Remote Sensing of Environment 115 (2011) 66–75

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery

Jennifer D. Watts [a,*], Scott L. Powell [b], Rick L. Lawrence [b], Thomas Hilker [c]

[a] Flathead Lake Biological Station, University of Montana, 32125 Bio Station Lane, Polson, MT, 59860, USA
[b] Department of Land Resources and Environmental Science, Montana State University, P.O. Box 173120, Bozeman, MT, 59717, USA
[c] Integrated Remote Sensing Studio, Department of Forest Resource Management, 2424 Main Mall, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

ARTICLE INFO

Article history:
Received 23 April 2010
Received in revised form 13 August 2010
Accepted 14 August 2010

Keywords:
Landsat
Random Forest
MODIS
STARFM
Tillage

ABSTRACT

Conservation tillage management has been advocated for carbon sequestration and soil quality preservation purposes. Past satellite image analyses have had difficulty in differentiating between no-till (NT) and minimal tillage (MT) conservation classes due to similarities in surface residues, and may have been restricted by the availability of cloud-free satellite imagery. This study hypothesized that the inclusion of high temporal data into the classification process would increase conservation tillage accuracy due to the added likelihood of capturing spectral changes in MT fields following a tillage disturbance. Classification accuracies were evaluated for Random Forest models based on 250-m and 500-m MODIS, 30-m Landsat, and 30-m synthetic reflectance values. Synthetic (30-m) data derived from the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) were evaluated because high frequency Landsat image sets are often unavailable within a cropping season due to cloud issues. Classification results from a five-date Landsat model were substantially better than those reported by previous classification tillage studies, with 94% total and ≥88% class producer's accuracies. Landsat-derived models based on individual image scenes (May through August) yielded poor MT classifications, but a monthly increase in accuracy illustrated the importance of temporal sampling for capturing regional tillage disturbance signatures. MODIS-based model accuracies (90% total; ≥82% class) were lower than in the five-date Landsat model, but were higher than previous image-based and survey-based tillage classification results. Almost all the STARFM prediction-based models had classification accuracies higher than, or comparable to, the MODIS-based results (>90% total; ≥84% class) but the resulting model accuracies were dependent on the MODIS/Landsat base pairs used to generate the STARFM predictions. Also evident within the STARFM prediction-based models was the ability for high frequency data series to compensate for degraded synthetic spectral values when classifying field-based tillage. The decision to use MODIS or STARFM-based data within conservation tillage analysis is likely situation dependent. A MODIS-based approach requires little data processing and could be more efficient for large-area mapping; however a STARFM-based analysis might be more appropriate in mixed-pixel situations that could potentially compromise classification accuracy.

1. Introduction

The need for improved terrestrial soil carbon management has received considerable attention in recent years. An estimated 42–78 Gt/carbon globally has already been lost from cropland soils after decades of cultivation (Lal, 2004). The adoption of cropland practices, including conservation tillage management, that increase plant residue deposition while also reducing soil disturbances have been advocated to help replenish lost soil carbon (Pacala & Socolow, 2004; Sperow et al., 2003). Benefits from soil carbon, in addition to climate change mitigation efforts, include added plant productivity due to water and nutrient retention and improved soil, air, and water quality (Reicosky, 2001).

Conservation tillage systems include those with minimal tillage (MT) and those without tillage, or no-till (NT). NT systems seed directly into the previous crop stubble and are allowed only 15–25% surface disturbance and residue removal (NRCS, 2006; NRCS, 2008). Fields under MT management include conservation practices that do not disturb more than ~67% of the surface residue (CTIC, 2010; Lal, 1997). Conservation tillage can reduce soil disturbance events that result in elevated carbon dioxide (CO_2) release due to the microbial decomposition of organic, carbon-rich materials. Conservation systems can also facilitate carbon sequestration in agricultural soils because preserved plant residues directly store organic carbon while helping to retain the soil moisture necessary to sustain plant photosynthesis.

* Corresponding author. Tel.: +1 406 243 6318.
E-mail address: jennifer.watts@ntsg.umt.edu (J.D. Watts).

0034-4257/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2010.08.005

National and localized conservation outreach efforts have helped to promote the adoption of conservation tillage, but documentation has been limited. Global NT adoption has been estimated at roughly 45 million ha (Derpsch & Friedrich, 2009). Past survey results in the U. S. have estimated NT adoption at ~24% (26 million ha) and MT at roughly 20 million ha (~19%), but these statistics were dependent on voluntary information from participating counties and are likely not representative of actual trends (CTIC, 2007).

The use of survey-based methods to collect tillage statistics is both expensive and time-intensive. The impracticality of survey-based approaches has often resulted in an absence of tillage data, or information that is out-dated and spatially limited. Satellite-based tillage mapping has been suggested as a possible means to facilitate the annual collection of large-area tillage statistics (Watts et al., 2009). The availability of these data would likely benefit future policy and conservation planning and would help to facilitate regional carbon sequestration estimates (Watts et al., Submitted for publication). Emphasis within agricultural carbon credit markets has been given to NT adoption, rather than to management types which incorporate tillage, as many studies have shown higher sequestration results with tillage absence (Lal, 2004). Satellite-based mapping might also be useful for monitoring tillage management compliance in fields under carbon sequestration programs, but has been hindered by past difficulties in separating NT from MT (Watts et al., 2009).

Previous image-based mapping studies have focused primarily on classifying NT from traditional tillage systems with a high percentage of surface disturbance (Bricklemyer et al., 2002; South et al., 2004) or conservation tillage (≥30% surface residue) from traditional tillage (Daughtry et al., 2006; Gowda et al., 2008; Sullivan et al., 2008; Viña et al., 2003). One study (Bricklemyer et al., 2006) attempted to distinguish NT from tillage (grouping conservation and intensive tillage into one class) by applying logistic regression to a 26-June Landsat Enhanced Thematic Mapper (ETM+) mosaic, but obtained only 29% accuracy in the tillage class. The low tillage accuracy was attributed to crop canopy interference as study sites included both cropped and fallow fields. Another study (Watts et al., 2009) also reported low tillage accuracy (<35%) after applying a binary (tillage or NT) split to both cropped and fallow fields, through an object-oriented Random Forest classification approach applied to May and August Landsat data. The study concluded that the class error had resulted from MT fields misclassified as NT due to similarities in surface residue cover, rather than bias from crop canopy cover.

The timing of image acquisition, in relation to tillage events, might be important for classification accuracy as spectral differences between the NT and MT classes are likely greater directly following a tillage event. Field-based spectroradiometer mapping has shown slight increases in spectral absorption within visible and infrared bands following a tillage disturbance event due to increased exposure of moist, carbon and oxide rich soil and a decrease in surface residues (Demattê et al., 2004; Haché et al., 2007). Tillage dates can vary greatly through a region; hence a high temporal frequency data set might increase the likelihood of capturing tillage-related disturbances in a greater percentage of fields. The incorporation of data sets with higher temporal resolution, instead of the one or two date approach used by many tillage classification studies, therefore, might better capture surface disturbances in MT fields and increase differentiation between the conservation tillage classes (Watts et al., 2009).

