A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery

Dennis C. Duro [a,*], Steven E. Franklin [a,b], Monique G. Dubé [c]

[a] *School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5C8*
[b] *Environmental and Resource Studies/Geography Department, Trent University, 1600 West Bank Drive, Peterborough, Ontario, Canada K9J 7B8*
[c] *Total E&P Canada Limited, Sustainability Division, #2900, 240-4th Ave SW, Calgary, Alberta, Canada T2P 4H4*

ARTICLE INFO

ABSTRACT

Pixel-based and object-based image analysis approaches for classifying broad land cover classes over agricultural landscapes are compared using three supervised machine learning algorithms: decision tree (DT), random forest (RF), and the support vector machine (SVM). Overall classification accuracies between pixel-based and object-based classifications were not statistically significant ($p > 0.05$) when the same machine learning algorithms were applied. Using object-based image analysis, there was a statistically significant difference in classification accuracy between maps produced using the DT algorithm compared to maps produced using either RF ($p = 0.0116$) or SVM algorithms ($p = 0.0067$). Using pixel-based image analysis, there was no statistically significant difference ($p > 0.05$) between results produced using different classification algorithms. Classifications based on RF and SVM algorithms provided a more visually adequate depiction of wetland, riparian, and crop land cover types when compared to DT based classifications, using either object-based or pixel-based image analysis. In this study, pixel-based classifications utilized fewer variables (15 vs. 300), achieved similar classification accuracies, and required less time to produce than object-based classifications. Object-based classifications produced a visually appealing generalized appearance of land cover classes. Based exclusively on overall accuracy reports, there was no advantage to preferring one image analysis approach over another for the purposes of mapping broad land cover types in agricultural environments using medium spatial resolution earth observation imagery.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The classification of land use and land cover (LULC) from remotely sensed imagery can be divided into two general image analysis approaches: i) classifications based on pixels, and ii) classifications based on objects. While pixel-based analysis has long been the mainstay approach for classifying remotely sensed imagery, object-based image analysis has become increasingly commonplace over the last decade (Blaschke, 2010). Whether pixels or objects are used as underlying units for the purposes of classifying remotely derived imagery, the information contained within - and among - these units (e.g., spectral, textural, etc.) can be subjected to a variety of classification algorithms. Previous comparative studies have been conducted that examine the relative performance of different classification algorithms using pixel-based, and/or object-based

image analysis. A brief summary of selected comparisons is provided below.

1.1. Algorithm comparisons using pixel-based or object-based classifications

Using pixel-based based image analysis on Landsat Thematic Mapper (TM) data, Huang et al. (2002) compared thematic mapping accuracies produced using four different classification algorithms: support vector machines (SVMs), decision trees (DTs), a neural network classifier, and the maximum likelihood classifier (MLC). Their results suggested that the accuracy of SVM-based classifications generally outperformed the other three classification algorithms. Pal (2005) compared the accuracies of two supervised classification algorithms using Landsat Enhanced Thematic Mapper (ETM+) data: SVMs and Random Forests (RFs) (Breiman, 2001), and found that they performed equally well. Gislason et al. (2006) compared a RF approach to a variety of decision tree-like algorithms using pixel-based image analysis of Landsat MSS data. They found that the selected tree-like algorithms tested performed similarly, but that the RF algorithm outperformed the standard implementation of Breiman et al.'s (1984) DTs;

* Corresponding author at: School of Environment and Sustainability, University of Saskatchewan, Room 323, Kirk Hall, 117 Science Place, Saskatoon, Canada SK S7N 5C8. Tel.: +1 705 748 1011x6111.
E-mail address: dennis.duro@usask.ca (D.C. Duro).

however, their findings also showed that the RF algorithm performed slightly less well than a modified DT algorithm (boosted 1R). Carreiras et al. (2006) examined several classification algorithms, which included standard DTs, quadratic discriminant analysis, probability-bagging classification trees (PBCT), and k-nearest neighbors (K-NN) using pixel-based analysis of spatially coarse (1 km pixels) SPOT-4 VEGETATION imagery. Their results, verified by 10-fold cross-validation, showed that the PBCT algorithm produced the best overall classification accuracy. Brenning (2009) compared eleven classification algorithms using a pixel-based image analysis, and Landsat ETM + imagery, for the detection of rock glaciers. This extensive study found that penalized linear discriminant analysis (PLDA) yielded significantly better mapping results as compared to all other classifiers, including both SVMs and RFs. Using Landsat TM and ETM + data, Otukei and Blaschke (2010) compared the MLC, SVM, and DT algorithms in a pixel-based approach, and found DTs performed better than MLC and SVM. In an earlier study, Laliberte et al. (2006) used an object-based approach on Quickbird imagery to compare K-NN with DT algorithms. Their study found that DTs produced better overall classification accuracies than the K-NN algorithm, but that the former was more difficult to implement as compared to the latter.

1.2. Algorithm comparisons between pixel-based and object-based classifications

Relatively recent comparisons between the results of pixel-based and object-based image analysis have also been conducted. For example, Yan et al. (2006) compared pixel-based image analysis using MLC and object-based image analysis using K-NN on Terra Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) imagery. In their study, the authors claimed that the overall accuracy of the object-based K-NN classification drastically outperformed the pixel-based MLC classification (83.25% and 46.48%, respectively). Yu et al. (2006) used high spatial resolution digital airborne imagery and compared a pixel-based classification based on MLC with an object-based classification using K-NN, using a DT as a mechanism for feature selection in both cases. Their study showed that the 1-NN object-based classification outperformed the pixel-based MLC classification by 17%, although calculation of the average classification accuracy of each of the 48 vegetation classes listed was only 51% for the object-based K-NN classification, and 61.8% for the pixel-based classification using MLC. Platt and Rapoza (2008) compared K-NN and MLC for both pixel-based and object-based classifications, with and without the addition of expert-based knowledge, using multispectral IKONOS imagery. Their results revealed that the object-based NN classification using expert knowledge had the best overall classification (78%), while the best pixel-based classification using MLC (without expert knowledge) achieved an overall accuracy of 64%. Castillejo-González et al. (2009) compared pixel-based and object-based classifications in agricultural environments using multispectral Quickbird imagery and a variety of classification algorithms. The best pixel-based classification used non-pan-sharpened imagery and the MLC algorithm, while the best purely object-based classification used pan-sharpened imagery and MLC, with both approaches achieving high overall accuracies of 89.6% and 93.69%, respectively. Their study also revealed that the two best results, using non-pan-sharpened imagery and MLC, showed a small difference in classification accuracy between pixel-based and object-based image analysis (89.60% and 90.66%, respectively); however, the difference between these same approaches grew considerably when using pan-sharpened imagery (82.55% and 93.69%, respectively). Myint et al. (2011) used Quickbird imagery to classify urban land cover. They compared results from a MLC pixel-based classification with an object-based classifier using K-NN and a series of fuzzy membership functions. The object-based classification (90.4%) outperformed the pixel-based classification (67.6%) in overall accuracy for their original image; however, in their test image, the differences between the object-based and

pixel-based approaches was reduced to less than 10% (95.2 and 87.8%, respectively). Finally, in a recent study, Dingle Robertson and King (2011) compared pixel-based and object-based image analysis for classifying broad agricultural land cover types for two time periods (1995 and 2005) using Landsat-5 TM imagery. They compared land cover maps produced using MLC (pixel-based) and K-NN (object-based) algorithms and found that the difference in overall accuracy between these classification approaches was not statistically significant. Despite these findings, an intensive visual analysis of their post-classification analysis revealed that the object-based classification using K-NN depicted areas of change more accurately than the pixel-based classification using MLC.

In general, the above comparisons between pixel-based and object-based classifications reveal that the latter typically outperform the former when comparing overall classification accuracy using a variety of remotely sensed imagery in settings ranging from agricultural to urban land cover classes. However, unlike the studies examining either pixel-based or object-based classifications in isolation, many comparison studies often rely on relatively simple classification algorithms (e.g., K-NN) for the object-based classification, and probabilistic based algorithms (e.g., MLC) for the pixel-based classification, the latter of which is less suited to datasets that are non-normally distributed, or that contain categorical data (Franklin & Wulder, 2002). The present study aims to bridge the gap between these previous comparisons by examining both pixel-based and object-based classification approaches, with a selection of relatively modern and robust supervised machine learning algorithms: decision trees (DTs), random forests (RFs), and support vector machines (SVMs). We conduct a visual and statistical assessment of the classification outputs using medium spatial resolution (10 m) multi-spectral imagery from the SPOT-5 HRG sensor. For the purposes of this study, six broad land cover classes were mapped in a riparian area undergoing intensive agricultural development in western Canada. We assessed each image analysis approach, and each of the selected machine learning algorithms, for their ability to accurately portray these selected land cover types. Recommendations are made in the context of operational mapping of agricultural landscapes for the purposes of general land cover mapping and monitoring in agricultural environments using medium spatial resolution earth observation imagery.

