). "Kriging with large data sets using sparse matrix techniques." Communications in Statistics-Simulation and Computation **26**(2): 619-629.

Beckers, J. M. and Rixen, M. (2003). "EOF calculations and data filling from incomplete oceanographic datasets." <u>Journal of Atmospheric and Oceanic Technology</u> **20**(12): 1839-1856.

Berterretche, M., Hudak, A. T., Cohen, W. B., Maiersperger, T. K., Gower, S. T. and Dungan, J. (2005). "Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest." <u>Remote Sensing of Environment</u> **96**(1): 49-61.

Bicheron, P., Leroy, M. and Hautecoeur, O. (1998). <u>LAI and fAPAR mapping at global scale by model inversion againstspaceborne POLDER data</u>. Proceedings of IEEE International Geoscience and Remote Sensing Symposium 1998: 1228-1230.

Borak, J. S. and Jasinski, M. F. (2009). "Effective interpolation of incomplete satellite-derived leaf-area index time series for the continental United States." <u>Agricultural and Forest Meteorology</u> **149**(2): 320-332.

Bretherton, F., Davis, R. and Fandry, C. (1976). "A technique for objective analysis and design of oceanographic experiments applied to MODE-73." <u>Deep-Sea Research</u> **23**: 559-582.

Brut, A., R¨udiger, C., Lafont, S., Roujean, J.-L., Calvet, J.-C., Jarlan, L., Gibelin, A.-L., Albergel, C., Moigne, P. L., Soussana, J.-F., Klumpp, K., Guyon, D., Wigneron, J.-P. and Ceschia, E. (2009). "Modelling LAI at a regional scale with ISBA-A-gs: comparison with satellite-derived LAI over southwestern France." <u>Biogeosciences,</u> **6**: 1389-1404.

Buermann, W., Wang, Y. J., Dong, J. R., Zhou, L. M., Zeng, X. B., Dickinson, R. E., Potter, C. S. and Myneni, R. B. (2002). "Analysis of a multiyear global vegetation leaf area index data set." <u>Journal of Geophysical Research-Atmospheres</u> **107**(D22).

Burrows, S. N., Gower, S. T., Clayton, M. K., Mackay, D. S., Ahl, D. E., Norman, J. M. and Diak, G. (2002). "Application of geostatistics to characterize leaf area index (LAI) from flux tower to landscape scales using a cyclic sampling design." <u>Ecosystems</u> **5**(7): 667-679.

Chao, Y., Li, Z. J., Farrara, J. D. and Hung, P. (2009). "Blending Sea Surface Temperatures from Multiple Satellites and In Situ Observations for Coastal

Oceans." <u>Journal of Atmospheric and Oceanic Technology</u> **26**(7): 1415-1426.

Chen, J., Jonsson, P., Tamura, M., Gu, Z. H., Matsushita, B. and Eklundh, L. (2004). "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter." <u>Remote Sensing of Environment</u> **91**(3-4): 332-344.

Chen, J. M. and Black, T. A. (1992). "Defining Leaf area index for non-flat leaves." <u>Plant Cell and Environment</u> **15**(4): 421-429.

Chen, J. M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S. G., White, H. P., Hall, R. J., Peddle, D. R., King, D. J., Trofymow, J. A., Swift, E., Van der Sanden, J. and Pellikka, P. K. E. (2002). "Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements." <u>Remote Sensing of Environment</u> **80**(1): 165-184.

Chen, J. M., Deng, F. and Chen, M. Z. (2006). "Locally adjusted cubic-spline capping for reconstructing seasonal trajectories of a satellite-derived surface parameter." <u>IEEE Transactions on Geoscience and Remote Sensing</u> **44**(8): 2230-2238.

Chen, M. Y., Shi, W., Xie, P. P., Silva, V. B. S., Kousky, V. E., Higgins, R. W. and Janowiak, J. E. (2008). "Assessing objective techniques for gauge-based analyses of global daily precipitation." <u>Journal of Geophysical Research-Atmospheres</u> **113**(D4).