Obtaining data sets with both high temporal and spatial resolution has long challenged the remote sensing community. Ultra-high spatial resolution sensors with 1–5 day collection intervals (i.e., IKONOS, QuickBird, RapidEye, and SPOT) are costly and can be disadvantageous for regional mapping due to their small footprint and often limited spectral resolution. Additionally, the high revisit rate can only be accomplished by pointing the sensor in an off nadir direction, and, as a result, full coverage over larger areas cannot be provided at high revisit rates. Many studies have instead opted to use mid-resolution sensors such as Landsat-5 Thematic Mapper (TM) and Landsat-7 Enhanced Thematic Mapper Plus (ETM+) (Cohen & Goward, 2004). The Landsat products are often appealing because of their 30-m spatial resolution, larger footprint (185 km), a combined 8-day collection interval, and low cost (basic products available at no-cost as of Dec. 2008). It is probable that exploiting fortnightly Landsat data sets for tillage mapping would help improve the classification of surface disturbances, but it is often difficult to obtain enough cloud-free scenes to produce a high frequency set. Restrictions in Landsat data availability due to clouds are likely when two or more cloud-free scenes are required per year, especially within a short time interval (Ju & Roy, 2008).

Daily-collected MODIS (Moderate Resolution Imaging Spectroradiometer) data have been used as an alternative to Landsat for mapping cropland systems. MODIS is comparable to Landsat in spectral coverage, but has a coarser spatial resolution (500 m–1 km; 250 m in red and near-infrared (NIR)). Several studies have incorporated MODIS data into cropland analyses (e.g., Lobell & Asner, 2004; Shao et al., 2010; Wardlow & Egbert, 2008, 2010). Only one study (Schroers et al., 2009) was identified that used MODIS data for tillage classification. In that analysis, logistic regression was applied to 8-day 500-m and 16-day 250-m MODIS composites to differentiate between 25 tilled and NT fields and resulted in 88% total model accuracy. While the temporal strength of MODIS might benefit cropland mapping, at least two studies have reported higher classification errors associated with smaller field sizes (Schroers et al., 2009; Shao et al., 2010). Another has expressed caution in using MODIS data for mapping in regions with high surface diversity, and consequently higher probabilities for mixed-pixel situations, without also incorporating high-resolution data to refine map estimations (Price, 2003).

The spectral blending of MODIS with higher-resolution Landsat data through the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) has been used to produce synthetic 30-m Landsat-like high temporal resolution data sets (Gao et al., 2006; Hilker et al., 2009a). The STARFM algorithm is based on a spatially weighted relationship between the reflectance of Landsat and MODIS, resampled to 30-m pixels, at Date-1; this relationship can then be used to predict high-resolution synthetic data at Date-2, after accounting for spectral differences between MODIS values at Date-1 and Date-2 (Gao et al., 2006). Studies have demonstrated STARFM's capability in accurately generating synthetic Landsat scene data (Gao et al., 2006; Hilker et al., 2009a), but have noted that better predictions generally occur in landscapes characterized by homogenous regions having spectrally "pure" MODIS-scale observations. STARFM has also shown difficulty in accounting for small-area spectral changes resulting from surface disturbance events (Hilker et al., 2009b). An extension to the STARFM model, STAARCH (Spatially and Temporal Adaptive Algorithm for Mapping Reflectance Change), has been developed to better detect small scale disturbances in forest ecosystems (Hilker et al., 2009b). STAARCH allows for STARFM-based predictions to be applied to MODIS pixels classified as "disturbed" via a 2-date Landsat change mask with the time of disturbance identified from MODIS series data sets. The STAARCH algorithm relies on a disturbance index derived from normalized tasseled cap components (Hilker et al., 2009b) to detect change events. The STAARCH tasseled cap-based disturbance index might not be the best approach for detecting small and temporally variable tillage disturbance events, given its reliance on a Landsat scene before and after a possible surface disturbance. This is due potential problems in obtaining cloud-free imagery. It is also questionable if spectral signature changes due to tillage would stay distinctive for long after the disturbance date as Watts et al. (2009) reported difficulty classifying tillage disturbance in high residue systems when using one Landsat image at either end of the cropping season, even with the incorporation of tasseled cap components, normalized difference vegetation index (NDVI) values,

per-band reflectance, and textural and neighborhood relationship components into an advanced decision tree-based classifier.

Our objective in this study was to examine the ability to classify NT fields from those characterized by MT management through the use of data sets with increased temporal and/or spatial resolution. We hypothesized that a high temporal data set would better capture subtle differences in spectral response between NT and MT fields within north central Montana. The timing of tillage events varies greatly within this region, therefore a high density of image dates within a cropping season would likely increase the probability of capturing spectral changes occurring within a tillage disturbance window. However, the temporal extent of this change within conservation systems is unknown as the degree of surface disturbance is relatively minimal and the decrease in reflectance is minimized over time due to surface weathering.

Tillage classification performance was evaluated for Random Forest models based on three categories of data, namely Landsat, MODIS, and STARFM-derived synthetic Landsat. The high temporal Landsat series provided a baseline for comparing model results with those from previous tillage studies that evaluated NT and MT classifications (Bricklemyer et al., 2006; Watts et al., 2009) and for evaluating the performance of MODIS and STARFM-based models. It was hypothesized that the added spatial resolution within the STARFM-based synthetic data sets might result in better classification performance than models based on the lower-resolution MODIS data, especially since field widths less than 500 m are common within this region.

2. Methods

2.1. Tillage data collection

Tillage management data were collected for MT and NT sites across north central Montana in 2009 (study center lat/long: 48.46, −110.84 degrees). This region is characterized by a semi-arid steppe climate (NRCS, 2007a). Textural properties for the mostly mollic soils can vary considerably throughout the region, ranging from clay to sandy loam (NRCS, 2007b). Mean annual precipitation is spatially variable and has typically ranged from 260 mm–380 mm (WRCC, 2009). Farmers within the region using MT management were identified through agricultural extension agents in Toole, Liberty, and Hill counties. The specified farmers were contacted to determine if MT practices had been used during the 2009 growing season (May–September) and to determine the specific field locations. Field locations were then cross-referenced with regional cadastral shapefile data to obtain field boundaries. Data for tillage dates, post-tillage surface residue percentages, and crop type were obtained during the farmer interviews whenever possible. The ensuing information indicated that tillage disturbance events during the 2009 crop season resulted in anywhere from 30 to 60% total disturbance of crop surface stubble (based on visual estimates provided to us by farm managers).

The duration of MT use was at least three years for most fields, with one field having been recently converted from NT in spring 2009 due to pest management issues. Additional fields converted to MT in 2009 were not identified. NT field locations were primarily obtained through the North Dakota Farmers Union Carbon Credit Program, but also included sites that were specified by farmers during the interview process. The resulting data pool consisted of 75 NT and 50 MT fields for reference training and validation data (Fig. 1). Field crops included spring and winter wheat and some barley, and primarily incorporated a crop and fallow rotation system. Reported productivity in these systems ranged from less than 280 to 700 L of grain per ha. The timing of tillage disturbances spanned from early May to the end of August and was often dependent on soil moisture conditions and operator flexibility. The majority of fields had widths spanning 250–480 m.

