Remote Sensing of Environment 174 (2016) 258–265

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data☆

Yang Shao [a,*], Ross S. Lunetta [b], Brandon Wheeler [a], John S. Iiames [b], James B. Campbell [a]

[a] Virginia Tech, College of Natural Resources and Environment, Geography Department, 115 Major Williams Hall, Blacksburg, VA 24061, USA
[b] U.S. Environmental Protection Agency, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA

ARTICLE INFO

Article history:
Received 12 August 2015
Received in revised form 8 December 2015
Accepted 15 December 2015
Available online 29 December 2015

Keywords:
MODIS-NDVI
Multi-temporal analysis
Smoothing algorithms
Validation

ABSTRACT

In this study we compared the Savitzky–Golay, asymmetric Gaussian, double-logistic, Whittaker smoother, and discrete Fourier transformation smoothing algorithms (noise reduction) applied to Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) time-series data, to provide continuous phenology data used for land-cover (LC) classifications across the Laurentian Great Lakes Basin (GLB). MODIS 16-day 250 m NDVI imagery for the GLB was used in conjunction with National Land Cover Database (NLCD) from 2001, 2006 and 2011, and the Cropland Data Layers (CDL) from 2011 to 2014 to conduct classification evaluations. Inter-class separability was measured by Jeffries–Matusita (JM) distances between selected cover type pairs (both general classes and specific crops), and intra-class variability was measured by calculating simple Euclidean distance for samples within cover types. For the GLB, we found that the application of a smoothing algorithm significantly reduced image noise compared to the raw data. However, the Jeffries–Matusita (JM) measures for smoothed NDVI temporal profiles resulted in large inconsistencies. Of the five algorithms tested, only the Fourier transformation algorithm and Whittaker smoother improved inter-class separability for corn–soybean class pair and significantly improved overall classification accuracy. When compared to the raw NDVI data as input, the overall classification accuracy from the Fourier transformation and Whittaker smoother improved performance by approximately 2–6% for some years. Conversely, the asymmetric Gaussian and double-logistic smoothing algorithms actually led to degradation of classification performance.

1. Introduction

Harmful algal blooms (HABs) have been estimated to deteriorate freshwater supplies at a cost of $2.2 billion annually within the United States (US). These damages negatively affect ecosystem services including recreational water use, water front real estate value, biodiversity, and drinking water treatment costs (Lopez, Jewett, Dortch, Walton, & Hudnell, 2008). The rapid spread of cyanobacteria and associated cyanotoxins that can result from HAB blooms holds particular consequence: cyanotoxins are included in the Safe Drinking Water Act's Contaminant Candidate List and are correlated with the degradation of potable water, human respiratory irritation, and other human illnesses resulting from ingestion or skin exposure. The severity of complications which result from HABs warrant the establishment of a regularly updated cyanobacteria assessment network, which is currently developed under an interagency effort funded by NASA, USEPA, NOAA, and USGS (NASA, 2015).

Algal blooms result from a combination of factors including excess nutrients (Michalak et al., 2013), environmental conditions related to temperature, light, stratification (Paerl & Huisman, 2008), and changes in land-cover (LC) and land-use (LU) practices related to urbanization and modern agricultural processes related to sediment and nutrient buildup in watersheds (Shao, Lunetta, Macpherson, Luo, & Chen, 2013). A method for determining future trends in cyanobacteria outbreaks utilizing LCLU data to determine correlations could be valuable. However, existing reference data may be inadequate due to issues in temporal resolution. For example, the application of National Land Cover Database (NLCD) reference data for understanding and monitoring HAB trends is problematic due to the 5–7 year time lapse between updates. An approach using time-series Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) data, available in 16-day composites, for LCLU characterization may augment research capabilities by filling in previous gaps.

Time-series MODIS-NDVI data have been proven to be useful for cover type characterization (Friedl et al., 2002; Xiao et al., 2006) and

☆ **Notice:** The U.S. Environmental Protection Agency funded and partially conducted the research described in this paper. Although this work was reviewed by EPA and has been approved for publication, it may not necessarily reflect official Agency policy. Mention of any trade names or commercial products does not constitute endorsement or recommendation for use. This research was conducted in support of the National Aeronautics and Space Administration (NASA) Ocean Biology and Biogeochemical Program grant NNH15AZ42I.

* Corresponding author.
E-mail address: yshao@vt.edu (Y. Shao).

http://dx.doi.org/10.1016/j.rse.2015.12.023
0034-4257/

change detection analysis (Lunetta, Knight, Ediriwickrema, Lyon, & Worthy, 2006). For LC characterization, monthly NDVI composite data derived from the Advanced Very High Resolution Radiometer (AVHRR) sensor have been used as a primary input to generate the 1.0 km resolution global LC database (Loveland et al., 2000). More recently, global and regional mapping efforts have focused on using time-series NDVI or other vegetation indices from MODIS (Friedl et al., 2002; Knight, Lunetta, Ediriwickrema, & Khorram, 2006). Compared to the traditional single image 'snap-shot' classification approach, the use of time-series remote sensing data or multi-temporal image classification often improves classification accuracy by incorporating both spectral and temporal profiles (Shao & Lunetta, 2011). Standard global MODIS-derived LC products (Friedl et al., 2002), however, are provided at relatively coarse spatial resolution (500 m) and focus on broad/general cover types. MODIS time-series data have also been successfully used for crop-specific mapping in which corn and soybean can be delineated by examining NDVI temporal profiles from the crop growing season (Wardlow, Egbert, & Kastens, 2007; Lunetta, Shao, Ediriwickrema, & Lyon, 2010). MODIS-based crop-specific mapping has yet to be generalized to national scale for operational use.