Chou, K. (1991). A stochastic modeling approach to multiscale signal processing. <u>Dept. EECS</u>, MIT. **Ph.D.**

Christakos, G. (1990). "A Bayesian maximum entropy view to the spatial estimation problem." <u>Mathematical Geology</u> **22**(7): 763-777.

Christakos, G. (2000). <u>Modern Spatiotemporal Geostatistics</u>. New York, Oxford University Press.

Christakos, G., Kolovos, A., Serre, M. L. and Vukovich, F. (2004). "Total ozone mapping by integrating databases from remote sensing instruments and empirical models." <u>IEEE Transactions on Geoscience and Remote Sensing</u> **42**(5): 991-1008.

Cohen, W. B., Maiersperger, T. K., Yang, Z. Q., Gower, S. T., Turner, D. P., Ritts, W. D., Berterretche, M. and Running, S. W. (2003). "Comparisons of land cover and LAI estimates derived from ETM plus and MODIS for four sites in North America: a quality assessment of 2000/2001 provisional MODIS products." Remote Sensing of Environment **88**(3): 233-255.

Cohen, W. B., Maiersperger, T. K., Turner, D. P., Ritts, W. D., Pflugmacher, D., Kennedy, R. E., Kirschbaum, A., Running, S. W., Costa, M. and Gower, S. T. (2006). "MODIS land cover and LAI collection 4 product quality across nine sites in the western hemisphere." IEEE Transactions on Geoscience and Remote Sensing **44**(7): 1843-1857.

Cressie, N. and Johannesson, G. (2008). "Fixed rank kriging for very large spatial data sets." Journal of the Royal Statistical Society Series B-Statistical Methodology **70**: 209-226.

Cressman, G. P. (1959). "An operational objective analysis system." Monthly Weather Review **87**(10): 367-374.

de Vyver, H. V. and Roulin, E. (2009). "Scale-recursive estimation for merging precipitation data from radar and microwave cross-track scanners." Journal of Geophysical Research-Atmospheres **114**.

Deng, F., Chen, J. M., Plummer, S., Chen, M. Z. and Pisek, J. (2006). "Algorithm for global leaf area index retrieval using satellite imagery." IEEE Transactions on Geoscience and Remote Sensing **44**(8): 2219-2229.

Ding, Y. Z., Wei, Z. H., Mao, Z. H., Wang, X. F. and Pan, D. L. (2009). "Reconstruction of incomplete satellite SST data sets based on EOF method." Acta Oceanologica Sinica **28**(2): 36-44.

Douaik, A., Van Meirvenne, M. and Toth, T. (2005). "Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data." Geoderma **128**(3-4): 234-248.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grunwald, T., Hollinger, D., Jensen, N. O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B. E., Meyers, T., Moncrieff, H., Moors, E., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala,

T., Wilson, K. and Wofsy, S. (2001). "Gap filling strategies for defensible annual sums of net ecosystem exchange." <u>Agricultural and Forest Meteorology</u> **107**(1): 43-69.

Fang, H. L. and Liang, S. L. (2005). "A hybrid inversion method for mapping leaf area index from MODIS data: experiments and application to broadleaf and needleleaf canopies." <u>Remote Sensing of Environment</u> **94**(3): 405-424.

Fang, H. L., Liang, S. L., Townshend, J. R. and Dickinson, R. E. (2008). "Spatially and temporally continuous LAI data sets based on an integrated filtering method: Examples from North America." <u>Remote Sensing of Environment</u> **112**: 75-93.

Fernandes, R., Butson, C., Leblanc, S. and Latifovic, R. (2003). "Landsat-5 TM and Landsat-7 ETM+ based accuracy assessment of leaf area index products for Canada derived from SPOT-4 VEGETATION data." <u>Canadian Journal of Remote Sensing</u> **29**(2): 241-258.

