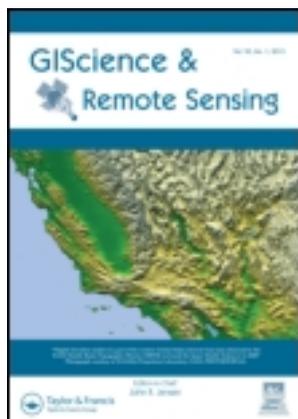


This article was downloaded by: [Montana State University Bozeman], [John A. Long]  
On: 25 August 2013, At: 15:04  
Publisher: Taylor & Francis  
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered  
office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## GIScience & Remote Sensing

Publication details, including instructions for authors and  
subscription information:

<http://www.tandfonline.com/loi/tgrs20>

### Object-oriented crop classification using multitemporal ETM+ SLC-off imagery and random forest

John A. Long <sup>a</sup>, Rick L. Lawrence <sup>a</sup>, Mark C. Greenwood <sup>b</sup>, Lucy  
Marshall <sup>a</sup> & Perry R. Miller <sup>a</sup>

<sup>a</sup> Department of Land Resources and Environmental Sciences,  
Montana State University, P.O. Box 173120, Bozeman, MT,  
59715-3120, USA

<sup>b</sup> Department of Mathematical Sciences, Montana State  
University, Bozeman, MT, 59715-3120, USA

Published online: 23 Jul 2013.

To cite this article: John A. Long, Rick L. Lawrence, Mark C. Greenwood, Lucy Marshall & Perry R. Miller (2013) Object-oriented crop classification using multitemporal ETM+ SLC-off imagery and random forest, *GIScience & Remote Sensing*, 50:4, 418-436

To link to this article: <http://dx.doi.org/10.1080/15481603.2013.817150>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## Object-oriented crop classification using multitemporal ETM+ SLC-off imagery and random forest

John A. Long<sup>a\*</sup>, Rick L. Lawrence<sup>a</sup>, Mark C. Greenwood<sup>b</sup>, Lucy Marshall<sup>a</sup> and Perry R. Miller<sup>a</sup>

<sup>a</sup>*Department of Land Resources and Environmental Sciences, Montana State University, P.O. Box 173120, Bozeman, MT 59715-3120, USA;* <sup>b</sup>*Department of Mathematical Sciences, Montana State University, Bozeman, MT 59715-3120, USA*

(Received 20 May 2013; final version received 14 June 2013)

The utility of Enhanced Thematic Mapper Plus (ETM+) has been diminished since the 2003 scan-line corrector (SLC) failure. Uncorrected images have data gaps of approximately 22% and gap-filling schemes have been developed to improve their usability. We present a method to classify a northeast Montana agricultural landscape using ETM+ SLC-off imagery without gap-filling. We use multitemporal data analysis and employ an object-oriented approach to define objects, agricultural fields, with cadastral data. This approach was assessed by comparison to a pixel-based approach. Results indicate that an ETM+ SLC-off image can be classified with better than 85% overall accuracy without gap-filling.

**Keywords:** remote sensing; agriculture; classification; multitemporal; multispectral; object-oriented; random forest; Enhanced Thematic Mapper Plus; Landsat

### Introduction

The Landsat series of satellites have provided researchers with a 40+ year archive of multispectral, moderate-resolution observations of the Earth. These observations comprise the world's longest continuously acquired collection and have been used by various educational, governmental, and industrial organizations, both within the United States and the international community (USGS 2012). Landsat-derived data are essential to researchers in diverse fields such as land cover mapping (e.g., Fuller, Groom, and Jones 1994), ecosystem and vegetation dynamics (e.g., Savage and Lawrence 2010), biophysical modeling (e.g., Cohen et al. 2003), and classification of urban landscapes (e.g., Lu and Weng 2005).

The United States Geological Survey (USGS) ceased acquisition of Landsat 5 Thematic Mapper (TM) imagery in November 2011 after detecting a failing amplifier in the satellite's downlink (Clark 2011). The loss of Landsat 5 left Landsat 7 as the sole operational Landsat satellite until May 2013 when Landsat 8 became operational. Landsat 7 carries the Enhanced Thematic Mapper Plus (ETM+) sensor; however, its scan-line corrector (SLC) failed permanently on 31 May 2003, leaving Landsat 7 with degraded performance (Markham et al. 2004). The SLC compensates for the satellite's forward motion during the cross-track scanning of the ETM+ sensor, thereby keeping the forward and reverse sweeps parallel to each other. The primary consequence of the SLC failure is gaps in coverage that range from none at the center of a scene to 14 pixels at

---

\*Correspondence author. Email: [john.long5@msu.montana.edu](mailto:john.long5@msu.montana.edu)

the scene's lateral edges. In total, data gaps comprise approximately 22% of an image (Markham et al. 2004). Images with these data gaps are known as SLC-off images, whereas those acquired prior to the SLC failure are designated SLC-on images (Arvidson et al. 2006). Data gaps are not identical for all spectral bands, and slight shifts in the position of the gaps result in the occasional pixel with valid data for some bands, but not others (Zhang 2007).

The radiometric performance and geometric characteristics of the ETM+ sensor have not been affected by the SLC failure (Markham et al. 2004). Researchers have continued to rely on SLC-off images, despite the availability of concurrent TM images, because of the superior quality of the ETM+ sensor. Improvements in quality include enhanced radiometric precision because of the addition of a two-gain state (low-gain and high-gain) to optimize sensitivity without detector saturation. Other improvements include better geodetic accuracy, reduced levels of spatial noise (e.g., banding and striping) due to improved electronics and detector calibration, a more reliable radiometric calibration, the addition of a panchromatic band, and substantial improvements in the spatial resolution of the thermal band (Masek et al. 2001). Research efforts that require Landsat imagery from 2012 must use SLC-off imagery as there is no alternative. We emphasize that complications arising from missing data are not restricted to ETM+ SLC-off imagery and might occur in imagery from any sensor due to a variety of causes such as sensor noise, clouds, or shadows. Consequently, recovering the missing data through various gap-filling schemes has become an increasing important endeavor.

Gap-filling schemes can be divided into three general approaches: (1) direct prediction of the missing data based on interpolation of the surrounding pixels; (2) predictions based on multitemporal Landsat data; and (3) predictions based on data from other sensors (Zeng, Shen, and Zhang 2013). Interpolation approaches attempt to infer missing values by using the values from nearby pixels in conjunction with an empirically derived formula or by using a geostatistical interpolation scheme such as kriging (Zhu, Liu, and Chen 2012). These approaches can improve the usability of SLC-off imagery, but under restrictive conditions. Landscapes in which pixel values can potentially change rapidly over a small distance, such as heterogeneous landscapes and landscapes with small or linear features, are difficult to predict when using approaches that rely on surrounding pixel values (Maxwell 2004; Zhu, Liu, and Chen 2012). Gap-filling is most useful in homogeneous areas with little landscape structural changes (Zeng, Shen, and Zhang 2013) or when the missing data represents small spatial gaps (Pringle, Schmidt, and Muir 2009). It also is unclear if interpolation approaches based on geostatistical methods meet the intrinsic stationarity assumption (Pringle, Schmidt, and Muir 2009).

Missing values have also been filled with a multitemporal Landsat approach by using an SLC-on image of the same area to fill the gaps in a SLC-off image (e.g., Storey et al. 2005; Maxwell, Schmidt, and Storey 2007); however, this method is unsuitable for applications that require spectral data from closely spaced dates. A good example is the classification of specific crops because dates more than 2 weeks apart might represent different stages in phenological development. An alternative is a multi-sensor approach in which data from a different sensor on the same, or nearly the same, day is used to predict the missing data. The multi-sensor approach to filling ETM+ gaps has been partially successful with a variety of auxiliary sensors. The Advanced Land Imager (ALI) aboard the Earth Observer One (EO-1) satellite (Bolorani, Erasmi, and Kappas 2008), the Moderate Resolution Imaging Spectroradiometer (MODIS) (Roy et al. 2008), the Charge Coupled Device (CCD) camera aboard the China Brazil Earth Resources

Satellite (CBERS-02B) (Chen, Tang, and Qiu 2010), and the Linear Imaging Self-Scanning System aboard the Indian Remote Sensing Satellite (IRS/1D) (Reza and Ali 2008) have all been used. Nonetheless, concerns regarding spectral compatibility, spatial resolution, clouds, and financial constraints remain (Zeng, Shen, and Zhang 2013).