2.2. MODIS and Landsat data processing

Twenty-six MODIS images were selected for the 2009 (May–September) cropping season (Table 1). MODIS images were excluded if a substantial proportion (>30%) of the study area was cloud,

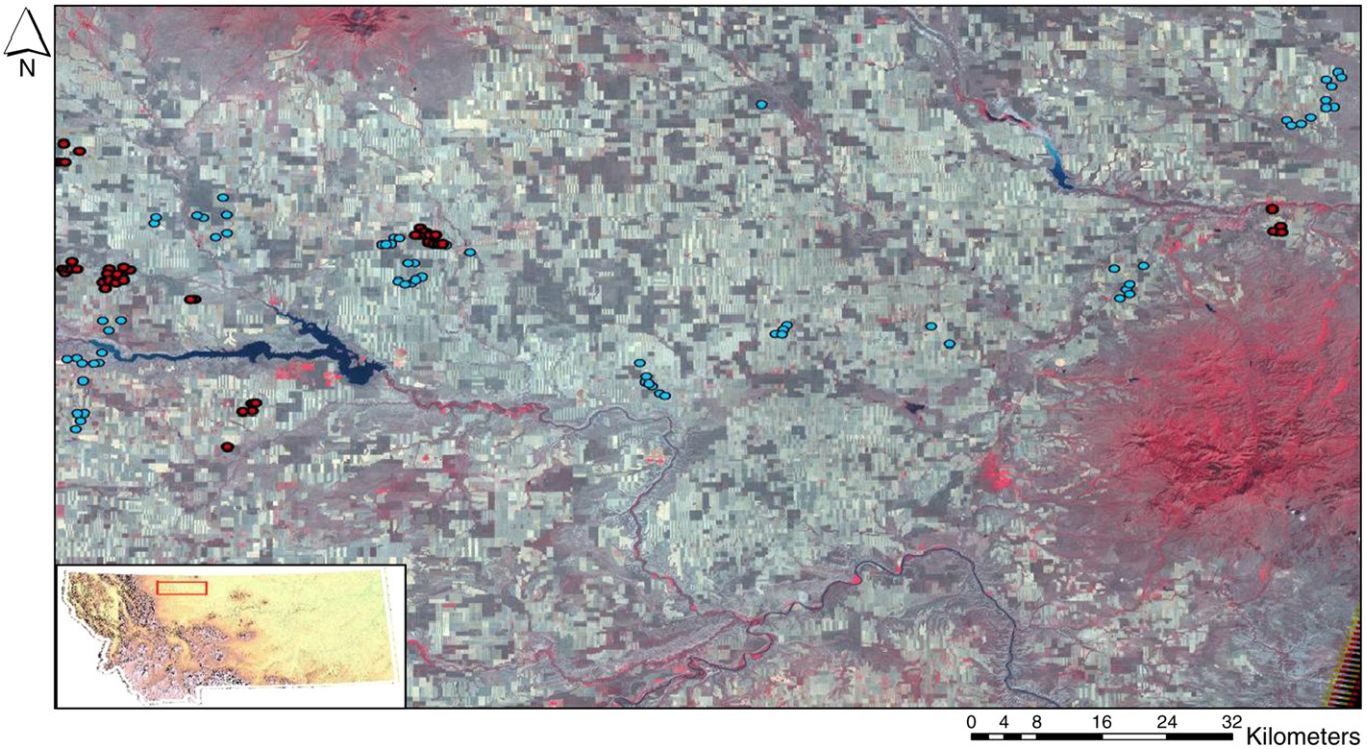

Fig. 1. Spatial locations for MT (red circles) and NT fields (blue circles) used for tillage classification, within an August 2009 Landsat TM image. The study area is indicated by the red outline (lower left) within the Montana insert.

Table 1
Schematic depicting the 2009 image scenes used in the Random Forest classifications, including multiple Landsat (column 1) and MODIS (column 2) dates. STARFM-based synthetic data were also generated for each MODIS image date (STARFM prediction sets 1-3). Each STARFM prediction scene within a set was based on the relationship between a base date Landsat and MODIS image. The base pair corresponding to each individual STARFM-based prediction at a particular MODIS date (three predictions were made for each MODIS scene) is indicated by the base pair symbol in that particular cell.

Image dates (2009)		STARFM prediction sets				Image dates cont.		STARFM cont.			
Landsat		MODIS	1	2	3	Landsat		MODIS	1	2	3
		2-May	o	o	•						
		18-May	o	o	•			17-Jul	‡	•	•
		19-May	o	o	•			19-Jul	‡	•	•
		22-May	o	o	•			21-Jul	‡	•	•
		23-May	o	o	•	23-Jul	‡	22-Jul	‡	•	•
28-May	o	26-May	o	o	•			25-Jul	‡	•	•
								1-Aug	‡	•	•
		30-May	†	o	•			2-Aug	‡	•	•
		3-Jun	†	o	•						
		4-Jun	†	o	•			11-Aug	•	•	•
13-Jun	†	13-Jun	†	o	•			20-Aug	•	•	•
								21-Aug	•	•	•
		24-Jun	Δ	o	•	24-Aug	•	25-Aug	•	•	•
		25-Jun	Δ	o	•			27-Aug	•	•	•
29-Jun[a]	Δ	27-Jun	Δ	•	•						
		30-Jun	Δ	•	•						

[a] Paired with 27-Jun instead of 30-Jun MODIS due to amplified cloud coverage within the combined 29-Jun Landsat/30-Jun MODIS data set.

shadow, or haze contaminated, as identified through a visual and histogram-based examination of pixel outliers. Five Landsat image-pairs (TM and ETM+; path/row 39/26 and 39/27), each pair occurring on the same date, were also obtained for the 2009 cropping period and were mosaicked to create one continuous image that spanned the spatial extent of the study region (Table 1). The Landsat image-pairs were georectified to a base image (RMSE<0.35 pixel) to minimize pixel spatial error.

The compiled MODIS surface reflectance product data (MOD09) were reprojected to UTM coordinates and clipped to the spatial extent of the Landsat image-pairs. The 250-m red and near-infrared MODIS data were incorporated with the 500-m blue, green, and middle-infrared MODIS data for each image scene. The MODIS data were re-scaled to a 30-m pixel resolution through a nearest-neighbor approach for incorporation into the STARFM algorithm.

The 30-m Landsat image data were converted from sensor-based digital number (DN) values to exo-atmospheric reflectance using standard processing techniques (Chander et al., 2007; Chavez, 1996; SDH-L7, 2006). This procedure adjusted for between-band scaling differences, earth–sun distance, and sun elevation angle. The Landsat surface reflectance data were formatted as unsigned 16 bit with a scale factor of 10,000 to coincide with the MODIS data format.

2.3. STARFM-based synthetic data

Synthetic Landsat data were generated using the STARFM algorithm developed by Gao et al. (2006), in a windows executable format (Hilker et al., 2009a). The STARFM algorithm predicts Landsat-like (30 m) spectral pixel values at *Date-2*, using MODIS data values at *Date-2* and spatially and spectrally-weighted differences between baseline Landsat and MODIS image data acquired at *Date-1* (Gao et al., 2006):

$$L\left(x_{w/2}, y_{w/2}, Date{-}2\right) = \sum_{i=1}^{w}\sum_{j=1}^{w} W_{ij} * \left(M\left(x_i, y_j, Date{-}2\right) + L\left(x_i, y_j, Date{-}1\right) - M\left(x_i, y_j, Date{-}1\right)\right),$$

where L is a Landsat reflectance pixel value $(x_{w/2}, y_{w/2})$ in the center of a moving window (here spanning 1500 m^2, Hilker et al., 2009a). W_{ij} is the spatial weighting function, $M\ (x_i\ ,\ y_j,\ Date\text{-}2)$ is the MODIS reflectance at the window location (x_i, y_j) observed at *Date-2*, while $L\ (x_i\ , y_j, Date\text{-}1)$ and $M(x_i, y_j\ Date\text{-}1)$ are the corresponding Landsat and MODIS reflectance values observed at *Date-1*, respectively (Gao et al., 2006).