2. Study area

The study area is located along the South Saskatchewan River approximately 90 km east of the provincial borders of Alberta and Saskatchewan (Fig. 1). Approximately 80 sq. km, the study area is a subset of a much larger drainage basin selected for a long-term study of land cover change and land use practices typical of the southern half of the western Prairie Provinces of Canada. Similar large drainage areas have been previously selected by others to assess potential impacts caused by development on aquatic ecosystems over time (e.g., Squires et al., 2009), and represent an appropriate scale and unit of measurement for conducting cumulative environmental effects assessments on aquatic ecosystems (Dubé, 2003; Duinker & Greig, 2006; Noble, 2008; Seitz et al., 2011). Indeed, over the past century, environmental impacts in the region due to agricultural development has replaced much of the native vegetation and has filled an estimated 40% of small wetland areas (Huel, 2000), facilitating the gradual introduction of crops and improved pasture lands that dominate much of the prairies today. The selected study area is typical of agricultural activity conducted near riparian and wetland environments in the region. Such environments have been linked to a range of species and environmental processes, the flow of nutrients between terrestrial and aquatic ecosystems, and are the focus of best management practices for protecting water quality in agricultural environments (Cooper et al., 1995; Gordon et al., 2008; Gregory et al., 1991; Naiman & Décamps, 1997; Thompson & Hansen, 2001; US EPA, 2005). Climate in the Prairie Ecozone is characterized by long and cold winters, with summers being
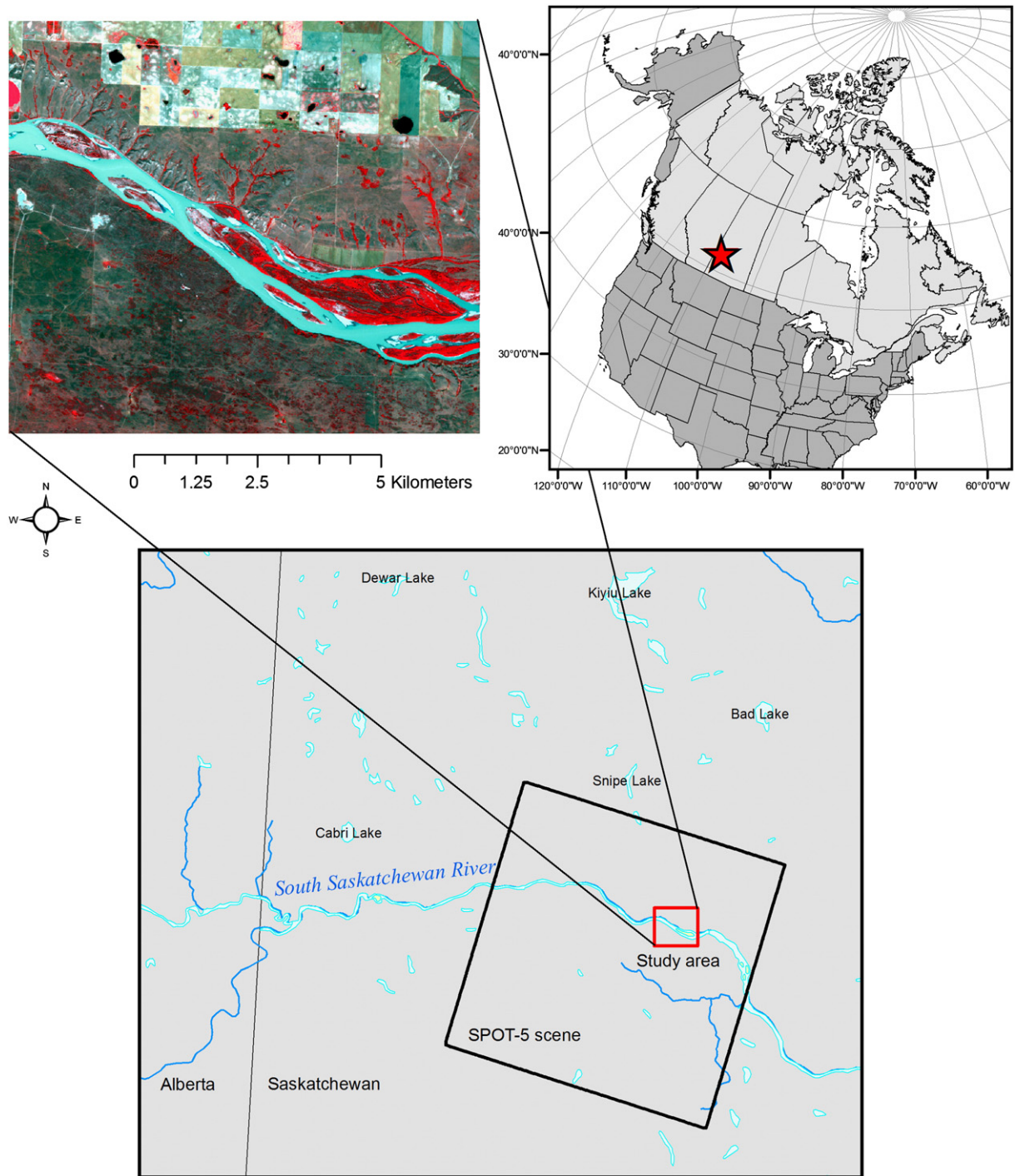
Fig. 1. Study area situated over the South Saskatchewan River (Saskatchewan, Canada). Inset shows SPOT-5 10 m HRG false color image of study area (R = NIR, G = Red, B = Green).

relatively short, but often very warm. The region receives little precipitation and is relatively dry as a result, with semi-arid regions existing in the southern portions of the province (e.g., The Great Sand Hills).

3. Methods

3.1. Data sets and processing

3.1.1. Ancillary datasets

Several tiles of the Canadian Digital Elevation Data (CDED) digital elevation model (DEM) were downloaded from the GeoBase online spatial data portal (www.geobase.ca). At latitudes of less than 68°

N, the CDED DEM has a horizontal post spacing of approximately 23 m (North–south) × 16–11 m (East–west). After projection into Albers Equal Area Conic and nearest-neighbor resampling, the CDED DEM was converted to square 16 × 16 m pixels. An Albers-Equal Area Conic was selected as the final projection for all data used in this study due to known area and shape preserving characteristics of this projection, and because using a standard Universal Transverse Mercator projection would have spanned multiple zones, introducing potential projection-related errors in the final map output. Together with elevation above sea level, slope and aspect, topographic features (e.g., ridge, channel, plane) (Pike, 2000) were calculated from the CDED DEM and included as variables in the classification process.

Table 1
Image layers used in pixel-based classifications.

Spectral bands	Vegetation index	Landscape variables	Texture measure[a]
Green	NDVI	Elevation	Green
Red		Slope (degrees)	Red
NIR		Aspect (degrees)	NIR
SWIR		Topgraphic class[b]	SWIR
			NDVI
			DEM

[a] – "Angular second moment" texture calculated for the listed image layers.
[b] – Topographic classes: Plain, Ridge, Channel (Pike, 2000).

Other ancillary datasets (e.g., road networks, geodetic monuments, administrative boundaries, etc.) were downloaded from the GeoSask online spatial data portal (www.geosask.ca), and used as reference layers for geometric and orthographic corrections of satellite imagery.

3.1.2. Remotely sensed imagery

Panchromatic (2.5 m) and multispectral (10 m) imagery from the Système Pour l'Observation de la Terre (SPOT-5) satellite was obtained from the Alberta Terrestrial Imaging Corporation (www.imagingcenter.ca). The SPOT-5 imagery was collected on August 28, 2005. High resolution digital color aerial orthoimagery (60 cm pixels) obtained in the same year as the SPOT-5 imagery was downloaded from the Saskatchewan Geospatial Imagery Collaborative (www.flysask.ca) online data portal. The panchromatic imagery was orthorectified using a rational polynomial coefficient model of the SPOT-5 sensor and the CDED DEM mosaic, in conjunction with ground control points obtained from ancillary layers (road network and geodetic monuments). Image-to-map registration yielded a root-mean-square error (RMSE) of 0.3 pixels using a 1st order polynomial transformation. The multispectral imagery was then registered to the panchromatic imagery, achieving an RMSE of less than 0.5 pixels using a 1st order polynomial transformation. A visual assessment confirmed that all image sources were aligned with ancillary data layers of higher spatially accuracy (e.g., road network, quarter section plots, etc.). The multispectral SPOT-5 scene was examined for a suitably representative study area, and a 630×553 pixel subset (348,390 pixels) of the full SPOT-5 scene was then selected for analysis (Fig. 1).

Radiometric processing was applied to the SPOT-5 multispectral imagery, and the Normalized Difference Vegetation Index (NDVI) layer was computed and included in the analysis (Song et al., 2001). Calibrated digital numbers (DNs) were first converted to top-of-atmosphere reflectance following procedures outlined by Chander et al. (2009) with updated sensor calibration coefficients for both SPOT-5 HRG sensors provided by the Centre National d'Études Spatiales (CNES, 2009), and updated exoatmospheric solar irradiance coefficients using the Thuiller spectrum (Thuillier et al., 2003) provided by G. Chander (personal communication, Sept. 2010). Absolute atmospheric correction of the imagery was not performed due to the lack of simultaneously acquired ground based spectral data or appropriate meteorological data available in the study area. Instead, a relative correction using the Dark-Object Subtraction (DOS) method was used to alleviate atmospheric scattering effects (Chavez, 1988). The second angular moment texture measure, from computed co-occurrence matrices, was calculated for each of the SPOT-5 multispectral bands and NDVI layer. Texture measures have been found to increase overall classification accuracies using SPOT imagery (Franklin & Peddle 1990), and have been shown to improve the quality of the image segmentation process (Ryherd & Woodcock, 1996).

The four bands of SPOT-5 multispectral imagery were placed in a single data set along with the calculated NDVI layer, texture measures, DEM, and related landscape variables. This combined data set, or "image stack", consisted of 15 individual layers, or predictor variables (Table 1). Pixel-based variables were selected from this stack based on previous experience in classifying land cover types in our study area. The object-based classification used several layers from the pixel-based image stack as input to the image segmentation process, and as input layers for the calculation of "object features" (see Section 3.1.3 for details).

3.1.3. Image segmentation and object feature selection

Image segmentation represents a fundamental first step in object-based image analysis, as the image objects (sensu stricto "image segments") resulting from this process form the basis of an object-based image classification (Castilla & Hay, 2008). In this study, image segmentation was performed using the multi-resolution segmentation (MRS) algorithm found in the 64-bit version of eCognition Developer 8 (Trimble, 2010a). The MRS algorithm uses a "bottom-up" image segmentation approach that begins with pixel sized objects which are iteratively grown through pair-wise merging of neighboring objects based on several user-defined parameters (scale, color/shape, smoothness/compactness) that are weighted together to define a homogeneity criterion; together, these parameters define a "stopping threshold" of within-object homogeneity based on underlying input layers, and thus define the size and shape of resulting image objects (Baatz & Schäpe, 2000; Benz et al., 2004; Trimble, 2010b).

Of the parameters used by the MRS algorithm, the selection of an appropriate value for the "scale" parameter is considered the most important, as this value controls the relative size of the image objects, which has a direct effect on the classification accuracy of the final map (Kim et al., 2008; Myint et al., 2011; Smith, 2010). In general, smaller values for the scale parameter produce relatively smaller image objects, while larger values produce correspondingly larger objects. An examination of the available literature reveals that a quantitative, semi-automated approach for the selection of optimum values for image segmentation parameters using genetic algorithms exists (e.g., Bhanu et al., 1995), but that such semi-automated methods are not yet fully implemented in mainstream image segmentation software (e.g., Definiens' eCognition; but, see Costa et al., 2008; Drăguţ et al., 2010). In this study, the selection of appropriate input layers and values for individual parameters used by the MRS algorithm was guided by previous experience and by using an iterative "trial-and-error" approach often employed by others conducting object-based image analysis (Dingle Robertson & King, 2011; Yan et al., 2006; Mathieu et al., 2007; Myint et al., 2011; Yu et al., 2006). The values for image segmentation parameters used in this study are found in Table 2.