Gap-filling approaches predict the missing values of individual pixels; however, not all remote sensing tasks require pixel-level data. Classification algorithms, for example, can operate on individual pixels (pixel-based classifiers) or on groups of contiguous pixels (object-oriented (O-O) classifiers). Classification using an O-O approach is most appropriate when the landscape consists of areas that can be delineated into meaningful homogeneous regions, such as agricultural fields (e.g., Forster et al. 2010). The O-O approach classifies at the object-level; the number of pixels in an object is largely irrelevant if the object is reasonably homogeneous. An O-O approach to classification, therefore, does not require estimation of the missing data in SLC-off imagery provided there are a sufficient number of extant pixels in the objects (Turker and Arıkan 2005).

Classification of agricultural regions at the pixel level presents a challenge in that monoculture fields are often classified as multiple classes. The Cropland Data Layer (CDL), a geo-referenced, crop-specific data layer derived from a fusion of moderate-resolution satellite imagery (USDA 2012), contains numerous examples. A number of studies conclude that O-O approaches produce better results (e.g., Pedley and Curran 1991; Janssen and van Amsterdam 1996). There are several reasons why O-O methods outperform pixel-based methods. A fundamental reason is that classification accuracy decreases with increasing spectral variability (e.g., Ashish, McClendon, and Hoogenboom 2009). Variability is high in heterogeneous landscapes and lower in homogeneous landscapes, such as agricultural fields. One compelling reason supporting the O-O approach in the classification of crops is that management decisions are generally made on a per-field basis (Forster et al. 2010). The O-O approach can incorporate ancillary information such as spatial relationships and management practices that, while not unique to the O-O approach, can be advantageous. The use of agricultural field boundaries defines the spatial relationship and identifies a particular object. Good results have resulted from merging images with a suitable vector layer, such as cadastral data (e.g., Wu et al. 2007; Watts et al. 2009).

The use of multitemporal imagery to classify vegetation is well-documented and has improved crop classification accuracies by 6–9% (e.g., El-Magd and Tanton 2003; Conrad et al. 2011). The main advantage is that this approach uses images from different dates within a single growing season to better capture the spectral diversity in crops due to differences in phenological stages (Gómez-Casero et al. 2010). The data are fused into a single classification scheme for the growing season. Multitemporal analysis is particularly attractive in crop classification for several reasons: (1) the spectral characteristics of a crop varies with phenological stage; (2) agricultural areas are spatially dynamic in that adjacent fields might be in different phenological stages, thereby increasing heterogeneity; and (3) criteria based on phenological stage have ecological significance and are easier to generalize and more robust than statistical criteria derived from a specific dataset (Simonneaux et al. 2008). The use of multitemporal imagery raises obvious questions regarding quantity and seasonal timing. Research suggests that classification accuracy increases with the number of images and that images acquired at peak growth are more useful than those from low-growth periods (Pax-Lenney and Woodcock 1997). Others have presented similar findings (e.g., Oetter et al. 2000). Multitemporal classifications have produced accuracies exceeding 84% (e.g., Pax-Lenney and Woodcock 1997; Wardlow, Egbert, and Kastens 2007; Simonneaux et al. 2008; Gómez-Casero et al. 2010).

Random forest (RF) is a tree-based ensemble method useful for classification when there are many weak explanatory variables (Breiman 2001) and is well-established in the remote sensing literature. RF classification is often superior to standard classification approaches, performing well with large numbers of classes (e.g., Rodrigues-Galiano et al. 2012), with difficult to classify data (e.g., Lawrence, Wood, and Sheley 2006), and routinely achieving classification accuracies well above 90% (e.g., Chapman et al. 2010). It generates an internal accuracy assessment that has been shown to be equivalent to an independent accuracy assessment (assuming no bias in the data), thereby, eliminating the need for separate validation data (Lawrence, Wood, and Sheley 2006). One of the advantages of RF is that it can accommodate high dimensional and redundant data, negating the need to reduce the dimensionality of the data or to transform the data to orthogonal components (Breiman 2001). This property permits, for example, the use of NDVI as a predictor simultaneously with bands 3 and 4. RF has been used in agricultural studies to classify agricultural practices (e.g., Watts et al. 2009), species-specific crop discrimination (e.g., Adam et al. 2012; Chapman et al. 2010), and multi-seasonal land cover classification (e.g., Rodrigues-Galiano et al. 2012).

We present here a method to classify a northeast Montana agricultural landscape into the region's dominant classes using SLC-off ETM+ imagery from the 2012 growing season – without using gap-filling schemes. Crops in the cereal class (Cereal) are familiar grain crops and include spring wheat (*Triticum aestivum*), durum wheat (*T. durum*), and oat (*Avena sativa*). The pulse crops (Pulse) are leguminous annual crops that include dry pea (*Pisum sativum*) and lentil (*Lens culinaris*). The other class (Other) refers to agricultural land that is either in fallow or is growing something other than cereal or pulse crops. This method is part of a larger project that seeks to investigate the spatio-temporal patterns of adoption for cereal-pulse crop rotation in northeast Montana during 1994–2012 using geospatial techniques. Two images, one from mid-season and one from late-season, are used to capture phenological variations. We employ an O-O approach and define objects, agricultural fields, with regional cadastral data. Spectral data are extracted from the extant pixels in an object and used in a random forest classification to identify the class of crop in a particular field. The usefulness of the O-O approach was assessed by comparison to a pixel-based approach, each with a random forest classifier.

## Methods

### Study area

The study area is in northeast Montana and includes the counties of Daniels, Sheridan, and Roosevelt, and the eastern portion of Valley (Figure 1). The region is bounded by Canada to the north, North Dakota to the east, the Missouri River to the south, and public lands such as the Bitter Creek Wilderness Study Area along much of the western border. The area is in the Missouri Plateau (glaciated) region of the Great Plains Physiographic Province and is characterized by low relief. The climate is semi-arid with low humidity. Average total annual precipitation is just over 31 cm, which is primarily rain during April to September (WRCC 2012). Average daily maximum temperatures in July are 31°C, while January temperatures average –10°C (WRCC 2012). Agriculture and shortgrass prairie are the dominant land cover types. Agricultural practices in the study area are primarily dry land cropping systems (Tanaka et al. 2010). Regional agriculture is largely cereal crops, primarily spring wheat, and increasingly substantial acreages of pulse crops (NASS 2012).

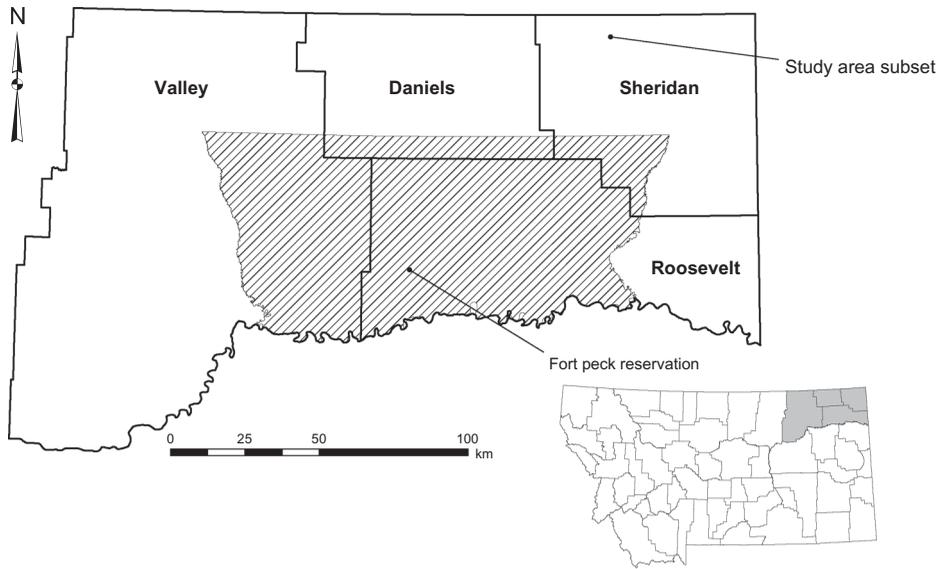


Figure 1. The study area is located in northeastern Montana and consists of four counties: Daniels, Roosevelt, Sheridan, and Valley; it also encompasses the Fort Peck Reservation. The study area subset is located in northwest Sheridan County.