One of the main challenges in time-series remote sensing data analysis is dealing with image noise such as pseudo-hikes and pseudo-lows caused by cloud and shadow issues, weather impacts, and sensor-introduced noises (Goward, Markham, Dye, Dulaney, & Yang, 1991; Lunetta et al., 2006). The presence of such outliers may add uncertainty in LC mapping and change detection efforts, therefore necessitating data cleaning; the process of filtering and smoothing anomalous time-series data (Holben, 1986; Lunetta et al., 2006). A large number of data smoothing algorithms have been developed to reduce noise in remote sensing time-series data. For AVHRR-NDVI data, Ma and Veroustraete (2006) developed a smoothing method using a mean-value iteration filter. Reed (2006) and Swets, Reed, Rowland, and Marko (1999) provided examples of smoothing using a weighted least-squares approach. For MODIS-NDVI data, Sakamoto et al. (2005) developed smoothing methods using wavelet and Fourier transforms. Bruce, Mathur, Byrd, and John (2006) developed a new wavelet-based feature extraction technique. Chen et al. (2004) modified the Savitzky–Golay filter to adapt to the upper envelope of the vegetation index data; an adaptation especially important because NDVI signals are often negatively biased. Atzberger and Eilers (2011a) showed that the Whittaker smoother (WS) can perfectly balance fidelity to the original data and smoothness of the fitted curve while being easy to implement with fast processing times (Eilers, 2003). Lunetta et al. (2006) applied inverse Fourier transformation to estimate new NDVI values, while retaining the original NDVI values for cloud-free and good quality pixels. Additionally, Jönsson and Eklundh (2002, 2004) integrated Savitzky–Golay, asymmetric Gaussian, and double-logistic algorithms into a TIMESAT package which can be applied for smoothing time-series NDVI data from a variety of sensors, including AVHRR, MODIS, and Medium Resolution Imaging Spectrometer (MERIS).

Although a variety of algorithms have been examined and implemented for smoothing time-series data, comparisons of the relative effectiveness of each algorithm is difficult due to lack of in-situ reference data (Hird & McDermid, 2009) and standard statistical measures (Atzberger & Eilers, 2011a). Most published studies compare different smoothing algorithms for how well they derive phenological metrics such as start of season (SOS) and end of season (EOS) (e.g., Atkinson, Jeganathan, Dash, & Atzberger, 2012; Beck, Atzberger, Høgda, Johansen, & Skidmore, 2006, Hird & McDermid, 2009). Little research has been conducted for the purpose of multi-temporal image classification. In a recent paper, Atzberger and Eilers (2011a) proposed possible measures for evaluating the effectiveness of smoothing algorithms including both increased inter-class separability, and reduced within-class variability of pseudo-invariant targets. For example, Jeffries–Matusita distance measures can be used to estimate inter-class separability, and Euclidean distance measures can be used to estimate within-class variability. Atzberger and Eilers (2011a), however, did not conduct actual comparison of different smoothing algorithms because the main purpose of their paper was to highlight potential evaluation methods. Few published studies have evaluated how different data smoothing algorithms affect actual classification performance using accuracy statistics such as overall accuracy, kappa statistics, and class-specific accuracy.

The main objective of this study was to compare five smoothing algorithms for multi-temporal land cover mapping applications in the Laurentian Great Lakes Basin (GLB) region (US portion). The five algorithms chosen were the Savitzky–Golay algorithm, the asymmetric Gaussian model, the double-logistic model, the Whittaker smoother, and a Fourier transform approach. The first three algorithms have been widely used for remote sensing time-series data analysis (Jönsson & Eklundh, 2002, 2004). The Whittaker smoother has been successfully used by Atzberger and Eilers (2011a, 2011b) to improve overall quality of time-series NDVI data and the signal-to-noise ratio. Previous reported LC classification and change detection efforts have demonstrated high potential of the Fourier transformation approach (Lunetta et al., 2006; Shao, Lunetta, Ediriwickrema, & Iiames, 2010). Our study objectives were to: (i) evaluate within-class variability for dominant LC types across the study region; (ii) quantify inter-class separability for cover type pairs; and (iii) determine how smoothing algorithms actually affect image classification performance. Evaluations incorporated error matrices, overall accuracy, and kappa statistics (Congalton, 1991). Both general cover types (e.g., deciduous forest, cropland and wetland) and crop-specific (e.g., corn and soybean) cover types were considered. We hypothesized that optimizing multi-temporal MODIS smoothing algorithm performance would result in better classification results for the GLB region, thus provide a better understanding of LU and LC change.

1.1. Study area

The US GLB contains the entire state of Michigan, portions of seven other states and Canadian Province of Ontario (not analyzed) (Fig. 1). The GLB is an important region in North America from both an economic and ecological standpoint. The ecosystem supported by the GLB features a range of habitats: forests, grasslands, and prairies, as well as dunes, marshes, and wetlands along the lakeshores. Lakes Superior, Michigan, Huron, Ontario, and Erie together make up the largest total surface area of freshwater on Earth, providing drinking water for over 40 million people as well as 56 billion gallons of water each day for industrial, municipal, and agricultural use. Almost a third of the area of the GLB is devoted to farmland, producing approximately 7% of the US and 25% of the Canadian total crop yield, respectively.

2. Methods

MODIS MOD13Q1 Terra vegetation index data from 2001 to 2014 were obtained from NASA Reverb website (http://reverb.echo.nasa.gov/) for the GLB. The MOD13Q1 product includes 250 m resolution NDVI and Enhanced Vegetation Index (EVI) data, and quality assessment (QA) information for both. MOD13Q1 data is calculated from atmospherically-corrected surface reflectance values and delivered as a 16-day composite image. For each composite period, the initial data consisted of four separate MODIS scenes to cover the large study area.

The NLCD data for years 2001, 2006 and 2011 were acquired from the Multi-Resolution Land Characteristics Consortium (http://www.mrlc.gov/). All NLCD datasets (30 m) were developed with Landsat data as the primary input to derive a 16-class LC classification scheme, using a decision-tree model to support the classification. The overall accuracy of NLCD 2001 classes (Anderson Level I) was assessed to be around 85% (Wickham, Stehman, Fry, Smith, & Homer, 2010). The accuracy of NLCD 2011 is currently being assessed.

To provide crop-specific validation reference data, we obtained the CDL data from CropScape (http://nassgeodata.gmu.edu/CropScape/), developed by the United States Department of Agriculture National