The W_{ij} value determines the degree to which spectrally similar (Gao et al., 2006) neighboring pixels within a moving window (w) contribute spectral information for a predicted central pixel. W_{ij} is defined by a normalized reverse distance, where the inverse of combined spatial, temporal, and spatial distance values, $C_{ij,}$ for a pixel location is divided by the area-based inverse for C_{ij} from the moving window (Gao et al., 2006):

$$W_{ij} = \left(1 / C_{ij}\right) / \sum_{i=1}^{w}\sum_{j=1}^{w}\left(1 / C_{ij}\right),$$

the reverse distance measure provides a decreased weighting for more extreme values, while the w-based normalization constrains the localized C_{ij} value according to neighborhood-based C_{ij} values. A logistic-based formula for C_{ij} was incorporated for a reduced sensitivity to spectral differences between L and M baseline pixel data, and temporal differences between M pixel values at *Date-1* and *Date-2* (Gao et al., 2006). C_{ij} is defined as:

$$C_{ij} = \ln\left(S_{ij} * B + 1\right) * \ln\left(T_{ij} * B + 1\right) * D_{ij},$$

where S_{ij} is spectral distance, T_{ij} is temporal distance, and D_{ij} is a spatial distance function that assigns a higher weighting to pixels within a moving window that are closer to the predicted central pixel. B is a scale reflectance value (10,000 for MODIS). S_{ij} is defined simply as the absolute value of the difference between a baseline Landsat and MODIS pixel value at *Date-1*:

$$S_{ij} = \left|L\left(x_i, y_j, Date{-}1\right) - M\left(x_i, y_j, Date{-}1\right)\right|.$$

T_{ij} is defined as the absolute value of the difference between a MODIS pixel at *Date-1* and the predicted date, *Date-2*:

$$T_{ij} = \left|M\left(x_i, y_j, Date{-}1\right) - M\left(x_i, y_j, Date{-}2\right|.$$

The STARFM algorithm was used to generate synthetic Landsat data for three different temporal prediction data sets (Table 1). The first set incorporated five separate Landsat–MODIS base pairs and represented a "best-case" scenario high temporal synthetic series that could optimize the potential for surface reflectance changes (i.e., tillage disturbance) to be captured by the algorithm and within the resulting data. MODIS images paired with the Landsat scenes were chosen according to their proximity to the Landsat image date (Table 1). The second set represented a more realistic scenario where only early and late season Landsat imagery was available due to area-wide cloud cover during the growing season. This set utilized Landsat–MODIS base pairs from May and August. The third scenario represented a "worst-case" scenario where area-wide cloud contamination was only absent in the late season and used the August base pair for all image predictions.

The suitability of the resulting synthetic data series for tillage classification purposes was evaluated based on tillage classification model accuracies. Pixel value differences between Landsat data and their predicted values were also examined based on random pixel samples in cropland fields for ~10% (40,000 pixels) of the study area, following the methods of Hilker et al. (2009a).

2.4. Random Forest classification models

A MODIS spectral data set was obtained for each MT and NT location by visually selecting per field the MODIS pixel with the highest percentage of spatial overlap. The MODIS pixels were chosen in this manner as the field areas were often smaller than the MODIS pixel extent. The Landsat and resulting synthetic STARFM-based datasets were obtained by taking the average pixel value within each field boundary.

Four overall categories of data were evaluated within tillage classification models and included all six spectral bands from each input image (blue through the second mid-infrared): MODIS, Landsat, MODIS and Landsat, and STARFM-based synthetic Landsat. The MODIS evaluation incorporated 26 image dates within the classification model. The Landsat evaluation included six model scenarios, with one including all five image dates and the other five scenarios being individual scene-based. The MODIS and Landsat evaluation included all 26 MODIS scenes and an August Landsat scene. The STARFM-based evaluation included four scenarios: (1) data that resulted from five MODIS/Landsat base pairs that were used to generate predictions for 26 MODIS dates (Table 1, prediction set 1); (2) data from only the five predicted scenes that directly corresponded to the five MODIS/Landsat base pair dates, and mimicked the five-date Landsat scenario; (3) data from a set based on a May and an August base pair predictions for the 26 MODIS dates (Table 1, prediction set 2); and (4) data from a set based on August base pair predictions for the 26 MODIS dates (Table 1, prediction set 3).

The tillage classification models were generated using the Random Forest package (S-Plus®). Random Forest is superior to many other tree-based classifiers (Watts et al., 2009), and has often outperformed logistic regression (Cutler et al., 2007; Guo et al., 2004). The Random Forest classifier utilizes a bagging-based approach (random sampling with replacement) to build a forest of classification trees (Breiman, 2001; Cutler et al., 2007; Lawrence et al., 2006). Each classification tree (~500 trees in a typical forest) is constructed from a randomly sampled set consisting of ~63% of the full data (Cutler et al., 2007). A random selection of variables also occurs at each node during tree partitioning. The Gini index of node impurity is used to determine, from the sampled variable set, the best binary split at each node and provides a measure of which predictor variable and associated splitting value best decreases class heterogeneity. A larger Gini index indicates greater class heterogeneity (higher class impurity among cases) while a low Gini index indicates increased homogeneity (low class impurity). Tree splitting at each node is determined by comparing the total impurity of child nodes resulting from a possible split at a parent node to the parent Gini value. A split is successful when the Gini of the children is less than that of a parent. Tree splitting terminates upon reaching a Gini index of zero, where only one class is present at each terminal node.

Classification model accuracies were determined using the Random Forest out-of-bag measure, which is an advantageous by-product of bagging and is inherent within the Random Forest algorithm (Breiman, 2001; Lawrence et al., 2006). The out-of-bag measure utilizes training data samples that were withheld during the tree-building process. The withheld sets are run through their associated classification trees and the predicted class for an observation is determined by a plurality vote. Model and class-related accuracies are determined through an analysis of out-of-bag observation prediction results from all trees. The out-of-bag approach has been shown equivalent to external accuracy measures, where a portion of data is withheld completely from model building (Lawrence et al., 2006), and is often ideal for smaller data sets as it allows for all information to be included within classification tree construction. Kappa values were also generated to quantitatively measure a model's predictive ability while accounting for chance agreement (Congalton & Green, 2009; Emam, 1999).

3. Results

3.1. STARFM predictions

The relative accuracies of the STARFM-based predictions incorporated into tillage modeling were evaluated to better understand STARFM's predictive capability within a cropland setting. This examination also helped to assess the classification performance of the synthetic data-based models, as compared to those based on the MODIS or Landsat sets. The accuracy assessments were both visual and regression-based.

A high level of spatial detail was observed within the predicted scenes as had been reported by previous studies (Gao et al., 2006; Hilker et al., 2009a,b). However, some smoothing of fine spatial details was also noted and was especially apparent within regions having distinctive crop-fallow strip patterns (Fig. 2). A pixel-based comparison of 40,000 randomly sampled cropland pixels showed significant ($\alpha = 0.05$, $p = 0.001$) correlations ($0.7 \leq r^2 \leq 0.93$) between the reference 24-August Landsat and the predicted 25-August data, with the highest correlations observed in the red, near-infrared (NIR), and second mid-infrared (MIR2) bands (Fig. 3). An analysis of the slope coefficients indicated close to a 1:1 unit increase between the August Landsat and predicted data, but the STARFM-predicted mean values were generally lower than Landsat (Table 2). Results from a two-sided t-test showed significant difference ($\alpha = 0.05$, $p = 0.001$) in means between the reference August Landsat and August predicted values in all bands, although these results are to be expected given the large sample size.