### *Data acquisition*

The study area required three Landsat scenes: Path 36, Row 26 (36/26); Path 36, Row 27 (36/27); and Path 35, Row 26 (35/26). Due to the positions of these three overlapping scenes relative to the study region, data loss due to SLC-off gaps was estimated at 15%–18%. SLC-off ETM+ images from the 2012 growing season were obtained from the USGS's Earth Resources Observation and Science Center (EROS) and were used to create two mosaics. The mid-season mosaic combined scenes from 11 July and 18 July, whereas the late-season mosaic used scenes from 12 August and 19 August. The disparity in dates for the mosaics was because the study area required imagery from two adjacent paths; thus, an 8-day difference. We chose these dates because they were the only usable images from the 2012 growing season; images from the remaining dates had excessive cloud coverage. All images from Path 36 were cloud-free, whereas the images from Path 35 had occasional cumulus clouds obscuring small portions of the study area.

Ancillary vector data were also used: the Montana cadastral framework and the Montana land cover framework. The cadastral framework is a continuously updated geodatabase of private (taxable) and public (exempt) lands (MSL 2012). The cadastral data are maintained by the Montana Department of Revenue and are available from the Montana State Library (MSL) Geographic Information Services as a vector layer. The land cover framework is a geodatabase that consists of land cover classes at three levels of detail. The definitions of the land cover classes follow those in the National Land Cover Database 2001 (NLCD2001) and the Northwest Regional Gap Analysis Project (NWGAP) (MSL 2010). The land cover data are maintained by the Montana Natural Heritage Program and are available from the MSL as a vector layer. We used the June 2012 version of the cadastral framework and the May 2010 version of the land cover framework.

Ground reference data were collected from 525 randomly chosen sites within the study area during July 2012. Sites were required to be agricultural in nature and viewable from public rights-of-way. We recorded the appropriate class (Cereal, Pulse, or Other) for each site (some producers in the area manage fields by alternating strips of a crop and fallow; these fields were classified as the appropriate crop – cereal or pulse). Some reference sites were not used because they: (1) were not accessible; (2) were not actually agriculture; (3) were duplicate observations of the same land parcel; or (4) had clouds in both the mid-season and the late-season mosaics. The final analysis used data from 434 of the ground reference sites; 278 (64%) were Cereal, 70 (16%) were Other, and Pulse comprised 86 sites (20%). These percentages are similar to those reported by the National Agricultural Statistics Service for the region (NASS 2012).

### ***Image and vector data processing***

All images were level-one terrain-corrected (L1T) products. Invalid pixels, caused by slight shifts in the position of the data gaps per band, were removed by masking out values less than 2. The mid-season mosaic was created in several steps. The two images from Path 36 (36/26 and 36/27) were directly combined because they were from the same day. This combined image is hereafter referred to as the “western” image. The third image (35/26), hereafter the “eastern” image, could not be combined with the western image as it was from 8 days earlier. We combined the eastern and western images as follows: (1) cloud-free and shadow-free regions from the overlapping areas of both images were used to derive a linear regression model with the western image as the reference; (2) the coefficients from the regression model were used to normalize the eastern image to the western image; (3) the western image and the normalized eastern image were combined into a mosaic such that missing data pixels in the western image were automatically filled, if available, by pixel values from the eastern image; (4) any remaining pixels in the final mosaic with invalid data were masked out; and (5) the mosaic was clipped to agriculture-only regions using the land cover vector data. The late-season mosaic was created using the same procedure.

Small clouds and shadows were a persistent problem in both mosaics, primarily in the southeastern corner of the study area. Clouds and shadows in each mosaic were classified with a supervised algorithm and a binary (cloud/shadow or no cloud/shadow) mask created from each mosaic. These binary masks were intersected to identify areas with clouds or shadows in both mosaics and a final binary (clouds or shadows in both mosaics or not) mask created. We used this final binary mask to eliminate pixels that had clouds or shadows in *both* mosaics. Pixels with clouds or shadows in a single mosaic were not removed as long as the corresponding pixels in the other mosaic were clear.

We sought a minimum field size of 16.2 ha (40 acres) to ensure that our analyses focused on the region’s major producers – the median agricultural field in the study area is approximately 64.7 ha (160 acres). The cadastral layer was consequently modified by removing all land parcels less than 16.1 ha (39.8 acres) in area; this cutoff allowed for small variations in minimum-sized fields. Finally, we examined the cadastral layer with respect to SLC-off gaps. Some land parcels, particularly the smaller ones, were entirely in a data gap, while others had only a portion of the parcel within a gap. We eliminated land parcels with less than 40 pixels of data to obtain a large enough sample size from a field that any statistical measures (e.g., mean spectral value of the blue band for the field) would closely approximate the entire field. This resulted in 13,143 usable agricultural fields in the study area.

### ***Image segmentation and data extraction***

The O-O approach required aggregating the pixels into objects – agricultural fields in this case. Both mosaics were segmented solely on the cadastral layer; land parcels were now objects. Pixels without spectral data (the gaps) were defined as not valid. Finally, we extracted the mean and standard deviation of each spectral band for each object using only the valid pixels (e.g., if an object/field consisted of X pixels, of which Y pixels lacked data because of data gaps, the mean and standard deviation was based on the X–Y valid pixels). These values were exported to a spreadsheet for subsequent analysis. Data for the pixel-based analysis were extracted from the mosaics using a shape file for each class as a spatial filter and exported to a spreadsheet for analysis.

### ***Classification***

The default classification procedure used a RF classifier on multitemporal data for regions of the study area that were free of clouds and shadows on both image dates. This was the case for both the O-O and the pixel-based approach. Adjustments were necessary for areas that were free of clouds or shadows only on a single date. We first describe the multitemporal procedure. RF models were developed from data extracted from the reference fields from the mid- and late-season mosaics. The number of potential variables depended on the approach. The multitemporal O-O approach had 45 potential variables: (1) the mean value of each spectral band for each of the two dates (14 variables); (2) the standard deviation of each spectral band for each date (14 variables); (3) the normalized difference vegetation index (NDVI) for each date (2 variables); and (4) the date-wise change in all variables, for example the difference between mid-season NDVI and late-season NDVI (15 variables). Standard deviations are not applicable at the pixel-level. Consequently, a separate RF model was developed for the multitemporal pixel-based approach. This model used only 24 variables: (1) the mean value of each spectral band for each date (14 variables); (2) NDVI for each date (2 variables); and (3) the date-wise changes (8 variables).

Portions of the study area obscured by clouds or shadows on one date were classified with RF models developed from the single clear image; areas obscured in the mid-season mosaic used data from the late-season mosaic and vice versa. Consequently, the RF model for the single-date O-O approach had 15 potential variables: 7 mean spectral values, 7 standard deviations, and NDVI. The single-date pixel-based approach used only 8 variables: 7 mean spectral values and NDVI. Summarizing, we developed a total of *six* RF models based on approach and the status of clouds. We predicted class membership based on data from both dates when possible; single-date classifications comprised less than 1.4% of the cases.

The data, regardless of approach or number of variables, were classified with a RF model as cereal crops, pulse crops, or other. Adjustable parameters included the number of trees, the number of variables to try at each split, and sample size for each class. We developed RF models with 1000 trees, though error rates were asymptotic with substantially fewer trees. Optimal values for the number of variables per split and class sample sizes were found iteratively with the objective of minimizing the overall classification error rate while maintaining approximately equal class error rates. We used the resultant RF models to predict class membership for the entire study area. This gave predicted classes for complete agricultural fields when using the O-O approach, despite data gaps in the SLC-off imagery, whereas the pixel-based approach can only predict the portion with extant data. Classification using the O-O approach produced vector-based maps, whereas classification based on pixels yielded raster maps.

We computed classification accuracy for all RF models via an error matrix (Congalton and Green 2009). The RF algorithm uses a random subset of approximately two-thirds of the data (sampling with replacement) to build multiple classification trees. The remaining data, the out-of-bag (OOB) observations, are used to compute a classification accuracy for each tree in the RF model, with the results from all trees combined to form the final OOB accuracy assessment (Breiman 2001). Overall and class accuracies based on OOB estimates have been shown to be reliable under the assumption that the reference data are unbiased (Lawrence, Wood, and Sheley 2006). We guarded against bias by randomly selecting our reference fields. Cohen's kappa statistic ( $\hat{\kappa}$ ) was calculated for each model as a measure of its predictive ability while accounting for correct classifications due to chance agreement (e.g., Congalton and Green 2009). We also tested whether the difference in kappa statistics between the O-O approach and the pixel-based approach was statistically significant. Finally, we extracted variable importance from each RF model to identify which variables contributed most to decreasing the overall error rate.