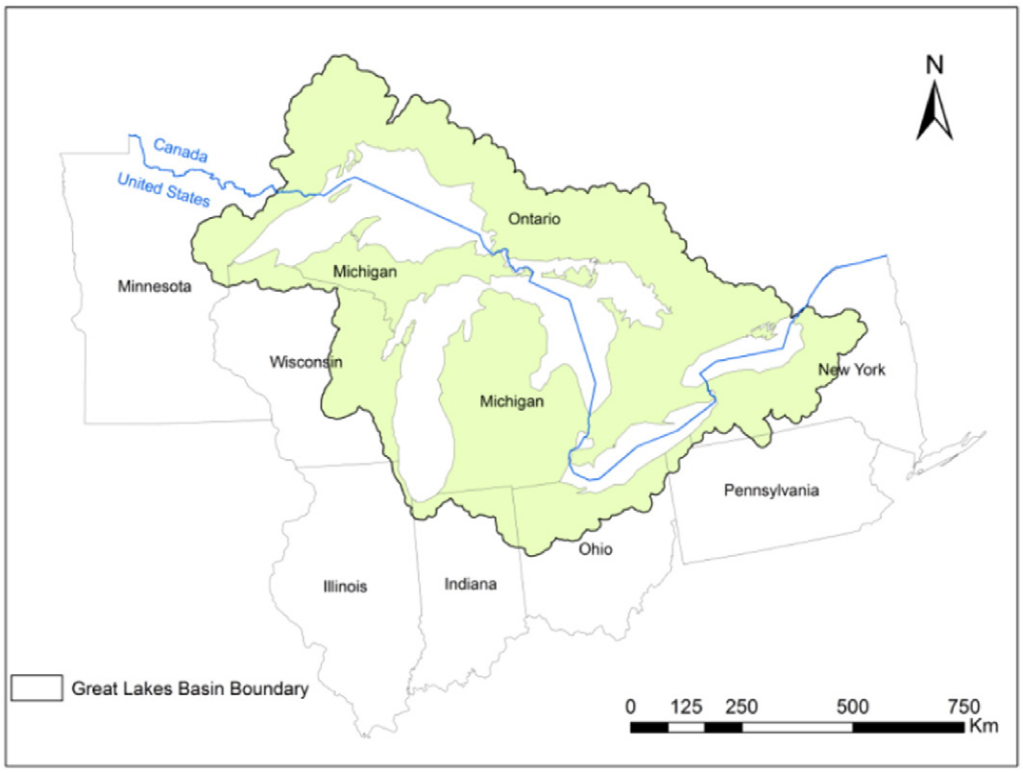

Fig. 1. Laurentian Great Lake Basin study area was limited to the US portion of the basin.

Agricultural Statistics Service (USDA-NASS). Complete coverage for the US is available from 2008 to 2014. The spatial resolution of the CDL varies from 30 to 56 m, depending on the imagery source (Boryan, Yang, Mueller, & Craig, 2011). Specific crop types including as corn, soybean, and wheat were mapped at an annual interval, with classification accuracies exceeding 80% for corn and soybean in most states. We obtained CDLs from 2011 to 2014, since all four years' data have consistent 30 m spatial resolution.

2.1. Data pre-processing and smoothing

The MODIS Reprojection Tool (MRT) was used to generate NDVI mosaics spanning the four MODIS scenes for each 16-day composite. The resultant mosaics were re-projected to an Albers Equal Area Conic (AEA) projection. Each set of reference/supplemental data (NLCD 2001, NLCD 2006, NLCD 2011, and CDL 2011–2014) also had the same AEA projection. For the NLCD data, each LC type of interest was extracted at the nominal 30 m resolution, then degraded to provide cover type proportions within 250 m MODIS-NDVI grids. A threshold value of 50% proportional cover was used to identify pixels with at least one dominant cover type within the MODIS grid. The same procedure was applied to CDLs for 2011–2014 to derive corn and soybean pixels. For CDLs, we focused on corn and soybean because these two summer crops have a similar crop calendar, thus they are difficult to classify even with multi-temporal MODIS data (Shao et al., 2010). Creating this diverse library of reference data allowed us to evaluate each smoothing algorithm for image classification performance, for both general and specific cover types, for multiple study years.

We implemented three smoothing algorithms through the TIMESAT software package including the: adaptive Savitzky–Golay, asymmetric Gaussian, and double-logistic function (Jönsson & Eklundh, 2002). The Savitzky–Golay approach applies a moving window to a given time-series dataset. Within a moving window (e.g., $2n + 1$ points, n is a user defined window width), a quadratic polynomial function is used to fit all points and then the value of the central point is replaced by the fitted value. The adaptive Savitzky–Golay approach considers that noises from cloud/cloud shadow typically reduce the original NDVI value. Thus, the weight for each point can be re-assigned to favor points located above the initial polynomial fit. A new quadratic polynomial function can then be applied to derive NDVI values adapted to the upper envelope of the time-series data.

The asymmetric Gaussian algorithm relies mainly on five parameters to fit time-series data including the time of the minimum or maximum NDVI, the width and flatness of the right side of the function, and the width and flatness of the left side of the function. The double logistic function estimates four parameters to determine the left inflection point, the right inflection point, and rates of changes at two inflection points. Both the asymmetric Gaussian algorithm and the double logistic function are modifications of local model functions, which have been proven to be effective in capturing phenological cycle events defined by NDVI curves (Atkinson et al., 2012).

We first used TIMESAT in the MATLAB environment and visually compared raw NDVI data and smoothed curves. The MODIS Reliability Index (RI) was used to identify pixels with quality issues (e.g., RI > 1). Such pixels were assigned with weight of zero, thus not affecting the subsequent smoothing. TIMESAT also incorporates an automated data pre-processing step to remove additional spikes and outliers within the time-series data. For example, if NDVI value is substantially different from the median value of a pre-defined temporal window (e.g., seven data points), it is considered as a spike/outlier before data smoothing. We also adjusted additional TIMESAT parameters such as envelope iterations and adaptation strength to fine-tune the upper envelope fitting. The final configurations were saved as TIMESAT setting files that allow automated data smoothing for a large MODIS-NDVI dataset. The Whittaker smoother was also implemented in MATLAB and relied on the smoothing parameter (λ) to control (larger the smoother) the NDVI temporal profile smoothness (Eilers, 2003; Atzberger & Eilers, 2011a, 2011b). Similar to Atzberger and Eilers (2011a), we used an automated pixel-by-pixel cross-validation (good quality NDVI values

only) approach to search the best smoothing parameter and generate the final smoothed NDVI temporal profiles.

For the Fourier transformation approach, pixels with quality issues (e.g., RI > 1) were first removed from the time-series. Based on previous work (Lunetta et al., 2006), NDVI values show sudden decrease or increase (±15%) and then return to the previous value were also considered as noises and were removed from the time-series. Fourier transformation was then applied to the filtered NDVI stack to separate the signal and noise (Lunetta et al., 2006; Roberts, Lehár, & Dreher, 1987). Specifically, the 14 year NDVI temporal profiles were transformed into frequency domain using a discrete Fourier transformation and the power spectrum was then used to separate signal and noise. For each pixel in the frequency domain, power density value above a predefined threshold (i.e., 2× mean density value) was considered as signal. Thresholds were empirically determined using the 4–6 harmonic series to represent NDVI profiles. The harmonic series number was minimized to avoid the overfitting or spurious oscillations, a common problem when high order Fourier series is used (Hermance, 2007). A nonlinear deconvolution method was then used to estimate new NDVI values for the removed NDVI data points. For our Fourier transformation approach, we note that the original NDVI values for cloud-free and good quality pixels remain unchanged in the NDVI time-series stack.