Decreased correlations between the Landsat and STARFM-based synthetic data were observed when analyzing image predictions for dates further from the 24/25-August base pair (Table 2), and approached zero when comparing the 27-June predictions with their corresponding Landsat date values. A regression between 24-August Landsat data and 27-June predictions, performed to observe the influence of baseline Landsat data on the resulting STARFM-based predictions, showed substantial increase in band-wise correlations when compared to the 29-June Landsat and 27-June predicted regression (Table 2). Temporal variability in correlation strength was observed between August, July, June, and May-predicted data (taken from the same predicted scenes used within the 5-Date STARFM classification model; Table 3) and comparative Landsat data (Table 2). The highest correlations were seen between the 22-July predicted and 23-July Landsat ($r^2 > 0.98$, all bands) data. The June and May synthetic/Landsat comparisons showed generally poorer correlations ($0.4 \leq r^2 \leq 0.77$), with the strongest observed in the red, NIR, and MIR bands.

3.2. Classification accuracies

Tillage accuracies resulting from the Random Forest models showed potential for distinguishing MT from NT in dryland settings within north central Montana using satellite image data. Considerable differences in total and/or class accuracies were observed. However, statistically significant ($\alpha = 0.05$) results from Z-tests were not observed when the stronger (>84% total accuracy) classification models were analyzed against each other.

The Landsat-based model using all five image dates (All-data model) yielded the highest classification accuracies (in the Landsat series) with 94% total accuracy (kappa = 0.88) and class (producer's) values of 88% MT and 99% NT (Table 3). This full-set, best case, scenario would be unrealistic in most years due to the high frequency of cloud cover within the growing season, but was included for comparative purposes. An analysis of models generated using individual Landsat scene dates illustrated a key relationship between temporal sampling and tillage classification accuracy. The May through August scene-specific Landsat models yielded total accuracies

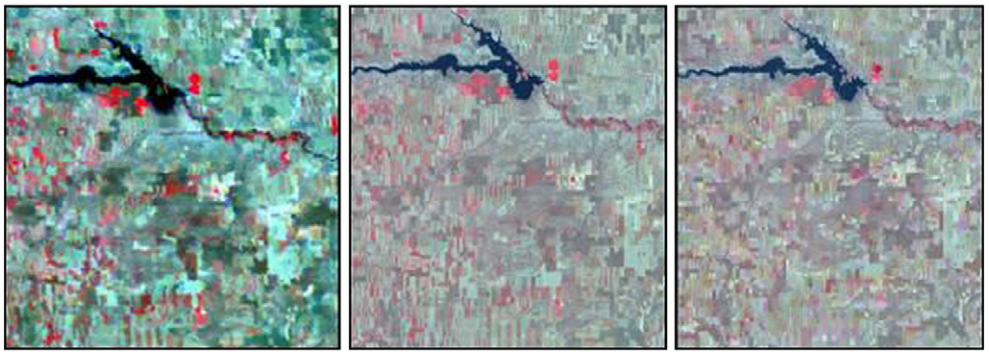

Fig. 2. Visual comparative for 07/22/2009 MODIS (left), 07/23/2009 Landsat (middle), and a STARFM-based reflectance prediction for 07/22/2009 using an August MODIS/Landsat base pair.

ranging from 71% (May) to 87% (August) with a step-wise increase in accuracy by month (Table 3). The producer's accuracies for these models ranged from 60% MT and 79% NT for the May-based model to 78% MT, 93% NT in the August-based model.

The MODIS-based tillage classification resulted in 90% total model accuracy (Table 3). The corresponding kappa value (0.8) indicated that the model had substantial strength in providing class agreements that were greater than chance occurrence (Landis & Koch, 1977). The MODIS-based model producer's accuracies were 82% (18% omission) and 93% (7% omission) for MT and NT, respectively. A lower producer's accuracy resulted in the MT category due to the misclassification of MT fields as NT. Lowered classification accuracies (86% total, 76% MT producer's) resulted when only the 250-m red and NIR MODIS data from each date were included into the model. An evaluation of the model using the 26 MODIS dates and August Landsat scene (referred to as the MODIS and Landsat model), which had been expected to add an increased spatial component to the high temporal data set, did not show improvement over the exclusively MODIS-based model.

The All-data STARFM-based model (Table 3), where five MODIS/Landsat base pairs were used to generate synthetic data for MODIS dates closest to each pair, resulted in the highest overall model accuracy (95%; kappa = 0.89) and had class accuracies similar to the All-data Landsat model. Again, the availability of five Landsat scenes for inclusion as base pairs in STARFM predictions is typically unrealistic but this step was taken for purposes of comparison and to help illustrate the importance of temporal resolution, aided by high spatial resolution, in detecting small-area surface tillage disturbances. The classification model (Table 3) using only the five STARFM-predicted dates within closest proximity to (on the same day of, or +/−2 days) the Landsat scene dates (taken from the abovementioned data, but excluding the full temporal set) had a 4.8% decrease in total classification accuracy and a substantial 12% decrease in MT producer's accuracy when compared to the All-data STARFM model.

The 26-date August STARFM base pair classification model (referred to previously as the likely “worst-case” scenario) had 93% total accuracy and a 0.85 kappa (Table 3; Table 1 illustrates MODIS/Landsat base pairs and corresponding predictive dates). Producer's

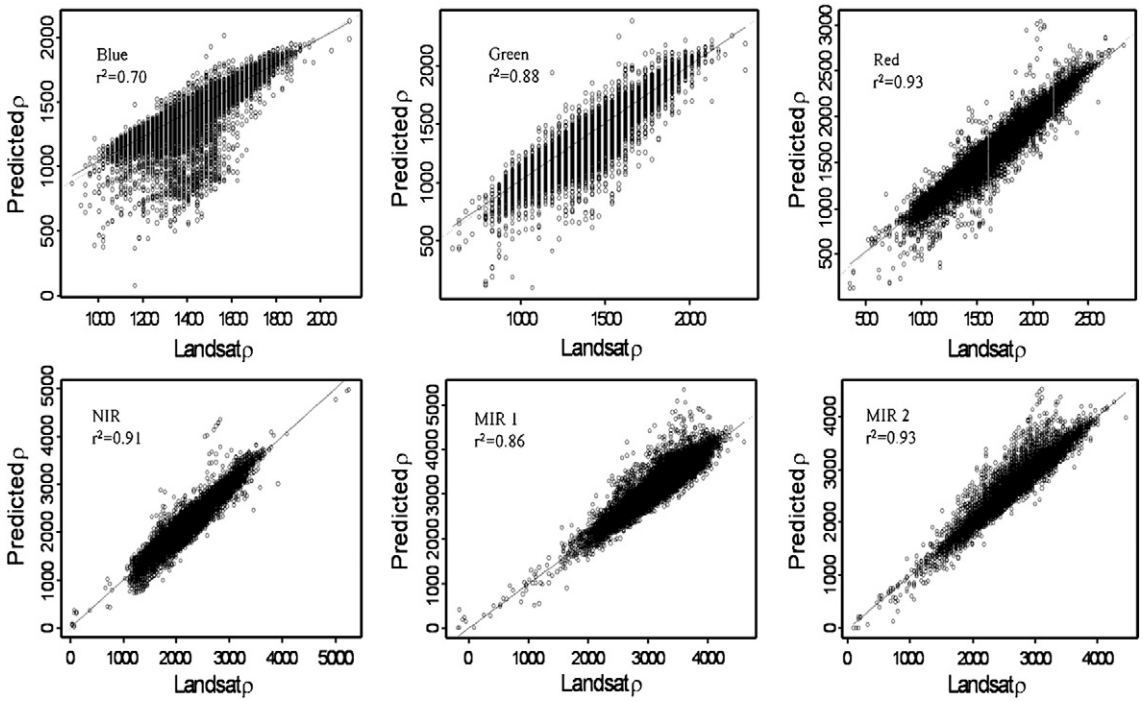

Fig. 3. Band-wise correlations (significant at α = 0.05, p = 0.001) between 24-Aug-2009 Landsat reflectance values and STARFM-predicted values derived from a 24/25-Aug Landsat/MODIS image pair, based on random pixel samples (40,000) taken from cropland. The strongest correlations were observed in the red, near-infrared (NIR), and second middle-infrared (MIR2) bands.