## Results

### *Object-oriented approach to classification*

The multitemporal approach was able to classify the object-level data from SLC-off ETM+ imagery, despite data gaps, and had an overall accuracy of 85.5% with a  $\kappa$  statistic of 0.75 (Table 1). Producer's accuracies (errors are due to omission) were

Table 1. Error matrices for object-oriented classifications.

Classified data	Reference data			User's accuracy <sup>a</sup>
	Cereal	Other	Pulse	
<b>Multitemporal</b>				
Cereal	187	4	8	94.0%
Other	21	54	5	67.5%
Pulse	7	5	59	81.9%
Producer's accuracy <sup>b</sup>	87.0%	84.4%	81.9%	
Overall accuracy	85.5%			
Kappa	0.75			
<b>Mid-season only</b>				
Cereal	180	10	6	91.8%
Other	34	49	8	53.9%
Pulse	15	7	62	73.8%
Producer's accuracy <sup>b</sup>	78.6%	74.2%	81.6%	
Overall accuracy	78.4%			
Kappa	0.63			
<b>Late-season only</b>				
Cereal	210	16	19	85.7%
Other	25	42	9	55.3%
Pulse	19	9	49	63.6%
Producer's accuracy <sup>b</sup>	82.7%	62.7%	63.6%	
Overall accuracy	75.6%			
Kappa	0.55			

Notes: <sup>a</sup>User's accuracy is measured with respect to errors of commission.

<sup>b</sup>User's accuracy is measured with respect to errors of omission.

approximately equal for all classes, ranging from 81.9% to 87.0%. There was more disparity in the user's accuracies (errors are due to commission), which ranged from 67.5% for the Other class to 94.0% for the Cereal class. Confusion between Cereal and Other accounted for slightly more than one-half of the errors, primarily where Cereal fields were classified as Other.

Single-date classifications (which were applied to less than 1.4% of the study area) with the mid-season data and the late-season data were less accurate than the multitemporal classification. The mid-season model had an overall accuracy of 78.4% with a  $\kappa$  statistic of 0.63. Producer's accuracies ranged from 74.2% to 81.6%, while user's accuracies ranged from 53.9% to 91.8%. The late-season data were slightly less accurate with overall accuracy of 75.6% and a  $\kappa$  statistic of 0.55. Producer's accuracies ranged from 62.7% to 82.6%, and user's accuracies ranged from 55.3% to 85.7%. Confusion between Cereal and Other continued to represent the majority of errors. Testing the significance of the differences in kappa statistics suggests that the multitemporal approach gives a statistically better classification ( $\alpha = 0.05$ ) than either of the single-date classifications (Table 3).

The variables that contributed most to decreasing the overall error rate differed according to the specific RF model (Figure 2 and Table 4). We refer hereafter to the following ETM+ spectral bands by band number: 1 = blue (0.45–0.515  $\mu\text{m}$ ), 2 = green (0.525–0.605  $\mu\text{m}$ ), 3 = red (0.63–0.69  $\mu\text{m}$ ), 4 = near infrared (0.75–0.90  $\mu\text{m}$ ), 5 = middle infrared (1.55–1.75  $\mu\text{m}$ ), 6 = thermal (10.4–12.5  $\mu\text{m}$ ), 7 = middle infrared (2.09–2.35  $\mu\text{m}$ ). The multitemporal O-O classification relied most on mid-season values from bands 2 and 4 for overall classification. Differences between mid-season and late-season values were very important as 6 of the top 8 variables as ranked by importance were differenced values, while standard deviation variables appeared only once among the top 8 variables. The single-date O-O classifications, in contrast, did not have differenced variables available and relied primarily on mean spectral band values rather than the standard deviations. NDVI and data from band 6 tended to be less important and were not among the top 8 most important variables.

The variables contributing the most to decreasing the error rate for specific classes also varied according to model, as well as class. The most important variables in the O-O multitemporal model (Table 4) for the Cereal and Pulse classes, as in the overall

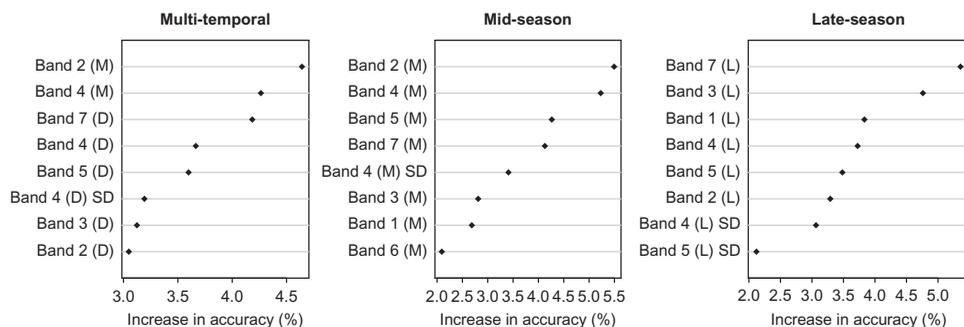


Figure 2. The top 8 variables in the O-O classification as ranked by order of importance. Band 1–7 refers to the appropriate ETM+ spectral band: 1 = blue (0.45–0.515  $\mu\text{m}$ ), 2 = green (0.525–0.605  $\mu\text{m}$ ), 3 = red (0.63–0.69  $\mu\text{m}$ ), 4 = near infrared (0.75–0.90  $\mu\text{m}$ ), 5 = middle infrared (1.55–1.75  $\mu\text{m}$ ), 6 = thermal (10.4–12.5  $\mu\text{m}$ ), 7 = middle infrared (2.09–2.35  $\mu\text{m}$ ). Parenthetical annotations indicate the source: mid = values from the mid-season mosaic, late = values from the late-season mosaic, D = differenced values. SD = standard deviation.

classification, were bands 2 and 4. Discrimination between Cereal and Pulse was based, in part, on differenced values from bands 2 and 7. The Other class relied most on differenced values from bands 5 and 7, while contributions from the mid-season band 2 dropped substantially in importance.

The most important variables for overall accuracy of the mid-season O-O classification (Table 5) were bands 2, 4, 5, and 7. The same variables contributed the most to classifying the Cereal and the Other classes, but with slight changes in order. The most important variables for the Pulse class were bands 2 and 4, while the standard deviation of bands 3 and 4 were ranked higher than in either of the other classes or in the overall classification. In contrast, none of the late-season O-O classifications (Table 6) were especially dependent on bands 2 and 4. The most important variables in the late-season classifications were instead bands 1, 3, and 7. These were the top three variables for the overall classification and each class with the exception of Pulse, in which band 3 was of lesser importance and band 1 was not among the top 8 variables.

### *Pixel-based approach to classification*

The multitemporal approach was able to classify the pixel-level data, except where data were missing due to SLC-off issues, and had an overall accuracy of 89.4% with a  $\kappa$  statistic of 0.81 (Table 2). These accuracies were approximately 5% higher than the corresponding O-O models. Producer's accuracies were nearly equal for all classes,

Table 2. Error matrices for pixel-based classifications.

Classified data	Reference data			User's accuracy <sup>a</sup>
	Cereal	Other	Pulse	
<b>Multitemporal</b>				
Cereal	206615	5286	5183	95.2%
Other	16257	57308	3171	74.7%
Pulse	7540	1267	63232	87.8%
Producer's accuracy <sup>b</sup>	89.7%	89.7%	88.3%	
Overall Accuracy	89.4%			
Kappa	0.81			
<b>Mid-season only</b>				
Cereal	193927	7348	9183	92.2%
Other	22991	54114	3877	66.8
Pulse	13494	2399	58526	78.6
Producer's accuracy <sup>b</sup>	84.2%	84.7%	81.8%	
Overall accuracy	83.8%			
Kappa	0.71			
<b>Late-season only</b>				
Cereal	188841	7090	8745	92.3%
Other	20317	52609	5473	67.1%
Pulse	21254	4162	57368	69.3%
Producer's accuracy <sup>b</sup>	82.0%	82.4%	80.1%	
Overall accuracy	81.7%			
Kappa	0.68			

Notes: <sup>a</sup>User's accuracy is measured with respect to errors of commission.

<sup>b</sup>User's accuracy is measured with respect to errors of omission.

Table 3. Tests for statistically significant differences between kappa statistics.