The data volume for all of the above time-series smoothing approaches was large due to the expansive geographic coverage of our study area and study period (14-years) examined. With this data volume, the application of the smoothing algorithms is computationally-intensive. Therefore, we used the Virginia Tech Advanced Research Computing's ITHACA cluster for the image analysis. The initial time-series data set was divided into 10 smaller study regions in order to take advantage of the parallel processing offered by ITHACA. The resultant sets of smoothed files were then merged into a single time series file providing full datasets of the region.

2.2. Within-class data variability analysis

For a given cover type (e.g., deciduous forest), we expected to see relatively similar NDVI temporal profiles. A simple Euclidean distance measure can be used to characterize the similarity of two randomly selected pixels in the spectral–temporal domain:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

where p_i is NDVI value at time i for p location and q_i is NDVI value at time i for q location. The pair-wise Euclidean distance can be grouped together if more randomly selected pixels are involved in the analysis. The mean and standard deviation of such Euclidean distance measures were then used as a criteria to compare the similarity of NDVI temporal profiles from raw (i.e., unsmoothed), adaptive Savitzky–Golay, asymmetric Gaussian, double-logistic function, Whittaker smoother, and Fourier transformation-derived NDVI products. Our hypothesis is that the smoothing algorithm providing the lowest mean and standard deviation Euclidean distance measurement is likely to be the most reliable option for the study area.

We focused on four main LC types: deciduous forest, cultivated crops, hay/pasture, and wetlands. According to NLCD 2011, these four cover types account for 24.5%, 22.9%, 9.2%, and 14.2% of total land area in the GLB. Using NLCD 2001, 2006, and 2011 as reference datasets, we first identified no-change (or pseudo-invariant) reference pixels from 2001 to 2011 for deciduous forest, cultivated crops, hay/pasture, and wetlands, respectively. Within the no-change area, we randomly selected 5000 MODIS pixels for each cover type, per study year. Selected pixels (n = 5000) represented approximately 0.8% of overall sample (~630,000). Edge-pixels were avoided due to potential spectral mixture problem. From these randomly selected pixels, we calculated pair-wise Euclidean distance and derived mean and standard deviation of distance measure for each smoothing algorithm. The random sampling and distance calculation procedures were repeated for 10 times to reduce potential sampling uncertainties. Then for each study year, the distance statistics were compared for different smoothing algorithms. In addition to the above four general cover types, we evaluated within-class variability for crop-specific classes including corn and soybean. Corn and soybean were the two dominant crop types in the GLB and previous studies have shown that these two crop types often have similar NDVI temporal profiles, thus, they are not easy to delineate (Shao et al., 2010; Wardlow et al., 2007). We used CDL data from years 2011 to 2014 to derive reference data, corresponding MODIS NDVI temporal profiles from various smoothing algorithms were analyzed and compared using Euclidean distance measures.

2.3. Cover type inter-class separability

Inter-class separability was evaluated using the Jeffries–Matusita (JM) distance measure. The JM distance has been widely used in remote sensing to measure the average distance between two class density functions (Richards & Jia, 2006; Wardlow et al., 2007). JM distance expresses separability between two classes where the lower-bound (0) indicates identical and impossible to separate two classes, and the upper-bound (1.41) represents that two classes can be perfectly separated. Similar to the above within-class variability analysis, we derived JM measures for both general land cover types and crop-specific types. The four cover types in the GLB would lead to six possible LC pairs for JM measures: deciduous forest-cultivated crops, deciduous forest–wetlands, deciduous forest–hay/pasture, cultivated crops–wetlands, cultivated crops–hay/pasture, and wetlands–hay/pasture. JM measurements were computed for 2001, 2006, and 2011, respectively. For each year, we conducted principal component analysis (PCA) to the NDVI temporal profiles to reduce the linear dependence of time-series data. The first 10 PCA components (approximately 80% variance) were used for JM distance calculation.

2.4. Performance evaluations

We further assessed image classification performance using various smoothed NDVI temporal profiles as image classification input datasets. For general LC classification, we continued with the four main cover types of deciduous forest, cultivated crops, pasture/hay, and wetlands. All other cover types were masked out using NLCD data as references. For each general cover type, we randomly selected 5000 MODIS pixels for image classification training and the remaining MODIS pixels were used for accuracy assessment. The training and accuracy assessments were conducted for raw NDVI profiles (RAW), the Savitzky–Golay smoothed profiles (SG), asymmetric Gaussian profiles (GA), double-logistic profiles (DL), Whittaker smoother profiles (WS), and discrete Fourier transformation profiles (FT), independent to each other. A maximum likelihood classifier was used for the image classification task since its performance is closely linked to JM distance measures (Richards & Jia, 2006). Error matrix, overall accuracy, and kappa statistics (Congalton, 1991) were generated and compared for NDVI input datasets corresponding to years 2001, 2006, and 2011. Similar image classification and accuracy assessments were conducted for corn–soybean separation using CDLs from years 2011 to 2014 as reference data.

3. Results

3.1. Within-class variability

Fig. 2 compares NDVI temporal profiles from five smoothing algorithms and the original NDVI data for an example data point. NDVI temporal profiles derived from SG, GA, DL, and WS algorithms were clearly much smoother compared to the original data. NDVI values appeared to be adapted to the upper envelope of the time-series data. The FT

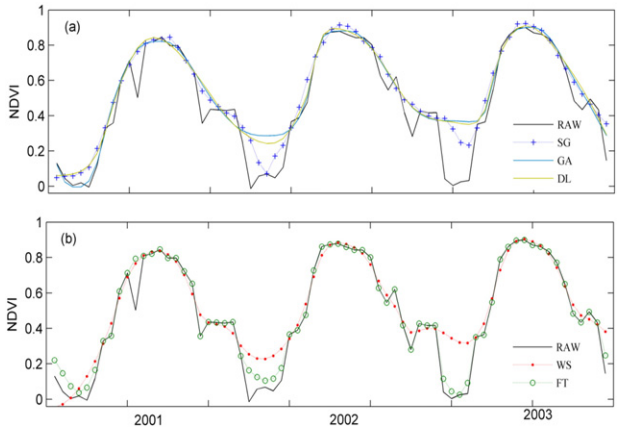

Fig. 2. Visual comparison of smoothing algorithm effects: (a) original NDVI value (RAW), the Savitzky–Golay (SG), asymmetric Gaussian (SG), and double-logistic (DL); (b) original NDVI value (RAW), Whittaker smoother (WS), and Fourier transformation method (FT).

algorithm, on the other hand, kept most original NDVI values and estimated new values only for outliers and points with data quality issues, thus it is less aggressive in terms of data smoothing and maintains subtle profile features that are potentially lost with the other algorithms tested.