Table 2
Pixel-based (n = 40,000) regression results from Landsat and STARFM-predicted synthetic data.

		Blue	Green	Red	NIR	MIR1	MIR2
24-August Landsat and	r^2	0.7	0.88	0.93	0.91	0.86	0.93
25-August Predicted[a]	Intercept	−258	−201	−147.6	−319	−330	−9.6
	Slope	1.12	1.08	1.04	1.07	1.06	1
22-July Landsat and	r^2	0.42	0.34	0.44	0.65	0.32	0.61
23-July Predicted[a]	Intercept	721	596	630	508	1727	957
	Slope	0.56	0.48	0.62	0.83	0.47	0.64
29-June Landsat and	r^2	0.01	0.01	0.03	0.06	0.03	0
27-June Predicted[a]	Intercept	1140	957	885	2259	2260	1628
	Slope	0.03	0.04	0.1	0.08	0.09	−0.02
24-August Landsat and	r^2	0.41	0.36	0.03	0.75	0.27	0.51
27-June Predicted[a]	Intercept	186	100	701	760	618	−202
	Slope	0.64	0.58	0.14	0.71	0.52	0.66
22-July Landsat and	r^2	0.99	0.99	0.99	0.99	0.99	0.99
23-July Predicted[b]	Intercept	7.49	0.38	12.1	26.25	17.09	12.78
	Slope	0.99	0.99	0.99	0.98	0.99	0.99
29-June Landsat and	r^2	0.41	0.54	0.65	0.45	0.68	0.99
27-June Predicted[c]	Intercept	654	491	337	980	796	4.64
	Slope	0.74	0.94	1.16	0.59	1.17	0.98
28-May Landsat and	r^2	0.59	0.57	0.77	0.73	0.61	0.41
26-May Predicted[d]	Intercept	−108	−235	66	351	−404	186
	Slope	1.03	1.19	0.84	0.77	1.11	0.64

[a] Predictions based on an August Landsat/MODIS base pair.
[b] Predictions based on a July base pair.
[c] Predictions based on a June base pair.
[d] Predictions based on a May base pair.

accuracies resulting from this model were 86% and 97% for MT and NT, slightly higher than the MODIS-based results, but substantially higher than accuracies from the use of a single 26-August Landsat image model. The May and August STARFM-based model (Table 3), where both May and August base pairs were used to predict synthetic data for the 26 scenes, showed decreased classification strength when compared to August-only base pair model.

Field-specific trends in tillage classification accuracy, as influenced by dataset inputs into each model, were not apparent with exception of one field that changed from NT to MT management in 2009. This field had been misclassified as NT in all models except for the five day Landsat and all of the STARFM-based models.

4. Discussion

While past remote sensing studies have successfully classified NT from high disturbance tillage, or intensive tillage from conservation tillage, there has been difficulty in distinguishing between the two primary categories of conservation tillage, MT and NT, due to surface residue similarities (Bricklemyer et al., 2006; Watts et al., 2009). Both these studies incorporated advanced classification algorithms, but were limited to using one or two image dates within a cropping season. This study analyzed tillage management within the same region of north central Montana as the abovementioned studies, but included high temporal resolution data sets within the classification process to ascertain if an increased temporal frequency would improve tillage class accuracy.

4.1. Landsat-based models

The importance of spectral sampling throughout a cropping season, spaced in a manner that maximized the likelihood of capturing tillage-influenced surface changes, was evident when evaluating the Landsat-based classification model accuracies. The All-data Landsat model yielded a MT class producer's accuracy that was substantially better (≥54%) than tillage accuracies reported by Bricklemyer et al. (2006) and Watts et al. (2009) where only June or May/August data had been used. Both studies had examined a binary split between NT and all other forms of tillage, although Watts et al. (2009) attributed the low tillage class accuracy to the misclassification of MT fields as NT. This problem was rectified in our study by including an increased number of image scenes, collected throughout the cropping season, into the classification model-building process. The availability of five quality Landsat scenes is unrealistic in most years due to cloud contamination. Results from the All-data Landsat model, therefore, were used more as a standard of reference for comparing the performance of individual Landsat scene models, and high temporal frequency MODIS and STARFM-based models, than as a practical approach to future mapping.

The monthly increase in classification accuracy observed in the single-date Landsat models demonstrated the importance of temporal sampling when mapping region-wide conservation tillage. Tillage timing is greatly dependent on soil moisture conditions, weed or pest-related soil management requirements, and machine operator availability. Spring tillage is often location dependent and might be nonexistent in some fields due to precipitation events or gradual soil drainage, hence the lower MT accuracies observed within the May Landsat-based model. In contrast, tillage had occurred in all the examined MT fields at least once by the end of August. The August model, consequently, produced the highest accuracies when compared to the other single-date Landsat models but still had only 78% classification success in the MT class. It is likely that spectral changes resulting from earlier summer tillage events had already diminished by the August scene date, which contributed to classification accuracy lower than in the All-data model.

4.2. MODIS-based models

Total and class accuracies for the MODIS-based model were lower than the All-data Landsat model, but the MT producer's accuracy was slightly higher (4%) than the more realistic August-based Landsat model. The resulting MT producer's accuracy (82%) was slightly lower than the often used 85% mapping standard (Wulder et al., 2006) and 5% lower than USGS National Land Cover class accuracy for cropland (Wickham et al., 2010), but still was substantially better than previous conservation tillage results (Bricklemyer et al., 2006; Watts et al., 2009). The MODIS-based classifications were also better than those obtained through visually-based transect surveys where human errors in discerning surface residue amounts reduced tillage accuracies to below 60% (Thoma et al., 2004).

Table 3
Random Forest model tillage classification using MODIS, Landsat, and STARFM-based synthetic data sets.

	Model accuracy		Classification results (n = 125)			Class accuracies (%)	
	Total (%)	Kappa and 95% CI		MT	NT	Producer's	User's
MODIS							
All-data (26 Dates)	90.4	0.8 (0.69–0.91)	MT	41	9	82	93
			NT	3	72	93	89
Landsat							
All-data (5 Dates)[a]	94.4	0.88 (0.8–0.97)	MT	44	6	88	98
			NT	1	74	99	93
May	71.2	0.39 (0.23–0.56)	MT	30	20	60	65
			NT	16	59	79	75
June 13	73.6	0.44 (0.28–0.61)	MT	32	18	64	68
			NT	15	60	80	77
June 29	77.6	0.53 (0.38–0.68)	MT	35	15	70	73
			NT	13	62	83	81
July	84	0.66 (0.52–0.8)	MT	37	13	74	84
			NT	7	68	91	84
August	87.2	0.73 (0.6–0.85)	MT	39	11	78	89
			NT	5	70	93	86
MODIS and Landsat							
26-Date MODIS; Aug. Landsat	91.2	0.81 (0.71–0.92)	MT	41	9	82	95
			NT	2	73	97	89
STARFM							
1. All-data (26 Dates; 5 MODIS/Landsat pairs)[a]	95.2	0.89 (0.82–0.98)	MT	45	5	90	98
			NT	1	74	99	94
2. 26-Date (Aug. MODIS/ Landsat pair)[a]	93	0.85 (0.75–0.94)	MT	43	7	86	96
			NT	2	73	97	91
3. 5-Date (MODIS/ Landsat)[b]	90.4	0.79 (0.68–0.9)	MT	39	11	78	98
			NT	1	74	99	87
4. 26-Date (May and Aug. MODIS/Landsat pairs)	90.4	0.79 (0.69–0.9)	MT	42	8	84	91
			NT	4	71	95	90

[a] Used for comparative/discussion purposes (highlighted), but do not represent realistic data availability due to frequent cloud cover during the summer cropping season.
[b] The MODIS/Landsat date-pairs were used.