The image segmentation process was considered complete once image objects were produced that visually corresponded to meaningful real-world objects of interest. Image objects produced using the smallest scale parameter (Fig. 2B) were sufficiently small enough to delineate fine scale features of interest within the study area such as narrow channels of riparian vegetation, or fringes of wetland vegetation located around pools of water. The two additional, coarser image segmentation scales (Fig. 2C and D) were included in the object-based classification to depict larger objects of interest (e.g., crop fields). The use of image object information derived from multiple image segmentation scales has been shown elsewhere to produce

Table 2
Parameter values used in multi-resolution segmentation (MRS) algorithm.

Image segmentation parameters[a]				
Scale	Color/shape	Smoothness/compactness	# of objects	Median area of objects (sq. m)
5	0.9/0.1	0.5/0.5	6583	9401
15	0.9/0.1	0.5/0.5	937	69243
30	0.9/0.1	0.5/0.5	273	241434

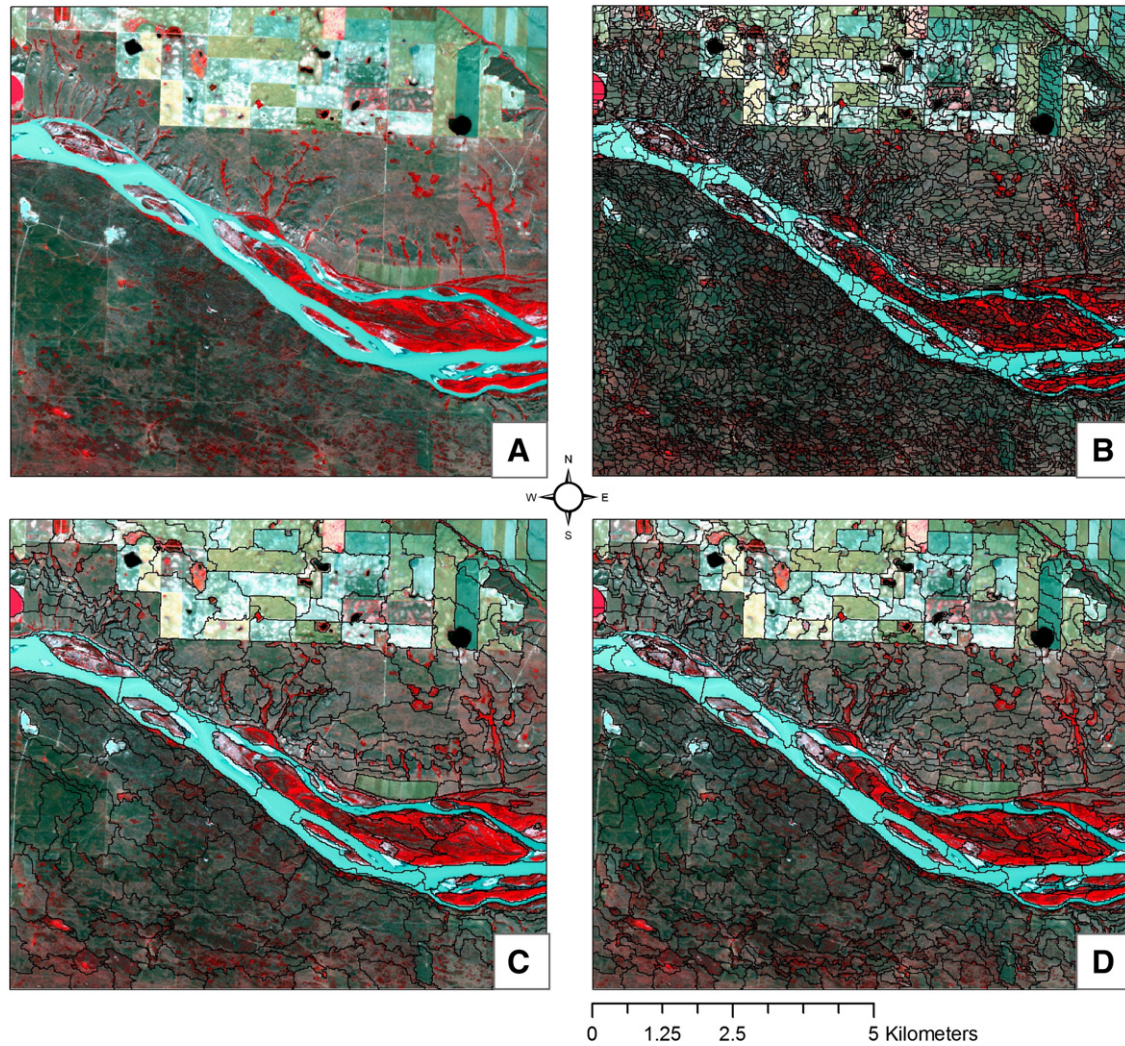[a] Image layers used: NDVI, DEM, and slope (weighted equally).

Fig. 2. Comparison of image segmentation levels used in object-based classification: A) SPOT-5 10 m HRG false color image of study area (R—NIR, G—Red, B—Green); B) Image segmentation (MRS scale 5); C) Image segmentation (MRS scale 15); D) Image segmentation (MRS scale 30).

Table 3
Object features used in object-based classifications (adapted from Trimble, 2010a).

Object features [a]	
Object layer features	Description
Mean	Mean value of an image object
Standard deviation	Standard deviation of image object
Mean difference to neighbors	The difference between mean values of an image object and neighboring image objects.
Mean difference to scene	The difference between the mean input layer value of an image object and the mean input layer value of the scene
Mean difference to super-objects	The difference between the mean input layer value of an image object and the mean input layer value of its superobject. Distance of 1.
Std. dev. difference to super-object	The difference of the std. dev. input layer value of an image object and the std. dev. input layer value of its super-object. Distance of 1.
Object texture features	Description
Mean of sub-objects	Standard deviation of the different input layer mean values of the sub-objects. Distance of 1.
Avrg. mean diff to neighbors of sub-objects	The contrast inside an image object expressed by the average mean difference of all its sub-objects for a specific input layer. Distance of 1.

[a] Object features listed were calculated for each of the 15 image layers listed in Table 1.

better overall classification accuracies (Smith, 2010), and better classification accuracies for individual land cover classes (Myint et al., 2011). Image objects produced at the finest image segmentation scale served as the underlying building blocks, or "image segments" (Castilla & Hay, 2008), used for the object-based classification, although information obtained from image objects produced at all three image segmentation scales (Figs. 2B–D) was utilized in the object-based classifications.

Following the image segmentation process, variables were selected for use in the object-based classification. The object-based image analysis software used in this study refers to such variables as "object features" (Trimble, 2010a), which is a term adopted throughout the rest of the text when referring to variables used by object-based classifications. Object features allow for contextual relationships between image objects to be incorporated into the object-based image analysis. For example, relationships between several smaller sub-objects (e.g., groups of individual crops) contained within a single image object (e.g., crop field) produced using a larger image segmentation scale, can be used for discriminating between land cover types (Myint et al., 2011). In such cases, the information being considered represents an "object texture feature" (see Table 3). Several types of object features are available within the Definiens eCognition software and are described elsewhere (Trimble, 2010a).

Selecting object features for use in object-based image analysis can be a subjective process based on past experience and user knowledge (e.g., Laliberte et al., 2007), or one driven by a feature selection algorithm prior to final classification (e.g., Yu et al., 2006; Van Coillie et al., 2007). In this study, we relied on our past experience with conducting object-based classifications in the study area to guide our selection of object features (Table 3).

The total number of object-features considered in a multi-scale object-based classification can be considerable since information is calculated per image object, and can be calculated at each segmentation scale for each of the input layers. In this study, information based on all 15 input layers (Table 1), 3 image segmentation scales (Table 2), and 8 object features (Table 3) were used in the object-based classification. The total number of object features considered (360) in the object-based image analyses was reduced to 300 as the calculation of values for certain object features required that certain conditions are met. For example, in this study, "object texture features" (Table 3) were selected that calculate values for an individual image object based on their underlying sub-objects, which are created at lower image segmentation scales. However, image objects produced at the finest image segmentation scale represent the finest level of detail, and therefore cannot be used to calculate sub-object information.

The total number of object features available to the object-based classifications greatly outnumbers the number of variables used in pixel-based classifications (300 versus 15, respectively). The ability to utilize and link information from image objects delineated at multiple scales inherent in the underlying imagery is often presented as one of the advantages of object-based image analysis (Blaschke, 2010). As such, multiple image segmentation scales were used for object-based classifications, which is an approach that has been adopted in several other recent studies comparing pixel-based and object-based classifications (e.g., Yan et al., 2006; Myint et al., 2011; Whiteside et al., 2011). While utilizing disparate numbers of potential predictor variables may hamper a strict comparison between image analysis approaches, it nonetheless represents a more typical comparison, as object-based classifications often utilize multiple image segmentation scales even if a single object-feature type is utilized (e.g., mean layer value; see Table 3).

3.1.4. Sampling data, accuracy assessment, and map comparison

In this study, high spatial resolution aerial orthophotos and panchromatic satellite imagery were used to collect ground reference data, as contemporaneous field-based samples were not available within the selected study area. A stratified random sampling approach was utilized in order to adequately sample land cover classes of interest (e.g., narrow channels of riparian vegetation) that were relatively underrepresented within the study area. An initial land cover map produced using an unsupervised ISODATA clustering algorithm was created to provide an initial stratification of the study area. Four multispectral bands from the SPOT-5 imagery were used to produce the initial stratified classification using 20 spectral classes. Six broad land cover classes were selected for the purposes of this comparison study: crop land, mixed grasslands, exposed rock/soil, riparian and wetland vegetation, and water (cloud and shadow were not present in the study area). The 20 spectral classes produced by the ISODATA algorithm were grouped into the six selected land cover types. Spectral classes remaining from the ISODATA classification that did not clearly fit into the selected six land cover types were classified as "no data" and excluded from further analysis. The generalized ISODATA classification was then converted into a polygon based map and imported into a GIS for further analysis.