Models compared <sup>a</sup>	Test statistic (Z)	p-value
<b>O-O multitemporal</b> vs. O-O mid-season	2.43	<0.01
<b>O-O multitemporal</b> vs. O-O late-season	4.00	<0.01
<b>Pixel-based multitemporal</b> vs. Pixel-based mid-season	68.79	<0.01
<b>Pixel-based multitemporal</b> vs. Pixel-based late-season	91.04	<0.01
O-O multitemporal vs. <b>Pixel-based multitemporal</b>	1.83	0.03
O-O mid-season vs. <b>Pixel-based mid-season</b>	2.21	0.01
O-O late-season vs. <b>Pixel-based late-season</b>	3.36	<0.01

Note: <sup>a</sup>The models in bold are those with the larger kappa statistic.

ranging from 88.3% to 89.7%; the user's accuracies ranged from 74.7% to 95.2%. Confusion between Cereal and Other, as found in models using the object-level data, comprised over one-half of the errors.

Single-date classifications with the mid-season and late-season data were less accurate than the multitemporal classification, the same pattern found with the O-O data. The mid-season model had an overall accuracy of 83.8% with a  $\kappa$  statistic of 0.71. Producer's accuracies ranged from 81.8% to 84.7%, and user's accuracies ranged from 66.8% to 92.2%. The late-season model had an overall accuracy of 81.7% and a  $\kappa$  statistic of 0.68, slightly lower than the mid-season model. Producer's and user's accuracies were similar to mid-season values with the exception that the user's accuracy for Pulse was substantially lower in the late-season model than in the mid-season model. Confusion between Cereal and Other comprised the majority of errors. The pixel-based multitemporal approach produced a statistically better classification ( $\alpha = 0.05$ ) than either of the single-date classifications, and the pixel-based approaches produced statistically better classifications than corresponding O-O approaches (Table 3).

The most important variables (Figure 3 and Table 4) for the overall accuracy of the pixel-based classifications differed by model, and were generally not the same variables important in the O-O classifications. Pixel-based multitemporal classifications relied most on values from bands 4 and 6. This is in direct contrast to the O-O classifications in which

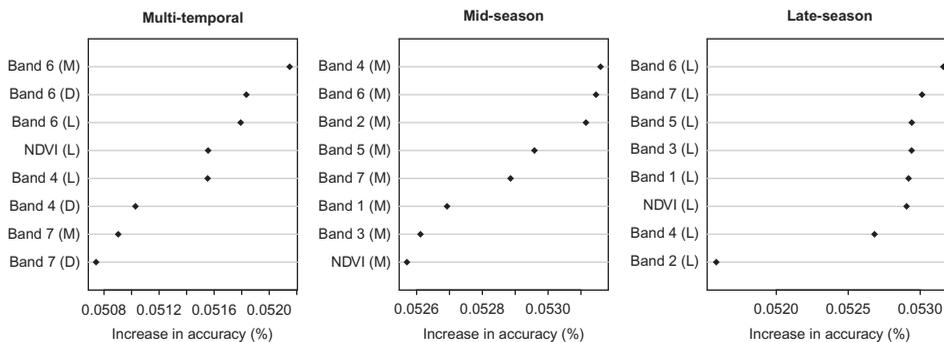


Figure 3. The top 8 variables in the pixel-based classification as ranked by order of importance. Band 1–7 refers to the appropriate ETM+ spectral band: 1 = blue (0.45–0.515  $\mu\text{m}$ ), 2 = green (0.525–0.605  $\mu\text{m}$ ), 3 = red (0.63–0.69  $\mu\text{m}$ ), 4 = near infrared (0.75–0.90  $\mu\text{m}$ ), 5 = middle infrared (1.55–1.75  $\mu\text{m}$ ), 6 = thermal (10.4–12.5  $\mu\text{m}$ ), 7 = middle infrared (2.09–2.35  $\mu\text{m}$ ). Parenthetical annotations indicate the source: mid = values from the mid-season mosaic, late = values from the late-season mosaic, D = differenced values. SD = standard deviation.

Table 4. The top 8 variables for multitemporal classification as ranked by importance<sup>a</sup>.

Rank	O-O multitemporal				Pixel-based multitemporal			
	Overall	Cereal	Other	Pulse	Overall	Cereal	Other	Pulse
1	B2 (M)	B2 (M)	B7 (D)	B2 (M)	B6 (M)	B6 (M)	B4 (M)	B4 (D)
2	B4 (M)	B4 (M)	B5 (D)	B4 (M)	B6 (L)	B6 (L)	B6 (L)	B7 (D)
3	B7 (D)	B4 (D)	B4 (M)	B7 (D)	B6 (D)	B6 (D)	B6 (D)	B6 (M)
4	B4 (D)	B2 (D)	B3 (D)	B4 (D)	B4 (L)	B2 (M)	B6 (M)	B6 (L)
5	B5 (D)	B7 (L)	B4 (D) SD	B7 (L)	NDVI (L)	B4 (L)	B4 (D)	B7 (M)
6	B4 (D) SD	B1 (M)	NDVI (D)	B4 (D) SD	B4 (D)	NDVI (L)	B7 (D)	B6 (D)
7	B3 (D)	B5 (D)	B4 (D)	B2 (D)	B7 (M)	B4 (M)	B7 (L)	B7 (L)
8	B2 (D)	B5 (L)	B1 (D)	B5 (L)	B4 (M)	B4 (D)	B5 (M)	B4 (L)

Notes: <sup>a</sup>B1–7 refers to the appropriate ETM+ spectral band: 1 = blue (0.45–0.515 μm), 2 = green (0.525–0.605 μm), 3 = red (0.63–0.69 μm), 4 = near infrared (0.75–0.90 μm), 5 = middle infrared (1.55–1.75 μm), 6 = thermal (10.4–12.5 μm), 7 = middle infrared (2.09–2.35 μm). Parenthetical annotations indicate the source: M = values from the mid-season, L = values from the late-season mosaic, D = differenced values. SD = standard deviation.

band 6 was much less important. Differenced values were less important than they were in the O-O approach, only 2 of the top 8 variables ranked by importance were differenced values. Pixel-based single-date classifications, with only 8 potential variables, relied primarily on the spectral band values, while NDVI ranked low in importance.

Variable importance was also class-dependent. In the pixel-based multitemporal classification (Table 5), contributions from the mid-season, late-season, and differenced values of band 6 remained strong for all classes. They accounted for 3 of the top 4 variables for all but the Pulse class, in which bands 4 and 7 were substantially more important. Mid-season band 4 values were especially useful for distinguishing the Other class, while band 4 differenced values were most important for the Pulse class.

Table 5. The top 8 variables for mid-season classification as ranked by importance<sup>a</sup>.

Rank	O-O mid-season				Pixel-based mid-season			
	Overall	Cereal	Other	Pulse	Overall	Cereal	Other	Pulse
1	B2 (M)	B2 (M)	B4 (M)	B4 (M)	B4 (M)	B2 (M)	B6 (M)	B4 (M)
2	B4 (M)	B5 (M)	B7 (M)	B2 (M)	B2 (M)	B6 (M)	B4 (M)	B2 (M)
3	B5 (M)	B7 (M)	B5 (M)	B4 (M)	B6 (M)	B4 (M)	B7 (M)	B6 (M)
4	B7 (M)	B4 (M)	B2 (M)	B3 (M)	B5 (M)	B5 (M)	B2 (M)	B5 (M)
5	B4 (M) SD	B3 (M)	B1 (M)	B3 (M)	B7 (M)	B7 (M)	B3 (M)	B1 (M)
6	B3 (M)	B1 (M)	B4 (M)	B7 (M)	B1 (M)	NDVI (M)	B1 (M)	B3 (M)
7	B1 (M)	B4 (M)	B7 (M)	NDVI (M)	NDVI (M)	B1 (M)	NDVI (M)	B7 (M)
8	B3 (M) SD	B6 (M)	NDVI (M)	B5 (M)	B3 (M)	B3 (M)	B5 (M)	NDVI (M)

Notes: <sup>a</sup>B1–7 refers to the appropriate ETM+ spectral band: 1 = blue (0.45–0.515 μm), 2 = green (0.525–0.605 μm), 3 = red (0.63–0.69 μm), 4 = near infrared (0.75–0.90 μm), 5 = middle infrared (1.55–1.75 μm), 6 = thermal (10.4–12.5 μm), 7 = middle infrared (2.09–2.35 μm). Parenthetical annotations indicate the source: M = values from the mid-season, L = values from the late-season mosaic, D = differenced values. SD = standard deviation.

Table 6. The top 8 variables for late-season classification as ranked by importance<sup>a</sup>.