Spectral noise from changing reflectance in adjoining fields due to crop phenology was undoubtedly present in the MODIS data; however the relatively high accuracy of the MODIS-based tillage classifications suggests that tillage spectral signals contained within these data were strong enough, in most cases, for the Random Forest algorithm to differentiate between the two classes. Total reliance on MODIS data for tillage mapping in regions with small field widths, however, might be inappropriate due to possible classification problems resulting from mixed-pixel effects. It is possible that the presence of stronger and more localized spectral influences by neighboring fields, resulting from residue burning or intensive tillage events, could compromise NT classifications. Alternatively, a MT field could be misclassified if a high percentage of reflectance from a neighboring NT field was present within the same MODIS pixel.

4.3. STARFM prediction-based models

The incorporation of higher spatial resolution data into analyses where MODIS data are used, as was suggested by Price (2003), might be advantageous when mapping conservation tillage classes in croplands where mixed-pixels are present. High tillage classification accuracies (MT ≥ 86%) were observed for almost all the synthetic-based models and were comparable to, or higher than, the MODIS-based results. It is also possible that the STARFM-based Random Forest models might be better suited to detecting first time tillage occurrences in NT fields. A field converted from NT to MT in 2009 was classified correctly by all of the STARFM-based models and the All-data Landsat model. The MODIS, MODIS and Landsat, and single-date Landsat models incorrectly classified the field as NT. This observation remains inconclusive, however, as further evaluation is necessary due to the single field sample.

The STARFM algorithm was seemingly appropriate for creating high temporal resolution synthetic image sets because of the 30-m Landsat-like output; nonetheless the resulting data predictions varied in their ability to capture localized spectral changes within highly heterogeneous cropland. Lowered accuracies were observed for the five-date synthetic model, compared to the All-data Landsat model, indicating some spectral inaccuracy within the STARFM predictions. Likewise, disparity in spectral reflectance values between the predicted data and Landsat comparatives was observed when examining the per-band regressions; correlations (r^2 values) between the two ranged from 0.41–0.99 and slope values were often less than one. Classification accuracies became comparable (~1% difference) to the five-date Landsat model when synthetic data from an additional 20 dates were incorporated into the classifier. This suggests that although the spectral patterns contained within data from a few synthetic dates were not strong enough for the Random Forest model to adequately separate MT from NT, the high frequency synthetic set contained enough information to compensate for the reduced accuracy.

A temporal variation in accuracy was observed within the synthetic STARFM-based predictions. The strongest correlations occurred between the July-based STARFM predictions and their Landsat comparatives, followed by the August predicted/Landat comparatives. Correlations were substantially weaker between May and June synthetic data and their Landsat comparatives. This variation in accuracy might be attributed to the spectrally mixed MODIS pixels that were incorporated into the STARFM model. STARFM predictions are based on the relationship between MODIS and Landsat base pair data at Date-1 and the relationship between baseline (Date-1) MODIS data and MODIS data at Date-2, for which the 30-m predictions are made. In the above-mentioned cases, STARFM predictions were made for the same MODIS date that was included into the base pair; consequently, the resulting predictions were solely products of the relationship between the 30-m Landsat and the 250-m and 500-m MODIS reflectance values. It is likely that better band-wise correlations occurred in late July because the senesced crop vegetation in neighboring fields had spectral reflectance values more similar to the surface crop residue present in both MT and NT fields, and hence there was less discrepancy between the "pure" Landsat and mixed MODIS data. By the August scene, however, most of the crops would have been harvested. Reflectance is often higher in newly cut fields than in fallow fields where the residue has weathered, so it is probable that the August MODIS mixed-pixel effect would have been more influential within the STARFM predictions.

Tillage classifications resulting from the August base pair STARFM model were slightly better (4% higher MT and NT producer's accuracy) than in the MODIS-based model, and class accuracies were only slightly lower (2%) than the All-data Landsat model. Although the regression analyses for the multi-date August-based predictions showed lowered correlations (r^2 0–0.65) for those further away from the base pair date, the spectral-temporal trends within the 26 day set were strong enough to adequately classify MT from NT. Again, this illustrates the ability for high temporal sampling to compensate for degraded field-based spectral values.

The importance of base-pair date selection for STARFM predictions, and resulting classification accuracies, was also observed in this study. The model generated from May and August base pair predictions resulted in lower classification accuracies than the August STARFM-based model. STARFM gives a high weighting to baseline Landsat data when computing prediction values, hence the resulting predictions have attributes similar to their Landsat base. This relationship between STARFM predictions and their baseline Landsat data was observed within the regression analyses, where June predicted data were more strongly correlated with their baseline August Landsat image than with data from a corresponding June date. Consequently, lowered tillage

accuracy should be expected when using a classification model generated from synthetic STARFM data based on a Landsat image that did not sufficiently capture regional disturbance events.

5. Conclusions

The adoption of conservation tillage management has been advocated globally and within the U.S. due to an increased awareness concerning cropland carbon sequestration potential and the importance of organic, carbon-rich, materials in improving soil quality for long-term food security (Lal, 2004). The collection of conservation tillage statistics has been greatly limited due to the need for an accurate and time-effective tillage mapping approach. Past satellite image-based classifications have had difficulties in differentiating between NT and other forms of conservation tillage (i.e., MT) due to similarities in surface residues. These studies, however, had included only one or two Landsat dates into their analyses. The results from this study demonstrated that adequate conservation tillage accuracy can be achieved by incorporating high temporal MODIS or STARFM-based synthetic data sets into classification model development when multi-date Landsat imagery is not available.

As the resulting classification accuracies were similar between MODIS and STARFM-based models, the decision to use one series over the other for tillage mapping is likely application dependent. MODIS data can be advantageous as they are easy to obtain and require little processing before inclusion into the classification model, although careful pixel sampling techniques are required when field dimensions are smaller than the MODIS pixels to ensure that each spectral sample is most representative for a given field. The incorporation of high frequency STARFM-based predictions into the classification process might be more appropriate in mixed-pixel situations due to spectral contributions from finer resolution Landsat data, which allow for added spectral differentiation between tillage classes. The small increases in tillage accuracy observed within the STARFM-based approach might become more critical when used for carbon-related mapping as any misclassification of MT as NT incorrectly allocates carbon credit (and subsequent monetary payments) to fields with higher CO_2 emissions.

Acknowledgements

This research was funded by the U.S. Department of Energy and the National Energy Technology Laboratory through 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.