Using image objects produced at the finest segmentation scale (Table 2), and the polygon-based ISODATA classification, a stratified random sample of image objects within the six land cover types was performed. A total of 690 image objects were selected (115 per land cover type). Image objects produced using the MRS algorithm

– even using small image segmentation scale values – can vary in size considerably (see Table 2), and may contain more than a single land cover type. As such, image objects were visually examined using a combination of SPOT-5 panchromatic and multispectral data, along with color aerial orthoimagery, to assess the homogeneity of the land cover types present within individual image objects. Those image objects that contained more than one of the six broad land cover types were rejected, leaving 679 samples in total. These samples were then split into training and testing set using proportional stratified random sampling, which allowed for both sets of data to retain the overall class distributions of the six selected land cover types present in the original data set. Approximately two-thirds of the samples (437) were used to train the machine learning algorithms, reserving approximately one-third (242) as a "hold-out" test set used exclusively for accuracy assessment and statistical comparisons between classifications. To clarify further, the test set was not used to train or tune parameters associated with the machine learning algorithms examined in this study. Model building and tuning of individual parameters used by the machine learning algorithms was accomplished through repeated k-fold cross-validation based on the training data set only (see Section 3.2).

In order to obtain training and testing samples for the pixel-based classification that were commensurate with training and testing image objects, a single point within each of the selected image objects was randomly selected. As each of the image objects used for training and testing were visually screened for land cover homogeneity, any point within the image object would correspond to the underlying land cover type already identified for the image object. This procedure ensured that both the object-based and pixel-based classifications used training and testing data gathered from the same locations.

Two measures for assessing the accuracy of thematic maps classified from remotely sensed imagery are commonly reported: i) overall accuracy and ii) the Kappa coefficient of inter-rater agreement (Congalton, 1991; Congalton & Green, 1998). Overall accuracy has the advantage of being directly interpretable as the proportion of pixels classified corresponds to probabilities related to a given thematic map's reported commission and omission accuracy (Stehman, 1997), while the Kappa coefficient has been used to assess statistical difference between classifications (Congalton, 1991). Studies often assess the performance of multiple classification algorithms utilizing the same testing and training samples (Foody, 2004). In such cases, the assumption that each classification was independently assessed is violated (Cohen, 1960) – i.e., that the number of the samples being compared are independent – and therefore, a statistical comparison using Kappa coefficient values is unwarranted (Foody, 2004). In such circumstances, it has been recommended that either a Monte Carlo permutation test of related κ coefficient values (McKenzie et al., 1996), or McNemar's test for paired-sample nominal scale data (Agresti, 2002; Zar, 2009), be used to assess whether statistically significant differences between classifications exists (Foody, 2004). The latter approach has been used by others to statistically compare object-based and pixel-based classifications (e.g., Dingle Robertson & King, 2011; Yan et al., 2006; Whiteside et al., 2011), and is therefore adopted here for comparability.

For each classification, a confusion matrix is presented, along with its overall accuracy (i.e., the percentage of correctly classified land cover types), and user's and producer's accuracy (Congalton & Green, 1998). As recommended by others, overall accuracy measures are reported using exact 95% confidence intervals (Morissette & Khorram, 1998; Foody, 2009). The McNemar test was used to assess the following goals of comparison: 1) whether a statistically significant difference exists between pixel-based and object-based classifications that utilize the same machine learning algorithm; and, 2) whether a statistically significant difference exists between different

machine learning algorithms when using either pixel-based or object-based image analysis. The McNemar test was run without Yates' correction for continuity for small sample sizes, as this correction is generally not recommended (Foody, 2004; Zar, 2009). Both the individual accuracy assessments and statistical comparisons are based on the "hold out" test set.

3.2. Tuning of machine learning algorithm parameters

Model building, tuning, and accuracy assessments of were performed using version 2.12 of the 64-bit version of R, a multi-platform, open-source language and software for statistical computing (R Development Core Team, 2010). Several add-on packages were used within R for creating each of the machine learning algorithms used in this study: decision tree (DT), random forest (RF), and the support vector machine (SVM). Classifications based on DT models used the Recursive PARTitioning or "rpart" package (Therneau & Ripley, 2010), which is largely based on the classification and regression tree (CART) algorithm originally developed by Breiman et al. (1984). The classifications built with the RF algorithm used the "randomForest" package (Liaw & Wiener, 2002), which is based on the original RF algorithm and software code developed by Breiman and Cutler (Breiman, 2001; Breiman & Cutler, 2007). Classifications using models based on the SVM algorithm (Cortes & Vapnik, 1995; Vapnik, 1998) used the "kernlab" package (Karatzoglou et al., 2004).

All classification models were developed using the "caret" package within R (Kuhn, 2008), which allowed for a single consistent environment for training each of the machine learning algorithms and tuning their associated parameters. A repeated *k*-fold cross-validation resampling technique was used to create and optimize classification models for both pixel-based and object-based classifications using all three machine learning algorithms. Resampling by *k*-fold cross-validation begins by partitioning a sample into k subsamples of roughly equal size, with k-1 subsamples used as a training set, and a single subsample left out as a test set. Using this approach, a classification model using each of the three machine learning algorithms is built using the training set and assessed against the single leftover test set. This process is repeated *k* times ("folds"), whereby each of the *k* subsamples serves a turn as a test set, ensuring that all subsamples are used as part of the training and testing sets. Results for each fold are then combined to select the model with the highest average accuracy. Similar cross-validation techniques have been used by others to compare the performance of multiple classifiers using earth observation imagery (e.g., Friedl & Brodley, 1997; Huang et al., 2002; Brenning, 2009, 2010).

Several adjustable "tuning parameters" used by each of the machine learning algorithms to optimize classification performance were examined using 10-fold cross validation, which is the number of folds recommended when comparing the performance of machine learning algorithms (Kohavi, 1995). "Optimal" values for tuning parameters were selected using three repetitions of a 10-fold cross-validation based on the original training data set, with the original test removed completely from the cross-validation process (i.e., the original test set was not used for training or tuning any of the classification models). Tuning parameters were considered optimized based on classification models that achieved the highest overall classification during the cross validation process. Specific details on tuning parameters used by the three machine learning algorithms examined in this study are listed in the following sections.

3.2.1. Decision Tree based models

For DT based classifications, several values were examined for the "maximum depth" tuning parameter, which controls the maximum depth of any single node in the tree. When using the "caret" package, an initial DT model is fit to all of the training data to obtain the

maximum depth of any node; this value is then used to obtain an upper bound on values considered during subsequent model building using cross validation (Kuhn, 2011). In general, using a larger maximum depth value will allow for a relatively complex tree to be built, with a potential increase in overall classification accuracy, whereas lower maximum depth values tend to build less complex trees, with potentially lower overall classification accuracies. By increasing the number of branching nodes (i.e., decision rules), the DT algorithm is capable of grouping a larger number of distinct observations present within a dataset. By default, the "rpart" package uses 10-fold cross validation of the training data to internally obtain classification error rates (Therneau & Ripley, 2010). When using "rpart" the appropriate sized tree is obtained using the "1 SE rule" established by Breiman et al. (1984), whereby the smallest-sized tree whose cross validation error is within 1 standard error of the minimum cross validation error is selected. The tree is then pruned using the "cost complexity" (*cp*) value that corresponds to the size of tree found using the "1 SE rule". The *cp* parameter controls the condition at which non-informative splits are pruned from the tree (Therneau & Ripley, 2010). Using the "caret" package, the default *cp* value (0.01) used by the "rpart" package was maintained, and only the maximum depth parameter was tuned for DT based classifications.

3.2.2. Random Forest based models

For random forest (RF) based classifications, the default number of trees (500) was selected since values larger than the default are known to have little influence on the overall classification accuracy (Breiman & Cutler, 2007). The other adjustable RF tuning parameter, the m_{try} parameter, controls the number of variables randomly considered at each split in the tree building process, and is believed to have a "somewhat sensitive" influence on the performance of the RF algorithm (Breiman & Cutler, 2007). For categorical classifications based on the RF algorithm, the default value for the m_{try} parameter is $\sqrt{p}$, where *p* equals the number of predictor variables within a dataset (Liaw & Wiener, 2002).

3.2.3. Support Vector Machine based models

Classifications based on the support vector machine (SVM) algorithm used the radial basis function (RBF) kernel. Other kernels were not considered in this study. The parameters used by the SVM algorithm have been shown to influence overall classification accuracy (Burges, 1998). The two model tuning parameters for SVM models using the RBF kernel in the "kernlab" package are "cost" (C) and "sigma" (σ). Increasing the former leads to larger penalties for prediction errors, which may produce an over-fitted model (Alpaydin, 2004); whereas increasing the latter parameter affects the shape of the separating hyperplane (Huang et al., 2002), which may also influence overall classification accuracy. An analytical method for directly estimating σ from the training data has been implemented in the kernlab package using the "sigest" function (Karatzoglou et al., 2004). The "caret" package estimates an appropriate value for the σ parameter using the sigest function by default; therefore, only the C parameter was tuned when running the SVM algorithm with the RBF kernel (Kuhn, 2011).

4. Results

4.1. Tuning of machine learning algorithm parameters

For DT-based classifications, values ranging from 1 to 8 were examined for the "maximum depth" tuning parameter. Based on the highest overall classification accuracy (i.e., the percentage of correctly classified samples) achieved by pixel-based and object-based models (85.4% and 83.3%, respectively) a maximum depth value of 8 was selected for both pixel-based and object-based classifications models. Several values for the m_{try} tuning parameter (2–4, 6–8, 10–12, 14)

were examined for the pixel-based RF classification. For the pixel-based RF classification, the highest classification accuracy value (91.1%) was obtained with an m$_{try}$ value of 7. A total of 10 m$_{try}$ parameter values (2, 35, 68, 101, 134, 167, 200, 233, 266, and 300) were examined for the object-based RF classifications. Based on the highest classification accuracy obtained (93.1%), an m$_{try}$ value of 68 was selected for the object-based RF classification. For the pixel-based and object-based classifications using the SVM algorithm, a total of 10 values for the C parameter (0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, and 128) were examined. The value for the σ parameter was held constant at 0.0928 for pixel-based classifications, and at 0.00361 for object-based classifications. Pixel-based and object-based classifications using the SVM algorithm (overall accuracy of 89.8% and 91.4%, respectively) were obtained using C parameter values of 8 and 1, respectively. Models with optimized tuning parameter values were used to produce the subsequent image classifications, associated accuracy assessments, and map comparisons.