Rank	O-O late-season				Pixel-based late-season			
	Overall	Cereal	Other	Pulse	Overall	Cereal	Other	Pulse
1	B7 (L)	B7 (L)	B3 (L)	B7 (L)	B6 (L)	B6 (L)	B6 (L)	B6 (L)
2	B3 (L)	B1 (L)	B7 (L)	B4 (L) SD	B7 (L)	B1 (L)	B3 (L)	B7 (L)
3	B1 (L)	B3 (L)	B1 (L)	B5 (L)	B1 (L)	B5 (L)	B1 (L)	B5 (L)
4	B4 (L)	B4 (L)	NDVI (L)	B3 (L)	B5 (L)	B3 (L)	B4 (L)	B3 (L)
5	B5 (L)	B5 (L)	B4 (L)	B4 (L)	B3 (L)	NDVI (L)	B7 (L)	B4 (L)
6	B2 (L)	B2 (L)	B7 (L) SD	B7 (L) SD	NDVI (L)	B4 (L)	B5 (L)	NDVI (L)
7	B4 (L) SD	B4 (L) SD	B2 (L)	B4 (L) SD	B4 (L)	B7 (L)	NDVI (L)	B1(L)
8	B7 (L) SD	B6 (L)	B6 (L) SD	B2 (L)	B2 (L)	B2 (L)	B2 (L)	B2 (L)

Notes: <sup>a</sup>B1–7 refers to the appropriate ETM+ spectral band: 1 = blue (0.45–0.515  $\mu\text{m}$ ), 2 = green (0.525–0.605  $\mu\text{m}$ ), 3 = red (0.63–0.69  $\mu\text{m}$ ), 4 = near infrared (0.75–0.90  $\mu\text{m}$ ), 5 = middle infrared (1.55–1.75  $\mu\text{m}$ ), 6 = thermal (10.4–12.5  $\mu\text{m}$ ), 7 = middle infrared (2.09–2.35  $\mu\text{m}$ ). Parenthetical annotations indicate the source: M = values from the mid-season, L = values from the late-season mosaic, D = differenced values. SD = standard deviation.

The most important variables for the overall accuracy of the pixel-based mid-season classification (Table 5) were bands 2, 4, 5, and 6; results similar to the O-O classifications. These bands were the top 4 variables for all but the Other class, in which band 7 was more important and band 5 decidedly less so. The most important variable for overall accuracy of the pixel-based late-season classification (Table 6) was band 6; this was the case for all classes as well. Bands 1 and 7 provided the distinctions between classes. Band 1 was of high importance for the Cereal class, intermediate for Other, and low for Pulse; band 7 showed the opposite pattern.

A comparison of the predicted maps with respect to the spatial distribution of the classes would provide a means to assess methodology independent of statistical measures. This comparison, however, requires knowledge of the true spatial distribution, which is not known. Nonetheless, a qualitative assessment was made by conferring with agricultural researchers active in the region (Personal communication with J. Long in June 2012 [Miller 2012]). We found that our predicted maps, both the object-level and the pixel-level, conformed well to the spatial distribution of the classes in the study area; pulse crops were predominant in Sheridan County while non-pulse or cereal crops tended to be located near the Missouri River.

## Discussion

Data gaps in ETM+ imagery have prompted a variety of approaches to fill the gaps with data interpolated from nearby pixels or to fill the gaps with data derived from a different date or a different sensor. Single-date images in which gaps are filled with data from a different date or sensor are unsatisfactory in crop mapping, which requires spectral data from the same or nearly the same day. Furthermore, interpolation schemes encounter predictive difficulties in heterogeneous landscapes or landscapes linear features (e.g., Maxwell 2004). Agricultural landscapes are often considered homogeneous, but this is a within-field condition. These landscapes can be heterogeneous at

field boundaries in which sharp changes in spectral characteristics might exist; for example, the border between adjoining wheat and fallow fields. Furthermore, linear features are characteristic features in agricultural landscapes. Gap-filling schemes, therefore, are unsuitable for crop mapping.

A number of studies have suggested that O-O approaches yield better classification accuracies for crops than classifications using traditional pixel-based approaches (e.g., Pedley and Curran 1991). Results from this study, however, indicate that models from the pixel-based approach produced overall classifications that were generally 5% better than equivalent models using the O-O approach. These differences in accuracy are significant at the 95% confidence level. We note that there was a considerable difference in the sample size of the training data between the O-O and the pixel-based approaches; 182 fields versus 15,000 pixels. The better classification accuracies of the pixel-based approach are obviated, however, by the inability to classify data gaps, potential errors in spatial pattern, location of field boundaries, and area.

A pixel-based model only classifies the extant pixels in an image, limiting its usefulness in a spatial context with ETM+ SLC-off imagery or in other cases where data is missing at the sub-field scale, despite higher accuracies. This was evident when we examined results at the field level (Figure 4). The classification based on object-level data is easily interpretable and the field boundaries are clear – even for fields in which data gaps obscure the boundaries. The O-O approach assumes that crop management decisions are made on a per-field basis. This is generally the case, yet some producer's in the study area continue to manage fields by strip cropping (typically alternating strips of

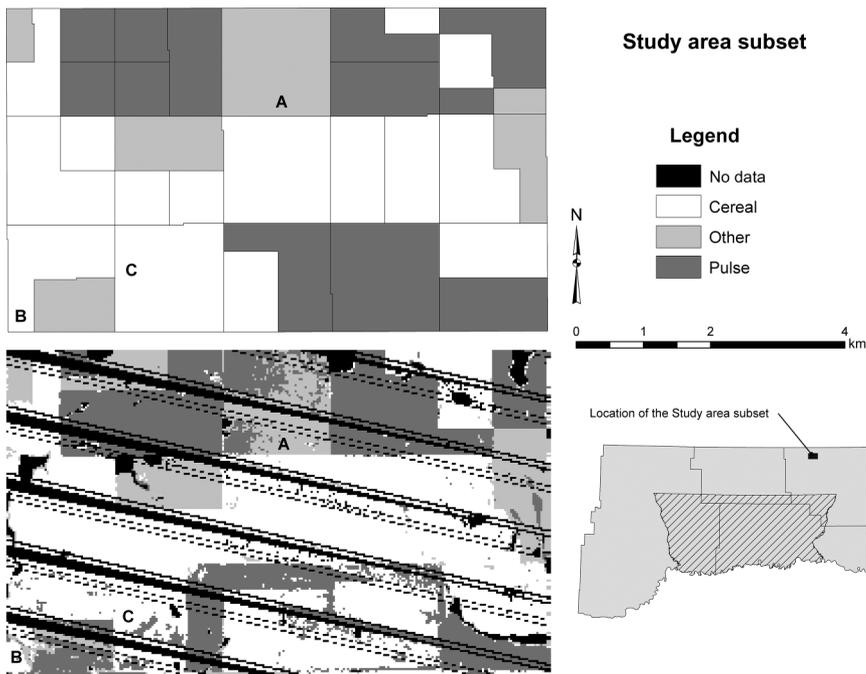


Figure 4. The study area subset is located in northwest Daniels County. The top-left portion of the figure is the subset as classified with the object-level data, while the bottom-left is the same area classified with pixel-level data. Locations A, B, and C are discussed in the text.

cereal and fallow). Fields managed by strip cropping include two classes, yet receive only one classification. This is an obvious disadvantage to the O-O approach; however, the number of fields managed by strip-cropping in the study area is quite small and was estimated at less than 4% of the productive agricultural land. We found no difference in the error rates of strip-cropped fields and monoculture fields. The classification based on the pixel-level data, in contrast, was distinctly different. Data gaps are obvious, and there are numerous fields with mixed classes in both the interior (e.g., fields A, B, and C in Figure 4) and along field boundaries (e.g., the southern and eastern boundaries of field B). Furthermore, the locations of many field boundaries are unknown. This is the case when the boundary is in a data gap and when adjoining fields are the same class (e.g., the western boundary of field C). Classification of the pixel-level data identified fields managed by strip-cropping very well; however, if the goal is to identify the crop growing in a field regardless of fallow or sparsely vegetated areas, as was the case for our study, the pixel-level data essentially failed in fallow areas of strip cropped fields to reflect the crops growing in those fields.

A logical consideration is to use the cadastral layer in a post-classification process, thereby capitalizing on the advantages of both approaches. This is certainly a viable option, but incorporating the cadastral layer post-classification does not allow object-level variables to be used in the classification process itself. Object-level variables, such as various shape attributes or relationships of the object to super- and sub-objects, can improve the reliability of classification if they are included *a priori* (Turker and Arıkan 2005). We used the standard deviation of the spectral bands for each field as a measure of variability in the O-O approach.