References

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5−32.
Bricklemyer, R. L., Lawrence, R. L., & Miller, R. R. (2002). Documenting no-till and conventional till practices using Landsat ETM + imagery and logistic regression. *Journal of Soil and Water Conservation*, *54*, 382−389.
Bricklemyer, R. S., Lawrence, R. L., Miller, P. R., & Battogtokh, N. (2006). Predicting tillage practices and agricultural soil disturbance in north central Montana with Landsat Imagery. *Agriculture, Ecosystems & Environment*, *114*, 210−216.
Chander, G., Markham, D. L., & Barsi, J. A. (2007). Revised Landsat 5 Thematic Mapper radiometric calibration. *IEEE Geoscience and Remote Sensing Letters*, *4*, 490−494.
Chavez, P. S. (1996). Image-based atmospheric corrections-revisited and revised. *Photogrammetric Engineering and Remote Sensing*, *62*, 1025−1036.
Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, *54*, 535−545.
Congalton, R. G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data.*Boca Raton, FL: CRC Press 184 pp.
CTIC (2007). *2007 amendment to the national crop residue management survey summary. National crop residue management survey*. West Lafayette, Indiana, USA: Conservation Technology Information Center.
CTIC (2010). *What's conservation tillage? Core4 Information Fact Sheet.*West Lafayette, IN: Conservation Technology Information Center 2 pp.
Cutler, D. R., Edwards, T. C., Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*, 2783−2792.
Daughtry, C. S. T., Doraiswamy, P. C., Hunt, E. R., Jr., Stern, A. J., McMurtrey, J. E., III, & Prueger, J. H. (2006). Remote sensing of crop residue cover and soil tillage intensity. *Soil and Tillage Research*, *91*, 101−108.
Demattê, J. A. M., Campos, R. C., Alves, M. C., Fiorio, P. R., & Nanni, M. R. (2004). Visible-NIR reflectance: a new approach on soil evaluation. *Geoderma*, *121*, 95−112.
Derpsch, R., & Friedrich, T. (2009). Development and current status of no-till adoption in the world. *Proceedings of the 18th Triennial Conference on the International Soil Tillage Research Organization (ISTRO), 15-19 June, 2009, Izmir, Turkey* 13 pp.
Emam, K. E. (1999). Benchmarking kappa: interrater agreement in software process assessments. *Empirical Software Engineering*, *4*, 113−133.
Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance: predict daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 2207−2218.
Gowda, P. H., Howell, T. A., Evett, S. R., Chavez, J. L., & New, L. (2008). Remote sensing of contrasting tillage practices in the Texas Panhandle. *International Journal of Remote Sensing*, *29*, 3477−3487.
Guo, L., Ma, Y., Cukic, B., & Singh, H. (2004). Robust prediction of fault-proneness by Random Forests. *Proceedings of the 15th International Symposium on Software Reliability Engineering* (pp. 417−428).
Haché, C., Shibusawa, S., Sasao, A., Suhama, T., & Sah, B. P. (2007). Field-derived spectral characteristics to classify conventional and conservation agricultural practices. *Computers and Electronics in Agriculture*, *57*, 47−61.
Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., Gao, F., & White, J. C. (2009). A new fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, *113*, 1613−1627.
Hilker, T., Wulder, M. A., Coops, N. C., Seitz, N., White, J. C., Gao, F., Masek, J. G., & Stenhouse, G. (2009). Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, *113*, 1988−1999.
Ju, J., & Roy, D. P. (2008). The availability of cloud-free Landsat ETM + data over the conterminous United States and globally. *Remote Sensing of Environment*, *112*, 1196−1211.
Lal, R. (1997). Residue management, conservation tillage and soil restoration for mitigating greenhouse effect by CO_2 enrichment. *Soil and Tillage Research*, *43*, 81−107.
Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. *Science*, *304*, 1623−1627.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159−174.
Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, *100*, 356−362.
Lobell, D. B., & Asner, G. P. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, *93*, 412−422.
NRCS (2006). *Tillage practice guide: A guide to USDA-NRCS practice standards 329 no till/strip till/direct seed & 345 mulch till* 1 pp.
NRCS (2007a). *Major biomes map.*: Natural Resource Conservation Service Last Accessed 12 December 2009 <http://soils.usda.gov/use/worldsoils /mapindex/biomes.html>.
NRCS (2007b). *Interactive soil survey map.*: Natural Resources Conservation Service Last Accessed 4 August 2010 <http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx>.
NRCS (2008). *Tillage practice guide: A guide to USDA-NRCS practice standards 329 no-till/strip till/direct seed & 345 mulch till.*USA: Natural Resource Conservation Service, United States Department of Agriculture 1 pp.
Pacala, S., & Socolow, R. (2004). Stabilization wedges: Solving the climate problem for the next 50 years with current technologies. *Science*, *305*, 968−972.
Price, J. C. (2003). Comparing MODIS and ETM + data for regional and global land classification. *Remote Sensing of Environment*, *86*, 491−499.
Reicosky, D. C. (2001). Conservation agriculture: Global environmental benefits of soil carbon management. In L. Garcia-Torres, J. Benites, & A. Martinez-Vilela (Eds.), *Conservation Agriculture: A Worldwide Challenge* (pp. 3−12). Cordoba, Spain: XUL.
Schroers, R., Denham, R., & Witte, C. (2009). Investigating the potential for mapping fallow management practices using MODIS image data. In S. Jones, & K. Reinke (Eds.), *Innovations in Remote Sensing and Photogrammetry, Lecture Notes in Geoinformation and Cartography* (pp. 331−348). Berlin: Springer-Verlag.
SDH-L7 (2006). Science Data Users Handbook-Landsat 7 Last Accessed 10 December 2009 <http://landsathandbook.gsfc.nasa.gov/handbook/handbook.toc.html>.
Shao, Y., Lunetta, R. S., Ediriwickrema, J., & Liames, J. (2010). Mapping cropland and major crop types across the Great Lakes Basin using MODIS-NDVI data. *Photogrammetric Engineering and Remote Sensing*, *75*, 73−84.

South, S., Qi, J., & Lusch, D. P. (2004). Optimal classification methods for mapping agricultural tillage practices. *Remote Sensing of Environment*, *91*, 90−97.
Sperow, M., Eve, M., & Paustian, K. (2003). Potential soil C sequestration on U.S. agricultural soils. *Climatic Change*, *57*, 319−339.
Sullivan, D. G., Strickland, T. C., & Masters, M. H. (2008). Satellite mapping of conservation tillage adoption in the Little River experimental watershed, Georgia. *Journal of Soil and Water Conservation*, *63*, 112−119.
Thoma, D. P., Gupta, S. C., & Bauer, M. E. (2004). Evaluation of optical remote sensing models for crop residue cover assessment. *Journal of Soil and Water Conservation*, *59*, 224−233.
Viña, A., Peters, A. J., & Ji, L. (2003). Use of multispectral Ikonos imagery for discriminating between conventional and conservation agricultural tillage practices. *Photogrammetric Engineering and Remote Sensing*, *69*, 537−544.
Wardlow, B. D., & Egbert, S. L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. central Great Plains. *Remote Sensing of Environment*, *112*, 1096−1116.
Wardlow, B. D., & Egbert, S. L. (2010). A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas. *International Journal of Remote Sensing*, *31*, 805−830.
Watts, J. D., Lawrence, R. L., Miller, P. R., & Montagne, C. (2009). Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing. *Remote Sensing of Environment*, *113*, 1843−1852.
Watts, J. D., Lawrence, R. L., Miller, P. R., & Montagne, C. (Submitted for publication). An analysis of cropland carbon sequestration estimates for north central Montana. 54 pp.
Wickham, J. D., Stehman, S. V., Fry, J. A., Smith, J. H., & Homer, C. G. (2010). Thematic accuracy of the NLCD 2001 land cover for the conterminous United States. *Remote Sensing of Environment*, *114*, 1286−1296.
WRCC (2009). *Historical climate information.*: Western Regional Climate Center Last Accessed 15 January 2010 <http://www.wrcc.dri.edu/CLIMATEDATA.html>.
Wulder, M. A., Franklin, S. E., White, J. C., Linke, J., & Magnussen, S. (2006). An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data. *International Journal of Remote Sensing*, *27*, 663−683.