4.2. Visual examination of thematic maps

Pixel-based and object-based image classifications using the three examined machine learning algorithms are depicted in Figs. 3 and 4, respectively. Post classification clean up (e.g., pixel-based filtering, GIS-based adjustment of classes, etc.) of the final thematic maps was not performed. A visual overview of the pixel-based classifications is presented first, followed by object-based classifications, and a comparison of outputs produced using both image analysis approaches and all three machine learning algorithms.

4.2.1. Pixel-based classifications

For the pixel-based classifications (Fig. 3), the major visual difference interpreted between thematic maps produced by the three different algorithms was the amount of wetland or riparian land cover depicted in the southern quarter of the study area. For tree-based classifications (Fig. 3B and C), the south-western corner of the study
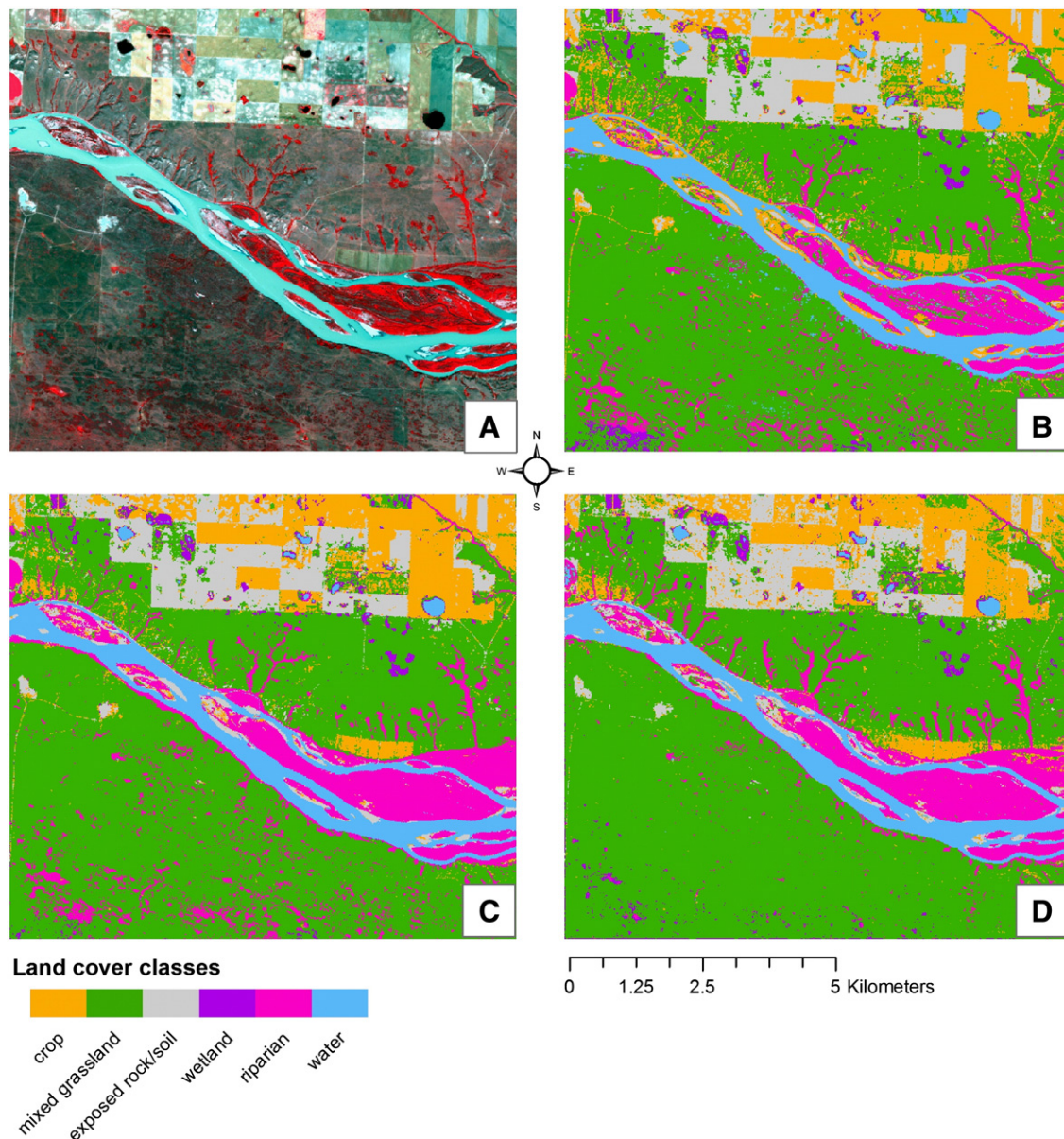

Fig. 3. Comparison of pixel-based classifications: A) SPOT-5 10 m HRG false color image of study area (R—NIR, G—Red, B—Green); B) Decision tree based classification; C) Random forest based classification; D) Support vector machine based classification.

Fig. 4. Comparison of object-based classifications: A) SPOT-5 10 m HRG false color image of study area (R—NIR, G—Red, B—Green);); B) Decision tree based classification; C) Random forest based classification; D) Support vector machine based classification.
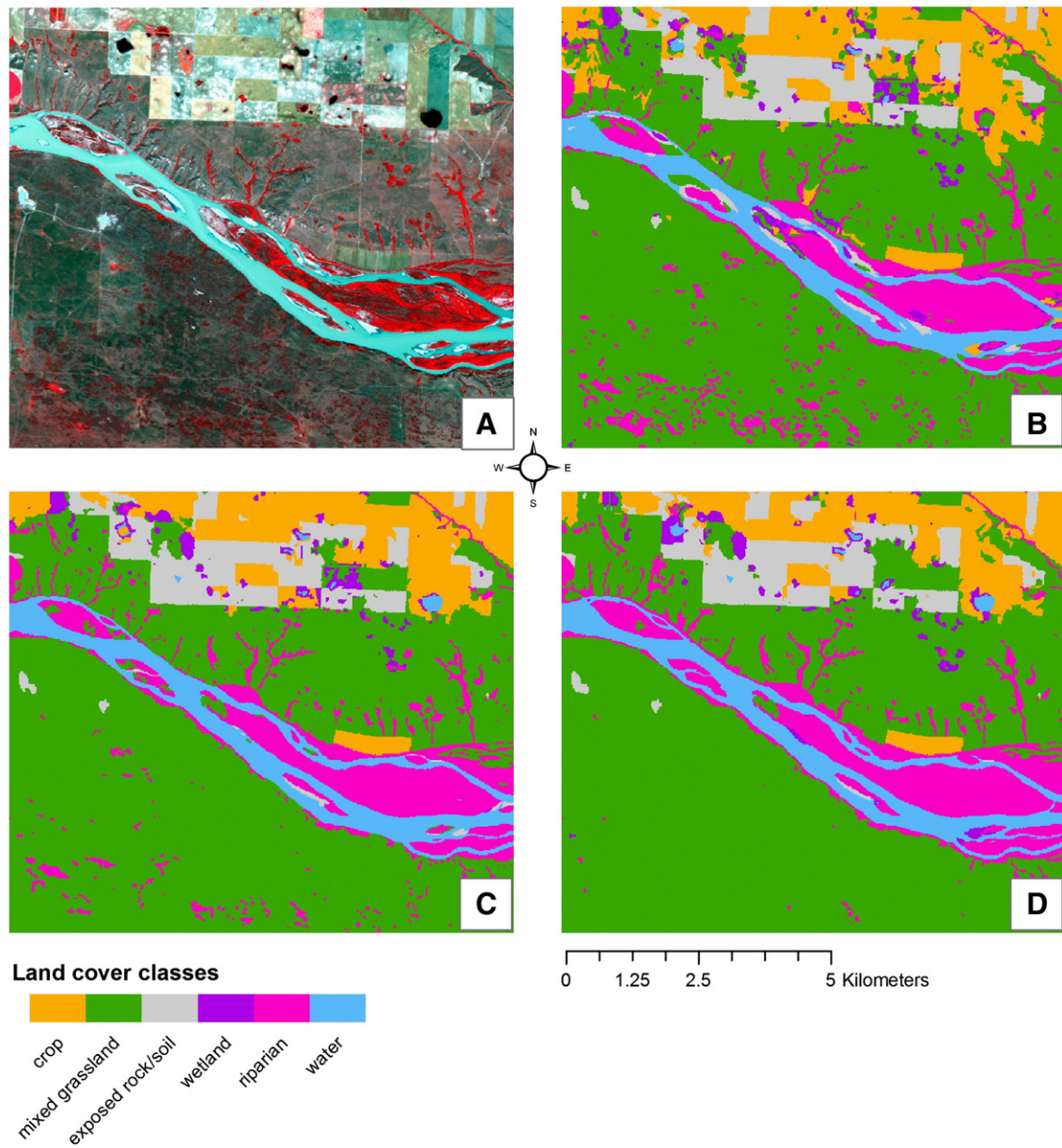
area depicts riparian vegetation, whereas the map produced by the SVM algorithm (Fig. 3D) depicts this area as dominated by mixed grasslands dotted primarily with wetlands. A visual inspection of this area using available high spatial resolution imagery and color orthoimagery revealed that this area is predominantly covered in vegetation typical of a mixed grasslands land cover type, although small stream channels can be seen filled with vegetation, indicating the presence of a riparian land cover class. Small areas of wetland vegetation are also present in the high resolution imagery. Two predominant patches of exposed rock/soil, shown as blue-white patches on the left portion of Fig. 3A, are best classified by the SVM algorithm, while both the RF and DT algorithms depict these areas with patches of crop land. In general, while all three pixel-based classifications produced a similarly speckled "salt-and-pepper" appearance, the DT and RF based classifications showed noticeably less of this speckle in the depiction of large crop land areas (e.g., see north-eastern corner of Fig. 3C). Overall, the pixel-based classification using the SVM algorithm (Fig. 3D) appears to contain less speckle compared to the DT

and RF classifications. The classification based on the SVM algorithm appears to show fewer errors of commission in the classification of mixed grassland vegetation along the north-western area, especially along channels containing riparian vegetation on the north side of the river.