Fig. 3 compares Euclidean distance measures for four selected general LC types for the years of 2001, 2006, and 2011. For all LC types, the smoothing algorithms reduced the mean Euclidean distance compared to the raw data. For example, for the wetland class of 2001, the mean Euclidean distance measures reduced from 0.849 to 0.637, 0.671, 0.706, 0.675, and 0.690 for SG, GA, DL, WS, and FT algorithms, respectively. Among the five smoothing algorithms, SG algorithm generated the smallest mean Euclidean distance measures for most cover types in years of 2001 and 2006. The FT filter led to the largest distance measures for all cover types in years 2006 and 2011. This result was expected because our FT method only estimated new values for NDVI outliers and pixels with quality issues. The resultant Euclidean distance measures thus would be more similar to those derived from raw NDVI data. The standard deviations of Euclidean distance measures (Fig. 3d–f) suggested that GA and DL had more scattered Euclidean distance measures compared to the other three smoothing algorithms.

For crop-specific classes (i.e., corn and soybean), we observed similar trends that all five smoothing algorithms reduced the mean Euclidean distance compared to the raw NDVI profiles and the WS and FT smoothers led to relatively larger Euclidean distance measures for all four study years from 2011 to 2014 (Fig. 4). Among the remaining three smoothers, the relatively ranks were inconsistent. For example, the SG smoother generated the smallest Euclidean distance measures for both corn and soybean classes in 2012. For 2013, DL smoother performed best although the differences among SG, GA, and DL smoothers were quite small. The standard deviations of Euclidean distance measures show that GA and DL had more scattered Euclidean distance measures, especially when compared to the WS and FT algorithms.

3.2. JM distance

JM distance measures for six general LC pairs were compared in Table 1. For the raw NDVI temporal profiles, the pair of cultivated crops and hay/pasture had the lowest JM distance (0.577 in 2001, 0.760 in 2006, and 0.390 in 2011), suggesting low separability between these two classes. Their separability increased substantially after all five different data smoothing algorithms. Results for other LC pairs were more complicated because we observed both increases and decreases of JM distance measures comparing raw NDVI and smoothing NDVI profiles. Among the five smoothing algorithms, the WS and FT performed best because they generated higher JM distance measures in most LC pairs compared to the raw NDVI data. For 2006, the SG algorithm generated higher JM distance measures for all six LC pairs compared to the raw NDVI data. However, it only outperformed raw data in one LC pair (cultivated crops–hay/pasture) in 2011. Additionally, the GA and the DL filter performed similarly, and both generated relatively low JM distances for Hay–Wetland LC pair.

Crop-specific JM measurements were compared in Table 2. Considering all study years from 2011 to 2014, only the WS and FT algorithms had consistently higher JM measurements compared to raw data. Results for other smoothing algorithms showed both increases and decreases of JM distance measures comparing raw NDVI. After evaluating all LC datasets for JM distance, it was determined that in both general LC and crop-specific analyses, the WS and FT algorithms out-performed the other smoothing algorithms in terms of class separability.

3.3. Multi-temporal classifications

Table 3 compares general classification performance using different MODIS input data. For year 2001, the overall classification accuracy was 71% (kappa = 0.60) using original NDVI data as input. The overall accuracy slightly increased to 73% for the SG, WS, and FT algorithms. The FT algorithm also slightly outperformed the raw data for 2006 image classification, however, the differences in overall accuracy were less than or equal to one percent and not likely meaningful in practice. We also

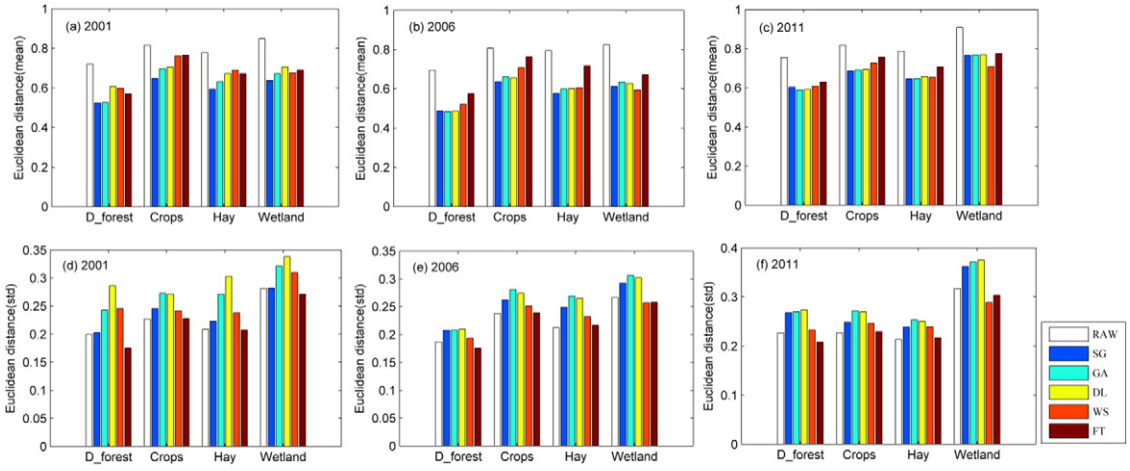

Fig. 3. Within-class distance measures (mean and standard deviation) for selected NLCD (2001, 2006, and 2011) cover classes: wetlands, hay/pasture, cultivated crops, and deciduous forest. Pixels were randomly sampled from each class and the results were averaged from 10 sampling trials.