Multitemporal models are reported to produce better classification accuracies than single-date models, because they can capture spectral diversity due to phenological differences. We reached the same conclusion. The overall accuracies from the analysis of the multitemporal data were higher than accuracies from the single-date data and were significant at the 95% confidence level. This conclusion is robust with respect to approach, O-O or pixel-based.

The largest source of errors is confusion between the Cereal and the Other classes. This is not surprising given the nature of these classes. As discussed earlier, discriminating between Cereal and Other was a challenging task. This was partially because each class in the study comprised several different crops and, therefore, had higher class variability than would have been expected in a class consisting of a single crop. Furthermore, the Cereal and Other classes had considerable overlap in spectral characteristics as some of the crops in the Other class are quite similar to the Cereal crops. Two notable examples are hay and fields in the Conservation Reserve Program (CRP) that have largely reverted to shortgrass prairie. We also note that much of the confusion between the Pulse and Other classes is likely due to similarities between alfalfa (Other), which is a member of the pea family, and peas (Pulse). Nonetheless, the RF classifier was largely successful by creating multiple paths, via differing variable choices, to the same class. For example, the spectral characteristics of an object consisting of mid-season wheat, late-season wheat, and wheat in a strip-cropped field are very different, but the random forest algorithm creates multiple paths to the correct classification.

Variable importance measures from a RF model are instructive, but a strict interpretation of those measures is not feasible and results should be considered speculative as currently available measures of variable importance have been shown to be problematic (Strobl and Zeileis 2008). Variable importance is not a particular focus of this paper; nonetheless, we present a brief discussion.

The usefulness of phenological differences, derived from multitemporal data, for crop classification is suggested by noting that in the multitemporal O-O model, 6 of the top 8 most important variables were differenced values. Only 2 of the top 8 variables are differenced values with the pixel-level data. Object-level data are based on a field's mean spectral values, which tends to smooth local (pixel-level) variations. Consequently, differenced values are more meaningful for fields than for individual pixels. The multitemporal model also relied heavily on mid-season variables, particularly bands 2 and 4, whereas only two of the late-season variables, bands 5 and 7, were important. Bands 2 and 4 are highly characteristic of green vegetation and the importance of these bands from the mid-season, when crops are close to maturity but are likely maturing at different rates, is reasonable. Bands 5 and 7 are associated with the amount of water in vegetation and the importance of these bands from the late-season, when crops are potentially water stressed, senescent, or have already been harvested, but again at phenologically different times, is likewise reasonable. Standard deviations of the spectral bands, applicable only to the object-level data, were infrequently among the most important variables, regardless of model or class. Standard deviation captures within-field variability, and to some extent texture. The lack of importance in classification for standard deviation suggests minimal differences in variability among classes.

The reliance of the mid-season O-O classification on bands 2, 4, 5, and 7 suggests that there are some key differences in crops that are not based on phenology (reflected only in multitemporal data) but on the crop themselves. These include differences in: (1) greenness (band 2); (2) leaf structure (band 4); and (3) the amount of plant-held water (bands 5 and 7). The Pulse class, for example, is distinguished by high importance for bands 2 and 4, while bands 5 and 7 are relatively low in importance. In contrast, these bands are the top four most important variables in classifying the Cereals and Other classes. Accordingly, plant-held water is an essential variable in separating the Cereal and Other classes from the Pulse class. We note that the importance of bands 5 and 7 for this particular separation does not necessarily imply that the Cereal and the Other classes have high amounts of plant-held water, only that the amounts differ in some important way. It is possible, and more likely, that the Pulse class has greater amounts of plant-held water. Bands 2, 4, 5, and 7 remained important for late-season O-O classification, in different relative orders, but bands 1 and 3 were of higher importance. The importance of band 1 is likely associated with harvesting (which does not typically occur for the Other class) since band 1 is associated with soil reflectance. Band 3 is crucial to distinguishing between green vegetation and senescent (or yellow) vegetation; mature cereal crops are yellow.

Patterns in variable importance for the pixel-based classification were generally similar to those found in the O-O classification. There was one key difference; band 6, unimportant with the object-level data, was important at the pixel-level. There were substantial differences among the classes with respect to mid-season band 6 values, suggesting different thermal characteristics among the classes at this time of the year.

The RF classifier is well-known and generally found to be superior to standard classification approaches (e.g., Lawrence, Wood, and Sheley 2006). The method performed well in this study and was able to classify the object-level data as well as pixel-level data with acceptable accuracies. One of the advantages of RF models is their ability to handle large datasets and missing values (Breiman 1996). We used RF to classify nearly 23 million pixels with 24 explanatory variables, and 13,143 fields using 45 variables; neither presented difficulties. The ability of RF to determine variable importance is beneficial, but not unique to RF. Variable importance facilitates an understanding of

which variables contribute the most to classification; but, current measures are problematic and results are speculative.

We sought a method to classify a northeast Montana agricultural landscape into the region's dominant classes: cereal crops, pulse crops, and other, using SLC-off ETM+ imagery – without using gap-filling schemes. Although the pixel-based approach gave higher overall classification accuracies than the O-O approach, it was unable to classify areas within data gaps and produced fields with mixed classifications. The O-O approach, in contrast, enabled the classification of fields partially located within data gaps; we were, of course, unable to classify fields completely within data gaps. The multitemporal approach encapsulated class-wise phenological variations and provided a statistically significant contribution to overall classification accuracies. RF modeling produced acceptable overall classification accuracies with approximately equal class error rates.

We did not investigate the minimum number of pixels required to classify an object. We used a minimum of 40 pixels under the supposition that this sample size was sufficiently large, according to the Law of Large Numbers (LLN), to obtain representative means and standard deviations. Research concerning the minimum number of pixels required to classify an object would be an appropriate next step.

### Acknowledgments

This article is based upon work supported by the U.S. Department of Energy and the National Energy Technology Laboratory under Award Number DE-FC26-05NT42587. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. We thank the anonymous reviewers for the thorough and thoughtful comments.

### References

- Adam, E. M., O. Mutanga, D. Rugege, and R. Ismail. 2012. "Discriminating the Papyrus Vegetation (*Cyperus papyrus* L.) and its Co-Existent Species using Random Forest and Hyperspectral Data Resampled to HYMAP." *International Journal of Remote Sensing* 33 (2): 552–569.
- Arvidson, T., S. Goward, J. Gasch, and D. Williams. 2006. "Landsat 7 Long-Term Acquisition Plan: Development and Validation." *Photogrammetric Engineering and Remote Sensing* 72 (10): 1137–1146.
- Ashish, D., R. W. McClendon, and G. Hoogenboom. 2009. "Land-Use Classification of Multispectral Aerial Images Using Artificial Neural Networks." *International Journal of Remote Sensing* 30 (8): 1989–2004.
- Bolorani, A. D., S. Erasmi, and M. Kappas. 2008. "Multi-Source Image Reconstruction: Exploitation of EO-1/ALI in Landsat 7/ETM+ SLC-Off Gap Filling." In *Proceedings of SPIE-IS&T: Image Processing: Algorithms and Systems VI*, edited by J. T. Astola, K. O. Egiazarian, and E. R. Dougherty, March 31, 6812: 681219-1–681219-12. Bellingham, WA: SPIE Press.
- Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24: 123–140.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