4.2.2. Object-based classifications

As with the pixel-based classification, the major visual difference interpreted between thematic maps produced using object-based image analysis (Fig. 4), is in the relative amount of wetland, riparian and mixed grassland land cover depicted in the southern half of the study area. For tree-based classifications (Fig. 4B and C), the southern half of the study area depicts larger patches of riparian vegetation, whereas the SVM algorithm (Fig. 4D) depicts this area as predominantly mixed grassland. The thematic maps based on DT and SVM algorithms (Fig. 4B and C) show several noticeable errors of commission, namely the misclassification of riparian land cover as wetland within the main river channel. All three object-based classifications

misclassify small areas of riparian and exposed/rock soil land cover located along the riverbank as mixed grasslands. The two object-based classifications using the RF and SVM algorithm show little indication of commission error when classifying crop land alongside riparian channels on the northern slope of the river channel, whereas several patches of misclassified crop land are present in this area of the object-based DT classification map. Wetland vegetation present in the northern part of the study area appears well defined by all three object-based classification algorithms, although several errors of commission are noticeable in large inundated fields.

4.2.3. Visual comparison of pixel-based and object-based classifications

In general, all land cover maps show a reasonably accurate visual depiction of the broad land cover types of interest in this area. When the same machine learning algorithm is compared, both pixel-based and object-based classifications showed similar patterns. For example, the predominance of mixed grassland areas in the southern portion of the study area was noticeably higher in pixel-based and object-based classifications that utilized the SVM algorithm when compared to classifications based on tree-based algorithms. Wetland and riparian areas were generally well defined, although different algorithms and image analysis approaches differed slightly in their specific depictions of these land cover types. Wetland areas appeared to be best represented by the SVM based classifications, particularly when using the object-based approach, which accurately portrayed vegetation encircling areas of open water, although this quality is present when using tree-based classifications to varying degrees. Likewise, the depiction of riparian vegetation was relatively consistent across approaches and algorithms, with pixel-based classifications producing the most visually accurate depictions along steep ridges and narrow channels. Crop land was best depicted by object-based classifications due to the generalized appearance, however the less speckled appearance of croplands using pixel-based RF and DT based classifications were also considered adequate. Pixel-based classifications based on RF and SVM algorithms produced more visually accurate depictions of sand bars (exposed rock/soil land cover type) in riparian areas than any of the object-based classifications.

4.3. Accuracy assessment and statistical comparisons

An accuracy assessment was performed for each classification produced in this study to evaluate how well predictions based on the optimized models, generated using repeated *k*-fold cross validation, compared against the "hold-out" test data. Table 4 contains detailed confusion matrices of classification accuracies based on the test data.

Overall, both pixel-based and object-based classifications performed similarly with respect to overall classification accuracy. In general, all land cover types achieved over 80% user's accuracy, with the exception of wetland land cover types, which scored below 80% when using pixel-based image analysis, or object-based image analysis using the DT algorithm. Producer's accuracy for several land cover types was relatively consistent for both pixel-based and object-based classifications, but specific differences between machine learning algorithms were apparent. For example, producer's accuracy for the crop land cover type was consistently over 80% for both pixel-based and object-based classifications, except when using the SVM classifier, where it decreased to 75% for both image analysis approaches. All pixel-based classifications achieved a producer's accuracy of 77.27% for wetland land cover types, while object-based classifications using the RF and SVM algorithm achieved over 95% for this class. Pixel-based classifications that utilized the DT algorithm had the lowest overall classification accuracy (87.6%), followed by SVM (89.26%), and RF (89.67%) classifications (Fig. 5). The same general trend was observed for object-based classifications, with the DT algorithm

obtaining the lowest overall classification accuracy (88.84%), followed by RF (93.39%) and SVM (94.21%) algorithms. Exact 95% confidence limits, calculated on the results obtained with the "hold-out" test data set, reveal a wide variability and overlap in overall accuracy reported between pixel-based and object-based classifications. Based on these results, the lowest performing classification model (pixel-based DT) potentially scored within the range of the best performing RF and SVM classifications (Fig. 5).

Based on a comparison between predictions made with optimized classification models built using repeated *k*-fold cross-validation (see Section 4.1) and the "hold-out" test data, the McNemar test indicated that the observed difference between pixel-based and object-based classifications was not statistically significant (p > 0.05) when the same machine learning algorithm was used (e.g., DT classification model using pixel-based or object-based image analysis). With pixel-based image analysis, the observed difference in classification accuracy between all three machine learning algorithms was not statistically significant (p > 0.05). For object-based classifications, a statistically significant difference (p = 0.05) in classification accuracy between models using DT and RF algorithms (p = 0.01162), and DT and SVM algorithms (p = 0.006714) was observed. The difference in overall classification accuracy between object-based classifications utilizing the RF and SVM algorithms was not statistically significant (p > 0.05).

5. Discussion

In general, classifications produced using either pixel-based or object-based image analysis created similar and visually acceptable depictions of the broad land cover classes present within the study area. As expected, compared to the pixel-based classifications, the object-based classifications offered a more generalized visual appearance and more contiguous depiction of land cover, which perhaps better represents how land cover interpreters and analysts actually perceive the landscape (Stuckens et al., 2000). In some cases, the generalized depiction of land cover classes produced by object-based image analysis may account for an apparent preference for object-based classifications over slightly better performing pixel-based classifications (e.g., Dorren et al., 2003). Nevertheless, additional processing of pixel-based imagery, either prior to or after classification, can also produce similar generalized representations of land cover, so such differences may in fact be largely trivial, at least when considering the use of medium spatial resolution imagery (10–30 m pixels).

When comparing overall classification accuracy (percentage of classes correctly predicted), there is an apparently consistent, but small (1–4%), improvement when using object-based image analysis over pixel-based image analysis (see Table 4 and Fig. 5). However, the large variability depicted by the exact 95% confidence intervals suggests that the sample size of the "hold-out" test data set (242) was too small for assessing such differences; therefore, any apparent trend reported here should be considered tentative. Deciding on a sampling effort that is economically feasible and logistically possible, with one that allows for statistically rigorous comparisons is a major consideration in operational settings where resources are often limited (Congalton, 1991). A sample size that is too large can waste valuable resources that provide unnecessary precision, whereas a sampling effort that is too small may not be capable of resolving any statistically meaningful differences when comparing classification accuracies (Foody, 2009).

Despite the low sample size of the test set and associated wider confidence limits, the McNemar test revealed that, when utilizing the same machine learning algorithm, the observed difference between pixel-based and object-based classification accuracy was not significant at the 5% level. The findings in this study suggest that, on the basis of achieving better overall classification accuracy for the application described in this study, there is no statistical basis for preferring pixel-based to object-based image analysis, when utilizing the

Table 4

Confusion matrices and associated classifier accuracies based on test data. A = crop land, B = mixed grasslands, C = exposed rock/soil, D = riparian, E = water, F = wetland; Oa = overall classification accuracy, Pa = producer's accuracy, Ua = user's accuracy, CI = confidence interval.

Pixel-based, decision tree

	A	B	C	D	E	F	Total	Ua
A	27	3	0	0	0	2	32	84.38%
B	1	60	1	5	0	3	70	85.71%
C	1	0	13	0	0	0	14	92.86%
D	3	4	0	72	0	0	79	91.14%
E	0	1	0	1	23	0	25	92.00%
F	0	1	0	4	0	17	22	77.27%
Total	32	69	14	82	23	22	242	
Pa	84.38%	86.96%	92.86%	87.80%	100.00%	77.27%		

Oa: 87.60%
Lower 95% CI: 82.78%
Upper 95% CI: 91.48%

Object-based, decision tree

	A	B	C	D	E	F	Total	Ua
A	26	0	1	0	0	0	27	96.30%
B	1	63	1	1	1	3	70	90.00%
C	1	0	12	0	1	1	15	80.00%
D	3	4	0	80	1	2	90	88.89%
E	0	0	0	0	18	0	18	100.00%
F	1	2	0	1	2	16	22	72.73%
Total	32	69	14	82	23	22	242	
Pa	81.25%	91.30%	85.71%	97.56%	78.26%	72.73%		

Oa: 88.84%
Lower 95% CI: 84.18%
Upper 95% CI: 92.52%

Pixel-based, random forest

	A	B	C	D	E	F	Total	Ua
A	27	2	0	0	0	0	29	93.10%
B	1	61	1	0	0	3	66	92.42%
C	1	1	13	0	0	0	15	86.67%
D	3	3	0	80	0	2	88	90.91%
E	0	0	0	0	19	0	19	100.00%
F	0	2	0	2	4	17	25	68.00%
Total	32	69	14	82	23	22	242	
Pa	84.38%	88.41%	92.86%	97.56%	82.61%	77.27%		

Oa: 89.67%
Lower 95% CI: 85.13%
Upper 95% CI: 93.20%

Object-based, random forest

	A	B	C	D	E	F	Total	Ua
A	27	1	0	0	1	0	29	93.10%
B	0	65	1	0	0	1	67	97.01%
C	1	0	13	0	0	0	14	92.86%
D	3	3	0	82	0	0	88	93.18%
E	0	0	0	0	18	0	18	100.00%
F	1	0	0	0	4	21	26	80.77%
Total	32	69	14	82	23	22	242	
Pa	84.38%	94.20%	92.86%	100.00%	78.26%	95.45%		

Oa: 93.39%
Lower 95% CI: 89.49%
Upper 95% CI: 96.17%

Pixel-based, support vector machine

	A	B	C	D	E	F	Total	Ua
A	24	2	1	1	0	1	29	82.76%
B	4	63	2	0	1	1	71	88.73%
C	1	1	11	0	0	0	13	84.62%
D	2	1	0	81	0	3	87	93.10%
E	0	0	0	0	20	0	20	100.00%
F	1	2	0	0	2	17	22	77.27%
Total	32	69	14	82	23	22	242	
Pa	75.00%	91.30%	78.57%	98.78%	86.96%	77.27%		