Fieguth, P., Menemenlis, D., Ho, T., Willsky, A. and Wunsch, C. (1998). "Mapping Mediterranean altimeter data with a multiresolution optimal interpolation algorithm." <u>Journal of Atmospheric and Oceanic Technology</u> **15**(2): 535-546.

Fieguth, P. W., Karl, W. C., Willsky, A. S. and Wunsch, C. (1995). "Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry." <u>IEEE Transactions on Geoscience and Remote Sensing</u> **33**(2): 280-292.

Fieguth, P. W. and Willsky, A. S. (1996). "Fractal estimation using models on multiscale trees." <u>IEEE Transactions on Signal Processing</u> **44**(5): 1297-1300.

Francois, C., Ottle, C. and Prevot, L. (1997). "Analytical parameterization of canopy directional emissivity and directional radiance in the thermal infrared. Application on the retrieval of soil and foliage temperatures using two directional measurements." <u>International Journal of Remote Sensing</u> **18**(12): 2587-2621.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C. (2002). "Global land cover mapping from MODIS: algorithms and early results." <u>Remote Sensing of Environment</u> **83**(1-2):

287-302.

Fuentes, M. (2007). "Approximate likelihood for large irregularly spaced spatial data." Journal of the American Statistical Association **102**(477): 321-331.

Furrer, R., Genton, M. G. and Nychka, D. (2006). "Covariance tapering for interpolation of large spatial datasets." Journal of Computational and Graphical Statistics **15**(3): 502-523.

Gandin, L. S. (1965). Objective analysis of meteorological fields. Jerusalem, Israel Program for Scientific Translations.

Gao, F., Morisette, J. T., Wolfe, R. E., Ederer, G., Pedelty, J., Masuoka, E., Myneni, R., Tan, B. and Nightingale, J. (2008). "An algorithm to produce temporally and spatially continuous MODIS-LAI time series." IEEE Geoscience and Remote Sensing Letters **5**(1): 60-64.

Garrigues, S., Lacaze, R., Baret, F., Morisette, J. T., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., Knyazikhin, Y. and Yang, W. (2008). "Validation and intercomparison of global Leaf Area Index products derived from remote sensing data." Journal of Geophysical Research-Biogeosciences **113**(G2).

Gregg, W. W. and Conkright, M. E. (2001). "Global seasonal climatologies of ocean chlorophyll: Blending in situ and satellite data for the Coastal Zone Color Scanner era." Journal of Geophysical Research-Oceans **106**(C2): 2499-2515.

Gu, J., Li, X., Huang, C. and Okin, G. S. (2009). "A simplified data assimilation method for reconstructing time-series MODIS NDVI data " Advances in Space Research **44**(4): 501-509.

Gu, Y. X., Belair, S., Mahfouf, J. F. and Deblonde, G. (2006). "Optimal interpolation analysis of leaf area index using MODIS data." Remote Sensing of Environment **104**(3): 283-296.

Gutman, G. G. (1999). "On the use of long-term global data of land reflectances and vegetation indices derived from the advanced very high resolution radiometer." Journal of Geophysical Research-Atmospheres **104**(D6): 6241-6255.

Hannachi, A., Jolliffe, I. T. and Stephenson, D. B. (2007). "Empirical orthogonal functions and related techniques in atmospheric science: A review." International Journal of Climatology **27**(9): 1119-1152.

Hu, J. N., Su, Y., Tan, B., Dong, H. A., Yang, W. Z., Schull, M., Bull, M. A., Martonchik, J. V., Diner, D. J., Knyazikhin, Y. and Myneni, R. B. (2007). "Analysis of the MISR LA/FPAR product for spatial and temporal coverage, accuracy and consistency." Remote Sensing of Environment **107**(1-2): 334-347.

Huang, H. C., Cressie, N. and Gabrosek, J. (2002). "Fast, resolution-consistent spatial prediction of global processes from satellite data." Journal of Computational and Graphical Statistics **11**(1): 63-88.