- Chapman, D. S., A. Bonn, W. E. Kunin, and S. J. Cornell. 2010. "Random Forest Characterization of Upland Vegetation and Management Burning from Aerial Imagery." *Journal of Biogeography* 37: 37–46.
- Chen, F., L. Tang, and Q. Qiu. 2010. "Exploitation of CBERS-02B as Auxiliary Data in Recovering the Landsat 7 ETM+ SLC-Off Image." In *Proceedings of the 18th International Conference on Geoinformatics*, edited by Y. Liu, and A. Chen, Piscataway, NJ, June 18–20, 6 pp. Beijing: Institute of Electrical and Electronics Engineers.
- Clark, S. 2011. "Veteran Landsat 5 Satellite on the Brink of Failure." Accessed February 20, 2012. <http://www.spaceflightnow.com/news/n1111/18landsat5/>
- Cohen, W. B., T. K. Maersperger, S. T. Gower, and D. P. Turner. 2003. "An Improved Strategy for Regression of Biophysical Variables and Landsat ETM+ Data." *Remote Sensing of Environment* 84: 561–571.
- Congalton, R. G., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd ed. Boca Raton, FL: CRC Press.
- Conrad, C., R. R. Colditz, S. Dech, D. Klein, and P. L. G. Vlek. 2011. "Temporal Segmentation of MODIS Time Series for Improving Crop Classification in Central Asian Irrigation Systems." *International Journal of Remote Sensing* 32 (23): 8763–8778.
- El-Magd, I. A., and T. W. Tanton. 2003. "Improvements in Land Use Mapping for Irrigated Agriculture from Satellite Sensor Data Using a Multi-Stage Maximum Likelihood Classification." *International Journal of Remote Sensing* 24 (21): 4197–4206.
- Forster, D., T. W. Kellenberger, Y. Buehler, and B. Lennartz. 2010. "Mapping Diversified Peri-Urban Agriculture – Potential of Object-Based Versus Per-Field Land Cover/Land Use Classification." *Geocarto International* 25 (3): 171–186.
- Fuller, R. M., G. B. Groom, and A. R. Jones. 1994. "The Land-Cover Map of Great-Britain – An automated Classification of Landsat Thematic Mapper data." *Photogrammetric Engineering and Remote Sensing* 60: 553–562.
- Gómez-Casero, M. T., I. L. Castillejo-González, A. García-Ferrer, J. M. Peña-Barragán, M. Jurado-Expósito, L. García-Torres, and F. López-Granados. 2010. "Spectral Discrimination of Wild Oat and Canary Grass in Wheat Fields for Less Herbicide Application." *Agronomy for Sustainable Development* 30: 689–699.
- Janssen, L. L. F., and J. D. van Amsterdam. 1996. "An Object-Based Approach to the Classification of Remotely Sensed Images." In *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS '91)*, edited by J. Putkonen, 2192–2195. New York: IEEE.
- Lawrence, R. L., S. D. Wood, and R. L. Sheley. 2006. "Mapping Invasive Plants Using Hyperspectral Imagery and Breiman Cutler Classifications (RandomForest)." *Remote Sensing of Environment* 100: 356–362.
- Lu, D., and Q. Weng. 2005. "Urban Classification Using Full Spectral Information of Landsat ETM + Imagery in Marion County, Indiana." *Photogrammetric Engineering and Remote Sensing* 71 (11): 1275–1284.
- Markham, B. L., J. C. Storey, D. L. Williams, and J. R. Irons. 2004. "Landsat Sensor Performance: History and Current Status." *IEEE Transactions on Geoscience and Remote Sensing* 42 (12): 2691–2694.
- Masek, J. G., M. Honzak, S. N. Goward, P. Liu, and E. Pak. 2001. "Landsat ETM+ as an Observatory for Land Cover: Initial Radiometric and Geometric Comparisons with Landsat 5 Thematic Mapper." *Remote Sensing of Environment* 78: 118–130.
- Maxwell, S. 2004. "Filling Landsat ETM+ SLC-Off Gaps Using a Segmentation Model Approach." *Photogrammetric Engineering and Remote Sensing* 70 (10): 1109–1111.
- Maxwell, S. K., G. L. Schmidt, and J. S. Storey. 2007. "A Multi-Scale Segmentation Approach to Filling Gaps in Landsat ETM+ SLC-Off Images." *International Journal of Remote Sensing* 28 (10): 5339–5356.
- MSL (Montana State Library). 2012. "Montana Geographic Information Clearinghouse." Accessed October 12, 2011. <http://nris.mt.gov/gis/default.asp>
- NASS (National Agricultural Statistics Service). 2012. *Montana Agricultural Statistics, Volume XLIX*. Washington, DC: United States Department of Agriculture.
- Oetter, D. R., W. B. Cohen, M. Berterretche, T. K. Maersperger, and R. E. Kennedy. 2000. "Land Cover Mapping in an Agricultural Setting using Multi-Seasonal Thematic Mapper data." *Remote Sensing of Environment* 76 (2): 139–155.

- Pax-Lenney, M., and C. E. Woodcock. 1997. "Monitoring Agricultural Lands in Egypt with Multitemporal Landsat TM Imagery: How many Images are Needed?" *Remote Sensing of Environment* 59: 522–529.
- Pedley, M. L., and P. J. Curran. 1991. "Per-Field Classification: An Example Using SPOT HRV Imagery." *International Journal of Remote Sensing* 12 (11): 2182–2192.
- Pringle, M. J., M. Schmidt, and J. S. Muir. 2009. "Geostatistical Interpolation of SLC-Off Landsat ETM+ Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 64: 654–664.
- Reza, M. M., and S. N. Ali. 2008. "Using IRS Products to Recover 7ETM+ Defective Images." *American Journal of Applied Sciences* 5: 618–625.
- Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. "An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93–104.
- Roy, D. P., J. Ju, P. Lewis, C. Schaaf, F. Gao, M. Hansen, and E. Lindquist. 2008. "Multitemporal MODIS-Landsat Data Fusion for Relative Radiometric Normalization, Gap Filling, and Prediction of Landsat Data." *Remote Sensing of Environment* 112: 3112–3130.
- Savage, S. L., and R. L. Lawrence. 2010. "Vegetation Dynamics in Yellowstone's Northern Range: 1885–1999." *Photogrammetric Engineering and Remote Sensing* 76 (5): 547–556.
- Simonneaux, V., B. Duchemin, D. Helson, S. Er-Raki, A. Olioso, and A. G. Chehbouni. 2008. "The Use of High-Resolution Image Time Series for Crop Classification and Evapotranspiration Estimate over an Irrigated Area in Central Morocco." *International Journal of Remote Sensing* 29 (1): 95–116.
- Storey, J., P. Scaramuzza, G. Schmidt, and J. Barsi. 2005. "Landsat 7 Scan Line Corrector-Off Gap-Filled Product Development." In *ASPRS 2005 – Pecora 16 Conference Proceedings*. Sioux Falls, SD: American Society of Photogrammetry and Remote Sensing.
- Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn. 2007. "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution." *BMC Bioinformatics* 8: 25.
- Strobl, C., and A. Zeileis. 2008. "Danger: High power! – Exploring the Statistical Properties of a Test for Random Forest Variable Importance." Proceedings of the 18th International Conference on Computational Statistics, Porto, August 24–28.
- Tanaka, D. L., D. J. Lyon, P. R. Miller, S. D. Merrill, and B. G. McConkey. 2010. "Soil and Water Conservation in the Semiarid Northern Great Plains." In *Soil and Water Conservation Advances in the United States*, edited by T. M. Zobeck, and W. F. Schillinger, 81–102. Madison, WI: American Society of Agronomy.
- Turker, M., and M. Arikan. 2005. "Sequential Masking Classification of Multitemporal Landsat ETM+ Images for Field-Based Crop Mapping in Karacabey, Turkey." *International Journal of Remote Sensing* 26 (17): 3813–3830.
- USDA (United States Department of Agriculture). 2012. "USDA National Agricultural Statistics Service Cropland Data Layer (2012)." Accessed June 11, 2013. <http://nassgeodata.gmu.edu/CropScape/>
- USGS (United States Geological Survey). 2012. *Landsat: A Global Land-Imaging Mission*. Reston, VA: U.S. Geological Survey Fact Sheet 2012–3072.
- Wardlow, B. D., S. L. Egbert, and J. H. Kastens. 2007. "Analysis of Time-Series MODIS 250 m Vegetation Index Data for Crop Classification in the U.S. Central Great Plains." *Remote Sensing of Environment* 108 (3): 290–310.
- Watts, J. D., R. L. Lawrence, P. R. Miller, and C. Montagne. 2009. "Monitoring of Cropland Practices for Carbon Sequestration Purposes in North Central Montana by Landsat Remote Sensing." *Remote Sensing of Environment* 113 (9): 1843–1852.
- WRCC (Western Regional Climate Center). 2012. "Local Climate Summary for Glasgow, Montana." Accessed October 18, 2011. <http://www.wrcc.dri.edu/summary/ggw.mt.html>
- Wu, S., J. Silván-Cárdenas, and L. Wang. 2007. "Per-Field Urban Land Use Classification Based on Tax Parcel Boundaries." *International Journal of Remote Sensing* 28 (12): 2777–2800.
- Zeng, C., H. Shen, and L. Zhang. 2013. "Recovering Missing Pixels for Landsat ETM+ SLC-Off Imagery Using Multitemporal Regression Analysis and a Regularization Method." *Remote Sensing of Environment* 131: 182–194.
- Zhang, C., W. Li, and D. Travis. 2007. "Gaps-Fill of SLC-Off Landsat ETM+ Satellite Image Using a Geostatistical Approach." *International Journal of Remote Sensing* 28 (22): 5103–5122.
- Zhu, X., D. Liu, and J. Chen. 2012. "A New Geostatistical Approach for Filling Gaps in Landsat ETM+ SLC-Off Images." *Remote sensing of Environment* 124: 49–60.