Oa: 89.26%
Lower 95% CI: 84.66%
Upper 95% CI: 92.86%

Object-based, support vector machine

	A	B	C	D	E	F	Total	Ua
A	24	0	1	0	0	0	25	96.00%
B	3	68	1	0	0	0	72	94.44%
C	1	0	11	0	0	0	12	91.67%
D	3	1	0	82	0	0	86	95.35%
E	0	0	0	0	21	0	21	100.00%
F	1	0	1	0	2	22	26	84.62%
Total	32	69	14	82	23	22	242	
Pa	75.00%	98.55%	78.57%	100.00%	91.30%	100.00%		

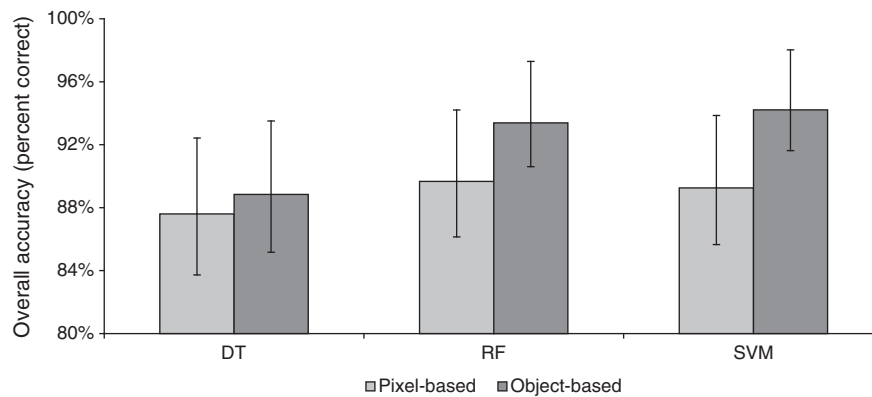Oa: 94.21%
Lower 95% CI: 90.40%
Upper 95% CI: 96.80%

Fig. 5. Comparison of overall classification accuracy (percent correct) of pixel-based and object-based classifications using three supervised machine learning algorithms: Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM). Results based on "hold-out" test set. Exact 95% confidence intervals plotted.

same machine learning algorithm. In addition, when using pixel-based image analysis, there was no statistically significant difference observed at the 5% level of significance between classification accuracies achieved by any of the machine learning algorithms. These findings are largely corroborated by the large overlap in confidence intervals depicted in Fig. 5. Nonetheless, when using object-based image analysis, statistically significant differences ($p < 0.05$) were observed for classification accuracies achieved by SVM and RF algorithms when compared to DT-based classifications. Unfortunately, the McNemar test as implemented here cannot be used for one-sided hypothesis testing (Foody, 2004), and the wide degree of overlap in the 95% confidence intervals for overall accuracy (Fig. 5) suggests that definitively asserting which classification algorithm or image analysis approach is capable of producing higher classification accuracies would be problematic based on the "hold-out" test set used in this study.

Other studies have indicated that both RF and SVM algorithms can achieve similar overall classification accuracies, which are typically greater than those obtained using DT based algorithms. For example, Pal (2005) found that both SVM and RF algorithms produced similar classification accuracies. Gislason et al. (2006) reported that RF based models achieved higher classification accuracies than those produced by standard DT (i.e., DTs that did not utilize bagged or boosting algorithms). These results differed from those reported by Otukei and Blaschke (2010) who found that DTs generally performed better than classifications produced using SVM. As with this study, the previous examples were based on medium- and relatively coarse-spatial resolution imagery (Landsat MSS, TM, ETM+) and used similar broad land cover classes; however, these comparisons relied on comparing overall classification accuracy values (i.e., the percentage of correctly classified samples) rather than using statistical comparison as employed here and elsewhere (e.g., Foody 2009).

When comparing overall accuracies between object-based and pixel-based classifications of Landsat-5 TM imagery, Dingle Robertson and King (2011) found no statistical difference between approaches. However, two studies (Yan et al., 2006; Whiteside et al., 2011) found that differences in overall classification accuracies produced using object-based image analysis were statistically significant ($p = 0.001$, and $p = 0.01$, respectively) than pixel-based image analysis, with both studies using medium spatial resolution EO imagery (ASTER and SPOT-5 HRG, respectively). Contrary to the side-by-side comparison conducted in this study, these previous studies compared different classifiers (e.g., MLC and K-NN) and image analysis methods, making direct comparisons difficult. Furthermore, as illustrated in this study, examination of confidence intervals around the overall classification accuracy assessments can reveal significant overlap in overall accuracies between image analysis approaches, confounding the interpretation of two-sided tests of significance such as McNemar's test (Foody, 2009), which have also been used

in previous comparisons (e.g., Dingle Robertson & King, 2011; Yan et al., 2006; Whiteside et al., 2011). Potential remedies include collecting a larger "hold-out" test sample to assess whether the large overlap in confidence intervals would remain, along with an appropriate means of testing a one-sided hypothesis for such a comparison. Unfortunately, the collection and use of an adequately sized "hold-out" test set might be prohibitive to assemble for logistical or financial reasons, and would represent an "inefficient use of data", as these data are, by definition, not utilized by the classifier (Kohavi, 1995). Implementing a repeated k-fold cross-validation, as illustrated in this study, with a larger dataset may provide statistically rigorous results without "wasting" data, while at the same time allowing for one-sided hypothesis testing to be performed (e.g., Kuhn, 2008).

From a practical production standpoint, the setup and execution of object-based classifications were more labor intensive as compared to their pixel-based counterparts. Much of the difference in execution time encountered was due to a lack of commercially available software for image analysis that implemented the machine learning algorithms examined in this study. This lack of a streamlined production environment multiplied the number of software packages needed and the amount of data transfers required. In addition, many of the present comparisons between pixel-based and object-based classifications of EO imagery in the available literature to date appear to rely on commercially available software solutions that provide relatively outdated and/or less advanced classification methods. The present study, along with others (e.g., Brenning, 2009, 2010), fill this void by providing a methodological basis for conducting statistically rigorous comparisons between classification outputs generated from EO imagery using freely available open-source software (e.g., R Development Core Team, 2010).

Regardless of which software packages are used, differences in execution time between pixel-based and object-based image analysis still remain. For example, the time spent selecting object-based variables (i.e., "object features") is roughly similar to that involved in selecting variables for a pixel-based classification; however, the additional time needed to select appropriate parameters for the underlying image segmentation is not trivial, especially if the tasks include mapping large overlapping scenes of imagery in an operational setting. Future development and adoption of more quantitative approaches for selecting optimal image segmentation parameters (e.g., Costa et al., 2008; Drăguţ et al., 2010) will hopefully reduce the time required for this important step, while at the same time producing superior results to the qualitative trial-and-error methods that are typically practiced now. In addition, faced with potentially hundreds of object features from which to select, the use of more advanced feature selection algorithms in object-based image analysis is gaining increasing attention (e.g., Yu et al., 2006; Chan & Paelinckx, 2008). Considered together, object-based image analysis will likely remain more labor intensive compared to pixel-based image analysis, which

is a factor that should be evaluated carefully when conducting image analysis of EO imagery in an operational environment.

While classification accuracy is an important attribute to consider, in circumstances where there are few overall statistical differences between image analysis approaches, other preferences may take precedence. For example, the node-based decision logic diagrams of DT based models may prove to be more preferable to users than achieving potentially higher overall classification accuracies using the RF algorithm. While statistically significant differences in overall classification accuracy were not observed in this study between pixel-based and object-based image analysis when utilizing the same machine learning algorithm, there may be other compelling reasons for selecting one image analysis approach over another. For example, object-based image analysis may prove to be more appropriate in situations that rely on the logic of updating and backdating image objects within a versatile GIS environment (e.g., Linke et al., 2009; Linke & McDermid, 2011). As previously mentioned, end users may prefer the generalized appearance of object-based classification maps as compared to pixel-based classification maps, even when pixel-based accuracy assessments are shown to be superior (Dorren et al., 2003). Such examples illustrate that the selection of an image analysis approach, or selection of an individual classification algorithm, may not always be driven by overall classification accuracy.

6. Conclusions

Classification of EO imagery using pixel-based and object-based image analysis was performed using three machine learning algorithms. No statistical difference between object-based and pixel-based classifications was found when the same machine learning algorithms were compared. When conducting object-based image analysis, RF or SVM algorithms produced classification accuracies that were statistically different compared to DT based algorithms. No statistical significant between pixel-based classifications were found. Based on visual assessments and interpretation of land cover distribution, all classifications were capable of depicting the broad land cover types selected for this study with similar, and acceptable, classification accuracies. More visually adequate overall depictions of riparian, wetland, and crop land cover types were attributed to RF and SVM based classifications, whereas DT based classifications contained noticeably more omission and commission errors in these classes. Object-based classifications were comparatively more time consuming to produce than their pixel-based counterparts. Based solely on overall classification accuracy, there appeared to be no advantage in selecting a particular image analysis approach.

References

Agresti, A. (2002). *Categorical data analysis.* : John Wiley and Sons.

Alpaydin, E. (2004). *Introduction to machine learning.* : MIT Press.

Baatz, M., & Schäpe, A. (2000). Multiresolution segmentation—an optimization approach for high quality multi-scale image segmentation. In J. Strobl, T. Blaschke, & G. Griesebner (Eds.), *Angewandte Geographische Informationsverarbeitung XII* (pp. 12–23). Wichmann-Verlag, Heidelberg.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing, 58*(3–4), 239–258.

Bhanu, B., Lee, S., & Ming, J. (1995). Adaptive image segmentation using a genetic algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, 25*(12), 1543–1567.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(1), 2–16.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32.

Breiman, L., & Cutler, A. (2007). *Random forests — Classification description.* : Random forests Available at:. http://stat-www.berkeley.edu/users/breiman/RandomForests/cc_home.htm [Accessed January 12, 2011]

Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and regression trees, Belmont, California, U.S.A.* : Chapman & Hall/CRC Available at:. http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0412048418 [Accessed January 12, 2011]

Brenning, A. (2009). Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. *Remote Sensing of Environment, 113*(1), 239–247.

Brenning, A. (2010). Land cover classification by multisource remote sensing: Comparing classifiers for spatial data. In H. Locarek-Junge, & C. Weihs (Eds.), *Classification as a tool for research* (pp. 435–443). Berlin, Heidelberg: Springer Berlin Heidelberg Available at: http://www.springerlink.com/content/x55n3g1314766146/ [Accessed October 4, 2011]

Burges, C. (1998). A tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery, 2*(2), 121–167.