Huang, H. C., Martinez, F., Mateu, J. and Montes, F. (2007). "Model comparison and selection for stationary space-time models." Computational Statistics & Data Analysis **51**(9): 4577-4596.

Jacquemoud, S. and Baret, F. (1990). "PROSPECT - a model of leaf optical properties spectra." Remote Sensing of Environment **34**(2): 75-91.

Johannesson, G., Cressie, N. and Huang, H. C. (2007). "Dynamic multi-resolution spatial models." Environmental and Ecological Statistics **14**(1): 5-25.

Jonsson, P. and Eklundh, L. (2002). "Seasonality extraction by function fitting to time-series of satellite sensor data." IEEE Transactions on Geoscience and Remote Sensing **40**(8): 1824-1832.

Jonsson, P. and Eklundh, L. (2004). "TIMESAT - a program for analyzing time-series of satellite sensor data." Computers & Geosciences **30**(8): 833-845.

Kang, S., Running, S. W., Zhao, M., Kimball, J. S. and Glassy, J. (2005). "Improving continuity of MODIS terrestrial photosynthesis products using an interpolation scheme for cloudy pixels." International Journal of Remote Sensing **26**(8): 1659-1676.

Kannan, A., Ostendorf, M., Karl, W. C., Castanon, D. A. and Fish, R. K. (2000). "ML parameter estimation of a multiscale stochastic process using the EM algorithm." IEEE Transactions on Signal Processing **48**(6): 1836-1840.

Kaufmann, R. K., Zhou, L. M., Knyazikhin, Y., Shabanov, N. V., Myneni, R. B. and Tucker, C. J. (2000). "Effect of orbital drift and sensor changes on the time series of AVHRR vegetation index data." IEEE Transactions on Geoscience and Remote Sensing **38**(6): 2584-2597.

Knyazikhin, Y., Martonchik, J. V., Diner, D. J., Myneni, R. B., Verstraete, M., Pinty, B. and Gobron, N. (1998). "Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data." Journal of Geophysical Research-Atmospheres **103**(D24): 32239-32256.

Kolovos, A., Christakos, G., Serre, M. L. and Miller, C. T. (2002) "Computational Bayesian maximum entropy solution of a stochastic advection-reaction equation in the light of site-specific information." Water Resources Research **38**(12), 1318 DOI: 10.1029/2001WR000743

Kondrashov, D. and Ghil, M. (2006). "Spatio-temporal filling of missing points in geophysical data sets." Nonlinear Processes in Geophysics **13**(2): 151-159.

Kwiatkowska, E. J. and Fargion, G. S. (2003). "Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data." IEEE Transactions on Geoscience and Remote Sensing **41**(12): 2844-2860.

Le Traon, P. Y., Nadal, F. and Ducet, N. (1998). "An improved mapping method of multisatellite altimeter data." Journal of Atmospheric and Oceanic Technology **15**(2): 522-534.

Liang, S. L. (2004). Quantitative remote sensing of land surfaces. Hoboken, New Jersey, John Wiley & Sons, Inc.

Liu, H. Q., Pinker, R. T. and Holben, B. N. (2005). "A global view of aerosols from merged transport models, satellite, and ground observations." Journal of Geophysical Research-Atmospheres **110**(D10).

Lorenc, A. C. (1986). "Analysis methods for numerical weather prediction." Quarterly Journal of the Royal Meteorological Society **112**(474): 1177-1194.

Lu, X. L., Liu, R. G., Liu, J. Y. and Liang, S. L. (2007). "Removal of noise by wavelet

method to generate high quality temporal data of terrestrial MODIS products." Photogrammetric Engineering and Remote Sensing **73**(10): 1129-1139.

Luettgen, M. (1993). Image processing with multiscale stochastic models. Department of electrical engineering and computer science, Massachusetts Institute of Technology.

Matheron, G. (1963). "Principles of geostatistics." Economic Geology **58**: 1246-1266.