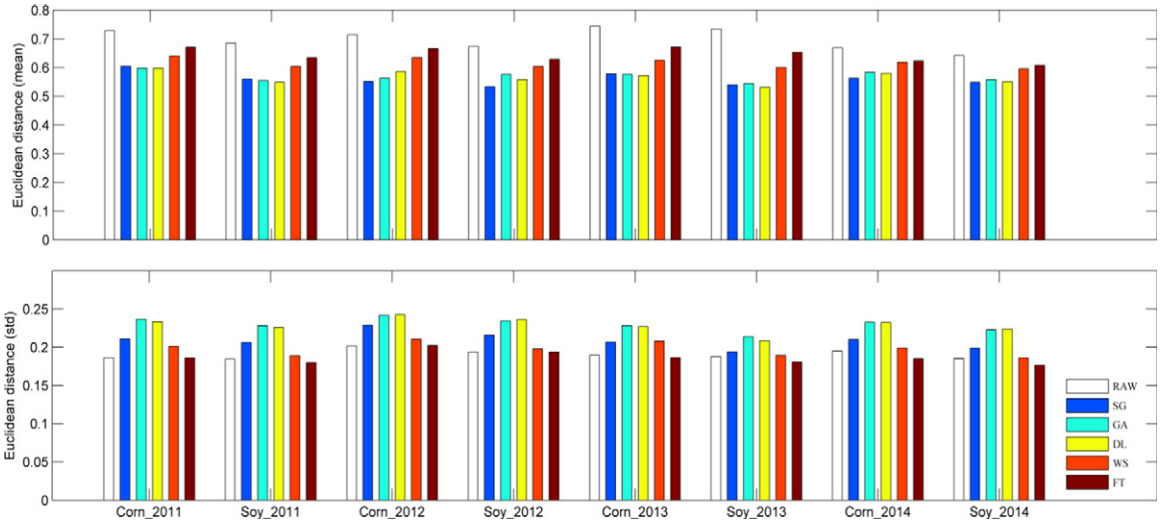

Fig. 4. Within-class distance measures (mean and standard deviation) for corn and soybean pixels (2011–2014). Pixels were randomly sampled from each class and the results were averaged from 10 sampling trials.

noted that the GA and DL algorithms resulted in a poorer classification performance when compared to the raw NDVI data. Using the GA smoothed NDVI, the overall accuracy was 67% for year 2001 classification, about 4% less than the overall accuracy of raw NDVI data. Such findings suggested that users should apply caution in selecting smoothing algorithms. Among the five selected smoothing algorithms, the FT, WS, and SG algorithms performed much better compared to the GA and DL algorithms.

Table 4 shows error matrices for 2001 classification results using raw NDVI data and FT-derived data as inputs. With raw NDVI data, deciduous forest and cultivated crops had relatively high classification accuracy of 80.9% and 89.1%, respectively. However, the hay/pasture class had very low user's accuracy of 38.4%. There was significant confusion between hay/pasture and cultivated crops. Wetland class had relatively low accuracy (65.6%) and errors could be attributed to confusion between wetland and deciduous forest. Using the FT filtered data, the classification accuracies for hay/pasture classes increased about 4.2%. The confusion between wetland and deciduous forest was also reduced. The FT performance (kappa = 0.63) was determined to be statistically significant ($p = 0.05$) compared to RAW (kappa = 0.60) based on the Z-statistic results (Z-score = 25.2).

For corn/soybean classification, the SG, WS, and FT smoothed NDVI data again outperformed raw NDVI input for most study years (Table 5). For year 2013, the overall classification accuracy was 65% (kappa = 0.30) using original NDVI data as input. The overall accuracy slightly increased to 71% and 70% for the WS and FT algorithm, respectively. Similar to the general LC classification, the GA and DL smoothed data led to decreased overall accuracy compared to raw NDVI data for almost all study years.

Table 1
Comparison of JM distance for general LC class pairs using different multi-temporal NDVI inputs including deciduous forest (FO), crops (CR), hay and pasture (HA), and wetlands (WE). Pixels were randomly sampled from each class and the results were averaged from 10 sampling trials.

	FO–CR	FO–HA	FO–WE	CR–HA	CR–WE	HA–WE
2001						
Raw	1.324	1.244	0.787	0.577	1.310	1.226
SG	1.304	1.157	0.802	0.939	1.274	1.090
GA	1.316	1.131	0.784	1.065	1.286	0.961
DL	1.307	1.128	0.801	1.081	1.281	0.981
WS	1.300	1.166	0.815	0.892	1.292	1.141
FT	1.317	1.250	0.808	0.918	1.292	1.137
2006						
Raw	1.290	1.131	0.698	0.760	1.267	1.035
SG	1.316	1.171	0.770	0.802	1.305	1.131
GA	1.328	1.206	0.802	0.924	1.310	1.081
DL	1.323	1.199	0.824	0.910	1.302	1.069
WS	1.337	1.220	0.756	0.770	1.336	1.212
FT	1.317	1.192	0.757	0.806	1.307	1.148
2011						
Raw	1.317	1.206	0.743	0.390	1.309	1.201
SG	1.301	1.170	0.733	0.824	1.281	1.141
GA	1.311	1.180	0.785	0.867	1.285	1.093
DL	1.300	1.169	0.803	0.852	1.271	1.071
WS	1.322	1.216	0.740	0.711	1.319	1.209
FT	1.333	1.233	0.762	0.656	1.318	1.219

4. Discussion

There are many factors and criteria involved in our MODIS NDVI data smoothing analysis. The algorithms chosen for comparison were from TIMESAT package, a Whittaker smoother, and a FT method developed by the USEPA, but our comparing methods may be applied to any smoothing algorithms for multi-temporal imagery datasets. The results of within-class variability using Euclidean distance as a measure provided evidence that in the GLB, smoothing can reduce mean Euclidean distances of NDVI profiles when dealing with both general LC types and crop-specific classes such as corn and soybean. This information, although it does not reveal actual improvement in terms of classification accuracy statistics, does show that pre-processing can potentially improve 'cleanness' of NDVI time-series data. The mean Euclidean distance measure alone should not be used as the only criterion in selecting smoothing algorithms. In this study, the FT algorithm generated the highest mean Euclidean distances among five selected algorithms, but it actually led to much better classification accuracies, especially compared to the GA and DL approaches. The standard deviations of Euclidean distance measures appeared to be better associated with classification performances.

Table 2
Comparison of JM distance for corn/soybean separability using different multi-temporal NDVI inputs. Pixels were randomly sampled from each class and the results were averaged from 10 sampling trials.

	2011	2012	2013	2014
RAW	0.738	0.711	0.502	0.708
SG	0.726	0.657	0.757	0.858
GA	0.784	0.718	0.604	0.656
DL	0.777	0.640	0.598	0.635
WS	0.981	0.919	0.693	0.865
FT	1.030	0.789	0.738	0.790

Table 3
Accuracy assessment statistics including percent correct and (kappa statistic) for general LC classification using different MODIS raw and smoothed input data.