Carreiras, J. M. B., Pereira, J. M. C., Campagnolo, M. L., & Shimabukuro, Y. E. (2006). Assessing the extent of agriculture/pasture and secondary succession forest in the Brazilian Legal Amazon using SPOT VEGETATION data. *Remote Sensing of Environment, 101*(3), 283–298.

Castilla, G., & Hay, G. J. (2008). Image objects and geographic objects. In Blaschke Thomas, S. Lang, & Geoffrey J. Hay (Eds.), *Object-based image analysis* (pp. 91–110). Berlin, Heidelberg: Springer Berlin Heidelberg Available at:. http://www.springerlink.com/content/g403k30318784w36/ [Accessed October 1, 2011]

Castillejo-González, I. L., López-Granados, F., García-Ferrer, A., Peña-Barragán, J. M., Jurado-Expósito, M., de la Orden, M. S, et al. (2009). Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Computers and Electronics in Agriculture, 68*(2), 207–215.

Chan, J., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment, 112*(6), 2999–3011.

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment, 113*(5), 893–903.

Chavez, P. S., Jr. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment, 24*(3), 459–479.

CNES (2009). SPOT image quality performances. Available at:. http://www.spotimage.com/automne_modules_files/standard/public/p551_29f05cbeaf21f085aab8a439d6fb4e14Performance_QI_Spot2009.pdf [Accessed January 11, 2011]

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37–46.

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

Congalton, R. G., & Green, K. (1998). *Assessing the accuracy of remotely sensed data: Principles and practices* (1st ed.). : CRC Press.

Cooper, A. B., Smith, C. M., & Smith, M. J. (1995). Effects of riparian set-aside on soil characteristics in an agricultural landscape: Implications for nutrient transport and retention. *Agriculture, Ecosystems & Environment, 55*(1), 61–67.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Costa, G. A. O. P., Feitosa, R. Q., Cazes, T. B., & Feijó, B. (2008). Genetic adaptation of segmentation parameters. In Thomas Blaschke, S. Lang, & Geoffrey J. Hay (Eds.), *Object-based image analysis* (pp. 679–695). Berlin, Heidelberg: Springer Berlin Heidelberg Available at:. http://www.springerlink.com/content/l7367p41j61715w5/ [Accessed July 27, 2011]

Dingle Robertson, L., & King, D. J. (2011). Comparison of pixel- and object-based classification in land cover change mapping. *International Journal of Remote Sensing, 32*(6), 1505–1529.

Dorren, L. K. A., Maier, B., & Seijmonsbergen, A. C. (2003). Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology and Management, 183*(1), 31–46.

Drăgut, L., Tiede, D., & Levick, S. R. (2010). ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science, 24*(6), 859.

Dubé, M. G. (2003). Cumulative effect assessment in Canada: a regional framework for aquatic ecosystems. *Environmental Impact Assessment Review, 23*(6), 723–745.

Duinker, P., & Greig, L. (2006). The impotence of cumulative effects assessment in Canada: Ailments and ideas for redeployment. *Environmental Management, 37*(2), 153–161.

Foody, G. M. (2004). Thematic map comparison: Evaluating the Statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing, 70*(5), 627–634.

Foody, G. M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing, 30*(20), 5273.

Franklin, S. E., & Peddle, D. R. (1990). Classification of SPOT HRV imagery and texture features. *International Journal of Remote Sensing, 11*(3), 551–556.

Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography, 26*(2), 173–205.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment, 61*(3), 399–409.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters, 27*(4), 294–300.

Gordon, L. J., Peterson, G. D., & Bennett, E. M. (2008). Agricultural modifications of hydrological flows create ecological surprises. *Trends in Ecology & Evolution, 23*(4), 211–219.

Gregory, S., Swanson, F., Mckee, W., & Cummins, K. (1991). An ecosystem perspective of riparian zones. *BioScience, 41*(8).

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing, 23*(4), 725–749.

Huel, D. (2000). Managing Saskatchewan Wetlands: A landowner's guide. Available at:. http://www.swa.ca/Publications/Documents/ManagingSaskatchewanWetlands.pdf [Accessed November 11, 2011]

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software, 11*(9), 1–20.

Kim, M., Madden, M., & Warner, T. (2008). Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. *Object-based image analysis* (pp. 293–307). Berlin, Heidelberg: Springer Berlin Heidelberg Available at:. http://www.springerlink.com/content/u7741201m568u327/ [Accessed January 11, 2011]

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International joint conference on artificial intelligence* (pp. 1137–1143). Available at:. http://citeseerx.ist.psu.edu/viewdoc/summary? doi=10.1.1.48.529

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal Of Statistical Software, 28*(5), 1–26.

Kuhn, M. (2011). The caret package. Available at:. http://cran.r-project.org/web/packages/caret/vignettes/caretTrain.pdf [Accessed October 3, 2011]

Laliberte, A. S., Fredrickson, E. L., & Rango, A. (2007). Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric Engineering and Remote Sensing, 73*(2), 197–207.

Laliberte, A.S., Koppa, J.S., Fredrickson, E.L., Rango, A. (2006). Comparison of nearest neighbor and rule-based decision tree classification in an object-oriented environment. In: IEEE international Geoscience and Remote Sensing Symposium Proceedings, July 31-August 4, 2006, Denver, Colorado.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18–22.

Linke, J., & McDermid, G. J. (2011). A conceptual model for multi-temporal landscape monitoring in an object-based environment. *Selected Topics in Applied Earth Observations and Remote Sensing. IEEE Journal of, 4*(2), 265–271.

Linke, J., McDermid, G. J., Laskin, D. N., McLane, A. J., Pape, A., Cranston, J., et al. (2009). A disturbance-inventory framework for flexible and reliable landscape monitoring. *Photogrammetric Engineering and Remote Sensing, 75*(8), 981–995.

Mathieu, R., Aryal, J., & Chong, A. K. (2007). Object-based classification of Ikonos imagery for mapping large-scale vegetation communities in urban areas. *Sensors, 7*(11), 2860–2880.

McKenzie, D. P., Mackinnon, A. J., Péladeau, N., Onghena, P., Bruce, P. C., Clarke, D. M., et al. (1996). Comparing correlated kappas by resampling: Is one level of agreement significantly different from another? *Journal of Psychiatric Research, 30*(6), 483–492.

Morissette, J. T., & Khorram, T. A. (1998). Exact binomial confidence interval for proportions. *Photogrammetric Engineering and Remote Sensing, 64*.

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment, 115*(5), 1145–1161.

Naiman, R., & Décamps, H. (1997). The ecology of interfaces: Riparian Zones. *Annual Review of Ecology and Systematics, 28*(1), 621–658.

Noble, B. F. (2008). Strategic approaches to regional cumulative effects assessment: a case study of the Great Sand Hills, Canada. *Impact Assessment and Project Appraisal, 26*, 78–90.

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation, 12*(Supplement 1), S27–S31.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217.

Pike, R. J. (2000). Geomorphometry —Diversity in quantitative surface analysis. *Progress in Physical Geography, 24*(1), 1–20.

Platt, R. V., & Rapoza, L. (2008). An evaluation of an object-oriented paradigm for land use/land cover classification. *The Professional Geographer, 60*(1), 87.

R Development Core Team (2010). R: A language and environment for statistical computing, Vienna, Austria. Available at:. http://www.R-project.org/

Ryherd, S., & Woodcock, C. (1996). Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering and Remote Sensing, 62*(2), 181–194.

Seitz, N. E., Westbrook, C. J., & Noble, B. F. (2011). Bringing science into river systems cumulative effects assessment practice. *Environmental Impact Assessment Review, 31*(3), 172–179.

Smith, A. (2010). Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm. *Journal of Spatial Science, 55*(1), 69.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and Change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sensing of Environment, 75*(2), 230–244.

Squires, A. J., Westbrook, C. J., & Dubé, M. G. (2009). An approach for assessing cumulative effects in a model river, the Athabasca River Basin. *Integrated Environmental Assessment and Management* (pp. 1)..

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment, 62*(1), 77–89.

Stuckens, J., Coppin, P. R., & Bauer, M. E. (2000). Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment, 71*(3), 282–296.

Therneau, T. M., & Ripley, B. A. (2010). rpart: Recursive Partitioning. Available at: http://CRAN.R-project.org/package=rpart

Thompson, W. H., & Hansen, P. L. (2001). Classification and management of riparian and wetland sites of the Saskatchewan prairie ecozone and parts of adjacent subregions. Available at:. http://www.swa.ca/Publications/Documents/ClassificationManagementRiparianWetlandSites.pdf [Accessed July 1, 2011]

Thuillier, G., Hersé, M., Labs, D., Foujols, T., Peetermans, W., Gillotay, D., et al. (2003). The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS and EURECA missions. *Solar Physics, 214*(1), 1–22.

Trimble (2010). eCognition® Developer 8.64.0 reference book. Available at:. http://www.definiens.com/ [Accessed January 11, 2011]

Trimble (2010). eCognition® Developer 8.64.0 user guide. Available at:. http://www.definiens.com/ [Accessed January 11, 2011]

US EPA (2005). National management measures to protect and restore wetlands and riparian areas for the abatement of nonpoint source pollution. Available at:. http://water.epa.gov/polwaste/nps/wetmeasures/index.cfm [Accessed July 26, 2011]

Van Coillie, F. M. B., Verbeke, L. P. C., & De Wulf, R. R. (2007). Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sensing of Environment, 110*(4), 476–487.

Vapnik, V. (1998). *Statistical learning theory.* : Wiley-Interscience Available at:. http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0471030031 [Accessed July 28, 2011].

Whiteside, T. G., Boggs, G. S., & Maier, S. W. (2011). Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation, 13*(6), 884–893.

Yan, G., Mas, J. F., Maathuis, B. H. P., Xiangmin, Z., & Van Dijk, P. M. (2006). Comparison of pixel-based and object-oriented image classification approaches-A case study in a coal fire area, Wuda, Inner Mongolia, China. *International Journal of Remote Sensing, 27*, 4039–4055.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., & Schirokauer, D. (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing, 72*(7), 799–811.

Zar, J. H. (2009). *Biostatistical analysis* (5th ed.). : Prentice Hall.