Menemenlis, D., Fieguth, P., Wunsch, C. and Willsky, A. (1997). "Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data." Journal of Geophysical Research-Oceans **102**(C5): 10573-10584.

Moody, E. G., King, M. D., Platnick, S., Schaaf, C. B. and Gao, F. (2005). "Spatially complete global spectral surface albedos: Value-added datasets derived from terra MODIS land products." IEEE Transactions on Geoscience and Remote Sensing **43**(1): 144-158.

Morisette, J. T., Privette, J. L. and Justice, C. O. (2002). "A framework for the validation of MODIS Land products." Remote Sensing of Environment **83**(1-2): 77-96.

Morisette, J. T., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J. E., Garrigues, S., Shabanov, N. V., Weiss, M., Fernandes, R. A., Leblanc, S. G., Kalacska, M., Sanchez-Azofeifa, G. A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A. N., Lewis, T. E., Iiames, J. S., Colombo, R., Meroni, M., Busetto, L., Cohen, W. B., Turner, D. P., Warner, E. D., Petersen, G. W., Seufert, G. and Cook, R. (2006). "Validation of global moderate-resolution LAI products: A framework proposed within the CEOS Land Product Validation subgroup." IEEE Transactions on Geoscience and Remote Sensing **44**(7): 1804-1817.

Myneni, R. B., Tucker, C. J., Asrar, G. and Keeling, C. D. (1998). "Interannual variations in satellite-sensed vegetation index data from 1981 to 1991." Journal of Geophysical Research-Atmospheres **103**(D6): 6145-6160.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R. and Running, S. W. (2002). "Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS

data." Remote Sensing of Environment **83**(1-2): 214-231.

Myneni, R. B., Yang, W. Z., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., Didan, K., Fu, R., Juarez, R. I. N., Saatchi, S. S., Hashimoto, H., Ichii, K., Shabanov, N. V., Tan, B., Ratana, P., Privette, J. L., Morisette, J. T., Vermote, E. F., Roy, D. P., Wolfe, R. E., Friedl, M. A., Running, S. W., Votava, P., El-Saleous, N., Devadiga, S., Su, Y. and Salomonson, V. V. (2007). "Large seasonal swings in leaf area of Amazon rainforests." Proceedings of the National Academy of Sciences of the United States of America **104**(12): 4820-4823.

Notaro, M., Liu, Z. and Williams, J. W. (2006). "Observed vegetation-climate feedbacks in the United States." Journal of Climate **19**(5): 763-786.

Nychka, D., Wikle, C. and Royle, J. A. (2002). "Multiresolution models for nonstationary spatial covariance functions " Statistical Modelling **2**(4): 315-331.

Ooba, M., Hirano, T., Mogami, J. I., Hirata, R. and Fujinuma, Y. (2006). "Comparisons of gap-filling methods for carbon flux dataset: A combination of a genetic algorithm and an artificial neural network." Ecological Modelling **198**(3-4): 473-486.

Parada, L. M. and Liang, X. (2004). "Optimal multiscale Kalman filter for assimilation of near-surface soil moisture into land surface models." Journal of Geophysical Research-Atmospheres **109**(D24).

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J. and Stenseth, N. C. (2005). "Using the satellite-derived NDVI to assess ecological responses to environmental change." Trends in Ecology & Evolution **20**(9): 503-510.

Piao, S., Mohammat, A., Fang, J. Y., Cai, Q. and Feng, J. M. (2006a). "NDVI-based increase in growth of temperate grasslands and its responses to climate changes in China." Global Environmental Change-Human and Policy Dimensions **16**(4): 340-348.

Piao, S. L., Fang, J. Y., Zhou, L. M., Ciais, P. and Zhu, B. (2006b). "Variations in satellite-derived phenology in China's temperate vegetation." Global Change Biology **12**(4): 672-685.