	RAW	SG	GA	DL	WS	FT
2001	71 (0.60)	73 (0.63)	67 (0.55)	67 (0.55)	73 (0.62)	73 (0.63)
2006	70 (0.59)	70 (0.59)	67 (0.55)	68 (0.56)	70 (0.59)	71 (0.60)
2011	69 (0.58)	69 (0.59)	66 (0.53)	67 (0.54)	70 (0.59)	71 (0.59)

JM distances have been proven to be useful for analyzing separability of one LC class from another. Atzberger and Eilers (2011a) found improved separability for almost all LC pairs by using a Whittaker smoother. In our study, the WS and FT algorithms appeared to be effective in improving class separability. The impacts of class separability from the remaining three smoothing algorithms, however, were inconclusive. A given smoothing algorithm may improve class separability of one specific class pair (e.g., cultivated crops vs. hay/pasture), while decrease separability of another class pair. Such class separability tradeoff makes comparison or selection of smoothing algorithm difficult. One smoothing algorithm should be favored (e.g., WS and FT algorithms here) only if a majority of class pairs showed improved class separability comparing to the raw NDVI dataset. We also note that we used PCA to reduce the linear dependence of time-series data before computing JM distances. For the original NDVI and smoothed data especially, there were strong linear dependences among the temporal NDVI signals (e.g., r = 0.94 for DOY 177 and DOY 193 in the 2001 NDVI GA-smoothed data). This linear dependency may lead to questionable JM measures. For example, using untransformed data, very high separability measures (JM > 1.30) might be generated for the class pairs corresponding to deciduous forest–wetland and crops–hay/pasture. However, the LC pairs are actually difficult to separate. Future research should be implemented to assess the use of JM distances, Euclidean distance, and other possible criteria for evaluating smoothing algorithms by allowing researchers to determine if there are positive relationships between those measures to classification accuracy statistics. One potential criterion is the signal-to-noise ratio (SNR) before and after data smoothing. Through a geostatistical analysis, Atzberger and Eilers (2011_b) observed an improved SNR after data smoothing. This method could be expanded to compare performances from multiple smoothing algorithms.

We used MODIS RIs as the primary data source to identify pixels with questionable NDVI values. However, MODIS RIs have their own reliability issue — good NDVI observations may be masked out while questionable observations may be kept in the time-series. Thus inaccurate RIs may negatively affect smoothing results. Tan et al. (2011) integrated MODIS surface reflectance QA information and land surface temperature products to define winter season and improved their NDVI smoothing performance. This procedure needs to be further examined for our multi-temporal LC classification research. The accuracy of LC reference data (e.g., NLCD and CDLs) is another important factor that may affect training data quality and classification performance evaluation. Both the NLCDs and CDLs used in this study had relatively high classification accuracy (e.g., overall accuracy > 85%), and their accuracy levels may be further increased when pixels are spatially-aggregated to coarser spatial resolutions.

Table 5
Accuracy assessment statistics percent correct and (kappa statistic) of crop-specific (corn and soybean) classification using different MODIS input data for raw data vs. smoothed products over four consecutive years.

	2011	2012	2013	2014
RAW	73% (0.45)	72% (0.44)	65% (0.30)	74% (0.47)
SG	75% (0.50)	74% (0.47)	69% (0.39)	73% (0.46)
GA	70% (0.41)	68% (0.32)	66% (0.31)	69% (0.38)
DL	69% (0.38)	63% (0.27)	63% (0.28)	70% (0.39)
WS	77% (0.53)	74% (0.47)	71% (0.41)	74% (0.48)
FT	75% (0.50)	73% (0.47)	70% (0.40)	74% (0.48)

One important finding from our study is that certain smoothing algorithms may actually reduce image classification performance for general LC and particularly crop-specific classification. For example, the overall accuracy of 2012 corn–soybean classification was 63% using DL filtered data compared to 72% accuracy using raw NDVI data. One possible reason is that aggressive smoothing process actually removed important temporal information and two LC classes may appear to be similar after the processing. Our FT data smoothing approach and the Whittaker smoother, on the other hand, have the advantage of maintaining 'good' original NDVI data (Lunetta et al., 2006; Atzberger & Eilers, 2011a). Note that our FT approach included one filtering step where new NDVI values were estimated for poor quality pixels while good quality pixels remain unchanged (the other four algorithms smoothed the entire time-series). Additionally, we further examined the other four other algorithms classification performance by including the same filtering step for corn–soybean classifications. Overall accuracies for SG algorithm improved to 76% and 70% for year 2011 and 2013, respectively (slightly outperforming the FT approach). There was no improvement for WS algorithm and it remained to be the best performer among five selected algorithms. The overall accuracies for GA and DL approaches improved for certain years (e.g., GA = 69% and DL = 68% for 2013), but they still ranked worse than other three algorithms. Among five selected algorithms, WS algorithm has another appealing

Table 4
Error matrices for time-series MODIS classification were generated using reference data derived from the NLCD 2001. Approximately 20% of the MODIS pixels were randomly selected for the accuracy assessment. Results from RAW NDVI data and FT filtered data were compared. LC classes included deciduous forest (FO), corn/soybean crops (CR), hay and pasture (HA), and wetlands (WE).

	Reference					
RAW	FO	CR	HA	WE	Total	% correct
FO	**149,689**	2673	2457	30,321	185,140	80.9
CR	5949	**171,594**	11,037	3939	192,519	89.1
HA	21,775	55,247	**55,826**	12,394	145,242	38.4
WE	32,337	3136	1884	**71,168**	108,525	65.6
Total	209,750	232,650	71,204	117,822	631,426	
% correct	71.4	73.8	78.4	60.4	Overall = 71% kappa = 0.60	
FT	FO	CR	HA	WE	Total	% correct
FO	**160,776**	6897	4777	28,288	200,738	80.1
CR	4629	**170,268**	10,075	3001	187,973	90.6
HA	12,972	49,905	**52,755**	8198	123,830	42.6
WE	31,373	5580	3597	**78,335**	118,885	65.9
Total	209,750	232,650	71,204	117,822		
% correct	76.7	73.2	74.1	66.5	Overall = 73% kappa = 0.63	

The diagonal elements (boldfaced) represent number of correctly classified pixels.

feature is that the smoothing parameter (λ) can be automatic adjusted using cross-validation of good quality NDVI data points (Eilers, 2003).

5. Conclusions

We designed this analysis to evaluate how smoothing algorithms for NDVI time-series data compare to one another, to increase the understanding of algorithm selection for pre-processing and potentially to improve the accuracy of multi-temporal LC classification efforts. We examined 14-years of MODIS NDVI data in the GLB using a variety of smoothing algorithms and LC reference datasets and used criteria to test which smoothing algorithms performed best for multi-temporal LC classifications. All smoothed data returned significantly reduced mean Euclidean distance measurements. The WS and FT algorithms, although generated the larger mean Euclidean distances, showed higher JM inter-class separability measure. Using different smoothed NDVI multi-temporal data as image classification inputs, we then compared accuracy statistics. Overall, the WS smoothing algorithm performed best and improved classification accuracy about 2% for general LC and about 6% for crop-specific classification for certain study years; which was statistically significant ($p = 0.05$). The GA and DL algorithms actually decreased classification performance compared to the use of raw NDVI data, suggesting the importance of pre-processing steps in the multi-temporal image classification.