Pottier, C., Garcon, V., Larnicol, G., Sudre, J., Schaeffer, P. and Le Traon, P. Y. (2006). "Merging SeaWiFS and MODIS/Aqua ocean color data in North and Equatorial Atlantic using weighted averaging and objective analysis." IEEE Transactions on Geoscience and Remote Sensing **44**(11): 3436-3451.

Pottier, C., Turiel, A. and Garcon, V. (2008). "Inferring missing data in satellite chlorophyll maps using turbulent cascading." Remote Sensing of Environment **112**(12): 4242-4260.

Preisendorfer, R. (1988). Principal Component Analysis in Meteorology and Oceanography, Elsevier.

Reynolds, R. W. and Smith, T. M. (1994). "Improved global sea surface temperature analyses using optimal interpolation." Journal of Climate **7**(6): 929-948.

Roujean, J. L., Leroy, M. and Deschamps, P. Y. (1992). "A bidirectional reflectance model of the earths surface for the correction of remote sensing data." Journal of Geophysical Research-Atmospheres **97**(D18): 20455-20468.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N. and Ohno, H. (2005). "A crop phenology detection method using time-series MODIS data." Remote Sensing of Environment **96**(3-4): 366-374.

Sapiano, M. R. P., Smith, T. M. and Arkin, P. A. (2008). "A new merged analysis of precipitation utilizing satellite and reanalysis data." Journal of Geophysical Research-Atmospheres **113**.

Schoellhamer, D. H. (2001). "Singular spectrum analysis for time series with missing data." Geophysical Research Letters **28**(16): 3187-3190.

Sellers, P. J., Tucker, C. J. and al., G. J. C. e. (1994). "A global 1° by 1° NDVI data set for climate studies. Part 2: The generation of global fields of terrestrial biophysical parameters from the NDVI." International Journal of Remote Sensing **15**: 3519-3545.

Serre, M. L. and Christakos, G. (1999). "Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge - the Equus Beds study." Stochastic Environmental Research and Risk Assessment **13**(1-2): 1-26.

Shabanov, N., Samanta, A., Myneni, R. B., Knyazikhin, Y., Votava, P. and Nemani, R.

(2007). "Collection 5 MODIS LAI and FPAR Products." <u>MODIS Land Collection 5/LTDR Workshop</u>, http://modis.gsfc.nasa.gov/sci_team/meetings/ c5meeting/pres/day1/shabanov.pdf (last accessed Oct 9, 2009).

Slatton, K. C., Crawford, M. M. and Evans, B. L. (2001). "Fusing interferometric radar and laser altimeter data to estimate surface topography and vegetation heights." <u>IEEE Transactions on Geoscience and Remote Sensing</u> **39**(11): 2470-2482.

Smith, T. M., Reynolds, R. W., Livezey, R. E. and Stokes, D. C. (1996). "Reconstruction of historical sea surface temperatures using empirical orthogonal functions." <u>Journal of Climate</u> **9**(6): 1403-1420.

Sylvander, S., Albert-Grousset, I. and Henry, P. (2003). <u>Geometrical performance of the VEGETATION products</u>. Proceedings of IEEE International Geoscience and Remote Sensing Symposium 2003: 573- 575.

Tian, Y. H., Zhang, Y., Knyazikhin, Y., Myneni, R. B., Glassy, J. M., Dedieu, G. and Running, S. W. (2000). "Prototyping of MODIS LAI and FPAR algorithm with LASUR and LANDSAT data." <u>IEEE Transactions on Geoscience and Remote Sensing</u> **38**(5): 2387-2401.

Townshend, J. R. G. and Justice, C. O. (2002). "Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing." <u>Remote Sensing of Environment</u> **83**(1-2): 351-359.

Tzeng, S. L., Huang, H. C. and Cressie, N. (2005). "A fast, optimal spatial-prediction method for massive datasets." <u>Journal of the American Statistical Association</u> **100**(472): 1343-1357.