Acknowledgments

The U.S. Environmental Protection Agency partially conducted the research described in this paper. Although this work was reviewed by EPA and has been approved for publication, it may not necessarily reflect official Agency policy. Mention of any trade names or commercial products does not constitute endorsement or recommendation for use. This research was partially funded by the U.S. EPA's Office of Research and Development and NASA's Ocean Biology and Biogeochemical Program under grant NNH15AZ42I.

References

Atkinson, P. M., Jeganathan, C., Dash, J., & Atzberger, C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, *123*, 400–417.

Atzberger, C., & Eilers, P. H. (2011a). Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. *International Journal of Remote Sensing*, *32*(13), 3689–3709.

Atzberger, C., & Eilers, P. H. (2011b). A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America. *International Journal of Digital Earth*, 365–386.

Beck, P. S., Atzberger, C., Høgda, K. A., Johansen, B., & Skidmore, A. K. (2006). Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, *100*(3), 321–334.

Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, *26*(5), 341–358.

Bruce, L. M., Mathur, A., Byrd, J., & John, D. (2006). Denoising and wavelet-based feature extraction of MODIS multi-temporal vegetation signatures. *GIScience & Remote Sensing*, *43*(1), 67–77.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, *91*(3), 332–344.

Eilers, P. H. (2003). A perfect smoother. *Analytical Chemistry*, *75*(14), 3631–3636.

Friedl, M. A., McIver, D. K., Hodges, J. C., Zhang, X., Muchoney, D., Strahler, A. H., ... Cooper, A. (2002). Global land cover mapping from MODIS: Algorithms and early results. *Remote Sensing of Environment*, *83*(1), 287–302.

Goward, S. N., Markham, B., Dye, D. G., Dulaney, W., & Yang, J. (1991). Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. *Remote Sensing of Environment*, *35*(2), 257–277.

Hermance, J. F. (2007). Stabilizing high-order, non-classical harmonic analysis of NDVI data for average annual models by damping model roughness. *International Journal of Remote Sensing*, 2801–2819.

Hird, J. N., & McDermid, G. J. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, *113*(1), 248–258.

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

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

Jönsson, P., & Eklundh, L. (2004). TIMESAT — A program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, *30*(8), 833–845.

Knight, J. F., Lunetta, R. S., Ediriwickrema, J., & Khorram, S. (2006). Regional scale land cover characterization using MODIS-NDVI 250 m multi-temporal imagery: A phenology-based approach. *GIScience & Remote Sensing*, *43*(1), 1–23.

Lopez, C. B., Jewett, E., Dortch, Q., Walton, B., & Hudnell, H. (2008). *Scientific assessment of freshwater harmful algal blooms.*

Loveland, T., Reed, B., Brown, J., Ohlen, D., Zhu, Z., Yang, L., & Merchant, J. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*(6–7), 1303–1330.

Lunetta, R. S., Knight, J. F., Ediriwickrema, J., Lyon, J. G., & Worthy, L. D. (2006). Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*, *105*(2), 142–154.

Lunetta, R. S., Shao, Y., Ediriwickrema, J., & Lyon, J. G. (2010). Monitoring agricultural cropping patterns across the Laurentian Great Lakes Basin using MODIS-NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, *12*(2), 81–88.

Ma, M., & Veroustraete, F. (2006). Reconstructing pathfinder AVHRR land NDVI time-series data for the Northwest of China. *Advances in Space Research*, *37*(4), 835–840.

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., ... Daloğlu, I. (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences*, *110*(16), 6448–6452.

NASA (National Aeronautical and Space Administration) (2015). *Cyanobacteria Assessment Network (CyAN) for freshwater systems: An early warning indicator for nuisance blumes using ocean color satellite.* NASA Ocean Biology and Biogeochemical Program (Grant NNH15AZ42I).

Paerl, H. W., & Huisman, J. (2008). Blooms like it hot. *Science-New York Then Washington*, *320*(5872), 57.

Reed, B. C. (2006). Trend analysis of time-series phenology of North America derived from satellite data. *GIScience & Remote Sensing*, *43*(1), 24–38.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46.

Richards, J. A., & Jia, X. (2006). *Remote sensing digital image analysis — Hardback.* Berlin/Heidelberg: Springer.

Roberts, D. H., Lehár, J., & Dreher, J. W. (1987). Time series analysis with clean — Part one — Derivation of a spectrum. *The Astronomical Journal*, *93*, 968.

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

Shao, Y., & Lunetta, R. S. (2011). Sub-pixel mapping of tree canopy, impervious surfaces, and cropland in the Laurentian Great Lakes Basin using MODIS time-series data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *4*(2), 336–347.

Shao, Y., Lunetta, R. S., Ediriwickrema, J., & Iiames, J. (2010). Mapping cropland and major crop types across the Great Lakes Basin using MODIS-NDVI data. *Photogrammetric Engineering & Remote Sensing*, *76*(1), 73–84.

Shao, Y., Lunetta, R. S., Macpherson, A. J., Luo, J., & Chen, G. (2013). Assessing sediment yield for selected watersheds in the Laurentian Great Lakes Basin under future agricultural scenarios. *Environmental Management*, *51*(1), 59–69.

Swets, D., Reed, B., Rowland, J., & Marko, S. (1999). A weighted least-squares approach to temporal smoothing of NDVI 1999 ASPRS Annual Conference, From Image to Information, Portland, Oregon, May 17–21, 1999. *Proceedings: Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing, CD-ROM, 1.*

Tan, B., Morisette, J. T., Wolfe, R. E., Gao, F., Ederer, G., Nightingale, J., & Pedelty, J. (2011). An enhanced TIMESAT algorithm for estimating vegetation phenology metrics from MODIS data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *4*(2), 361–371.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment*, *108*(3), 290–310.

Wickham, J., Stehman, S., Fry, J., Smith, J., & Homer, C. (2010). Thematic accuracy of the NLCD 2001 land cover for the conterminous United States. *Remote Sensing of Environment*, *114*(6), 1286–1296.

Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J. Y., Salas, W., & Moore, B. (2006). Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. *Remote Sensing of Environment*, *100*(1), 95–113.