Uz, B. M. and Yoder, J. A. (2004). "High frequency and mesoscale variability in SeaWiFS chlorophyll imagery and its relation to other remotely sensed oceanographic variables." <u>Deep-Sea Research Part II-Topical Studies in Oceanography</u> **51**(10-11): 1001-1017.

Vautard, R. and Ghil, M. (1989). "Singular spectrum analysis in nonlinear dynamics, with applictions to paleoclimatic time-series." <u>Physica D</u> **35**(3): 395-424.

Verger, A., Baret, F. and Weiss, M. (2008). "Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products."

Remote Sensing of Environment **112**(6): 2789-2803.

Verhoef, W. (1984). "Light-scattering by leaf layers with application to canopy reflectance modeling - the SAIL model." Remote Sensing of Environment **16**(2): 125-141.

Vermote, E. and 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.

Wang, D. and Liang, S. (2008). Singular Spectrum Analysis For Filling Gaps And Reducing Uncertainties Of MODIS Land Products. Proceedings of IEEE International Geoscience and Remote Sensing Symposium 2008: 558-561.

Weare, B. C. and Nasstrom, J. S. (1982). "Examples of extended empirical orthogonal function anaylyses." Monthly Weather Review **110**(6): 481-485.

Weiss, M., Baret, F., Garrigues, S. and Lacaze, R. (2007). "LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products." Remote Sensing of Environment **110**(3): 317-331.

Wikle, C. K. and Berliner, L. M. (2007). "A Bayesian tutorial for data assimilation." Physica D-Nonlinear Phenomena **230**(1-2): 1-16.

WMO (2006). Systematic observation requirements for satellite-based products for climate. GCOS-107. Geneva, Switzerland

Wolfe, R. E., Nishihama, M., Fleig, A. J., Kuyper, J. A., Roy, D. P., Storey, J. C. and Patt, F. S. (2002). "Achieving sub-pixel geolocation accuracy in support of MODIS land science." Remote Sensing of Environment **83**(1-2): 31-49.

Yang, L. M., Wylie, B. K., Tieszen, L. L. and Reed, B. C. (1998). "An analysis of relationships among climate forcing and time-integrated NDVI of grasslands over the US northern and central Great Plains." Remote Sensing of Environment **65**(1): 25-37.

Yang, W. Z., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., Privette, J. L., Huemmrich, K. F., Fensholt, R., Sandholt, I., Weiss, M., Ahl, D. E., Gower, S. T., Nemani, R. R., Knyazikhin, Y. and Myneni, R. B. (2006).

"MODIS leaf area index products: From validation to algorithm improvement." <u>IEEE Transactions on Geoscience and Remote Sensing</u> **44**(7): 1885-1898.

Young, P. C., Pedregal, D. J. and Tych, W. (1999). "Dynamic harmonic regression." <u>Journal of Forecasting</u> **18**(6): 369-394.

Zhang, B. L., Pinker, R. T. and Stackhouse, P. W. (2007). "An empirical orthogonal function iteration approach for obtaining homogeneous radiative fluxes from satellite observations." <u>Journal of Applied Meteorology and Climatology</u> **46**(4): 435-444.

Zhang, P., Anderson, B. and Barlow, M. (2004a). "Climate-related vegetation characteristics derived from Moderate Resolution Imaging Spectroradiometer (MODIS) leaf area index and normalized difference vegetation index." <u>Journal of Geophysical Research-Atmospheres</u> **109**(D20).

Zhang, X. Y., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C. and Huete, A. (2003). "Monitoring vegetation phenology using MODIS." <u>Remote Sensing of Environment</u> **84**(3): 471-475.

Zhang, X. Y., Friedl, M. A., Schaaf, C. B. and Strahler, A. H. (2004b). "Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data." <u>Global Change Biology</u> **10**(7): 1133-1145.

Zhang, X. Y., Friedl, M. A. and Schaaf, C. B. (2006). "Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements." <u>Journal of Geophysical Research-Biogeosciences</u> **111**(G4).